

Analítica del aprendizaje para la predicción en escenarios educativos heterogéneos

Pedro Manuel Moreno Marcos

Tesis depositada en cumplimiento parcial de los requisitos
para el grado de Doctor en

Ingeniería Telemática

Universidad Carlos III de Madrid

Directores:

Carlos Delgado Kloos
Pedro José Muñoz Merino

Tutor:

Pedro José Muñoz Merino

Julio de 2020

Esta tesis se distribuye bajo licencia “Creative Commons **Reconocimiento – No Comercial – Sin Obra Derivada**”.



*La paciencia es un árbol de raíz
amarga pero de frutos muy dulces.*

Proverbio persa

Agradecimientos

Durante estos últimos años, he vivido muchas experiencias y ha sido un largo camino, con mucho esfuerzo, días de luz, y días de sombra, para poder realizar todos los experimentos y poder llegar a escribir esta tesis doctoral. Sin embargo, en este largo viaje de la tesis, he estado bien acompañado, y antes de comenzar a presentar el texto, quiero dedicar unas palabras a todas aquellas personas que han estado conmigo durante estos años.

En primer lugar, quería agradecer a Pedro J. Muñoz, mi tutor de la tesis. Él fue quien me animó a comenzar esta tesis y el que me ha guiado durante todos estos años, y me ha dado la formación necesaria para llegar hasta aquí. También quiero agradecer a mi director Carlos Delgado por darme la oportunidad de unirme al Departamento y todas las oportunidades que me ha dado durante la investigación.

No puedo olvidar a todos los compañeros del Departamento, con los que he compartido todos estos años. De forma especial quiero agradecer a Carlos Alario, por todas sus aportaciones e ideas en los distintos trabajos en los que colaboramos y por ser mi guía en mis primeras experiencias docentes. Asimismo, quiero agradecer al resto de profesores del Grupo de Aplicaciones y Servicios Telemáticos, con los que he trabajado estos años: Iria Estévez, M. Carmen Fernández, M. Blanca Ibáñez, Raquel Crespo, Mario Muñoz, Luis Sánchez, Pablo Basanta, Marcos Mouriño, Jesús Arias, Florina Almenares, Daniel Díaz, Carlos García, Celeste Campo y Andrés Marín. También al resto de compañeros que hayan estado estos años en el departamento: Ramona Ruiz, M. Elena Alonso, Mario Sánchez, Jon Imaz, Dánae Martínez, Gabriel González, Salvador Benavent, José Antonio Ruipérez, Aarón Rubio, Álvaro Morante, Ainhoa Velacoracho, Lucía Uguina, Marta Moure, Nuria González, Raquel L. Pérez, Ana Martín, Cristina Márquez y muchos más.

Quiero también dar las gracias a la gente con la que pasé mis dos estancias de investigación. En primer lugar, a Katrien Verbert, por acogerme durante mi primera estancia en Lovaina. También a Tinne De Laet y Tom Broos, con los que colaboré durante la estancia y al resto del equipo, que me hicieron sentirme como en casa durante la estancia: Óscar Alvarado, Martijn Millecamp, Houda Lamqaddam, Robin De Croon, Nyi-Nyi Htun y Francisco Gutiérrez. Por otro lado, quiero dar las gracias a TC Pong, por acogerme en mi estancia en Hong Kong y por todo su feedback y apoyo durante la estancia. Ha sido un placer trabajar con él. También, al resto de miembros del equipo, y en especial a Albert Lo y Leo Fan, que me ayudaron a comprender mejor los datos y el MOOC con el que trabajé allí.

También quiero hacer mención a todos los compañeros de los proyectos de investigación en los que colaborado durante estos años, ya que de ellos también he aprendido y la colaboración ha sido muy enriquecedora. En especial, quiero nombrar a aquellos con los que he publicado o he preparado artículos de investigación en este tiempo: Yi-Shan Tsai, Jorge Maldonado, Mar Pérez, Dragan Gašević, Carolien Van Soom, Ioana Jivet, Maren Scheffel, Kairit Tammets, Kaire Kollom, Alexander Whitelock-Wainwright, Hendrik Drachsler, Adolfo Ruiz, Tobias Ley, Diego Rates,

Cristian Olivares, Eliana Scheihing, Margarita Ortiz, Alberto Jiménez y Vanessa Heredia.

Asimismo, quiero agradecer a todo el equipo de Santillana, a Javier Delicado Molina, Adrián García González, Carolina Acero Martín y Natasha Yushkevich Alexandrov, por su colaboración y comentarios, que han sido muy útiles para el desarrollo de esta tesis.

Un pilar esencial para realizar esta tesis doctoral ha sido la familia. Por ello, quiero agradecer en especial a mi padre, Pedro Manuel Moreno y mi madre, Rocío Marcos, por todo su apoyo y cariño durante todos estos años. Sin vuestro apoyo para hacer más llevadera la tesis, hubiera sido mucho más difícil. También, a mi tía, Nieves Marcos, y en especial, a mi abuelo, Lorenzo Marcos, que siempre me ha apoyado y seguro que le hace mucha ilusión cuando termine la tesis. No me puedo olvidar de Eufemia Arroyo, que se ha preocupado por mis avances en todo momento y ha sido un gran apoyo. También de Clara Martínez y María José Nache, que ya desde hace años creyó en mí y dijo que llegaría lejos. Por último, quiero extender el agradecimiento al resto personas, que a lo largo de mi carrera profesional me hayan apoyado de uno u otro modo, ya que sin la colaboración de todos no podría haber llegado hasta aquí.

¡Muchas gracias a todos!

Contenidos publicados y presentados

La tesis doctoral cubre parte de los contenidos publicados y presentados en varias contribuciones previas. A continuación, se muestra el listado de todas las publicaciones del autor. Todas estas publicaciones han sido mencionadas en la tesis, aunque algunas tienen mayor relación con la tesis y otras tienen una relación menor. En el listado se enuncian los autores, título, año de publicación y nombre de revista o conferencia; las partes de la tesis en las que se ubican partes de cada publicación; y el rol del autor en dicha publicación.

A. Publicaciones en revistas indexadas en el JCR

En primer lugar, se presentan las publicaciones indexadas en el Journal Citation Report (JCR).

1. **Pedro Manuel Moreno-Marcos**, Pedro J. Muñoz-Merino, Carlos Alario-Hoyos, and Carlos Delgado Kloos. 2020. Re-defining, Analyzing and Predicting Persistence using Student events in Online Learning. *Applied Sciences*, 10(5):1722. DOI: [10.3390/app10051722](https://doi.org/10.3390/app10051722). Factor de impacto 2018: 2.217. JCR-SCI (Q2), categoría: Physics, Applied.
 - Esta contribución está incluida parcialmente en la tesis, en los capítulos 2, 3, 5 y 6.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de la limpieza y procesamiento de los datos a partir de los eventos de los alumnos parcialmente preprocesados, la realización del software para calcular los indicadores, la realización de los modelos predictivos, la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
2. **Pedro Manuel Moreno-Marcos**, Ting-Chuen Pong, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. 2020. Analysis of the Factors Influencing Learners' Performance Prediction with Learning Analytics. *IEEE Access*, 8:5264-5282. DOI: [10.1109/ACCESS.2019.2963503](https://doi.org/10.1109/ACCESS.2019.2963503). Factor de impacto 2018: 4.098. JCR-SCI (Q1), categorías: Computer Science, Information Systems; Engineering, Electrical & Electronic; y Telecommunications.
 - Esta contribución está incluida parcialmente en la tesis, en los capítulos 3, 5, 6, 7 y 9.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de la limpieza y procesamiento de los datos a partir de los eventos de los alumnos, la realización del software para calcular los indicadores, la realización de

los modelos predictivos, la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.

- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

3. Yi-Shan Tsai, Diego Rates, **Pedro Manuel Moreno-Marcos**, Pedro J. Muñoz-Merino, Ioana Jivet, Maren Scheffel, Hendrik Drachslar, Carlos Delgado Kloos, and Dragan Gasevic. Learning Analytics in European Higher Education-Trends and Barriers. *Computers & Education*, 155, October 2020. DOI: [10.1016/j.compedu.2020.103933](https://doi.org/10.1016/j.compedu.2020.103933). Factor de impacto 2018: 5.627. JCR-SCI (Q1), categoría: Computer Science, Interdisciplinary Applications. JCR-SSCI (Q1), categoría: Education & Educational Research.

- Esta contribución está incluida parcialmente en la tesis, en el capítulo 4.
- El rol del autor de esta tesis en esta publicación fue la realización parcial de la transcripción y codificación de las entrevistas a líderes institucionales y expertos, y la contribución en parte en el análisis e interpretación de los resultados. También, el autor de esta tesis contribuyó en parte de la escritura del artículo.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

4. **Pedro Manuel Moreno-Marcos**, Pedro J. Muñoz-Merino, Jorge Maldonado-Mahauad, Mar Pérez-Sanagustín, Carlos Alario-Hoyos, and Carlos Delgado Kloos. 2020. Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Computers & Education*, 145(1), February 2020. DOI: [10.1016/j.compedu.2019.103728](https://doi.org/10.1016/j.compedu.2019.103728). Factor de impacto 2018: 5.627. JCR-SCI (Q1), categoría: Computer Science, Interdisciplinary Applications. JCR-SSCI (Q1), categoría: Education & Educational Research.

- Esta contribución está incluida parcialmente en la tesis, en los capítulos 2, 3, 5, 6, 7 y 8.
- El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización del software para calcular los indicadores a partir de datos parcialmente procesados, la realización de los modelos predictivos, la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

5. **Pedro Manuel Moreno-Marcos**, Tinne De Laet, Pedro J. Muñoz-Merino, Carolien Van Soom, Tom Broos, Katrien Verbert, and Carlos Delgado Kloos.

2019. Generalizing predictive models of admission test success based on online interactions. *Sustainability*, 11(18):4940. DOI: [10.3390/su11184940](https://doi.org/10.3390/su11184940). Factor de impacto 2018: 2.592. JCR-SCI (Q2), categoría: Environmental Sciences. JCR-SSCI (Q2), categorías: Environmental Studies, y Green & Sustainable Science & Technology.

- Esta contribución está incluida parcialmente en la tesis, en los capítulos 2, 3, 5, 6, 7 y 8.
- El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de la limpieza y procesamiento de los datos a partir de los eventos de los alumnos, la realización del software para calcular los indicadores, la realización de los modelos predictivos, la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

6. **Pedro Manuel Moreno-Marcos**, Carlos Alario-Hoyos, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. 2019. Prediction in MOOCs: A review and future research directions. *IEEE Transactions on Learning Technologies*, 12(3), 384-401. DOI: [10.1109/TLT.2018.2856808](https://doi.org/10.1109/TLT.2018.2856808). Factor de impacto 2018: 2.315. JCR-SCI (Q2), categoría: Computer Science, Interdisciplinary Applications. JCR-SSCI (Q1), categoría: Education & Educational Research.

- Esta contribución está incluida parcialmente en la tesis, en el capítulo 2.
- El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de las búsquedas en las bases de datos, el filtrado y clasificación de los artículos, la obtención de los datos en forma de tablas y visualizaciones tras procesar los artículos, y la contribución en parte en la extracción de resultados y conclusiones sobre lo realizado en el estado del arte y los posibles trabajos futuros del área. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

7. **Pedro Manuel Moreno-Marcos**, Carlos Alario-Hoyos, Pedro J. Muñoz-Merino, Iria Estévez-Ayres, and Carlos Delgado Kloos. 2018. A learning analytics methodology for understanding social interactions in MOOCs. *IEEE Transactions on Learning Technologies*. DOI: [10.1109/TLT.2018.2883419](https://doi.org/10.1109/TLT.2018.2883419). Factor de impacto 2018: 2.315. JCR-SCI (Q2), categoría: Computer Science, Interdisciplinary Applications. JCR-SSCI (Q1), categoría: Education & Educational Research.

- Esta contribución está incluida parcialmente en la tesis, en el capítulo 3.

- El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la contribución en parte en la propuesta de la metodología, el desarrollo del software acorde a la metodología propuesta, la aplicación de la metodología propuesta en datos de MOOCs, la elaboración de tablas y visualizaciones para ilustrar la metodología y la herramienta presentada, y la contribución en parte en el análisis e interpretación de los resultados. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

8. **Pedro Manuel Moreno-Marcos**, Pedro J. Muñoz-Merino, Carlos Alario-Hoyos, Iria Estévez-Ayres, and Carlos Delgado Kloos. 2018. Analysing the predictive power for anticipating assignment grades in a Massive Open Online Course. *Behaviour & Information Technology*, 37(10-11):1021-1036. DOI: [10.1080/0144929X.2018.1458904](https://doi.org/10.1080/0144929X.2018.1458904). Factor de impacto 2018: 1.429. JCR-SCI (Q3), categoría: Computer Science, Cybernetics. JCR-SSCI (Q2), categoría: Ergonomics.

- Esta contribución está incluida parcialmente en la tesis, en los capítulos 3, 5, 6 y 7.
- El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de la limpieza y procesamiento de los datos a partir de los registros de los alumnos, la realización del software para calcular los indicadores, la realización de los modelos predictivos, la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

B. Publicaciones en otras revistas

1. Yi-Shan Tsai, **Pedro Manuel Moreno-Marcos**, Ioana Jivet, Maren Schefel, Kairit Tammets, Kaire Kollom, and Dragan Gasevic. 2018. The SHEILA framework: informing institutional strategies and policy processes of learning analytics. *Journal of Learning Analytics*, 5(3):5-20. DOI: [10.18608/j-la.2018.53.2](https://doi.org/10.18608/j-la.2018.53.2).

- Esta contribución está incluida parcialmente en la tesis, en el capítulo 4.
- El rol del autor de esta tesis en esta publicación fue la realización parcial de la transcripción y codificación de las entrevistas a líderes institucionales y expertos, y la contribución en parte en el análisis e interpretación de los resultados. También, el autor de esta tesis contribuyó en parte de la escritura del artículo.

- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

C. Publicaciones en congresos

1. **Pedro Manuel Moreno-Marcos**, Dánae Martínez de la Torre, Gabriel González Castro, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. 2020. Should we consider Efficiency and Constancy for Adaptation in Intelligent Tutoring Systems? In *Proceedings of the 16th International Conference on Intelligent Tutoring Systems, Athens, Greece, June 2020 (ITS '20)*, 10 pages [aceptado] (Ranking: Core A).
 - Esta contribución está incluida parcialmente en la tesis, en los capítulos 2, 3, 5 y 6.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización parcial del procesamiento de los datos y cálculo de indicadores, la obtención de datos en forma de visualizaciones, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
2. Aarón Rubio-Fernández, **Pedro Manuel Moreno-Marcos**, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. 2020. An Initial Analysis of Prediction Techniques as a Support for the Flipped Classroom. In *Proceedings of the Learning Analytics Summer Institute Spain 2020, Valladolid, Spain, June 2020 (LASI SPAIN '20)*, 8 pages [aceptado].
 - Esta contribución está incluida parcialmente en la tesis, en el capítulo 10.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización del software de la herramienta de predicción que se comenta en el artículo y la contribución en parte en la escritura del artículo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
3. **Pedro Manuel Moreno-Marcos**, Pedro J. Muñoz-Merino, Carlos Alario-Hoyos, and Carlos Delgado Kloos. 2019. Analyzing Students' Persistence using an Event-Based Model. In *Proceedings of the Learning Analytics Summer Institute Spain 2019, Vigo, Spain, June 2019 (LASI SPAIN '19)*, 15 pages, CEUR-WS.org, 2415, pp. 56-70. ISBN: 978-84-16829-40-8. Enlace: <http://ceur-ws.org/Vol-2415/paper06.pdf>.
 - Esta contribución está incluida parcialmente en la tesis, en los capítulos 3 y 5.

- El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de la limpieza y procesamiento de los datos a partir de los eventos de los alumnos parcialmente preprocesados, la realización del software para calcular la persistencia, la realización de los modelos predictivos, la obtención de los datos en forma de visualizaciones tras realizar los análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
4. **Pedro Manuel Moreno-Marcos**, Tinne De Laet, Pedro J. Muñoz-Merino, Carolien Van Soom, Tom Broos, Katrien Verbert, and Carlos Delgado Kloos. 2019. Predicting admission test success using SPOC interactions. In *Companion Proceedings of the International Conference on Learning Analytics and Knowledge, Tempe, Arizona, USA, March 2019 (LAK '19)*, 11 pages, pp. 924-934. Enlace: [Companion Proceedings LAK 2019](#).
- Esta contribución está incluida parcialmente en la tesis, en los capítulos 3, 5, 6, 7 y 8.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de la limpieza y procesamiento de los datos a partir de los eventos de los alumnos, la realización del software para calcular los indicadores, la realización de los modelos predictivos, la obtención de los datos en forma de tablas tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
5. Jorge Maldonado-Mahauad, Mar Perez-Sanagustin, **Pedro Manuel Moreno-Marcos**, Carlos Alario-Hoyos, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. 2018. Predicting Learners' Success in a Self-Paced MOOC Based on Sequence Patterns of Self-Regulated Learning. In *Proceedings of the 13th European Conference on Technology Enhanced Learning, Leeds, UK, September 2018 (EC-TEL '18)*, 14 pages, pp. 355-369. DOI: [10.1007/978-3-319-98572-5_27](#).
- Esta contribución está incluida parcialmente en la tesis, en los capítulos 3, 5 y 6.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de los modelos predictivos, la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis contribuyó en parte de la escritura del artículo.

- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
6. **Pedro Manuel Moreno-Marcos**, Carlos Alario-Hoyos, Pedro J. Muñoz-Merino, Iria Estévez-Ayres, and Carlos Delgado Kloos. 2018. Sentiment Analysis in MOOCs: A case study. In *Proceedings of the IEEE Global Engineering Education Conference, Santa Cruz de Tenerife, Spain, April 2018 (EDUCON '18)*, 8 pages, pp. 1489-1496. DOI: [10.1109/EDUCON.2018.8363409](https://doi.org/10.1109/EDUCON.2018.8363409).
- Esta contribución está incluida parcialmente en la tesis, en el capítulo 3.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización de la limpieza y procesamiento de los datos a partir de los registros de los alumnos, la realización del software para calcular el sentimiento de los alumnos, la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis se encargó de liderar la escritura del artículo y fue el contribuidor principal de la escritura del mismo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
7. Yi-Shan Tsai, **Pedro Manuel Moreno-Marcos**, Kairit Tammets, Kaire Kollom, and Dragan Gasevic. 2018. SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. In *Proceedings of the International Conference on Learning Analytics and Knowledge, Sydney, Australia, March 2018 (LAK '18)*, 10 pages, pp. 320-329. DOI: [10.1145/3170358.3170367](https://doi.org/10.1145/3170358.3170367) (Seleccionado entre los mejores artículos del congreso).
- Esta contribución está incluida parcialmente en la tesis, en el capítulo 4.
 - El rol del autor de esta tesis en esta publicación fue la realización parcial de la transcripción y codificación de las entrevistas a líderes institucionales y expertos, y la contribución en parte en el análisis e interpretación de los resultados. También, el autor de esta tesis contribuyó en parte de la escritura del artículo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

D. Libros

1. Yi-Shan Tsai, Dragan Gasevic, Alexander Whitelock-Wainwright, Pedro J. Muñoz-Merino, **Pedro M. Moreno-Marcos**, Aarón Rubio Fernández, Carlos Delgado Kloos, Maren Scheffel, Ioana Jivet, Hendrik Drachsler, Kairit Tammets, Adolfo Ruiz Calleja, and Kaire Kollom. 2018. SHEILA: Supporting Higher Education to Intergrate Learning Analytics. ISBN: 978-1-912669-02-8. Retrieved from: <http://sheilaproject.eu/wp-content/uploads/2018/11/SHEILA-research-report.pdf>.

- Esta contribución está incluida parcialmente en la tesis, en el capítulo 4.
- El rol del autor de esta tesis en esta publicación fue la realización parcial de la transcripción y codificación de las entrevistas a líderes institucionales, expertos y de los grupos focales a profesores, y la contribución en parte en el análisis e interpretación de los resultados de los análisis de dichas entrevistas y grupos focales. También, el autor de esta tesis contribuyó en parte de la escritura del documento.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

E. Publicaciones enviadas en proceso de revisión

A continuación se enumeran los artículos que todavía no se encuentran publicados, pero que han sido enviados a revistas y se encuentran en proceso de revisión. También hay otros resultados de la tesis que no han sido publicados y que no se han incluido todavía como parte de ningún artículo, los cuales no se enumeran aquí.

1. Pedro J. Muñoz-Merino, **Pedro Manuel Moreno-Marcos**, Aaron Rubio-Fernández, Yi-Shan Tsai, Dragan Gašević, and Carlos Delgado Kloos. A Systematic Analysis of Learning Analytics using Multi-source Data in the Context of Spain (*enviado*).

- Esta contribución está incluida parcialmente en la tesis, en el capítulo 4.
- El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización parcial de la transcripción de las entrevistas a líderes institucionales, expertos y de los grupos focales a profesores, y la contribución en parte en el análisis e interpretación de los datos de las entrevistas y grupos focales (principalmente los relacionados con la predicción). También, el autor de esta tesis contribuyó en parte de la escritura del artículo.
- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

2. Kaire Kollom, Maren Scheffel, Kairit Tammets, Ioana Jivet, Yi-Shan Tsai, Pedro J. Muñoz-Merino, **Pedro Manuel Moreno-Marcos**, Alexander Whitelock-Wainwright, Adolfo Ruiz Calleja, Dragan Gasevic, Carlos Delgado Kloos, Hendrik Drachsler, Tobias Ley. A four-country cross-case analysis of academic staff expectations about learning analytics in higher education (*enviado*).

- Esta contribución está incluida parcialmente en la tesis, en el capítulo 4.
- El rol del autor de esta tesis en esta publicación fue la realización parcial de la transcripción y codificación de los grupos focales a profesores, y la contribución en parte en el análisis e interpretación de los resultados. También, el autor de esta tesis contribuyó en parte de la escritura del artículo.

- El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.
- 3. Cristian Olivares-Rodríguez, **Pedro Manuel Moreno-Marcos**, Eliana Scheihing, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. A configuration model to predict dropout based on academic trajectory features (*enviado*).
 - Esta contribución está incluida parcialmente en la tesis, en el capítulo 8.
 - El rol del autor de esta tesis en esta publicación fue la contribución en parte en la concepción y diseño, la realización parcial de la limpieza y procesamiento de los datos a partir de los eventos de los alumnos parcialmente preprocesados, la realización parcial del software para calcular los indicadores, la realización de los modelos predictivos, la contribución en parte en la obtención de los datos en forma de tablas y visualizaciones tras aplicar diferentes modelos y análisis, y la contribución en parte en el análisis e interpretación de los datos. También, el autor de esta tesis contribuyó en parte de la escritura del artículo.
 - El material de esta fuente incluido en la tesis no está señalado por medios tipográficos ni referencias.

F. Trabajos previos

El trabajo de la tesis doctoral toma como partida el trabajo realizado en un Trabajo Fin de Máster previo. A continuación, se proporciona la cita de dicho trabajo y se comenta sobre qué parte se continúa el trabajo en esta tesis doctoral.

- **Pedro Manuel Moreno Marcos**. Analysing the predictive power for anticipating assignment grades in a Massive Open Online Course. 2017. Trabajo Fin de Máster del Máster Universitario en Ingeniería Telemática. Universidad Carlos III de Madrid. Calificación: Matrícula de Honor (10).
 - En este trabajo se aborda la predicción de notas en un MOOC. El análisis desarrollado en este trabajo no es parte de la tesis, pero se comentan resultados porque son necesarios para entender otros resultados de la tesis que extienden la investigación desarrollada en este trabajo. En particular, en este trabajo se analiza uno de los seis escenarios de la tesis, y en la tesis se reutiliza este escenario para poder compararlo con otros escenarios y poder obtener conclusiones globales sobre la predicción. Los resultados de los análisis presentados en este trabajo aparecen parcialmente en los capítulos 5, 6 y 7. Cada vez que aparece algún resultado procedente de este trabajo, se indica explícitamente en el texto.

Otros méritos de investigación

Aparte de las publicaciones mostradas en la sección anterior, esta tesis doctoral ha contribuido en los siguientes proyectos de investigación:

1. LALA (Building Capacity to Use Learning Analytics to Improve Higher Education in Latin America)¹.
 - Referencia: 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP.
 - Proyecto europeo, financiado por la Comisión Europea a través del programa Erasmus+.
 - Duración: 36 meses (10/2017 a 10/2020).
 - Responsable en la Universidad Carlos III de Madrid: Pedro José Muñoz Merino.
 - Entidades participantes: Universidad Carlos III de Madrid (coordinador), Escuela Superior Politécnica del Litoral, Universidad de Cuenca, Pontificia Universidad Católica de Chile, Universidad Austral de Chile, Katholieke Universiteit Leuven y University of Edinburgh.
2. SHEILA (Supporting Higher Education to Integrate Learning Analytics)².
 - Referencia: 562080-EPP-1-2015-1-BE-EPPKA3-PI-FORWARD.
 - Proyecto europeo, financiado por la Comisión Europea a través del programa Erasmus+.
 - Duración: 30 meses (11/2015 a 04/2018).
 - Responsable en la Universidad Carlos III de Madrid: Carlos Delgado Kloos.
 - Entidades participantes: University of Edinburgh (coordinador), Brussels Education Services, Open University Netherlands, Tallinn University, Universidad Carlos III de Madrid, European Association for Quality Assurance in Higher Education y Erasmus Student Network.
3. SmartLET: Analítica del aprendizaje para mejorar el diseño y la orquestación en entornos inteligentes de aprendizaje escalables y ubicuos, enriquecidos con Internet de las Cosas³.
 - Referencia: TIN2017-85179-C3-1-R.
 - Proyecto nacional, financiado por el Ministerio de Ciencia, Innovación y Universidades, a través del programa Retos.
 - Duración: 36 meses (01/2018 a 12/2020).
 - Responsables en la Universidad Carlos III de Madrid: Carlos Delgado Kloos y Pedro José Muñoz Merino.

¹<https://www.lalaproject.org>

²<https://sheilaproject.eu/>

³<https://smartlet.gsic.uva.es/>

- Entidades participantes: Universidad Carlos III de Madrid (coordinador), Universidad de Valladolid y Universidad Pompeu Fabra.
4. SNOLA: Spanish Network of Learning Analytics⁴.
- Referencia: TIN2015-71669-REDT.
 - Proyecto nacional, financiado por el Ministerio de Economía, Industria y Competitividad, a través del programa Retos-Redes.
 - Duración: 24 meses (01/2016 a 12/2017).
 - Responsable en la Universidad Carlos III de Madrid: Pedro José Muñoz Merino.
 - Entidades participantes: Universidad de Deusto (coordinador), Universidad de León, Universidad del País Vasco, Universidad Politécnica de Madrid, Universidad de Valladolid, Universidad Carlos III de Madrid, Universidad de Vigo, Universidad Nacional de Educación a Distancia y Universidad de Salamanca.
5. RESET-UC3M: Reformulando ecosistemas escalables educativos⁵.
- Referencia: TIN2014-53199-C3-1-R.
 - Proyecto nacional, financiado por el Ministerio de Economía, Industria y Competitividad, a través del programa Retos.
 - Duración: 36 meses (01/2015 a 12/2017).
 - Responsable en la Universidad Carlos III de Madrid: Carlos Delgado Kloos.
 - Entidades participantes: Universidad Carlos III de Madrid (coordinador), Universidad de Valladolid y Universidad Pompeu Fabra.
6. eMadrid: Investigación y Desarrollo de tecnologías educativas en la Comunidad de Madrid⁶.
- Referencia: S2013/ICE-2715.
 - Proyecto regional, financiado por la Comunidad de Madrid.
 - Duración: 48 meses (10/2014 a 09/2018).
 - Responsable en la Universidad Carlos III de Madrid: Carlos Delgado Kloos.
 - Entidades participantes: Universidad Carlos III de Madrid (coordinador), Universidad Autónoma de Madrid, Universidad Complutense de Madrid, Universidad Politécnica de Madrid, Universidad Rey Juan Carlos y Universidad Nacional de Educación a Distancia.
7. e-Madrid-CM. Investigación y desarrollo de tecnologías educativas en la Comunidad de Madrid⁶.

⁴<https://snola.es/>

⁵<http://reset.gast.it.uc3m.es/>

⁶www.emadridnet.org

- Referencia: S2018/TCS-4307.
 - Proyecto regional, financiado por la Comunidad de Madrid.
 - Duración: 24 meses (01/2019 a 12/2020, aunque prorrogable a 48 meses si es aceptado el informe científico-técnico).
 - Responsable en la Universidad Carlos III de Madrid: Carlos Delgado Kloos.
 - Entidades participantes: Universidad Carlos III de Madrid (coordinador), Universidad Autónoma de Madrid, Universidad Complutense de Madrid, Universidad Politécnica de Madrid, Universidad Rey Juan Carlos y Universidad Nacional de Educación a Distancia.
8. Analítica visual, predicción y recomendación para aplicación educativa de Santillana.
- Proyecto artículo 83 (proyecto de colaboración con empresa).
 - Duración: 24 meses (07/2018 a 07/2020).
 - Responsables en la universidad Carlos III de Madrid: Carlos Delgado Kloos y Pedro José Muñoz Merino.
 - Entidades participantes: Santillana Educación S.L. y Universidad Carlos III de Madrid.

Además, cabe destacar que algunos resultados de la tesis provienen del trabajo en dos estancias de investigación:

- La primera estancia se realizó en la Katholieke Universiteit Leuven (KU Leuven), Bélgica, entre el 1 y 31 de agosto de 2018. La estancia se realizó dentro el grupo Augment, perteneciente al grupo de investigación HCI (Human-Computer Interaction, en español, interacción entre el hombre y el ordenador), bajo la supervisión de la profesora Katrien Verbert. En esta estancia, se realizaron modelos predictivos sobre SPOCs para predecir la nota de la prueba de acceso a medicina y odontología en Flandes. Como publicaciones relacionadas con esta estancia están un artículo en revista (A5, del listado de publicaciones) y un artículo en un workshop de un congreso (C3).
- La segunda estancia se realizó en la Hong Kong University of Science and Technology (HKUST), entre el 10 de junio y el 10 de septiembre de 2019. La estancia se desarrolló dentro del Departamento de Ciencias de la Computación e Ingeniería, y bajo la supervisión del profesor Ting-Chuen Pong. En esta estancia se trabajó la predicción de notas en MOOCs. Como publicaciones relacionadas con esta estancia está un artículo en revista (A2). Esta estancia estuvo financiada por el Ministerio de Ciencia, Innovación y Universidades, a través de la convocatoria para ayudas a la movilidad para estancias breves y traslados temporales (EST18/00554).

Por último, se hace constar que esta tesis obtuvo la financiación del Ministerio de Universidades a través de una ayuda para la Formación del Profesorado Universitario (FPU016/00526).

Resumen

Los entornos educativos permiten recoger una gran cantidad de información sobre la actividad de los estudiantes en los cursos. Esta información puede ser utilizada, por ejemplo, para predecir el comportamiento o los resultados de los estudiantes y detectar alumnos en riesgo, de modo que el profesor pueda identificar posibles dificultades y llevar a cabo intervenciones. A pesar de que ha habido muchos trabajos previos de predicción en educación, quedan aún muchos retos pendientes. Por ejemplo, es necesario conocer más sobre variables predictoras nuevas y su relación con otras ya utilizadas (p. ej., de vídeos o ejercicios). También, es importante poder determinar a partir de varios contextos cuál es el momento a partir del cual la predicción puede ser lo suficientemente precisa y cómo los modelos pueden generalizar a otros contextos. Asimismo, hace falta comprender cómo se relacionan diferentes variables de la predicción y si las predicciones de estas variables pueden combinarse para aportar más información.

Para contribuir en parte a la solución de estos problemas, esta tesis se centra en el estudio de la predicción en varios escenarios educativos en relación con varios aspectos, en concreto (1) el efecto de algunas variables predictoras, (2) el análisis temporal, (3) el problema de la generalización, y (4) el efecto de algunas variables a predecir. Se utilizan seis escenarios de aprendizaje, en los cuales se realiza un análisis de la predicción para posteriormente extraer conclusiones globales. Estos seis escenarios incluyen Massive Open Online Courses (MOOCs) de edX y Coursera, Small Private Online Courses (SPOCs) en Open edX como apoyo a cursos universitarios y en edX Edge para la preparación de pruebas de acceso universitarias, y datos de alumnos de primaria y secundaria (K-12) en una plataforma propietaria.

Las variables dependientes a predecir incluyen el abandono académico, la consecución de objetivos (éxito), y las notas, tras identificar el interés de estas variables, tanto en la revisión de la literatura, como en entrevistas y grupos focales llevados a cabo a líderes institucionales, expertos, profesores y estudiantes.

Los resultados muestran que las variables relacionadas con interacciones con ejercicios proporcionan un poder predictivo alto, y este resultado generaliza a los diferentes contextos. Asimismo, las notas sumativas previas consiguen mejorar el poder predictivo cuando están disponibles, y las variables sobre actividad en la plataforma e interacciones con vídeos normalmente ofrecen un buen poder predictivo y son útiles en los modelos. También, las variables de aprendizaje autorregulado pueden contribuir en los modelos y ofrecer un buen poder predictivo por sí mismas, aunque pueden tener menor efecto cuando otras variables principales, como las de ejercicios, están presentes, por su posible relación con las mismas. Sin embargo, otras variables como la actividad en el foro o comportamientos, como la eficiencia, constancia o persistencia no consiguen un alto poder predictivo o mejorar significativamente los modelos.

En cuanto al análisis temporal, los resultados muestran que en todos los contextos, independientemente del tipo de curso o si eran síncronos o asíncronos, pueden obtenerse unas buenas predicciones a partir del 13-40 % de la duración del curso. Aunque el contexto pueda influir significativamente en el poder predictivo, esto implica que pueden conseguirse predicciones relativamente tempranas.

Respecto a la generalización, se analizan tres aproximaciones. En la primera, se estudia la transferencia de un modelo de un curso a otro, y se obtienen buenos resultados cuando se transfiere un modelo a otro curso diferente (aunque con contexto similar) con los mismos estudiantes (mismo cohorte). Además, los resultados pueden ser aceptables al transferir a otra edición del mismo curso. Sin embargo, el poder predictivo puede reducirse considerablemente al cambiar de curso y estudiantes, aunque puede haber contextos concretos (como se identifican en MOOCs) donde no ocurra.

La segunda aproximación consiste en un modelo global, entrenado con datos de varios cursos, los cuales pueden tener bastantes diferencias entre ellos. Esta aproximación puede ser útil, por ejemplo, cuando hay cursos nuevos o cursos con muy pocos alumnos para generar modelos específicos. Sobre esta aproximación, se observa que los modelos pueden obtener un alto poder predictivo a nivel global, aunque pueden fallar en cursos concretos. En particular, en torno al 15-18 % de los cursos, el Área Bajo la Curva (AUC) disminuye en más de 0.1 y la Raíz del Error Cuadrático Medio (RMSE) es al menos 0.05 mayor. Por ello, debe analizarse en cada caso si el modelo global puede funcionar en cada curso concreto. No obstante, dado que el contexto puede afectar bastante en los modelos, una recomendación es reutilizar y adaptar los modelos a cada contexto, que es la tercera aproximación propuesta. En este caso, se toma un modelo ya existente y se readapta el código para añadir/eliminar variables si es necesario y generar un modelo para cada curso específico.

En torno a las variables a predecir, el análisis de los diferentes escenarios muestra que hay diversos aspectos que pueden condicionar la selección y definición de la variable. Por ejemplo, en un curso puede considerarse que un alumno abandona cuando está inactivo durante cierto tiempo, pero en otro, aunque un alumno se conecte a la plataforma, se puede considerar abandono si no está centrado en el curso y no hace las actividades (p. ej., en primaria/secundaria donde los alumnos pueden conectarse a la plataforma en clase, pero no realizar las actividades que se les piden). Además, en torno a las variables a predecir, se observa que se obtiene un mejor poder predictivo al predecir la nota final que la nota del examen final. También, se aprecian variaciones entre el abandono y la nota, ya que, por ejemplo, puede haber estudiantes con alta probabilidad de suspenso porque tengan problemas con los contenidos (y tengan baja probabilidad de abandono) o porque no estén comprometidos con el curso (y también tengan alta probabilidad de abandono). Estas variaciones pueden servir para identificar perfiles más detallados de estudiantes en función de sus dificultades (p. ej., posible falta de motivación, de comprensión, etc.).

Una limitación de este trabajo es que los resultados obtenidos pueden depender de los seis escenarios considerados, pudiendo variar en otros contextos. Además, dado que los contextos utilizados son diferentes, las variables utilizadas y los posibles filtrados varían, lo que podría afectar a algunos resultados. No obstante, los resultados de este trabajo pueden servir de ayuda para las futuras implementaciones de los modelos predictivos, de modo que los investigadores puedan conocer mejor los factores relacionados con la predicción, y los técnicos que implementen los sistemas puedan conocer mejor qué variables tomar, en qué momento los sistemas pueden dar resultados precisos o cómo generalizar los modelos. De este modo, se podrán diseñar

modelos que puedan ayudar a hacer recomendaciones a alumnos y a mejorar el diseño de los cursos, lo que puede tener un impacto positivo en el aprendizaje.

Palabras clave: predicción, analítica del aprendizaje, minería de datos educativos, cursos online, comportamiento de los estudiantes, indicadores de aprendizaje, aprendizaje máquina.

Abstract

Educational environments allow retrieving a vast amount of information about students' activity in the courses. This information can be used, for example, to predict students' behaviors and performance, and to detect students at risk, so that instructors can identify possible difficulties and carry out interventions. Despite the fact that there have been many previous works related to prediction in education, there are many challenges to be addressed. Among them, it is necessary to explore new predictor variables and their relationship with other well-known variables (e.g., variables related to interactions with videos and exercises). Furthermore, it is important to determine, based on several contexts, what the right moment to start predicting is so that predictions are good enough, and how predictive models can generalize to other contexts. Moreover, it is relevant to understand how dependent variables are related to and whether or not predictions of those variables can be combined to provide further information about students' behaviors.

In order to contribute in part to these issues, this thesis focuses in the analysis of different factors related to prediction in different scenarios. Particularly, this thesis analyzes (1) the effect of some predictor variables, (2) the temporal analysis, (3) the generalizability issue, and (4) the effect of some prediction outcomes. For the analyses, six scenarios are considered, so that they can be used to extract global conclusions from all of them. These scenarios include Massive Open Online Courses (MOOCs) in edX and Coursera, Small Private Online Courses (SPOCs) in Open edX to support university courses and SPOCs in edX Edge to support the preparation of university admission tests, and data from a proprietary platform that collects data from primary and secondary (K-12) students.

The dependent variables used in this thesis include dropout, academic achievement (success), and grades. These variables have been identified as relevant in the literature, and some interviews and focus groups that were carried out involving institutional leaders, experts, professors and students.

Results show that variables related to interactions with exercises can achieve a high predictive power, and this result generalizes to different contexts. Moreover, previous summative grades can improve the predictive power whenever they are available, and variables related to activity in the platform and interactions with videos normally offer a good predictive power and they are useful in the models. In addition, variables about self-regulated learning can contribute to the predictive models and they offer a high predictive power by themselves, although their relevance in the models is lower when other main variables (e.g., variables related to exercises) are included, probably because self-regulated learning variables can be related to other main variables. However, other variables, such as variables related to forum activity and students' behaviors, including efficiency, constancy and persistence, do not achieve a high predictive power and they cannot significantly improve the predictive models.

In relation to the temporal analysis, results show that it can be possible to obtain accurate predictions from 13-40 % of the course duration. This result is consistent in all the scenarios of this thesis, regardless the kind of course and the instruction mode (synchronous or asynchronous). Despite the fact that course context can significantly influence the predictive power, this result implies that it can be possible to achieve early predictions.

With regard to the generalizability issue, three different approaches are considered. In the first approach, predictive models are transferred from one course to another. Using this approach, accurate results can be obtained when the model is transferred to a different course with similar context and the same students (same cohort). Moreover, results can also be acceptable when transferring a model to a different edition of the same course. However, the predictive power is considerably reduced when changing the course and cohort, although there can be some specific cases (they are identified in MOOCs) where this issue does not apply.

The second approach consists on developing a global model trained with data from several courses, which do not necessarily share a similar context. This approach can be useful, for example, when models want to be used in new courses and courses with few students to generate specific models. Using this approach, it is observed that it is possible to achieve a high predictive power at global level, but the model can fail at some specific courses. Particularly, the Area Under the Curve (AUC) decreases more than 0.1 and the Root Mean Square Error (RMSE) increases more than 0.05 in about 15-18 % of the courses. Because of that, it is important to analyze each individual case to identify whether or not the global model fits each specific course. Nevertheless, given that the context can significantly affect the models, one recommendation is reusing and adapting models to each context, which is the third proposed approach. In this case, an existing model is taken, and software is adapted to add/remove variables if needed and generate an specific model for another course.

As for the variables to predict, the analyses in different scenarios show that there can be many aspects to consider when selecting and defining the dependent variable. For example, dropout can be considered as a lack of activity in a certain period in one course, while it can be considered a lack of completion of activities in another. For example, in primary/secondary education, students may access to the platform in class, but they may not be engaged with the exercises and they could be considered as dropout. Furthermore, in relation to the dependent variables, it is observed that better results can be achieved when predicting the final grade than when predicting the final exam grade. In addition, some variations between dropout and grades were identified. For example, there can be some students with high probability of failure because they do not understand the contents (but they do not necessarily have a high dropout probability), and there can be other students at risk of failure because they are not engaged (and they can also be at risk of dropout). These variations can be used to identify more detailed students' profiles depending on their possible difficulties (e.g., lack of motivation, understanding, etc.).

One limitation of this work is that the obtained results can depend on the six analyzed scenarios, and they could vary in other contexts. In addition, given that the contexts are different, the variables used and filtering criteria vary in each context, and that can affect some results. Nevertheless, results can serve to guide future implementations of predictive models, so that researchers can better understand factors related to prediction, and technicians can implement systems using these results to better select the variables, determine when accurate predictions can be provided and how models can be generalized. This way, models can help to provide recommendations to students and

improve to the design of the courses so as to achieve a positive impact on learning.

Keywords: prediction, learning analytics, educational data mining, online courses, students' behaviors, learning indicators, machine learning.

Índice

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	4
1.3. Estructura del documento	5
2. Estado del arte	7
2.1. Analítica del aprendizaje / Learning Analytics	7
2.2. Plataformas educativas y entornos de aprendizaje	11
2.3. Autorregulación del aprendizaje y comportamientos de los estudiantes	15
2.4. Predicción en el ámbito educativo	17
2.4.1. Variables predictoras	17
2.4.2. Variables a predecir	21
2.4.3. Algoritmos y modelos predictivos	23
2.4.4. Métricas de evaluación	26
2.4.5. Generalización	28
2.4.6. Conclusiones	29
3. Metodología	33
3.1. Metodología global de la investigación	33
3.2. Recolección de datos	35
3.3. Medidas	38
3.3.1. Presentación de las variables independientes	38
3.3.2. Presentación de las variables dependientes	48
3.4. Métodos analíticos y herramientas	49
4. Necesidad de la predicción en educación	53
4.1. Estado de adopción de la analítica del aprendizaje y la predicción . .	53
4.2. Análisis de la necesidad del uso de la predicción	56
4.2.1. Opinión de los líderes institucionales y expertos	57
4.2.2. Opinión de los profesores	58
4.2.3. Opinión de los estudiantes	59
4.2.4. Conclusiones globales	60
5. Descripción de los escenarios	63
5.1. Descripción general de los escenarios	63
5.2. Escenario 1. MOOC de programación en Java (UC3M)	64
5.3. Escenario 2. MOOC de programación en Java (HKUST)	66
5.4. Escenario 3. MOOCs en Coursera (PUC)	68
5.5. Escenario 4. SPOCs de apoyo en pruebas de acceso (KU Leuven) . .	71
5.6. Escenario 5. SPOCs de apoyo en cursos universitarios (UC3M)	74
5.7. Escenario 6. Plataforma para alumnos de primaria y secundaria . . .	76
5.8. Similitudes y diferencias entre los diferentes escenarios	79

6. Efecto de las variables predictoras	81
6.1. Variables sobre interacciones con vídeos, ejercicios y actividad en la plataforma	82
6.1.1. Análisis de los escenarios 1 y 2	83
6.1.2. Análisis del escenario 3	85
6.1.3. Análisis del escenario 4	86
6.1.4. Análisis del escenario 5	87
6.1.5. Análisis del escenario 6	89
6.2. Notas previas de la evaluación sumativa del curso	90
6.3. Variables de la actividad del alumno en el foro	92
6.4. Variables sobre la autorregulación del aprendizaje	94
6.5. Variables sobre comportamientos del alumno	97
6.5.1. Efecto de la persistencia en la predicción	98
6.5.2. Efecto de la eficiencia y constancia en la predicción	100
6.6. Tipo de metodología de recogida de datos	102
6.7. Lecciones aprendidas	105
7. Análisis temporal	109
7.1. Evolución temporal del poder predictivo en cursos síncronos	110
7.2. Evolución temporal del poder predictivo en cursos asíncronos	122
7.3. Lecciones aprendidas	124
8. Generalización de los modelos predictivos	129
8.1. Transferencia de modelos	130
8.2. Agregado de cursos	133
8.3. Reutilización y adaptación de los modelos	140
8.4. Lecciones aprendidas	144
9. Efecto de las variables a predecir	149
9.1. Relación y diferencias entre predicción de la nota final y la del examen final	150
9.2. Relación y diferencias entre la predicción del abandono y de la nota final	152
9.3. Selección de las variables dependientes	154
9.4. Lecciones aprendidas	157
10. Conclusiones / Conclusions	159
10.1. Versión en español	160
10.1.1. Síntesis de resultados	160
10.1.2. Aplicabilidad de los resultados	163
10.1.3. Limitaciones y trabajo futuro	165
10.2. English version	167
10.2.1. Summary of results	167
10.2.2. Applicability of the results	170
10.2.3. Limitations and future work	172
Referencias	175

Índice de figuras

2.1. Modelo del proceso de la analítica del aprendizaje, tomado y adaptado de [35]	9
3.1. Esquema de la metodología en la investigación	34
4.1. Nivel de adopción de analítica del aprendizaje en Europa, tomado de [248]	55
5.1. Diagrama de flujo con las reglas para determinar el abandono	70
5.2. Ejemplo de las cuatro primeras semanas (bloques) del curso para tres alumnos que comienzan en diferentes periodos	71
6.1. Importancia relativa de las variables en la predicción, utilizando persistencia en escenario 5	100
6.2. Relación entre la eficiencia, el inverso de la constancia y la nota final	101
7.1. Evaluación de los resultados de la predicción del éxito a lo largo del tiempo en el escenario 1	112
7.2. Evolución temporal de la predicción de notas usando datos de módulos previos en el escenario 1	113
7.3. Evolución temporal de la predicción de notas usando datos de módulos previos en el escenario 2	114
7.4. Evolución temporal de la predicción de la nota final en el escenario 2	115
7.5. Evolución temporal de la predicción en el escenario 5	117
7.6. Evolución temporal de la predicción del abandono en el escenario 6 .	118
7.7. Evolución temporal de la predicción de la nota en el escenario 6 . . .	119
7.8. Evolución temporal del poder predictivo de las variables al predecir el abandono en el escenario 6	120
7.9. Evolución temporal del poder predictivo de las variables al predecir la nota en el escenario 6	121
7.10. Evolución temporal de la predicción del abandono en el MOOC “Electrones en Acción”	123
7.11. Evolución temporal de la predicción del abandono en los tres MOOCs del escenario 3	124
8.1. Predicción utilizando variables de curso en el escenario 5	137
8.2. Diagramas de caja con la relación entre las variables de curso y las variables dependientes	138
9.1. Relación entre la nota del examen final y la nota final	150
9.2. Comparación entre los modelos predictivos para la nota final y el examen final en el escenario 2.	151
9.3. Relación entre la predicción del abandono y de la nota final	153

Índice de tablas

2.1.	Criterios de evaluación para los resultados del AUC, tomado de [192]	26
3.1.	Ejemplo de cálculo de la persistencia	47
3.2.	Resumen de las métricas utilizadas en los análisis predictivos	50
5.1.	Resumen y comparativas de los cursos en los escenarios 1 y 2	67
5.2.	Percentiles del periodo inactivo de los alumnos en el MOOC	69
5.3.	Resumen de los participantes en los SPOCs	72
5.4.	Resumen de los escenarios analizados en la tesis	80
6.1.	Predicción de notas con variables de ejercicios y vídeos en escenario 1	84
6.2.	Predicción de notas con variables de ejercicios y vídeos en escenario 2	85
6.3.	Predicción de abandono en el escenario 3	86
6.4.	Poder predictivo en los cursos del escenario 4	87
6.5.	Resultados de la predicción en el escenario 5	88
6.6.	Predicción de abandono y notas en el escenario 6	90
6.7.	Predicción de notas con notas previas en escenario 1	91
6.8.	Predicción de notas con notas previas en escenario 2	92
6.9.	Predicción de notas con variables de foro en escenario 1	93
6.10.	Predicción de notas con variables de foro en escenario 2	94
6.11.	Evaluación de predicción de notas y éxito en escenario 3	96
6.12.	Predicción con variables de SRL en escenario 3	97
6.13.	Predicción utilizando persistencia en escenario 5	98
6.14.	Modelos utilizados en la predicción en el escenario 5	99
6.15.	Predicción utilizando eficiencia y constancia en escenario 6	102
6.16.	Predicción según el modo de recogida de datos en escenario 1	104
6.17.	Predicción según el modo de recogida de datos en escenario 2	105
7.1.	Evolución temporal de la predicción en el escenario 4	116
7.2.	Resultados agregados del análisis temporal	125
8.1.	Transferencia de modelos en cursos del escenario 3	130
8.2.	Resultados de la transferibilidad de los modelos en el escenario 4 . . .	132
8.3.	Resultados de la predicción en cursos concretos en el escenario 5 . . .	135
8.4.	Resultados de la predicción en cursos concretos en el escenario 6 . . .	136

Índice de siglas

AUC Area Under the Curve.

BKT Bayesian Knowledge Tracing.

CFA Correct at First Attempt.

cMOOC Connectivist Massive Open Online Course.

CSV Comma-Separated Values.

CV Cross-Validation.

DSRM Design Science Research Methodology.

DT Decision Tree.

EDM Educational Data Mining.

GBM Gradient Boosting Machine.

GLM Generalized Linear Model.

HKUST Hong Kong University of Science and Technology.

IoT Internet of Things.

ITS Intelligent Tutoring System.

JSON JavaScript Object Notation.

KU Leuven Katholieke Universiteit te Leuven.

LA Learning Analytics.

LMS Learning Management System.

MAE Mean Absolute Error.

ML Machine Learning.

MOOC Massive Open Online Course.

NLP Natural Language Processing.

OrLA Orchestrating Learning Analytics.

PUC Pontificia Universidad Católica de Chile.

RF Random Forest.

RG Regresión.

RMSE Root Mean Square Error.

ROC Receiver Operating Characteristic.

SHEILA Supporting Higher Education to Integrate Learning Analytics.

SIS Student Information System.

SLE Smart Learning Environment.

SPOC Small Private Online Course.

SRL Self-Regulated Learning.

STEM Science, Technology, Engineering and Mathematics.

SVM Support Vector Machine.

TFM Trabajo Fin de Máster.

UACH Universidad Austral de Chile.

UC3M Universidad Carlos III de Madrid.

xMOOC eXtended Massive Open Online Course.

1

Introducción

Contenido

1.1. Motivación	1
1.2. Objetivos	4
1.3. Estructura del documento	5

El primer capítulo tiene como objetivo introducir cuáles son las motivaciones que llevan a desarrollar esta tesis, presentar cuáles son los objetivos del trabajo y mostrar la estructura que se utiliza a lo largo del documento. El capítulo se divide en tres secciones que tratan la motivación, los objetivos y la estructura del documento, respectivamente.

1.1. Motivación

Durante los últimos años, las tecnologías educativas, y en particular, las plataformas educativas han ido evolucionando. Entre estas evoluciones, se encuentra la mayor recogida de datos sobre las interacciones de los estudiantes [1]. Esta recogida de datos se puede llevar a cabo tanto en los cursos presenciales como a distancia, aunque tiene una especial relevancia en aquellos cursos a distancia o semipresenciales, en los que los alumnos trabajan en mayor medida con entornos educativos digitales en línea. Con estos datos, es posible mejorar el proceso educativo mediante la llamada analítica del aprendizaje, conocida por su término en inglés **Learning Analytics (LA)**. Este término apareció por primera vez en 2011 y fue definido como *“la medida, colección, análisis e informe de datos sobre los estudiantes y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que ocurre”* [2].

Entre los contextos donde existe una amplia recogida de datos y tiene sentido el uso de **LA** se encuentran los **Massive Open Online Courses (MOOCs)**. Los **MOOCs** son cursos abiertos en línea cuyo objetivo es proporcionar conocimiento a un amplio número de estudiantes [3]. Dado su carácter abierto, el número de estudiantes puede ser muy elevado, lo que conlleva a que la cantidad de datos puede ser muy alta. También destacan los **Small Private Online Courses (SPOCs)** [4], que son cursos,

que utilizan tecnologías similares a los MOOCs, pero que se imparten en un contexto cerrado, como el de un curso universitario. Estos cursos típicamente se pueden combinar con aprendizaje semipresencial y/o clase invertida [5], y permiten obtener una gran cantidad de datos sobre las interacciones de los estudiantes. Además, los sistemas de tutoría inteligente (del inglés, *Intelligent Tutoring Systems (ITSs)*) [6] y los sistemas de gestión del aprendizaje [7] (del inglés *Learning Management Systems (LMSs)*), que han existido desde hace bastante tiempo, permiten una recogida de datos detallada sobre los diferentes alumnos, si bien no suelen tener una cantidad de alumnos tan elevada como los MOOCs.

En todos estos entornos educativos, existe un creciente interés en aplicar LA, dado que ésta puede ayudar a los profesores, por ejemplo, a realizar un seguimiento personalizado de sus estudiantes y adaptar el curso para la mejora del aprendizaje. Este interés puede estar acrecentado por el avance de la Industria 4.0 [8] y la introducción del uso de inteligencia artificial, Big Data, Internet de las Cosas (conocido por su término en inglés *Internet of Things (IoT)*), y otros servicios de Internet, como la computación en la nube (*cloud computing*) [9]. Estas técnicas se están aplicando ya a otros campos, como la medicina [10] o la banca [11], lo que hace pensar que la educación también tiene un potencial para su mejora mediante la introducción de la tecnología [9] y en particular, de LA.

Como ejemplo de entorno educativo donde se aplica LA están los entornos de aprendizaje inteligente, comúnmente conocidos por su término en inglés *Smart Learning Environments (SLEs)*, que combinan diversas tecnologías educativas [12]. Uno de los objetivos de los SLEs es proporcionar información (p. ej., mediante indicadores y paneles de visualización [13]) sobre el comportamiento y rendimiento de los estudiantes, de modo que se puedan proporcionar experiencias de aprendizaje que sean personalizadas al alumno. Además, esta información puede servir para poder adaptar y coordinar el aprendizaje durante la ejecución de los cursos (mejorar la orquestación [14]), para mejorar el diseño de los cursos, y para detectar posibles dificultades de los estudiantes, como riesgo de abandono, falta de compromiso, motivación, etc., de modo que se puedan realizar intervenciones que puedan resolver o mitigar estas dificultades.

Para detectar dichas dificultades, dentro del área de investigación de LA, es posible desarrollar modelos predictivos que tomen los datos de las interacciones de los alumnos con las plataformas y escenarios educativos para predecir variables (p. ej., comportamientos o rendimiento del alumno) a futuro del estudiante. Un ejemplo bastante común es el de la predicción del abandono. Dado que las tasas de abandono pueden llegar a ser bastante altas, tanto en cursos como a nivel académico (programa) [15], existe un alto interés en predecir tanto el abandono como qué alumnos tendrán un bajo rendimiento (o suspenderán) el curso [16]. Sin embargo, puede haber muchas otras variables a predecir, tales como el compromiso [17], la personalidad de los estudiantes [18], etc.

El desarrollo de estos modelos predictivos puede ser muy beneficioso para los diferentes actores del aprendizaje [19]. A los profesores, les puede ser útil la información para poder anticipar los posibles problemas del curso, y adaptar la metodología docente, o los contenidos para mejorar la experiencia educativa [20]. Además, la detección temprana de los estudiantes con dificultades puede servir a los profesores para poder ejecutar intervenciones (de muy diferentes formas, según el contexto educativo, como proporcionar una ayuda personalizada) que ayuden a

solventar los problemas [21]. También, proporcionar información a los estudiantes directamente sobre cómo lo están haciendo les puede ayudar a reflexionar sobre su proceso de aprendizaje para mejorar [22] (aunque deba tenerse cuidado en cómo mostrar la información para no desmotivar). Asimismo, proporcionar información a líderes institucionales o gestores puede contribuir a una mejora de la gestión de recursos institucionales o de la planificación institucional y de los programas, para adaptarse a las necesidades de los estudiantes [19].

A pesar de los posibles beneficios de las técnicas de predicción en educación, existen diversos retos que abordar para el desarrollo de los mismos. Un aspecto fundamental es el aspecto temporal. Los modelos predictivos, al desarrollarse con vistas a predecir el futuro, deben proporcionar un poder predictivo adecuado con suficiente antelación para poder detectar dificultades o alumnos en riesgo, con la idea de poder tomar acciones lo antes posible. Si estos modelos se desarrollan con muy poco margen de antelación, la información que aportan puede perder valor. Por ejemplo, si se predice que un alumno está en riesgo de abandonar al principio de curso, se pueden tomar medidas para intentar mantenerle motivado y que no abandone, pero si esta predicción se realiza 15 días antes del final de curso, puede ser que no haya tiempo suficiente para reaccionar. En la literatura, existen varios trabajos que han abordado el análisis temporal (p. ej., [23], [24]), aunque estos trabajos generalmente tratan un conjunto limitado de cursos y es necesario analizar varios contextos para obtener conclusiones globales.

Además, hay varias variables dependientes que se pueden considerar en los modelos predictivos. Por ejemplo, puede haber varias definiciones de qué es abandono dependiendo del contexto [25]: puede ser el alumno que no entre en el curso durante cierto tiempo, el que no haga las actividades, etc. En ese caso, es importante ver qué definición se adapta mejor al contexto. También, es interesante conocer si distintas variables dependientes tienen cierta relación. Por ejemplo, una hipótesis es que el abandono y la nota en el curso están muy relacionadas. En relación con la predicción, otra hipótesis es que es más fácil/difícil predecir la nota final del curso que la del examen final. En la literatura, hay trabajos en los que las variables dependientes se han analizado de forma separada (p. ej., [26], [27]), pero el análisis de estas posibles hipótesis y de la relación entre las variables dependientes es relevante para conocer mejor estas variables, lo que puede servir para caracterizar mejor al estudiante y sacar más provecho de las predicciones.

Otro de los aspectos relevantes es las variables que se utilizan para predecir (variables predictoras). Dependiendo del contexto, es posible que unas variables estén disponibles y otras no, pero en general puede haber ciertas variables que tengan un alto poder predictivo y otras que aporten poco a los modelos. Es interesante conocer qué variables pueden aportar más a la predicción, cuáles aportan en ausencia de otras pero no aportan cuando otras están presentes, etc. En la literatura, ya se ha explorado el poder predictivo de algunas variables, como variables de ejercicios, vídeos o actividad en la plataforma (en trabajos como [23], [28]), pero faltan por incluir nuevas variables, como por ejemplo, variables sobre comportamientos obtenidas a partir de la transformación de otras de bajo nivel.

Por último, otro aspecto importante en los modelos predictivos es la generalización de los mismos [29]. Si los modelos predictivos funcionan bien para un solo curso y no pueden extrapolarse a otros contextos, o al menos a otras cohortes del mismo curso, la aplicabilidad de los mismos y su sostenibilidad se verán reducidos, ya que

se generarán con datos pasados y no podrán utilizarse posteriormente. Por ello, es muy importante analizar la generalización de los modelos predictivos y cómo pueden aplicarse en otro tipo de cursos. En la literatura, algunos trabajos como [30] han empezado a abordar el tema de la generalización y han presentado algunos análisis concretos (p. ej., la transferencia de modelos entre ediciones de un curso), pero faltan trabajos que analicen este problema con más detalle y mediante diferentes aproximaciones y contextos.

1.2. Objetivos

En relación con la motivación presentada anteriormente, el objetivo global de la tesis es obtener conclusiones sobre un conjunto de aspectos involucrados en la predicción en educación, mediante el análisis de diferentes escenarios de aprendizaje. De esta manera, se contribuye analizando cuál es la relación de las diferentes variables a predecir en diferentes contextos y cuál es la diferencia entre el poder predictivo de diversas variables dependientes, como la nota final y la nota del examen final. También, se contribuye mediante el análisis del efecto de diferentes variables predictoras, incluyendo nuevas variables en la predicción, como variables sobre la autorregulación del aprendizaje (del inglés *Self-Regulated Learning (SRL)*) o comportamientos de alto nivel obtenidos a partir de la transformación de variables de bajo nivel (como la persistencia del alumno). Asimismo, otra aportación está en el análisis de ver a partir de qué momento el poder predictivo es suficientemente bueno para predecir en diferentes contextos. Además, el análisis de diferentes aproximaciones y diferentes escenarios para la generalización de los modelos predictivos también aporta y es relevante para garantizar el funcionamiento de los modelos. Por último, es de destacar que en esta tesis, una contribución importante es que el análisis de los aspectos mencionados se realiza en diversos contextos. Esto permite extraer conclusiones globales a partir de los análisis de varios escenarios, lo cual es un aspecto que falta en trabajos previos de la literatura.

De cara a poder cumplir el objetivo global de la tesis, se han definido varios objetivos específicos, que se abordan en los distintos capítulos. Estos objetivos específicos son los siguientes:

1. Analizar la necesidad del uso de los sistemas predictivos en el ámbito educativo, para demostrar si los diferentes actores realmente aprecian el valor de estos modelos y valoran positivamente su uso en el contexto educativo.
2. Analizar variables predictoras ya utilizadas en trabajos previos en contextos nuevos y añadir nuevas variables predictoras no utilizadas, para saber cuáles son las variables predictoras que pueden aportar valor a los modelos predictivos. En particular, se proponen los siguientes análisis:
 - Corroborar el alto poder predictivo identificado en la literatura sobre variables de interacción con vídeos, de interacción con ejercicios y de actividad en la plataforma.
 - Analizar el efecto de las notas previas de la evaluación sumativa del curso.
 - Analizar el efecto de variables de la actividad de la alumno en el foro del curso.

- Analizar el efecto de variables sobre SRL.
 - Analizar el efecto de variables de comportamientos como la eficiencia, constancia y persistencia del alumno.
3. Analizar cuál es el mejor momento para empezar a predecir, teniendo en cuenta la evolución temporal del poder predictivo. En este punto se analizan los siguientes aspectos:
- Evolución temporal del poder predictivo en cursos síncronos.
 - Evolución temporal del poder predictivo en cursos asíncronos.
4. Analizar el grado de generalización de los modelos predictivos a otros contextos. En particular, se analizan los siguientes casos:
- Generalización mediante la transferencia de modelos predictivos a otros cursos o ediciones del mismo curso.
 - Generalización mediante el agregado de cursos.
 - Generalización mediante la reutilización y adaptación de los modelos.
5. Analizar el efecto de las variables a predecir en el desarrollo de los modelos predictivos. En particular, se analizan los siguientes aspectos:
- Relación y diferencias en el poder predictivo entre predecir la nota del examen final y el examen final del curso.
 - Relación y diferencias entre la predicción del abandono y de la nota final.
 - Selección de la variable dependiente en diferentes escenarios educativos.

1.3. Estructura del documento

Con el objeto de facilitar la lectura del documento e identificar las diferentes partes donde se abordan los objetivos mencionados con anterioridad, la tesis se ha dividido en 10 capítulos. A continuación, se enumeran y describen brevemente los diferentes capítulos.

- Capítulo 1 - Introducción. Este capítulo, de la que esta sección es parte, presenta la motivación que lleva al desarrollo de la tesis doctoral, la enumeración de sus objetivos y la estructura del resto de la memoria.
- Capítulo 2 - Estado del arte. Este capítulo presenta una visión sobre la analítica del aprendizaje, las plataformas y entornos educativos, el aprendizaje autorregulado, algunos comportamientos de los estudiantes que se abordan en la tesis (persistencia, eficiencia y constancia), y los trabajos más relevantes que se han realizado sobre la predicción en diferentes contextos de aprendizaje.
- Capítulo 3 - Metodología. Este capítulo describe la metodología global que se ha llevado a cabo para la elaboración de la investigación y para la recogida de datos. Asimismo, se detalla la forma de medir las variables independientes y dependientes, y los métodos analíticos y herramientas utilizadas para los análisis.

- Capítulo 4 - Necesidad de la predicción en educación. Este capítulo presenta un análisis para justificar por qué es importante el desarrollo de los modelos predictivos en el contexto de la educación superior. Para ello, se presenta el estado de adopción de la analítica del aprendizaje y opiniones de líderes institucionales, expertos, profesores y alumnos sobre la necesidad de incorporar modelos predictivos en el ámbito universitario.
- Capítulo 5 - Descripción de los escenarios. Este capítulo tiene como objetivo describir los contextos educativos en los que se realizan los análisis predictivos, detallando la información conocida sobre los cursos y/o metodología docente, y las variables dependientes e independientes que se consideran en cada uno de los contextos.
- Capítulo 6 - Efecto de las variables predictoras. Este capítulo analiza cómo afecta el uso de algunas variables predictoras en los resultados de la predicción, en ausencia y presencia de otras, y explica si ciertas variables pueden ser utilizadas con un buen poder predictivo.
- Capítulo 7 - Análisis temporal. Este capítulo presenta los resultados de los análisis de cómo varía y como evoluciona la predicción a lo largo del tiempo y debate aspectos como cuál es el mejor momento para empezar a predecir y cómo realizar la predicción temporal en función de la metodología del curso.
- Capítulo 8 - Generalización de los modelos predictivos. Este capítulo aborda el problema de la generalización de los modelos predictivos, incluyendo varias aproximaciones para ver cómo poder utilizar un modelo entrenado con uno o varios cursos para realizar predicciones en otros cursos u otras ediciones del mismo curso.
- Capítulo 9 - Efecto de las variables a predecir. Este capítulo tiene como objetivo dar indicaciones sobre cómo seleccionar las variables dependientes en diferentes escenarios, y analizar cómo puede variar el poder predictivo en función de la variable a predecir y cuál es la relación entre dichas variables.
- Capítulo 10 - Conclusiones. Este capítulo final detalla cuáles son las conclusiones obtenidas de esta tesis. Para ello, se presenta una síntesis de los resultados y una discusión sobre la aplicabilidad práctica de los mismos. Además, se detallan las limitaciones de la investigación llevada a cabo, posibles metodologías alternativas y posibles líneas de trabajo futuro. Este capítulo se presenta tanto en español como en inglés para cumplir con los requisitos establecidos por la Escuela de Doctorado para la obtención de la Mención Doctorado Internacional.

2

Estado del arte

Contenido

2.1. Analítica del aprendizaje / Learning Analytics	7
2.2. Plataformas educativas y entornos de aprendizaje	11
2.3. Autorregulación del aprendizaje y comportamientos de los estudiantes	15
2.4. Predicción en el ámbito educativo	17
2.4.1. Variables predictoras	17
2.4.2. Variables a predecir	21
2.4.3. Algoritmos y modelos predictivos	23
2.4.4. Métricas de evaluación	26
2.4.5. Generalización	28
2.4.6. Conclusiones	29

Este capítulo tiene como objetivo poner en contexto el área de investigación donde se ubica la tesis y revisar cuáles son los trabajos previos más relevantes sobre los aspectos que se abordan en esta tesis doctoral. El capítulo se divide en cuatro partes. En primer lugar, se presenta una introducción sobre **LA**, que es la disciplina de estudio global en la que se enmarca esta tesis. A continuación, se hace una revisión sobre las principales plataformas y entornos educativos. Después, se realiza una introducción sobre el aprendizaje autorregulado y algunos comportamientos de estudiantes que se abordan en la tesis, tales como la persistencia, eficiencia y constancia. Finalmente, se hace una revisión sobre los trabajos en el área de predicción.

2.1. Analítica del aprendizaje / Learning Analytics

El trabajo de esta tesis doctoral está enmarcado dentro de lo que se conoce como la disciplina de la analítica del aprendizaje, comúnmente conocida por su término en inglés, **Learning Analytics (LA)**. Este término apareció por primera vez en 2011

y fue definido como “*la medida, colección, análisis e informe de datos sobre los estudiantes y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que ocurre*” [2]. Esta disciplina tiene una gran relación con la minería de datos educativos, conocida por su término en inglés **Educational Data Mining (EDM)**. Este término fue definido por la sociedad de minería de datos EDM como “*una disciplina emergente, que se encarga de desarrollar métodos para explorar tipos de datos únicos que proceden de entornos educativos, y utilizar dichos métodos para entender mejor a sus estudiantes y los entornos en los que aprenden*” [31]. A la vista de ambas definiciones, es fácilmente apreciable que **LA** y **EDM** están muy estrechamente relacionados y que ambos tienen objetivos comunes de mejorar el aprendizaje de los estudiantes. Sin embargo, según se haga referencia a un término o a otro, existen algunas diferencias en las técnicas utilizadas y en el énfasis que pone cada una de ellas [32], [33]. **LA** busca un marco integral para entender los procesos de aprendizaje desde una perspectiva más global, mientras que **EDM** a menudo tiene una visión más reducida y se centra más en el propio análisis de los datos, incluyendo la búsqueda de patrones y el desarrollo de algoritmos de inteligencia artificial para el análisis. No obstante, la relación entre ambas disciplinas es muy estrecha y en muchas ocasiones un trabajo ser enmarcado tanto en una como en otra área temática.

Respecto a **LA**, algunos estudios han tratado de establecer marcos que caractericen sus procesos y elementos utilizados. Entre ellos, Elias [34] presenta un modelo del proceso de **LA**. Este proceso es cíclico y consta de varias fases. En primer lugar, se encuentra la fase de recogida de datos, que involucra la propia recogida y la selección de los datos. Posteriormente, se encuentra el procesamiento de la información, que incluye métodos como la agregación o predicción. Finalmente, se encuentra la aplicación del conocimiento del análisis de los datos, que incluye el uso de los resultados y refinamiento de los análisis. Tras esto el proceso se va repitiendo cíclicamente. Este proceso puede tener ciertas variaciones, pero es ampliamente utilizado. Por ejemplo, Steiner et al. [35] extendieron el modelo con etapas más detalladas. En su trabajo, el proceso pasa por (1) definir objetivos, (2) capturar los datos, (3) preprocesar y agregar los datos, (4) analizar los datos y proporcionar visualizaciones, (5) predecir eventos futuros, (6) tomar decisiones basados en las etapas previas y (7) refinar las soluciones planteadas. La Fig. 2.1 presenta un esquema del proceso propuesto en este modelo.

Por otro lado, Chatti et al. [36], aparte de presentar un ciclo similar del proceso de **LA**, describieron un modelo para **LA** basado en cuatro dimensiones, sobre las cuales identificaron algunos retos y oportunidades. Estas dimensiones se basaban en las preguntas ¿Qué? (¿Qué tipos de datos recoger y usar en el análisis?), ¿Quién? (¿Para quién se realiza el análisis?), ¿Por qué? (¿Por qué el sistema analiza los datos) y ¿Cómo? (¿Cómo se realiza el análisis?). De forma similar, un marco bastante conocido sobre **LA** es el de Greller y Drachsler [37]. Su marco contempla las siguientes dimensiones:

- **Interesados:** son aquellos beneficiarios del proceso de **LA**. Puede haber muchos tipos de actores en este grupo, pero los principales suelen ser el alumno (para la mejora de su aprendizaje), el profesor (para la mejora de las metodologías docentes) o la propia institución (para mejorar las tasas de abandono o de graduación o para mejorar las titulaciones) [37].



Fig. 2.1. Modelo del proceso de la analítica del aprendizaje, tomado y adaptado de [35]

- **Objetivos:** pueden ser muy variados, aunque en líneas generales, se pueden subdividir en los siguientes [36]: (1) monitorización y análisis de las actividades de los estudiantes, (2) predicción de comportamientos e intervención, (3) tutoría inteligente para ayudar en el desarrollo de las tareas, (4) evolución de resultados y retroalimentación para mejorar el proceso educativo, (5) personalización y recomendación para ayudar a cada alumno en su propio camino de aprendizaje y (6) reflexión, para poder comparar resultados y mejorar.
- **Datos:** son la fuente sobre la que realizar los análisis [37]. Estos datos, normalmente proceden de plataformas educativas (ver Sección 2.2) o entornos de aprendizaje. Pueden clasificarse en centralizados o distribuidos, dependiendo de si proceden de una fuente o si combinan datos de varias fuentes y/o plataformas, que pueden tener distintos formatos [36]. En la mayor parte de los casos, estos datos son privados, aunque existen movimientos que promueven la compartición de datos educativos [38].
- **Instrumentos:** incluye las tecnologías utilizadas para el análisis, los algoritmos, y las teorías educativas, entre otros [37]. Entre estos, algunas tecnologías destacadas son [39]: minería de datos, mediante algoritmos como máquinas de vectores soporte, **Support Vector Machine (SVM)**, árboles de decisión, regresión, etc.; el uso de estadística descriptiva; minería de texto y procesamiento del lenguaje natural, **Natural Language Processing (NLP)**, visualizaciones, análisis social (análisis de cómo se comunican diferentes actores del proceso educativo, p. ej., los estudiantes en el foro de un curso), análisis cualitativo (p. ej., análisis de entrevistas) y gamificación.

- Limitaciones internas: están relacionadas con las competencias para poder interpretar y comprender los resultados, y con la aceptación de los resultados obtenidos mediante el uso de LA [37].
- Limitaciones externas: se refiere a convenciones y normas que frenan el uso de LA [37]. En esta categoría se encuentran los problemas éticos y de privacidad y las normativas sobre la protección de datos de los estudiantes. También, en esta categoría se incluyen los problemas sobre la escala temporal, es decir, sobre si los análisis se realizan a posteriori o en tiempo real y cómo eso puede afectar en el beneficio que se espera de los resultados.

En línea con los objetivos y métodos mencionados en el esquema anterior, han surgido diversas líneas de investigación. Entre estas líneas de investigación, se encuentra la predicción, que consiste en el desarrollo de modelos que permitan determinar variables sobre el comportamiento de los alumnos o de sus resultados en el futuro. En esta línea se han desarrollado diversos trabajos con diversos fines, como predecir el abandono [24], las notas [27], etc.

Otra de las líneas está en torno a la tutoría. En esta línea el objetivo es el diseño de sistemas (p. ej., [40]) que puedan ayudar y guiar a los estudiantes a aprender los contenidos de una materia. Muy relacionado con esta línea está la de los sistemas de recomendación, cuyo objetivo es proporcionar recomendaciones individuales a cada estudiante que le sirvan como guía a la hora de tomar decisiones de elegir un item frente a otros [41]. Por ejemplo, un tipo de sistema de recomendación es el que se encarga de recomendar recursos de un curso a un estudiante en base a lo que hicieron otros estudiantes [42], de modo que estos sistemas puedan también generar un camino de aprendizaje [43] de forma adaptativa.

Otra línea relacionada está en torno a los sistemas de visualización. Esta línea consiste en ver analizar cómo proporcionar la información sobre las analíticas a los interesados (estudiantes, profesores, gestores), de manera que esta información sea comprensible para ellos y les aporte información. Esta información, generalmente se muestra en paneles de visualización (p. ej., VisMOOC [44], ANALYSE [45]), cuyas características fundamentales son que deben conseguir dar una visión general del curso que permita concienciarse, reflexionar, obtener respuestas sobre lo que ocurre y cambiar comportamientos [13]. Dentro de esta línea de investigación, algunos aspectos importantes son identificar qué datos son útiles, cómo presentar la información y evaluar el impacto de los paneles, aunque sobre esto último hace falta más trabajo [46].

Además, dentro de LA, hay trabajos en la línea de obtener datos y variables. Por un lado, existen trabajos cuyo propósito es definir nuevos indicadores de aprendizaje a partir de datos de bajo nivel (p. ej., [47]), pero también hay trabajos centrados en obtener nuevos tipos de datos y de diversas fuentes, incluyendo el uso de sensores y cámaras (lo que se conoce como analítica del aprendizaje multimodal [48]). Por ejemplo, es posible usar sensores para recoger información de simulaciones médicas que proporcionen retroalimentación a los estudiantes de enfermería sobre la realización de las tareas [49]. En torno al tipo de información analizada, también existen otros trabajos que se encargan de realizar minería de texto usando técnicas de NLP para obtener conclusiones sobre el comportamiento de los estudiantes sobre lo que escriben (por ejemplo, en los foros [50]). Además, entre la comunicación de los estudiantes, también existen trabajos que aplican análisis social para determinar

las relaciones entre los estudiantes y promover el aprendizaje colaborativo [51]. Asimismo, otros trabajos realizan análisis cualitativos a partir de los mensajes de los estudiantes, como por ejemplo, Ezen-Can et al. [52], que utilizaron técnicas de agrupamiento (clustering) para agrupar mensajes del foro de un MOOC para analizar los datos cualitativamente.

Por último, otra de las líneas relevantes de LA es la gamificación. La gamificación consiste en el uso de herramientas y mecánicas del juego para hacer el aprendizaje más atractivo y divertido [53]. De acuerdo con Khalil y Ebner [39], algunos ejemplos de gamificación incluyen el uso de medallas, puntos de recompensa o sistemas de puntuación gamificados [54]. También, en el ámbito de los juegos educativos, es posible recoger trazas sobre la actividad en el juego (p. ej., trazas cuando comienza/termina el juego, cuando hay algún cambio de fase, sobre el rendimiento en el juego, o sobre las acciones realizadas con los dispositivos de juego) para entender cómo interactúan los estudiantes con los juegos y mejorar la usabilidad de los mismos [55]. También, recientemente se ha combinado el uso de juegos educativos con la predicción y Alonso-Fernández et al. [56] desarrollaron modelos predictivos para determinar el conocimiento de los estudiantes tras jugar al juego, lo que muestra que los juegos educativos pueden utilizarse tanto para enseñar como para medir el conocimiento de los estudiantes.

De entre todas las líneas mencionadas, esta tesis se centra en una de ellas: la predicción, que es además uno de los principales objetivos de investigación en LA, de acuerdo con una revisión del estado del arte actual [57]. En esta línea de investigación, se están realizando bastantes trabajos actualmente (como se demuestra en secciones especiales recientes sobre predicción, como [58]), pero hay varios retos que requieren mayor investigación. En la sección 2.4.6 se tratan dichos retos y cómo esta tesis contribuye en ellos.

2.2. Plataformas educativas y entornos de aprendizaje

Tras presentar una breve introducción a LA, se ha visto que uno de los aspectos fundamentales son los datos. Estos datos, aunque pueden estar tomados directamente en la clase presencial (p. ej., asistencia o participación del alumno en clase [59]), se toman muy frecuentemente de las interacciones de los estudiantes con plataformas educativas o entornos de aprendizaje distribuidos. Esta sección trata de ver cuáles son los tipos de plataformas y de entornos de aprendizaje más relevantes.

En primer lugar, uno de los sistemas que llevan utilizándose desde finales del siglo XX son los sistemas de tutoría inteligente (*Intelligent Tutoring Systems* (ITSs)). Los ITSs son aplicaciones software diseñadas utilizando inteligencia artificial para proporcionar tutores que tengan en muchas ocasiones un comportamiento similar al humano, y puedan enseñar de la mejor manera posible, proporcionando retroalimentación inmediata y personalizada a cada estudiante [60]. Estos tutores se basan en la intersección entre las ciencias de la computación, la psicología cognitiva y las tecnologías educativas. Murray [61] definió las partes de un ITS y diferenció entre cuatro: la interfaz del estudiante, el modelo del dominio (que contiene representaciones del conocimiento del temario y de la experiencia para resolver los problemas [61]), el modelo de enseñanza (que toma las decisiones pedagógicas en función del conocimiento de los estudiantes [62]) y el modelo del estudiante (que representa el conocimiento de los estudiantes [62]). Un ejemplo de

ITS es el tutor Andes [40]. Este tutor inteligente tiene modelos para determinar el conocimiento de los estudiantes, de modo que el tutor es capaz de predecir las acciones de los estudiantes durante la resolución de los problemas y evaluar su conocimiento. En este tutor, cada vez que se resuelve alguna parte de un problema, se recibe retroalimentación sobre lo que está haciendo y, en caso de error, se reciben dos tipos de ayuda: una ayuda sobre cómo solucionar el error y otra ayuda que explique cuál es el siguiente paso a realizar. También, en muchos otros casos, los ITSs incorporan preguntas de andamiaje (p. ej., [63]) para ir guiando la resolución de los ejercicios a través de preguntas más sencillas.

En los tutores inteligentes, tradicionalmente se han realizado modelos para estimar el conocimiento de los estudiantes (p. ej., [64], [65]) y para ello se han utilizado técnicas, por ejemplo, basadas en espacios de conocimiento, redes bayesianas o teoría de respuesta al ítem (p. ej., [64], [66], [67], [68]). Estos modelos permiten una adaptación del aprendizaje a cada alumno, principalmente a partir de las respuestas de alumnos con ejercicios [64], y que suelen estar focalizadas en áreas muy concretas para las que hay un modelado de dominio (p. ej., geometría [69]). Sin embargo, más allá del modelado del conocimiento, las interacciones con los ITSs permiten analizar otros comportamientos tales como estados afectivos (p. ej., emociones [70]). También algunos ITSs permiten realizar predicciones (p. ej., éxito [71]), que será el foco de los ITSs en esta tesis.

Otro tipo de plataformas ampliamente utilizadas en el sector educativo son los sistemas de gestión del aprendizaje (del inglés *Learning Management Systems* (LMSs)). Los LMSs son portales web que permiten conectar a estudiantes y profesores, de modo que los profesores puedan compartir materiales, recursos o actividades con sus estudiantes [7]. Algunos ejemplos de LMSs son Moodle, .LRN, Canvas y Blackboard, entre otros. Un uso bastante extendido de estas plataformas es el de subir los materiales utilizados en las clases presenciales para que los estudiantes puedan revisarlos fuera del aula. Sin embargo, los LMSs también permiten tener foros de discusión de los estudiantes y realizar ejercicios. Estos ejercicios pueden estar diseñados de modo que el estudiante reciba cierta retroalimentación tras realizarlos, pero no suelen tener el grado de personalización y adaptación de los ITSs. A pesar de ello, los LMSs suelen tener un mayor número de funcionalidades que se pueden aplicar de manera genérica a cualquier curso (p. ej., capacidad de carga de contenidos estandarizados, foros o entrega de tareas) y no solo para dominios específicos. Además, los LMS permiten la recogida de trazas de la actividad de los estudiantes. Mediante el uso de estas trazas es posible realizar análisis sobre el comportamiento de los estudiantes. Por ejemplo, Romero et al. [72] calcularon varias variables sobre los estudiantes, principalmente sobre el tiempo invertido en la plataforma, la procrastinación y la participación en el foro de Moodle, y a partir de ello obtuvieron diferentes perfiles de estudiantes, que mostraban diferencias tanto en rendimiento como en estrategia (p. ej., había un grupo que se centraba principalmente en las tareas obligatorias). También, Quinn y Gray [73] utilizaron las trazas generadas por Moodle para predecir qué estudiantes aprobarían el curso y qué nota alfabética conseguirían.

Aparte de los ITSs y los LMSs, particularmente en el ámbito de la enseñanza a distancia, destaca el uso de los cursos online masivos y abiertos, *Massive Open Online Courses* (MOOCs). Los MOOCs, de acuerdo con Baturay [74] tienen tres propiedades, que son su carácter abierto, participativo y distribuido, dado que son

accesibles desde cualquier lugar de forma online, y promueven las contribuciones y la compartición de información entre compañeros. Además de estas características, es de destacar el aspecto masivo de estos cursos (el número de alumnos está en el orden de miles), lo que conlleva a una amplia variedad de estudiantes, de diferentes nacionalidades y perfiles [75].

En cuanto a la clasificación de los MOOCs, puede haber diferentes categorizaciones, aunque una extendida es la de Siemens [76], que separa entre eXtended Massive Open Online Course (xMOOC) y Connectivist Massive Open Online Course (cMOOC). Los xMOOCs son cursos en los que el profesor es el que proporciona los contenidos y conocimiento, y los alumnos se encargan de realizar el trabajo del curso y preguntar dudas cuando sea necesario [77]. En cambio, los cMOOCs tienen una aproximación diferente y en ellos, la figura principal es el alumno, que se encarga de construir y fomentar conocimiento mediante su participación en comunidades virtuales. También, otra clasificación relevante es de acuerdo con el modo de instrucción. De acuerdo con esta clasificación, existen los MOOCs síncronos y asíncronos [78]. Los MOOCs síncronos (también conocidos como a ritmo del instructor, *instructor-paced*) son cursos en los que el profesor va marcando el ritmo que deben llevar los alumnos y marca fechas concretas para el desarrollo de las tareas. En esta modalidad, los cursos se desarrollan en un momento determinado del tiempo y todos los alumnos realizan las tareas de los mismos módulos al mismo tiempo. Por otro lado, los MOOCs asíncronos (también conocidos como a ritmo propio, *self-paced*) son cursos bastante flexibles donde no hay unas fechas definidas y los materiales se dejan disponibles desde el inicio para que los alumnos puedan trabajar con ellos a su ritmo, en el momento y lugar que quieran.

Respecto a las plataformas donde se albergan estos MOOCs, también existen varias posibilidades. Entre las plataformas más populares destacan Coursera¹, edX², FutureLearn³, existiendo también muchas iniciativas, como MiríadaX⁴ en el ámbito iberoamericano. Estas plataformas, aunque cada una tenga sus propias políticas en los cursos (p. ej., requisitos para superar el curso, modo de conseguir certificados, etc.), tienen una serie de elementos en común. De acuerdo con Grainger [79], los elementos principales de un MOOC son (1) los vídeos, en los que se exponen los contenidos del curso, (2) las evaluaciones, que pueden ser muy variadas (tanto formativas como sumativas, aunque destacan las preguntas de respuesta múltiple, las preguntas de corrección mecanizada para cursos de programación, y la evaluación entre pares), (3) el foro del curso, (4) sesiones de vídeo en directo, (5) lecturas, (6) materiales de vídeo adicionales (p. ej., debates sobre el tema para reforzar contenidos) y (7) herramientas sociales, tales como Facebook o Twitter. A partir de estos elementos, y debido al carácter masivo de los MOOCs, es posible recoger datos de muchos estudiantes. Esto hace que los MOOCs sean bastante interesantes desde el punto de vista de la investigación. Además, el hecho de que todos los contenidos se consuman a través de una plataforma implica que haya una gran cantidad de eventos (sobre actividad en vídeos, ejercicios, foro, etc.) para los posibles análisis.

También, con gran relación a los MOOCs, se encuentran los cursos online cerrados y privados, Small Private Online Courses (SPOCs). Este término fue introducido por

¹<https://www.coursera.org>

²<https://www.edx.org/es>

³<https://www.futurelearn.co/>

⁴<https://miriadax.net>

Fox [4] para referirse a cursos cerrados que se imparten en una plataforma con el mismo formato que el de un MOOC. Por ejemplo, utilizando el software de Open edX, es posible crear un MOOC en la plataforma edX, o crear un curso cerrado (SPOC) en una instancia local de Open edX o en la versión de edX para cursos cerrados, edX Edge. En todos estos casos, el curso tendría la misma apariencia y funcionalidad, aunque dependiendo del ámbito en el que use se consideraría un MOOC o SPOC. Los SPOCs pueden ser usados en cursos universitarios que formen parte de una titulación (enseñanza reglada), y pueden ser de gran apoyo en las clases presenciales y en contextos de aprendizaje semipresencial (*blended learning*). Sobre el uso de estos SPOCs, Moreno-Marcos et al. [80] identificaron tres usos fundamentales: (1) SPOCs cuya realización es necesaria para aprobar el curso, con un posible peso significativo en la nota final, (2) SPOCs que son parte del curso (a menudo combinados con la clase invertida), pero que no forman parte de la evaluación sumativa, y (3) SPOCs que aportan material complementario de apoyo al curso, sin ser de carácter obligatorio. La finalidad del SPOC puede ser bastante relevante y las interacciones de los alumnos pueden variar en función de eso. Sin embargo, dado que el formato de los SPOCs es muy similar al de los MOOCs, la forma de recopilar los datos es similar (p. ej., sobre interacciones en vídeos y ejercicios).

Aparte del uso de las plataformas mencionadas de forma individual, en el ámbito educativo es posible hablar de entorno educativo, que es un término más amplio para referirse a los contextos en los que se realiza el aprendizaje. Las plataformas anteriores son ejemplos de entornos educativos, pero en la actualidad están apareciendo nuevos entornos distribuidos donde se combinan diversas plataformas, servicios, etc. Por ejemplo, es posible tener plataformas orientadas a servicios que pueden integrar diversos servicios externos [81] (como servicios con actividades de aprendizaje, juegos externos, blogs, wikis, etc. [82]) y combinar diferentes dispositivos, incluyendo ordenadores, smartphones, tabletas, etc.

También, dentro de los entornos distribuidos destacan los entornos de aprendizaje inteligente (del inglés *Smart Learning Environments (SLEs)*), que combinan diversas tecnologías, incluyendo LA, para proporcionar más información sobre el comportamiento y rendimiento de los estudiantes, de modo que sea posible, por ejemplo, proporcionar un aprendizaje personalizado y adaptativo [83], [84]. En cuanto a las tecnologías utilizadas en los SLEs, de acuerdo con una revisión sistemática del estado del arte [85], destaca el uso de IoT, lo que incluye sensores, balizas electrónicas, cámaras y placas computadoras, y el uso de realidad virtual y/o aumentada.

Aunque la combinación de diversas tecnologías pueda ser positiva, para su correcta integración en los SLEs es importante garantizar la interoperabilidad entre los diferentes formatos y estándares de las diferentes tecnologías y recursos [86]. Para ello, pueden usarse estándares para el almacenamiento de datos en e-learning, tales como xAPI [87], CAM [88], IMS Calliper [89] y Marsupial [90]. Mediante el uso de dichos estándares, es posible combinar datos de distintas fuentes y utilizar técnicas de Big Data y LA para, entre otras cosas, detectar problemas en los estudiantes y ofrecer apoyo personalizado y adaptativo. Entre los usos de LA, es posible utilizar modelos predictivos y, tal y como se comentó en la revisión del estado del arte mencionada [85], hace falta más trabajo en torno a la predicción en los SLE. Los resultados de la predicción de esta tesis pueden contribuir en el diseño de los SLEs.

2.3. Autorregulación del aprendizaje y comportamientos de los estudiantes

En la tesis doctoral, algunas variables que se utilizan como variables predictoras son sobre la autorregulación del aprendizaje o de comportamientos de los estudiantes. De cara a comprender mejor ese tipo de variables cuando se presenten, esta sección presenta una visión general sobre este tipo de variables.

En primer lugar, se trata la autorregulación del aprendizaje [91]. La autorregulación del aprendizaje (del inglés, *Self-Regulated Learning* (SRL)) es el proceso por el cual los estudiantes toman control de su proceso de aprendizaje para conseguir los objetivos propuestos en el aprendizaje [92]. Por tanto, los estudiantes con capacidades de autorregulación son capaces de iniciar procesos cognitivos, metacognitivos, afectivos y emocionales [93]. Esto es importante, ya que por ejemplo, en los MOOCs, un problema es que los estudiantes suelen procrastinar sus tareas, lo que aumenta la probabilidad de abandono [94], y en particular en los MOOCs asíncronos, donde los estudiantes deben organizarse para ir trabajando en el curso [95]. Sin embargo, si los estudiantes son capaces de establecer sus objetivos, planificar su trabajo y reflexionar sobre su aprendizaje, es decir, tienen habilidades de SRL, tienen mayor probabilidad de tener éxito [96].

En esta línea, trabajos previos han demostrado que algunas estrategias, como la gestión del tiempo, la motivación o el autocontrol tienen un impacto en el éxito de los estudiantes (p. ej., [97], [98]). Broadbent [99] encontró una relación positiva entre las habilidades de SRL y las notas del MOOC, y señaló la importancia de la gestión del tiempo y la elaboración (capacidad para combinar y construir conocimiento de múltiples fuentes [100]). Hood et al. [101] estudiaron la relación cómo el SRL afecta a los resultados del aprendizaje y concluyeron que el contexto tiene un efecto importante en la forma en la que los estudiantes se autorregulan y se comprometen con el curso. También, Papamitsiou et al. [102] mostraron que los estudiantes con mejores habilidades para la gestión del tiempo conseguían mejores resultados. Además, Tempelaar et al. [103] concluyeron que los estudiantes que utilizaban ejercicios con solución de forma esporádica conseguían mejores resultados, lo cual se relaciona con la estrategia de saber pedir ayuda cuando se tienen dificultades.

En relación con el abandono, también ha habido análisis empíricos para demostrar que las estrategias de SRL pueden tener un impacto en el abandono. Lee et al. [104] exploraron las habilidades de SRL y encontraron diferencias entre las habilidades metacognitivas (si los estudiantes reflexionan en lo que aprenden) entre los estudiantes que abandonaban y los que no abandonaban. Además, Sun et al. [105] encontraron una relación positiva entre la autoeficacia (creencia del estudiante sobre su capacidad de conseguir los objetivos [106]), las habilidades de pedir ayuda y el éxito académico. También, Kizilcec et al. [107] concluyó que la gestión del tiempo y la regulación del esfuerzo eran habilidades esenciales para que los estudiantes consigan sus objetivos. Dada la importancia que se presenta en la literatura sobre el SRL, una hipótesis que se plantea en la tesis es que las variables de SRL pueden ser obtener un alto poder predictivo en la predicción.

Tras comentar sobre SRL, uno de los comportamientos que se tratan en la tesis respecto a la predicción es la persistencia [108]. La persistencia, a menudo referida como perseverancia, es la habilidad de un estudiante de continuar trabajando con esfuerzo en una tarea (p. ej., un ejercicio) a pesar de las dificultades (p. ej., tras

obtener una respuesta incorrecta) [109]. Dependiendo de la tarea específica y las posibles dificultades, puede haber varias formas de calcular la persistencia. Silvervarg et al. [110] realizaron un estudio con alumnos de 10 a 12 años que utilizaban un juego educativo. Este juego estaba diseñado de modo que había ejercicios que eran muy improbables que fueran resueltos correctamente. Cada vez que fallaban un ejercicio, se les presentaban varias opciones incluyendo continuar el ejercicio, realizar un ejercicio más fácil, tomar un descanso para jugar a un juego, etc. El estudio concluyó que los estudiantes con mayor persistencia alcanzaban niveles de mayor dificultad. También, Eley et al. [111] analizaron los perfiles de personalidad de estudiantes de medicina y concluyeron que el 60 % tenían un perfil con alta persistencia, lo cual es importante para tener éxito en medicina.

En cuanto a rendimiento académico, Scherer y Gustafsson [112] analizaron la relación entre la persistencia y el rendimiento a partir de datos autorreportados de tres países (Australia, Noruega y Singapur), y encontraron una correlación positiva (de en torno a 0.3, con mínima diferencia entre países) entre la persistencia y el rendimiento. Además, Farrington et al. concluyeron que la persistencia tenía una relación directa con las notas, aunque observaron diferencias en la relación dependiendo del periodo en que se medían ambas variables. En particular, la relación era mayor cuando ambas variables se medían de forma concurrente.

A pesar de que algunas contribuciones han analizado la relación entre la persistencia y otras variables, una limitación es que la mayoría utilizan datos autorreportados (p. ej., [109], [111], [112]) y muy pocas (p. ej., [110], [113]) han utilizado datos de eventos. Por ejemplo, Ventura et al. [113] midieron la persistencia como el tiempo invertido en los ejercicios que no habían sido resueltos correctamente. En esta tesis, se contribuye con el análisis del efecto de esta variable en los modelos predictivos.

Otro de los comportamientos a considerar en esta tesis es la eficiencia [114]. En la literatura, muchos trabajos han tratado la efectividad de los estudiantes, que significa cuánto han aprendido los estudiantes [115], pero el concepto de eficiencia no solo está relacionado con los resultados de aprendizaje, sino que también considera cómo de rápido los estudiantes consiguen los resultados. Esto es importante porque puede haber estudiantes que consigan resolver correctamente los ejercicios, pero necesiten mucho tiempo (no sean eficientes). En estos casos, los instructores pueden identificar a dichos estudiantes e intentar ver si tienen alguna dificultad que les haga progresar más lentos.

En torno a la eficiencia, previamente se presentaron medidas relacionadas en [47] (un estudiante era menos eficiente si visualizaba un vídeo o realizaba un ejercicio varias veces y si empleaba más tiempo de lo normal en resolver los ejercicios). En un análisis posterior de los mismos autores, se observó que a pesar de que el progreso en la plataforma puede ser similar para varios estudiantes, la eficiencia podía ser diferente [116]. Pardos et al. [117] mencionaron como eficiencia el número de clicks por rendimiento, de manera que un alumno era más eficiente si conseguía un rendimiento mayor con menor número de clicks. Sin embargo, no evaluaron al detalle esta variable. Estas dos definiciones, a pesar de que miden la eficiencia, no consideran la relación entre tiempo y rendimiento, por lo que proporcionan otra perspectiva de la eficiencia. No obstante, en general, la eficiencia no se ha tratado ampliamente en la predicción, y por ello, en esta tesis se analiza su efecto en los modelos predictivos.

Por último, se considera la variable sobre la constancia. La constancia indica en qué medida los estudiantes pasan un tiempo similar cada día realizando ejercicios en la plataforma [114]. Este parámetro es relevante porque puede haber diferencias en el rendimiento dependiendo de si los estudiantes interactúan en la plataforma de forma regular y en intervalos similares de tiempo o no. Esta variable fue mencionada como una variable relevante para su visualización en un trabajo previo de la literatura [116]. En otro trabajo, se indicó específicamente una manera para calcular esta variable [47] (utilizando la varianza del tiempo invertido en la plataforma cada día del curso). En otros trabajos, como en [118], se menciona el término constancia, pero luego los autores lo que analizan son el número de ingresos al curso por fecha, pero no calculan una métrica global, como se propone aquí. En esta tesis, se contribuye con el análisis del efecto de esta variable en la predicción.

2.4. Predicción en el ámbito educativo

Como se ha mencionado en la sección anterior, las plataformas educativas permiten recoger una gran cantidad de datos sobre los estudiantes. Mediante el uso de LA, es posible obtener bastante información sobre el comportamiento de los alumnos en estas plataformas y detectar posibles dificultades. En particular, siguiendo los objetivos de LA mencionados en la sección 2.1, es posible predecir variables a futuro sobre el resultado o comportamiento de los estudiantes en las plataformas educativas y entornos de aprendizaje. Esta predicción es el objeto fundamental de estudio de esta tesis doctoral, y por ello, esta sección trata de revisar los trabajos más relevantes que se han realizado sobre predicción en el ámbito educativo.

Esta revisión se ha estructurado en diversas subsecciones que detallan cada uno de los aspectos fundamentales que tratan los artículos sobre predicción en el ámbito educativo. Estos aspectos son los siguientes: (1) variables predictoras, (2) variables a predecir, (3) algoritmos y modelos predictivos, y (4) métricas de evaluación. Estos aspectos fueron identificados en una revisión sistemática de la literatura realizada sobre predicción en MOOCs [16]. Esta revisión fue elaborada al inicio de la tesis y cubre los trabajos más relevantes hasta final de 2017. En esta sección, se toman los aspectos más relevantes considerados en dicha revisión, aunque se extienden para cubrir los artículos publicados de 2018 en adelante y aquellos publicados en plataformas diferentes a los MOOCs. De este modo, esta revisión del estado del arte cubre transversalmente el tipo de plataforma, y se incluyen varios tipos, principalmente de MOOCs, SPOCs, ITSs y LMSs. Aparte del análisis de estos aspectos, se incluye una subsección para cubrir el problema de la generalización de la predicción, dada su relevancia y su foco en esta tesis. Finalmente, se presentan las conclusiones y las necesidades de investigación que se han identificado sobre la predicción, y se conectan con los objetivos de la tesis.

2.4.1. Variables predictoras

Uno de los aspectos fundamentales en los modelos predictivos son las variables predictoras. Las variables predictoras son a menudo indicadores asociados a los alumnos aunque también pueden ser por ejemplo características de un curso. Estas variables predictoras son las variables independientes del modelo predictivo, que capturan información diversa y sirven para predecir la variable dependiente. En los

trabajos de predicción, existe una gran variedad de variables predictoras y en un mismo trabajo es habitual que se combinen variables de distinta naturaleza (p. ej., combinar interacciones de vídeos y ejercicios). Sin embargo, no todas las variables están disponibles en todos los sistemas, por lo que los modelos deben desarrollarse en cada contexto teniendo en cuenta la información disponible. Teniendo esto en cuenta, a continuación se presentan diferentes grupos de variables que han sido utilizadas como variables predictoras en trabajos previos, de los cuales se menciona también en qué plataformas estas variables son más frecuentes, y qué poder predictivo se obtiene con ellas.

- Variables demográficas: están relacionadas con las características generales del estudiante (p. ej., edad, género, lugar de origen, etc.) y pueden estar presentes en cualquier tipo de plataforma educativa. Por ejemplo, en el ámbito universitario, Lykourentzou et al. [119] predijeron el abandono utilizando variables demográficas (género, variable binaria sobre si el alumno reside en la capital o no, experiencia laboral, nivel educativo y nivel de inglés) en dos cursos de la Universidad de Atenas. Sin embargo, el poder predictivo no era bueno cuando se utilizaban solo estas variables y se concluyó que estas variables no aportaban en la predicción. En el contexto de los MOOCs, también se utilizan este tipo de variables (p. ej., [120], [121]), y en muchas ocasiones son recogidas a través de una encuesta. Dada la variabilidad de perfiles en los MOOCs, las variables demográficas pueden ser más amplias como país de origen, lengua materna, familiaridad con el tema. Sin embargo, al igual que se comentó en el ejemplo anterior, este tipo de variables no siempre aportan un alto poder predictivo. Por ejemplo, Brooks et al. [122] encontraron un bajo poder predictivo en variables como género, edad, raza, capacidades de lenguaje, localización geográfica, tipo de suscripción (pago o no) al curso y fecha de inscripción cuando predijeron quién iba a aprobar en un MOOC. Este resultado además también fue observado por Ritter et al. [123] al añadir variables demográficas en modelos que tenían otras variables en un ITS. Por tanto, este tipo de variables pueden no aportar un alto poder predictivo a la predicción, aunque esto pudiese variar en algún contexto determinado.
- Variables sobre la interacción con vídeos: están relacionadas con las interacciones de los alumnos cuando visualizan vídeos en la plataforma. Este tipo de variables son más propias de los MOOCs y los SPOCs, ya que es típico el uso de este tipo de recursos. Por ejemplo, Brinton et al. [124] consideraron las interacciones de varios eventos de vídeo, como reproducir, pausar, rebobinar, avanzar hacia adelante y cambiar la velocidad de reproducción, para predecir si los estudiantes acertarían las preguntas en su primer intento, lo que se conoce como *Correct at First Attempt (CFA)*. De forma similar, Yang et al. [125] combinaron las notas de las tareas con variables de interacción con vídeos, como el porcentaje de vídeo que el estudiante había visualizado, el ratio medio de la velocidad de reproducción, el número de rebobinados, etc. En dicho trabajo, se observó que esta combinación de variables (incluyendo vídeos) podía mejorar el poder predictivo. En el ámbito de los SPOCs, Ran et al. [126], utilizaron el número de vídeos vistos para predecir el rendimiento de los estudiantes en tres SPOCs del ámbito de las ciencias de la computación, y concluyeron que esta variable tenía un alto impacto en los resultados. En

general, y en vista de estos trabajos, este tipo de variables pueden ofrecer un buen poder predictivo y es bueno utilizarlas cuando están disponibles.

- Variables sobre la interacción con ejercicios: están relacionadas con los eventos que generan los alumnos cuando interaccionan con ejercicios (p. ej., cuestionarios, entregables, exámenes) en la plataforma. Este tipo de variables pueden medir tanto la actividad (p. ej., porcentaje de ejercicios intentados) como el rendimiento (p. ej., calificaciones en los ejercicios). Por ejemplo, Ren et al. [28], desarrollaron modelos para predecir la nota de la siguiente tarea calificable en un MOOC y obtuvieron que el número de cuestionarios intentados era la variable con mayor correlación. Del mismo modo, Ruipérez-Valiente et al. [23] desarrollaron modelos para predecir quién iba a obtener el certificado en un MOOC y concluyeron que el mejor predictor era el progreso en problemas (nota media en las tareas realizadas). En el ámbito universitario, Ashenafi et al. [127] predijeron las notas en dos cursos de programación a partir de las tareas. Aparte de variables como el número de intentos utilizados o tareas completadas, usaron variables sobre la dificultad, relevancia e interés de preguntas que proponían los propios alumnos para que otros las resolvieran, siendo la dificultad un buen predictor para el rendimiento. En el ámbito de los ITSs, suelen aparecer otro tipo de variables, como el número de pistas, que en otras plataformas no está disponible. En este contexto, por ejemplo, Feng et al. [128] desarrollaron un modelo para predecir la nota de una prueba estandarizada (con buenos resultados) y usaron variables como el tiempo medio por ejercicio, el número de pistas por problema, y el número de veces que el alumno llega a la última pista (que revela la solución), entre otros. En general, las variables de esta categoría pueden ofrecer un alto poder predictivo, y en muchos casos (p. ej., [23], [28], [127]) son las mejores variables para predecir, por lo que es bueno incorporar este tipo de variables siempre que sea posible.
- Variables sobre el uso del foro: estas variables están relacionadas con la actividad en el foro. Principalmente aparecen en contextos de MOOCs, ya que allí es donde el foro tiene una mayor relevancia para la comunicación (p. ej., en cursos presenciales, el alumno puede comunicarse por correo electrónico o en persona con el profesor). Existe una gran cantidad de variables posibles, ya que a parte de la información sobre la actividad en el foro, pueden obtenerse variables a partir del texto y de las relaciones sociales entre usuarios. Por ejemplo, Klüsener y Fortenbacher [129] únicamente utilizaron variables de foro para predecir el éxito de los estudiantes, y usaron variables como el número de votos positivos y negativos, número de palabras y número de mensajes publicados, entre otros. Arguello y Shaffer [130] también trabajaron en esta línea y obtuvieron variables sobre el sentimiento de los estudiantes y medidas relacionadas con el grado de seguridad con la que los estudiantes se expresaban. Sin embargo, a pesar de que trabajos como [129] reportan buenos resultados con este tipo de variables, existen otros trabajos de predicción (como Cobos y Palla [131]) donde las variables del foro no aportan poder predictivo. Por ello, es posible realizar más análisis para analizar esta variable.
- Variables sobre la actividad en la plataforma: esta categoría incluye variables generales sobre las acciones que un usuario realiza en la plataforma o cuándo

se realizan las acciones, independientemente del tipo de acciones (interacción con vídeos, ejercicios, etc.) que sean. Por ejemplo, Liang et al. [132] incluyeron variables como: número de accesos, número total de clics, número de semanas en las que el alumno ha accedido al curso y el intervalo entre accesos del curso, entre otros. Este tipo de variables pueden tener un buen poder predictivo, como fue demostrado por Alamri et al. [133] que obtuvieron exactitudes de entre un 0.82 y 0.92 al predecir abandono utilizando únicamente el tiempo invertido para completar cada contenido (artículos, imágenes y vídeos) y el número total de accesos a los contenidos. En el ámbito de SPOCs, Wan et al. [134] analizaron un SPOC sobre ciencias de la computación para predecir la nota final del curso. En dicho trabajo, utilizaron algunas variables como el tiempo medio que pasa entre que la tarea se publica y el alumno comienza a realizarla, el número de accesos tras fallar un problema, la duración media entre envíos, el tiempo invertido en todos los recursos y la varianza entre la distribución del tiempo durante la semana. Éstas dos últimas eran además las dos variables con mayor poder predictivo en el análisis. En general, este tipo de variables pueden ofrecer un buen poder predictivo, como se muestra en los ejemplos anteriores, y su uso es recomendable en la predicción.

- Variables sobre comportamientos: son variables de alto nivel obtenidas a través de los eventos y que aportan un significado sobre un determinado comportamiento o aspecto de la personalidad del estudiante. Actualmente, no hay muchos trabajos que utilicen este tipo de variables y lo que es más común es que el comportamiento sea la variable a predecir, pero existen algunos ejemplos. Entre ellos, Xie et al. [135] obtuvieron varias variables sobre estrategias de aprendizaje (aprender de errores, cambiar de un tema a otro, revisar temas aprendidos) para predecir la nota al final del semestre, utilizando datos de un ITS, y obtuvieron resultados moderados (*Area Under the Curve (AUC)* de 0.77). Además, San Pedro et al. [136] utilizaron variables de comportamientos en un ITS para predecir qué estudiantes elegirían estudios relacionados con la ciencia, tecnología, matemáticas o ingeniería (*Science, Technology, Engineering and Mathematics (STEM)*). Entre dichas variables, destacó el descuido (probabilidad de que el alumno falle, incluso dominando las habilidades para responder), aunque se utilizaron otros estados afectivos como la frustración, aburrimiento o confusión de los estudiantes. No obstante, se observó que los estados afectivos utilizados no ofrecían un poder predictivo muy alto. Respecto a esta categoría, dado que pueden existir muchos tipos de comportamientos, el poder predictivo puede depender de cada tipo de comportamiento y más trabajos pueden realizarse para analizar nuevos comportamientos.
- Variables autorreportadas: son variables en las que el alumno aporta la información, sin ser obtenida a través de eventos, por lo que la información en muchos casos no puede verificarse (el alumno puede por diversos motivos poner información incorrecta). Se obtienen normalmente a través de cuestionarios o encuestas y frecuentemente tratan temas como el interés de los estudiantes, su compromiso con el curso o sus expectativas. En el ámbito educativo, se han utilizado muy frecuentemente, sobre todo para conocer las características de los estudiantes y predecir su motivación, aunque cada vez más se prefiere el análisis

de eventos porque los cuestionarios pueden consumir bastante tiempo y pueden no ser fiables [137]. En el ámbito de los MOOCs, por ejemplo, sí que se utilizan para conocer datos del estudiante que de otra manera no pueden ser conocidos. En este ámbito, por ejemplo, Zhong et al. [138] predijeron estilos de aprendizaje (activos, pasivos, o ambos) en un MOOC. Para ello utilizó una encuesta con preguntas sobre las impresiones sobre el aprendizaje mediante un MOOC (p. ej., “¿Cuáles son las ventajas de aprender en los MOOCs”). Además, Hicks et al. [139] combinaron información sobre el acceso a los recursos de un MOOC con una encuesta para clasificar a los estudiantes en diferentes categorías de compromiso en el curso (muy comprometidos, consistentes, alumnos de dos semanas en el curso, de una, o esporádicos). Sin embargo, los resultados de su análisis mostraron que no había indicadores con alto poder predictivo, por lo que la encuesta no era muy efectiva. Este motivo puede hacer que este tipo de variables no sean las preferidas respecto a otros tipos. No obstante, también podrían ser útiles, como demostraron Pardo et al. [140] al obtener mejores resultados de la predicción del resultado final combinando variables de la plataforma y variables sobre SRL autorreportadas. Por ello, el poder predictivo de estas variables puede depender del contexto.

- Otras variables: esta categoría incluye variables que no pueden ser categorizadas en el resto de los grupos. Por ejemplo, Hong et al. [141] utilizaron variables como la tasa de abandono del usuario (p. ej., número de cursos que había abandonado previamente) o de la clase (p. ej., tasa de abandono del curso) para predecir el abandono en 39 MOOCs. En el área social, Chen et al. [142] recogieron tweets utilizando geolocalización, de modo que solo se recogieran aquellos en un radio de 1.3 millas del campus. En ese trabajo, se utilizó dicha información para predecir problemas del estudiante, como alta carga de trabajo, falta de interacción social, pensamientos negativos o problemas de sueño. En el ámbito de los ITSs, Crossley et al. [71] predijeron éxito en matemáticas en alumnos de entre 8-12 años aproximadamente. Para su modelo, utilizaron NLP sobre los mensajes enviados por el alumno al agente virtual pedagógico del ITS y obtuvieron variables sobre la cohesión del texto, la sofisticación léxica y el sentimiento. Entre los resultados, vieron que los alumnos más exitosos producían textos más elaborados, aunque la variable con mejor poder predictivo era el nivel educativo (curso concreto al que pertenecía el estudiante). Dado que esta categoría puede incluir muchas variables, el poder predictivo puede depender de cada variable en concreto.

2.4.2. Variables a predecir

En el ámbito de la predicción en educación, puede haber muchas posibilidades de variables a predecir, aunque destacan las variables relacionadas con el rendimiento y comportamiento de los estudiantes. A continuación, se muestra una categorización de las variables a predecir encontradas.

- Éxito: el propósito de esta variable es predecir si el estudiante conseguirá los objetivos del curso y será exitoso o no, entendido como una variable binaria (es decir con solo dos posibles valores). Nótese que por ejemplo, esta categoría es diferente por ejemplo de la predicción de la nota, ya que aunque predecir una

nota buena sea sinónimo de éxito, aquí solo se considera la predicción binaria del éxito. Para esta categoría, el éxito académico generalmente se traduce en aprobar el curso. En el caso de los MOOCs, es típico encontrar el éxito como obtener el certificado del curso [143], aunque dado que la mayoría de las plataformas ya no emiten ningún tipo de certificado sin pagar, el éxito en muchos casos se indica simplemente como aprobar o suspender (p. ej., [144]). En este ámbito, Cobos y Macías Palla [131], por ejemplo, desarrollaron una herramienta interactiva con visualizaciones sobre el resultado de diferentes modelos para predecir quién obtendría el certificado en los MOOCs. En el ámbito universitario, Marbouti et al. [145] predijeron si los alumnos iban a aprobar en un curso de ingeniería. En dicho trabajo, únicamente utilizaron información de rendimiento dentro del curso, incluyendo la nota del examen parcial, de algunos cuestionarios y de algunos objetivos de aprendizaje (tras realizar un proceso de selección de variables, que mejoró el poder predictivo). En el ámbito de los ITSs, Kelly et al. [146] desarrollaron un modelo para la predicción de una prueba estandarizada a través de las interacciones en un ITS. Su modelo clasificaba entre los estudiantes dependiendo de si su resultado era el máximo (avanzado) o no.

- Abandono: el objetivo es predecir si un estudiante dejará el curso antes de completarlo o no, por lo que el abandono es un problema de clasificación binario. Esta variable es una de las más relevantes en la predicción en las diferentes plataformas. En los MOOCs, es una de las variables de mayor interés por la alta tasa de abandono (pueden estar incluso por encima del 90 % [147]) y por ello, existen numerosos trabajos en esta línea (p. ej., [148], [149], [150]). Por ejemplo, Halawa et al. [26] predijeron el abandono usando datos de MOOC y concluyeron que había una alta relación entre tiempo de inactividad de más de tres semanas y el abandono. Para dicho trabajo, el abandono se definía como una ausencia de actividad superior a un mes o un visualización de vídeos inferior al 50 %. En el ámbito universitario, Burgos et al. [151], predijeron el abandono en varios cursos de una universidad española a distancia utilizando regresión logística y datos de Moodle, y utilizando un plan a partir de dichas predicciones, consiguieron reducir el abandono en un 14 %. En dicho caso, el abandono se definía como completar menos del 50 % de los puntos que se podían obtener con actividades. Por otro lado, Pereira et al. [152] definieron abandono como un nivel de asistencia inferior al 75 % en un curso de programación. Para sus modelos predictivos, utilizaron variables relacionadas con el código que los estudiantes realizaban (p. ej., número de líneas de cada código, número de pruebas unitarias realizadas o la latencia de pulsación de teclas cuando programaban en un entorno de desarrollo integrado).
- Notas: el propósito es predecir el resultado que obtendrá el estudiante en alguna prueba calificable o en el curso. La variable a predecir normalmente es continua (por ejemplo, una nota en el rango de 0 a 10), aunque a veces también puede ser una variable discreta con un cierto número de valores (por ejemplo, calificación alfabética de la A a la F). Por ejemplo, Ding et al. [153] predijeron la nota obtenida en cada módulo de un MOOC a partir de los eventos de bajo nivel proporcionados por edX y redes neuronales. En el caso de los SPOCs, Yu [154] combinó la regresión lineal y redes neuronales para predecir la nota final

en un curso sobre computación. En los ITSs, la predicción de la nota también es muy común. En muchos casos la predicción de la nota no es la de los ejercicios del sistema, sino la de una prueba estandarizada [155]. Por ejemplo, Fancsali et al. [156] predijeron la nota de un examen estatal de matemáticas a partir de variables obtenidas en el ITS como el tiempo invertido o las habilidades adquiridas.

- Comportamientos de los estudiantes: el objetivo es predecir características de los estudiantes relacionadas con su comportamiento en el curso. Por ejemplo, Jena [157] realizó un estudio para predecir estilos de aprendizaje (si los estudiantes son reflexivos, visuales, intuitivos, secuenciales, etc.) en estudiantes de postgrado de la India. Asimismo, Tato [158] analizó la predicción de la atención de los estudiantes de primaria mientras resolvían problemas de matemáticas en una plataforma educativa. En un entorno de tutores inteligentes, Peters et al. [159] se enfocaron en detectar los alumnos que estaban jugando con el sistema, es decir, quiénes se estaban aprovechando de las reglas del sistema para conseguir resolver los ejercicios. Entre sus conclusiones, vieron que tanto las características del ejercicio (p. ej., número de alumnos que intentan jugar con el sistema en el problema X) como del alumno tenían un poder predictivo similar. En ámbito de los MOOCs, Chen et al. [18] realizaron un estudio sobre la predicción de la personalidad de los estudiantes. En particular, en dicho estudio se clasificó la personalidad de acuerdo a las cinco grandes dimensiones de la personalidad (apertura a la experiencia, responsabilidad, extraversión, agradable y neuroticismo) [160].
- Estados afectivos: el objetivo es predecir un estado emocional o los sentimientos de un estudiante en el curso. En los MOOCs, en una revisión del estado del arte, se analizaron 46 artículos sobre predicción de estados afectivos entre 2012 y 2018. De esta revisión, se concluyó que la mayoría de los trabajos utilizaban LA y técnicas de aprendizaje máquina o análisis de sentimientos (p. ej., [161]), y en pocos casos se realizaba un reconocimiento de señales fisiológicas [162]. Sin embargo, en el ámbito de los ITSs hay un gran interés recientemente en este tipo de variables. Por ejemplo, Jacques et al. [163] investigaron la utilidad de utilizar datos de seguimiento ocular para predecir el aburrimiento y la curiosidad de los estudiantes.
- Otros: esta última categoría engloba a cualquier trabajo que no se pueda clasificar en las categorías anteriores. En esta categoría se encuentran, por ejemplo, trabajos sobre la predicción de si un estudiante va a volver a publicar mensajes en el foro [164], sobre la predicción de la preferencia de los estudiantes en cuanto a preguntas del foro [165], la clasificación de mensajes del foro en función de la relación con los temas del curso y el conocimiento que muestran los estudiantes en los mensajes [166], y la predicción de la habilidad de los estudiantes (que se predice frecuentemente en los ITSs) [167], entre otros.

2.4.3. Algoritmos y modelos predictivos

Las variables predictoras, como se ha visto, pueden usarse para obtener la variable a predecir. Sin embargo, para realizar dicha predicción, es necesario un modelo que relacione las variables predictoras y a predecir. Estos modelos se

diseñan utilizando algoritmos de aprendizaje máquina (en inglés **Machine Learning (ML)**), que tras entrenarse con datos pasados, generan un modelo predictivo. A continuación, se muestran los algoritmos de predicción más utilizados en la literatura, de acuerdo con [16].

- **Regresión (RG)**: Makhabel [168] definió el objetivo de la regresión como encontrar la mejor curva en un espacio multidimensional que se ajuste a los datos. La regresión es uno de los métodos más sencillos, lo que hace que sean bastante utilizados (aunque sea únicamente para una exploración inicial). Por ejemplo, Wise et al. [169] utilizaron regresión logística regularizada L2 para clasificar los hilos del foro de varios MOOCs dependiendo de si estaban relacionados o no con los temas del curso. Por otro lado, Ruipérez-Valiente et al. [170] utilizaron regresión lineal para predecir la ganancia de aprendizaje (diferencia entre la nota de un post-test y la de un pre-test inicial) en cursos preparatorios de (SPOCs) de física y química para estudiantes de nuevo ingreso a la universidad. En este estudio, consiguieron un R^2 de 0.68 con tan solo seis variables, incluyendo la nota del pre-test, el número medio de intentos, el número de ejercicios resueltos sin pistas y en el primer intento, el tiempo medio empleado cada día que interacciona el alumno, el número de vídeos que alumno comienza y no termina, y una variable que engloba comportamientos negativos (p. ej., comportamientos irreflexivos o de no pedir pistas).
- **Máquinas de vectores soporte (SVM)**: su objetivo es encontrar un hiperplano que maximice el margen entre las diferentes clases para realizar la clasificación [171]. En el ámbito de los MOOCs, Fei y Yeung, por ejemplo, utilizaron SVMs para predecir el abandono en dos cursos de Coursera y edX [172], y Macina et al. [173] las utilizaron para predecir si los estudiantes responderían o no a las preguntas del curso. También, Huang y Fang [174] predijeron la nota final en un curso de ingeniería (de nivel de grado) utilizando varios modelos. Aunque una de sus conclusiones fue que el modelo predictivo no era el factor que más influía en los resultados, obtuvieron que el SVM les proporcionaba un mayor número de alumnos cuyo error en la nota predicha era inferior al 10% de la nota.
- **Árboles de decisión (en inglés Decision Tree (DT))**: son modelos de aprendizaje supervisado cuyo objetivo es realizar la predicción a partir de simples reglas de decisión a partir de los indicadores [175]. Es uno de los algoritmos más utilizados en la predicción [176], y en muchos casos se usan por su simplicidad y facilidad de comprensión [25]. En el ámbito de los MOOCs, Xing et al. [177], por ejemplo, combinaron una red bayesiana con un árbol de decisión para predecir la semana en la que el alumno abandonaría el curso (si lo abandonaba). Para ello utilizaron variables como el número de veces que entraban al foro, hacían ejercicios, veían recursos, el número de días activos o el grado del alumno en la red social generada por los mensajes del foro (usuarios con los que el alumno se comunicó). Por otro lado, Pardo et al. [178] desarrollaron modelos predictivos con DT para predecir la nota del examen parcial y del examen final en un curso de ingeniería de primer año, y obtuvieron un **Root Mean Square Error (RMSE)** de en torno a 0.2 para el parcial y 0.14 para el examen final.

- **Random Forest (RF)**: es un conjunto de árboles creados a partir de muestras bootstrap del conjunto del entrenamiento y a partir de la selección aleatoria de variables para la creación de las reglas de decisión de los árboles [179]. En el ámbito de los MOOCs, Laveti et al. [180] utilizaron varios algoritmos, incluyendo RF para predecir el abandono, y obtuvieron un AUC de 0.88 con dicho algoritmo. No obstante, combinando varios algoritmos consiguieron mejorar el AUC hasta 0.91, aunque aumentando considerablemente el coste computacional. El RF, en general, es uno de los algoritmos que puede obtener uno de los mejores poderes predictivos, como concluyó Ye et al. [181], al comparar este algoritmo con otros, como regresión logística, SVMs y DT para predecir el abandono en MOOCs. En entornos de enseñanza presencial, también el RF puede obtener muy buenos resultados, como en el caso de Cortez y Silva [182], cuando predijeron el resultado de los estudiantes en dos cursos de secundaria sobre matemáticas y portugués. En este caso, el RF fue el mejor algoritmo en la mayoría de los análisis, seguido del DT.

- **Redes Neuronales**: son sistemas de computación paralelos para la optimización y aprendizaje basados en la estructura y funcionalidad del cerebro [183]. En el ámbito de los MOOCs, han cobrado una especial importancia a partir de 2017 [16] y, como ejemplo, Pérez-Lemonche et al. [184] predijeron notas en un curso sobre programación en Android en edX y obtuvieron un Mean Absolute Error (MAE) de en torno 0.13 con redes neuronales, que es razonablemente bueno. El uso de las redes neuronales también se ha extendido a otros contextos y por ejemplo, Joshi et al. [185] las utilizaron para predecir el resultado de los estudiantes a partir de grabaciones de vídeo mientras realizaban problemas de matemáticas en un ITS.

- **Otros**: esta categoría incluye cualquier algoritmo de predicción no incluido en las categorías anteriores. Entre ellos están el Naïve Bayes, Gradient Boosting Machine (GBM), redes bayesianas, máquinas de factorización, K-NN y procesos gaussianos, entre otros. Por ejemplo, Lu et al. [186] utilizaron varios algoritmos, incluyendo Naïve Bayes para predecir si un alumno se iba a presentar o no al examen final de un MOOC. Utilizaron datos de tres cursos y aunque el mejor modelo variaba de un curso a otro, el Naïve Bayes obtuvo la mejor exhaustividad en uno de ellos. En el ámbito de los ITSs, un ejemplo bastante común es el del modelado de la habilidad de los estudiantes, para el cual es frecuente utilizar técnicas de redes bayesianas, y en particular es bastante frecuente usar Bayesian Knowledge Tracing (BKT) (p. ej., [65], [187]).

De entre estos algoritmos, Gardner y Brooks [188] indican en su revisión sobre predicción del éxito en MOOCs que los modelos más utilizados son los basados en árboles (como DT y RF), los modelos de regresión y SVM (lo que también coincide con [16]). Según dicho estudio, el uso de algoritmos basados en árboles es debido a que estos algoritmos pueden funcionar con diversos tipos de datos, son menos susceptibles a la multicolinealidad, son rápidos y sencillos, pueden proporcionar la importancia de las variables, etc. No obstante, una de las conclusiones, es que no se identifica ningún algoritmo que funcione mejor que el resto [188].

2.4.4. Métricas de evaluación

Una vez desarrollados los modelos predictivos, teniendo en cuenta tanto las variables predictoras, las variables a predecir y el algoritmo de predicción, es importante evaluar los resultados que proporciona el modelo predictivo. Para medir la bondad de los resultados, existen diferentes métricas, aunque algunas han sido mayormente utilizadas en la literatura. A continuación, se describen las métricas más relevantes para la evaluación de los modelos predictivos.

- **Area Under the Curve (AUC):** esta métrica indica el área de la curva **Receiver Operating Characteristic (ROC)**. Esta curva indica la relación entre las tasas de verdaderos y falsos positivos para diferentes umbrales y su área indica la probabilidad de clasificar con la clase positiva una muestra aleatoria con una probabilidad mayor que una de la clase negativa [189]. Esta métrica es ampliamente utilizada para problemas de clasificación en diferentes contextos. Por ejemplo, Qiu et al. [190] predijeron quién iba a obtener el certificado de varios MOOCs y obtuvieron un AUC de 0.96, y Botelho et al. [191] predijeron cuando el alumno iba a presentar comportamientos de “giro de la rueda” (comportamiento de ser persistente intentando una tarea pero siendo incapaz de dominar los conceptos en un periodo razonable). En este último trabajo, se observó cómo el poder predictivo llegaba a 0.89 al predecir el comportamiento en la tarea actual, pero disminuía a 0.6 cuando se predecía en las tareas sucesivas. En cuanto a los umbrales, Mezaour [192] estableció unos umbrales para interpretar los valores del AUC. Estos umbrales se utilizarán como referencia a lo largo de la tesis, y aparecen en la Tabla 2.1.

TABLA 2.1
CRITERIOS DE EVALUACIÓN PARA LOS RESULTADOS DEL AUC, TOMADO DE [192]

Valor del AUC	Calidad
$0.9 < AUC \leq 1.0$	Excelente
$0.8 < AUC \leq 0.9$	Bueno
$0.7 < AUC \leq 0.8$	Moderado
$0.6 < AUC \leq 0.7$	Pobre
$0.0 < AUC \leq 0.6$	Malo

- **Exactitud (del inglés, *accuracy*):** se utiliza en problemas de clasificación e indica el número de casos en los que el clasificador acierta (ya sean positivos o negativos) y el número total de casos. Esta medida es ampliamente utilizada, aunque en algunos casos puede no ser muy representativa. Por ejemplo, si en un MOOC abandona el 90% de los alumnos y se tiene un clasificador que dice que todos los estudiantes abandonan, la exactitud sería de un 90%, aun cuando el clasificador no aporta demasiado valor (simplemente elige la clase más frecuente). Como ejemplos de uso, Coleman et al. [193] utilizaron predijeron quién iba a conseguir el certificado en un MOOC de 18 semanas y consiguiendo una exactitud de 0.81 utilizando únicamente los registros de la primera semana. Del mismo modo, Kostopoulos et al. [194] predijeron si los estudiantes iban a aprobar un curso anual en un grado a distancia, combinando información proporcionada por el tutor y por el LMS. Para medir los resultados utilizaron tanto la exactitud como el F-Score y recalcaron que si bien utilizaban

la exactitud, se debía poner mayor énfasis en el F-score por los problemas mencionados con esta métrica.

- Métricas relacionadas con la matriz de confusión: en esta categoría se incluyen las métricas que se pueden calcular a partir de los siguientes parámetros en la clasificación: (1) verdaderos positivos, (2) falsos positivos, (3) falsos negativos y (4) verdaderos negativos. Las medidas más relevantes de esta categoría son la precisión (en inglés *precision*), la exhaustividad (en inglés *recall*) y el F-score. Como ejemplos, Rasmesh et al. [195] clasificaron mensajes del foro de MOOCs en cuanto a los aspectos del curso a los que se referían (videos, cuestionarios, certificados) y en cuanto al sentimiento (en mensajes positivos, negativos y neutros). En este trabajo, se alcanzó un F-score de 0.71 en aspectos del curso relacionados con cuestionarios, pero solo 0.19 en la detección de sentimientos positivos. Por otro lado, Baneres et al. [196] predijeron la probabilidad de suspenso en cursos universitarios de una universidad a distancia, y utilizaron F-score como métrica. En un análisis temporal, pudieron ver cómo el poder predictivo evolucionaba de 0.53 a 0.84 a lo largo de las semanas del curso.
- Root Mean Square Error (RMSE): esta métrica es principalmente utilizada para problemas de regresión, es decir, problemas donde la variable a predecir es continua [197] (aunque también se usa en algunos problemas de clasificación, como el modelado de habilidad [198]). El RMSE se calcula como la desviación típica de la diferencia entre el valor real y el que proporciona la predicción [16], y se caracteriza porque se expresa en las mismas unidades que los datos. Por ejemplo, Elbadrawy et al. [27] utilizaron el RMSE cuando predijeron las notas de los entregables de un MOOC, teniendo en cuenta las notas de las actividades previas que había realizado el alumno; y Vázquez et al. [199] lo utilizaron cuando predijeron el rendimiento de los estudiantes en un ITS.
- Otras: esta categoría incluye otras métricas utilizadas que no están en las categorías anteriores. Por ejemplo, para problemas de regresión, en algunas ocasiones se utiliza la MAE (p. ej., [128]). También, en problemas de clasificación, se ha utilizado el kappa (p. ej., [200]) y el R^2 , que indica el porcentaje de variabilidad que explica el modelo [16] (p. ej., [201]).

Respecto al uso de una métrica u otra, en la literatura no hay un acuerdo y los resultados aparecen en métricas diferentes [16], [188]. Sin embargo, Pelánek [202] realizó un estudio y obtuvo algunas conclusiones sobre qué métricas usar. Entre sus conclusiones, se indica que la MAE no debe ser utilizada ya que está sesgada hacia el resultado mayoritario. En cambio, el RMSE sí puede ser una buena métrica [203], en particular para el modelado de la habilidad de los estudiantes. En cuanto al AUC, se indica que puede ser una buena métrica para problemas de clasificación relacionados con comportamientos de los estudiantes o estados afectivos (por lo que puede ser buena métrica para el abandono), aunque no debe usarse para el modelado de la habilidad de los estudiantes, al igual que las métricas relacionadas con la matriz de confusión. Por último, se indica que la exactitud se ha utilizado pero no es una métrica apropiada. Uno de los motivos es que presenta algunos problemas cuando se tienen conjuntos de datos no balanceados [204]. Por ejemplo, si en un MOOC abandona el 90 % de los alumnos, un modelo que siempre dice abandono tendría un 90 % de exactitud, sin tener en cuenta las variables de cada alumno.

2.4.5. Generalización

En las subsecciones anteriores, se han presentado una gran cantidad de trabajos que abordan la predicción desde diferentes perspectivas, incluyendo diferentes variables, algoritmos y métricas. Dichos modelos predictivos se pueden diseñar con datos del pasado para poder utilizarse en cursos nuevos. Sin embargo, dadas las diferencias que pueden producirse debido al contexto, un problema importante es cómo asegurar que los modelos obtenidos en un contexto pueden ser transferidos a nuevos cursos con un alto poder predictivo [205]. Este problema fue analizado como parte del estado del arte en un artículo previo [29] y se reutiliza en parte para esta sección. En dicho análisis, se vio que el problema de la generalización ya se ha identificado y por ejemplo, Ocumpaugh et al. [206] observaron cómo sus modelos para predecir estados afectivos no generalizaban a poblaciones diferentes a las utilizadas en el entrenamiento. También, Olivé et al. [207] desarrollaron modelos predictivos con redes neuronales y observaron que el poder predictivo variaba dependiendo del uso que los profesores y las instituciones hacían de los contenidos en los LMSs. Este último hecho es una limitación importante, dado que si los modelos pueden variar significativamente, es necesario comprobar al detalle la validez y la transferencia de cada modelo, lo que puede dificultar la adopción de LA [208].

Sin embargo, a pesar de que es un problema identificado, no existen demasiados trabajos que hayan abordado la generalización de los modelos predictivos. Algunos simplemente han mencionado este problema o lo han señalado como una limitación del artículo, pero no lo han analizado explícitamente (p. ej., [209], [210]). Entre los artículos que sí han abordado este problema, Boyer y Veeramachaneini [30] analizaron la transferencia de los modelos predictivos y obtuvieron una pérdida del AUC de 0.1 cuando se predecía utilizando un modelo entrenado con datos de una edición anterior del mismo curso. Sin embargo, He et al. [211] obtuvieron una buena transferibilidad de un modelo entrenado en una edición del curso y utilizado en la siguiente. Además, Gitinabard et al. [212] analizaron la generalización en cuatro cursos y vieron que se podían obtener resultados aceptables al transferir los modelos, si bien la generalización era mejor cuando se utilizaba otra edición del mismo curso. También, Hung et al. [213] propusieron tres modelos predictivos en contextos de educación primaria y secundaria, y encontraron importantes diferencias tanto en el contexto como en el poder predictivo de las variables. Un resultado similar fue obtenido por López-Zambrano et al. [214]. En su trabajo, se analizó la generalización utilizando datos de Moodle de 24 cursos y se realizaron experimentos de cursos similares en cuanto a titulación o nivel de uso. Como resultados, se concluyó que solamente se podían transferir los modelos con resultados aceptables en algunas circunstancias (p. ej., utilizando datos con todas las variables categóricas y transfiriendo en cursos del mismo grado).

Además, Kidzinsk et al. [215] analizaron la generalización de los modelos y su conclusión es que debe haber un balance entre la especificidad y la generalización. Es decir, para obtener un mejor poder predictivo en un conjunto pequeño de cursos, es mejor utilizar variables que dependen del contexto, pero el uso de esas variables también puede afectar la generalización. Para intentar resolver el problema de la generalización, Kizilcec y Halawa [216] tomaron una aproximación diferente a la de transferir de un curso a otro. En su aproximación, se entrenaron los modelos predictivos utilizando datos de 20 MOOCs y se concluyó que el uso de muchos cursos para entrenar el modelo podía mejorar la generalización (en este caso, se obtuvo un

AUC de 0.92 en el modelo). Sin embargo, dados los trabajos anteriores mencionados que muestran lo contrario (p. ej., [30], [206]), hace falta más investigación para intentar analizar bajo qué circunstancias es posible transferir los modelos.

Dada la incertidumbre sobre la generalización, algunos autores, como Bote-Lorenzo y Gómez-Sánchez [217] han propuesto lo que se conoce como modelos *in-situ*, que son modelos que se generan utilizando únicamente datos del curso sobre el que se realiza la predicción y van añadiendo datos a lo largo del tiempo (p. ej., para predecir en la semana 2, se utilizan datos de la semana 1, para predecir en la semana 3 se utilizan datos de las dos primeras semanas, y así sucesivamente). De este modo, al usar siempre el mismo curso, se evitan los problemas derivados del contexto y la metodología del curso. Whitehill et al. [218] también trabajaron en los modelos *in-situ*. Sus conclusiones fueron que los modelos *post-hoc* (modelos que utilizan datos de cursos pasados) pueden dar mejor poder predictivo en el curso que se analizan, pero los modelos *in-situ* pueden conseguir un mayor rendimiento global. Sin embargo, la limitación de estos modelos no pueden ser utilizados cuando la variable dependiente solo está disponible al final del curso, como la nota final. No obstante, pueden ser utilizados para predecir, por ejemplo, el compromiso de los estudiantes [17], que puede medirse semana a semana.

Teniendo en cuenta lo visto anteriormente, de esta sección se concluye que existe una necesidad de profundizar en qué casos pueden generalizar los modelos, bajo qué condiciones y con qué poder predictivo, de modo que la aplicación práctica de los modelos predictivos pueda ser extensible y escalable en más cursos.

2.4.6. Conclusiones

A lo largo de esta sección, se han presentado varios trabajos que abordan la predicción en educación utilizando diferentes variables y técnicas. Sin embargo, hay varios aspectos que requieren un mayor análisis y que no han sido cubiertos por el estado del arte. En torno a las variables predictoras, se ha visto que las variables de vídeos, ejercicios y actividad en la plataforma pueden ser buenas variables para predecir (como se ha visto, por ejemplo en [23], [126], [133]). Sin embargo, hay otras variables, como las variables del foro en las cuales su poder predictivo no es tan claro. Del mismo modo, se ha identificado la necesidad de explorar otros tipos de variables para ver si pueden aportar valor a los modelos. Entre esos tipos de variables, hay necesidad de analizar variables de alto nivel sobre comportamientos de alumnos (como aspectos de la personalidad del estudiante) obtenidas a partir de la transformación de otros tipos de variables. También, es relevante analizar cómo puede afectar la autorregulación del estudiante en la predicción. Por ello, en esta tesis, uno de los objetivos es analizar el efecto de nuevas variables a predecir, como variables de SRL o variables de comportamientos (tales como la persistencia eficiencia o constancia del alumno) y analizar más en profundidad el efecto de otras variables conocidas, como las variables del foro.

En torno a las variables a predecir, se ha identificado que la mayoría de los trabajos abordan variables relacionadas con el desempeño del estudiante en el curso, como el abandono, el éxito o las notas. Sin embargo, estos trabajos generalmente analizan una variable a predecir y no se conecta el resultado de diferentes predicciones ni se analiza la relación entre dichas variables. En esta tesis, se contribuye con el análisis de la relación entre diferentes variables, para ver si hay diferencias en la predicción entre las mismas. Por ejemplo, si cambia el poder

predictivo de predecir la nota del examen final o la nota final, o si hay relación entre el abandono y la nota.

También, en la literatura, se ha identificado que en muchas ocasiones, los artículos tienen diferentes definiciones de las variables. Por ejemplo, para Halawa et al. [26], un alumno se considera que ha abandonado cuando tiene un periodo de inactividad superior a un mes y/o visualiza menos del 50 % de los vídeos. Sin embargo, para Burgos et al. [151], la condición de abandono se cumple cuando no se completa al menos el 50 % de los puntos de las actividades (en caso de que todas tuvieran el mismo peso, equivaldría a realizar la mitad de ellas), y para Pereira et al. [152] significa tener un nivel de asistencia inferior al 75 %. Uno de los motivos de esta variación puede ser el contexto específico de cada entorno educativo. Para profundizar en este aspecto, en esta tesis se utilizan diversos escenarios educativos y se proporciona una justificación de qué variable puede ser más apropiada en cada contexto.

En relación con lo presentado de los algoritmos, una de las conclusiones de la literatura es que no hay un algoritmo que destaque sobre el resto [188]. En esta tesis, se utilizan diversos algoritmos de predicción, escogidos entre los más utilizados (según lo comentado previamente) para corroborar o refutar dicha conclusión.

Además, al revisar los valores de las métricas de predicción, uno de los efectos identificados es la evolución del poder predictivo a lo largo del tiempo. Por ejemplo, Baneres et al. [196] obtuvieron una evolución del F-score de 0.53 a 0.84 a lo largo del curso. Aunque esta evolución es normal ya que al añadir datos las predicciones son más precisas, un aspecto relevante es determinar a partir de qué momento el poder predictivo es lo suficientemente bueno. Este aspecto es importante ya que si se proporcionan predicciones demasiado tempranas, pueden ser poco precisas y se podría perder credibilidad en los modelos. En trabajos previos de la literatura, se ha abordado este análisis temporal (p. ej., [23], [24]), aunque generalmente se ha realizado en un curso (o varios cursos de la misma plataforma). En esta tesis se contribuye con el análisis temporal en diversos escenarios heterogéneos de diferentes plataformas, incluyendo MOOCs (síncronos y asíncronos) y SPOCs. También, estos escenarios comprenden diferentes niveles educativos, incluyendo cursos universitarios, de preparación al acceso a la universidad, y de primaria y secundaria (enseñanza K-12), que es un nivel educativo sobre el que falta investigación en predicción (de acuerdo con una revisión actual sobre la predicción del rendimiento [219]). Estos análisis permiten obtener una visión global a partir de diferentes entornos educativos sobre cuál es el momento a partir del cual las predicciones pueden ser lo suficientemente precisas, lo cual puede utilizarse para futuras implementaciones de los modelos predictivos.

Por último, en la literatura se ha identificado que uno de los problemas de los modelos predictivos es la generalización de los mismos (como también se identificó en [220]). En la literatura existen algunos trabajos que han comenzado a abordar esta cuestión y han planteado, principalmente, el problema de transferir un modelo a otro curso o una edición del mismo (p. ej., [30], [211]). Sin embargo, faltan obtener conclusiones sobre en qué casos se puede conseguir la generalización, con qué condiciones y con qué poder predictivo. Esta tesis contribuye en esta dirección y presenta varias aproximaciones para lograr la generalización en los modelos predictivos, y analiza varios casos en diferentes escenarios. Entre estos casos, se analiza la transferencia a cursos con los mismos estudiantes, a otra edición del

mismo curso y a cursos diferentes con otros estudiantes. También, se analiza el uso de modelos globales entrenados con un conjunto de cursos diferentes para analizar en qué grado pueden ser útiles para resolver el problema de la generalización.

3

Metodología

Contenido

3.1. Metodología global de la investigación	33
3.2. Recolección de datos	35
3.3. Medidas	38
3.3.1. Presentación de las variables independientes	38
3.3.2. Presentación de las variables dependientes	48
3.4. Métodos analíticos y herramientas	49

Este capítulo tiene como objetivo presentar la metodología de investigación de forma global, así como para los experimentos llevados a cabo en cada uno de los contextos educativos que se han analizado, englobando la recolección de datos, variables dependientes e independientes utilizadas, y los métodos analíticos y herramientas. Para ello, el capítulo se divide en cuatro secciones. La primera presenta la metodología global de investigación. La segunda trata la recogida de datos en las diferentes plataformas educativas utilizadas para los experimentos y el formato de los datos en crudo. La tercera cubre las medidas de las variables utilizadas para los análisis y presenta tanto las variables independientes como dependientes. La cuarta y última sección presenta los métodos analíticos y herramientas utilizadas para el desarrollo de los experimentos.

3.1. Metodología global de la investigación

En esta investigación, se ha seguido la metodología [Design Science Research Methodology \(DSRM\)](#) (en español, Metodología de Investigación de la Ciencia del Diseño). Esta metodología ha sido presentada y utilizada en trabajos previos de la literatura [\[221\]](#), [\[222\]](#), y consta de varias fases. Estas fases se estructuran en un orden nominalmente secuencial, aunque no siempre es necesario proceder en orden [\[221\]](#). Las seis fases de las que consta la metodología son: (1) identificar el problema y motivar, (2) definir objetivos de una solución, (3) diseño y desarrollo, (4) demostración, (5) evaluación y (6) comunicación. El punto de partida depende

de la aproximación tomada para la investigación, aunque en este caso, se comienza por la actividad 1, lo que implica que se toma una aproximación centrada en resolver un problema, que en este caso es conocer más acerca de los factores que afectan a la predicción en la educación. La Fig. 3.1 representa esquemáticamente las diferentes fases, que se presentan con más detalle a continuación.

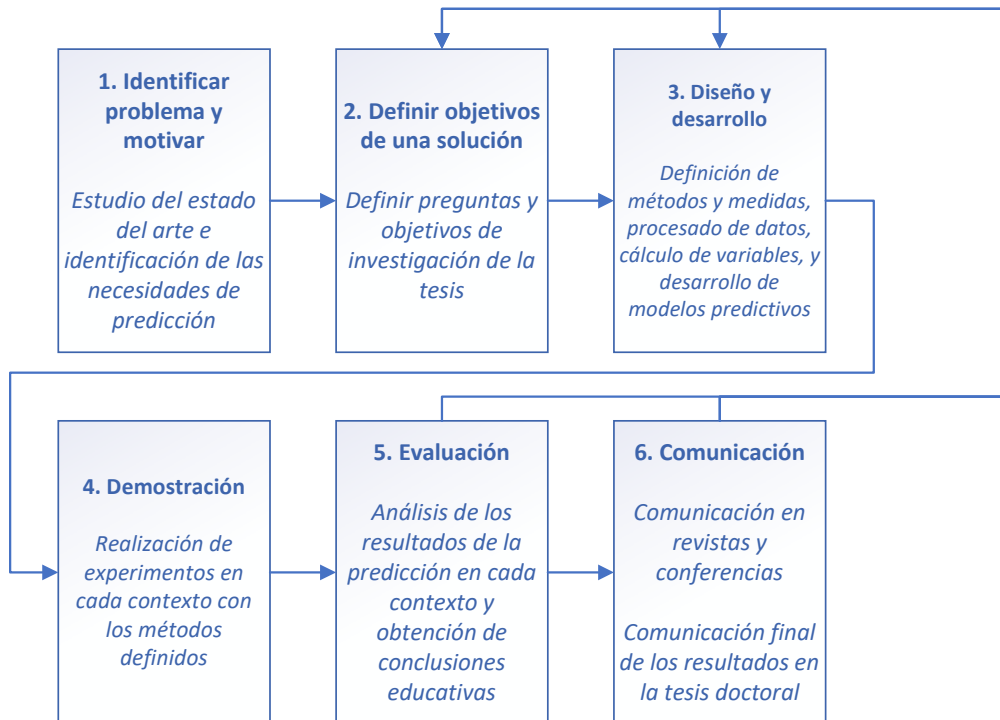


Fig. 3.1. Esquema de la metodología en la investigación

1. Identificar problema y motivar: esta primera fase incluye la identificación del problema que va a abordar la tesis, concretando qué aspectos dentro de la predicción en educación se van a tratar. Para ello, una tarea dentro de esta fase es la revisión del estado del arte, para conocer los trabajos previos sobre la predicción en el ámbito educativo, y poder identificar qué temas ya se han tratado, qué se conoce y qué falta por abordarse. También otra tarea de esta fase es motivar la relevancia de resolver el problema y en concreto identificar las necesidades de los interesados (líderes institucionales, expertos, profesores y estudiantes) respecto a la predicción para alinear la investigación con sus demandas reales.
2. Definir objetivos de una solución: esta fase consiste en definir cuáles son los objetivos y preguntas de investigación que se van a abordar durante la tesis. Dado que la metodología tiene flexibilidad en las iteraciones, estos objetivos se pueden definir parcialmente y luego añadir nuevas preguntas de investigación, de una forma iterativa.
3. Diseño y desarrollo: esta fase tiene como objetivo crear y diseñar los artefactos que se van a utilizar en los análisis. Por ello, en esta fase, se definen los

métodos analíticos que se van a usar para los análisis de la predicción, los contextos educativos y las variables (medidas). Dado que en una primera instancia pueden no contemplarse todos los escenarios educativos, es posible realizar un diseño general y luego ir adaptando el diseño a cada escenario. Por ejemplo, se puede elaborar una lista de posibles variables a utilizar (en base a variables ya conocidas o nuevas variables) y luego analizar para cada contexto cuáles pueden ser utilizadas. Posteriormente, se realiza el desarrollo necesario para poder realizar los análisis en cada contexto educativo. En particular, se realiza el procesado y limpieza de los datos, el cálculo obtención de las variables independientes y dependientes y el desarrollo de los modelos predictivos para su posterior análisis. Dado que existen varios contextos, es posible pasar por esta fase varias veces para el desarrollo de cada contexto.

4. Demostración: esta fase consiste en el uso de los métodos y modelos definidos previamente para obtener los resultados acerca de la predicción. Por tanto, dentro de esta fase, se realizan diversos experimentos con los modelos, por ejemplo, predecir en diferentes instantes de tiempo, transferir un modelo de un curso a otro, etc., para obtener resultados sobre las preguntas de investigación definidas. Durante el desarrollo de la tesis, esta fase se repite varias veces para realizar experimentos con cada contexto educativo.
5. Evaluación: esta fase consiste en la evaluación de los resultados obtenidos mediante los diferentes experimentos y la obtención de conclusiones educativas. En primera instancia, se analiza cada contexto educativo por separado y se obtienen unas conclusiones para dicho contexto. Finalmente, se toman todas las conclusiones para obtener conclusiones globales para presentar en la tesis. Tras realizar esta fase, es posible que se identifiquen posibles mejoras en los modelos, para lo cual se puede volver a la fase de diseño para el refinamiento del diseño y posterior reevaluación, en un proceso iterativo.
6. Comunicación: esta fase consiste en la comunicación de los resultados. Dicha comunicación se ha realizado en revistas y conferencias fundamentalmente.

3.2. Recolección de datos

El análisis experimental de esta tesis requiere el análisis de datos educativos de diferentes contextos. Para poder analizar esos datos, hay que realizar un proceso de recolección de datos y transformación de los mismos. Esta sección explica el proceso de recogida de datos y cómo es el formato de entrada de cada uno de ellos.

Para esta tesis, se utilizan datos procedentes de MOOCs, SPOCs y de una plataforma educativa propietaria. Aunque el detalle de cada contexto se presenta en el capítulo 5, a continuación se indica de forma general cuál es la procedencia de los datos.

- MOOC sobre programación en Java albergado en edX (desarrollado por la Universidad Carlos III de Madrid (UC3M)).
- MOOC sobre programación en Java albergado en edX (desarrollado por Hong Kong University of Science and Technology (HKUST)).

- Tres MOOCs albergados en Coursera (desarrollados por la Pontificia Universidad Católica de Chile (PUC)).
- SPOCs de preparación a las pruebas de acceso a la universidad en Flandes, albergados en edX Edge (desarrollados por Katholieke Universiteit te Leuven (KU Leuven)).
- SPOCs de apoyo a las clases presenciales universitarias, albergados en Open edX (desarrollados por la UC3M).
- Datos de alumnos de primaria y secundaria que interactúan en una plataforma educativa propietaria (datos omitidos por confidencialidad).

Estos datos se han recogido de diversas instituciones, como se ha mencionado en el listado anterior, aunque de cara al análisis, solo es necesario considerar tres formatos de datos, ya que varias de ellas comparten el mismo formato. En primer lugar, se considera el formato de la plataforma Open edX. Este formato se utiliza tanto para MOOCs como SPOCs de edX y sus variantes (Open edX y edX Edge). Además, se considera el formato de los datos de Coursera para los MOOCs de PUC. Por último, la plataforma propietaria utiliza el formato de traza Marsupial [90], que es un formato utilizado para comunicar los LMSs con los contenidos proporcionados por las editoriales de los libros digitales.

Respecto al caso de Open edX, la plataforma utiliza un formato común para todos sus cursos, independientemente de si están alojados en la instancia oficial de MOOCs, edX, de si están en la versión para SPOCs, edX Edge, o de si se ha utilizado una instancia del software de Open edX. Este hecho es bastante provechoso ya que permite reutilizar gran parte del código de limpieza de datos y obtener los indicadores entre un contexto y otro con el mismo formato. Además, para comprender cómo son los datos, edX proporciona una detallada guía al investigador [223] que indica los diferentes campos y eventos utilizados.

De cara al análisis, se debe tener en cuenta que edX proporciona dos versiones de datos: las exportaciones de la base de datos y el registro de eventos (conocido como *clickstream*). En cuanto a las exportaciones de la base de datos, contienen información sobre la estructura del curso y sobre el estado del alumno en los diferentes elementos. Estos datos son proporcionados en un fichero denominado `{org}-{date}.zip`, siendo *org* el nombre de la organización en edX (p. ej., UC3M se denomina UC3Mx) y *date* la fecha a la que están referidos los datos. Dentro de este fichero comprimido, existen varios ficheros de datos, de los cuales se utiliza el siguiente subconjunto, que se describe a continuación [223]. Nótese, que en la nomenclatura de los ficheros, *course* indica id del curso, *run* la edición (p. ej., 2º trimestre de 2014) y *site* el lugar donde se alberga el curso (toma valor “*prod*” para edX y “*edge*” para edX Edge).

- `{org}-{course}-{run}-course_structure-{site}-analytics.json`: contiene la estructura del curso, incluyendo todos sus elementos de evaluación. Este fichero es utilizado para conocer el número de vídeos y ejercicios del curso. Asimismo, proporciona información sobre la apertura de los materiales del curso, lo que sirve para controlar en qué momento los estudiantes deben interactuar con cada una de las partes del curso.

- `{org}-{course}-{run}-courseware_studentmodule-{site}-analytics.sql`: contiene el estado de los estudiantes en cada elemento del curso. Por estado, se entiende a cómo se dejó un elemento la última vez que se interactuó con él. Por ejemplo, en un vídeo, indica la posición del cursor de reproducción y de un ejercicio, indica el número de intentos, nota obtenida en el ejercicio, entre otros aspectos. Respecto a vídeos, permite saber cuántos ejercicios se han abierto, ya que si no se han abierto, no aparece una entrada en la base de datos, pero no se puede saber cuánto se ha reproducido. Del mismo modo, en cuanto a ejercicios, se puede saber cuál es la nota del ejercicio y el número de intentos, pero no la evolución en cada uno de los intentos.
- `{org}-{course}-{run}-{site}.mongo`: contiene los mensajes publicados en el foro del curso y la información sobre las interacciones en el foro.

Toda esta información proporcionada por el fichero permite conocer múltiples aspectos de las interacciones de los alumnos en la plataforma. Sin embargo, es limitado, ya que al conocer la última interacción no se conoce todo el proceso y de algunos aspectos, como los vídeos, no se puede tener un detalle sobre el visionado. Para tener detalles más finos sobre la actividad de los alumnos, existe el registro de eventos, que guarda una entrada para cada acción (*click*) que realiza el estudiante en la plataforma. Este registro, se almacena en un fichero con formato `{org}-{size}-events-{date}.log.gz.gpg`.

Entre los eventos proporcionados, para el análisis se considera un subconjunto que permita obtener variables de más alto nivel sobre las interacciones de los estudiantes. En particular, se han considerado seis tipos de eventos: (1) *problem_check*, (2) *problem_show*, (3) *play_video*, (4) *pause_video*, (5) *seek_video*, y (6) *stop_video*. Los dos primeros sirven para obtener información sobre las interacciones con los ejercicios. En particular, *problem_check* indica el resultado de cada ejercicio una vez que el alumno envía su solución y guarda el número de intento, y *problem_show* indica cuándo el estudiante pide ver la explicación de la solución para un ejercicio concreto. Los eventos (3) a (6) indican, respectivamente, cuando un alumno comienza a reproducir un vídeo, lo pausa, lo avanza o rebobina, o lo para. Estos indicadores sirven para poder obtener información sobre cuántos vídeos se han visualizado, qué porcentaje se ha visualizado de cada vídeo y cuántas veces se repite cada vídeo, entre otros.

Además de la exportación de la base de datos y el registro de eventos, también existe el informe de notas (fichero `{course_id}-grade_report-{datetime}.csv`), que contiene las notas del alumno en cada elemento calificable del curso.

Sin embargo, debido a políticas o acuerdos de datos con edX o políticas de datos en la institución, no siempre es posible acceder a todos estos ficheros, lo que puede limitar las variables que se pueden obtener en cada contexto. En el capítulo 5 se especifica qué información se dispone para cada uno de los contextos, teniendo en cuenta las posibles limitaciones.

Respecto a Coursera, los datos contienen el registro de varios eventos, incluyendo: (1) comenzar una sesión, (2) comenzar a reproducir un vídeo, (3) completar un vídeo, (4) revisar un vídeo, (5) intentar una tarea de evaluación, (6) completar una tarea de evaluación), (7) revisar una tarea de evaluación y (8) terminar una sesión. Estos eventos sirven para obtener información de la actividad e interacción con vídeos y

ejercicios. Además, estos datos sirvieron para obtener información sobre patrones de SRL.

Sin embargo, en este caso, los datos utilizados no fueron los datos en crudo de Coursera, sino unos datos preprocesados por el equipo de PUC. En particular, desde el equipo de PUC, se obtuvieron patrones de SRL basados en eventos a partir de los registros de Coursera, que se pudieron combinar con otras variables de actividad e interacciones con vídeos y ejercicios. Estos patrones fueron obtenidos mediante técnicas de minería de procesos utilizando el método PM², junto con clústering jerárquico basado en el método de Ward [224]. En particular se obtuvieron seis patrones, que se detallan cuando se explican las variables independientes asociadas a ellos, en la sección 3.3.1.

Por último, respecto a la última plataforma educativa propietaria, los datos se encontraban con el formato de traza Marsupial [90]. Marsupial define su propio formato sobre cómo se deben almacenar las interacciones de los estudiantes con los ejercicios. En este contexto, solo se dispone de datos de interacciones con ejercicios y para el análisis se utilizan algunos de los campos almacenados en la traza Marsupial. Respecto al tiempo, se conoce la marca de tiempo (*timestamp*) de la hora de inicio y la duración en segundos. Respecto a la calificación, la traza Marsupial proporciona información sobre la nota obtenida y la máxima y mínima nota posible del ejercicio para poder escalar la nota si fuera necesario.

3.3. Medidas

En la sección anterior, se explicó el proceso de recogida de datos y el formato de los datos en crudo. Sin embargo, para el análisis predictivo, es necesario habitualmente transformar esos datos iniciales en variables de más alto nivel que por ejemplo aporten un significado educativo. Además, es necesario establecer cómo se va a medir aquello que se quiere predecir con los modelos, ya que eso tiene un efecto importante en el modelo. Por ello, esta sección detalla cuáles son las variables independientes y dependientes utilizadas. En esta sección, se realiza una presentación de variables de forma general y posteriormente, en el capítulo 5 se indica qué variables se han utilizado específicamente en cada escenario educativo, de acuerdo a las características del curso, los datos disponibles en cada contexto y los objetivos específicos.

3.3.1. Presentación de las variables independientes

En esta subsección, se indican las variables independientes utilizadas en los análisis predictivos a lo largo de la tesis.

A. Variables sobre interacciones con vídeos

Estas variables proporcionan una información global sobre el grado de interacción del alumno con los vídeos de la plataforma. Variables iguales o similares a las propuestas han sido utilizadas en trabajos previos de la literatura (p. ej., [47], [125], [218]). A continuación, se enumeran las variables utilizadas.

1. *perc_vabiertos*: indica el porcentaje de vídeos que ha abierto el alumno sobre el total de vídeos del curso. Un vídeo se considera abierto cuando el alumno

ha comenzado a visualizarlo (independientemente de que lo haya completado o no).

2. *perc_vtotal*: indica el porcentaje visualizado respecto al total de contenido, sin contar repeticiones de un mismo fragmento de vídeo. Por ejemplo, si el conjunto de vídeos tienen una duración de 10 horas y el alumno visualiza 7 horas (de contenido diferente), este indicador tendrá un valor de 0.7.
3. *perc_compl*: indica el porcentaje de vídeos que el alumno ha completado de visualizar respecto al total de vídeos del curso.
4. *avg_rep*: indica la relación entre el tiempo invertido viendo vídeos en la plataforma y el tiempo total de contenido diferente visto. Por ejemplo, si un alumno visualiza 3 horas de contenido, pero solo 2 horas de contenido diferente, eso implica que repite $3/2 = 1.5$ veces el contenido.
5. *avg_pausas*: indica el número medio de veces que un alumno pausa el vídeo durante su reproducción, teniendo en cuenta los vídeos abiertos. Por ejemplo, si el usuario abre dos vídeos y realiza 10 pausas entre los dos vídeos, este indicador valdría 5.
6. *vl_complete*: indica el número de veces que un alumno ha completado un vídeo en su totalidad por primera vez. Según esta definición, un vídeo solo puede contar una vez por cada alumno, a pesar de que se visualice varias veces. Por tanto, el valor máximo es el número de vídeos.
7. *vl_begin*: indica el número de veces que un alumno ha comenzado a ver vídeos sin terminarlos. Para que un evento cuente en este caso, el vídeo no debe haberse completado con anterioridad. Para esta variable, no hay un valor máximo, ya que un alumno puede iniciar vídeos sin terminarlos un número indefinido de veces.
8. *vl_review*: indica el número de veces que un alumno ha revisado vídeos, una vez que los ha completado previamente. Para esta variable, no hay un valor máximo, ya que un alumno puede revisar los vídeos un número indefinido de veces.
9. *perc_vlreview*: indica el porcentaje de vídeos que el alumno ha revisado una vez los ha completado.

B. Variables sobre interacciones con ejercicios

El segundo bloque de variables se refiere a la actividad que realizan los estudiantes con los ejercicios. Este tipo de variables pueden indicar principalmente dos aspectos: (1) actividad en los ejercicios (p. ej., cuántos ejercicios se hacen, con cuántos intentos, etc.) y (2) rendimiento, es decir, cómo de bien se realizan los ejercicios. Este tipo de variables son muy utilizadas en la literatura, y variables similares o iguales han sido utilizadas en trabajos previos (p. ej., [17], [23], [124], [170], [218], [224], [225]). A continuación, se presentan las variables utilizadas.

1. *nota_media*: indica la nota media del alumno en los ejercicios de la plataforma, teniendo en cuenta todos los ejercicios y todos los intentos. Por tanto, los ejercicios no intentados cuentan como 0.
2. *media_intentados*: es la nota media del alumno en los ejercicios intentados, considerando todos los intentos.
3. *media_primer*: es la nota media de los ejercicios de la plataforma, teniendo en cuenta todos los ejercicios y únicamente el resultado del primer intento. Los ejercicios no intentados cuentan como 0.
4. *media_intentados_pr*: es la nota media de los ejercicios intentados en la plataforma, utilizando únicamente el resultado del primer intento.
5. *perc_ej_abiertos*: indica el porcentaje de ejercicios a los que el alumno ha accedido en la plataforma, independientemente de si ha llegado a enviar una solución o no.
6. *perc_intentados*: indica el porcentaje de ejercicios que se han intentado sobre el total de ejercicios disponibles.
7. *avg_intentos*: indica el número medio de intentos utilizados en los ejercicios intentados.
8. *perc_correctos*: indica el porcentaje de ejercicios que el alumno ha conseguido resolver de forma correcta (al 100 %) sobre los ejercicios intentados. Para esta variable, se contabilizan todos los intentos.
9. *perc_correctos_total*: indica el porcentaje de ejercicios que el alumno ha conseguido resolver de forma correcta (al 100 %) sobre el total de ejercicios disponibles. Para esta variable, se contabilizan todos los intentos.
10. *CFA*: indica el porcentaje de ejercicios que el alumno ha conseguido resolver de forma correcta (al 100 %) en el primer intento, teniendo en cuenta todos los ejercicios.
11. *racha_ej*: indica la racha más larga de ejercicios consecutivos correctos (al 100 %) que el alumno ha conseguido.
12. *nshow*: indica el número de veces que el alumno pide ver la solución de un ejercicio (evento *problem_show* de Open edX).
13. *tiempo_medio*: indica el tiempo medio invertido en resolver cada ejercicio, de entre los intentados. Este indicador se calcula como el ratio entre el tiempo total invertido en todos los ejercicios en la plataforma y el número de recursos diferentes, por lo que se contabilizan todos los intentos.
14. *naccesos*: indica el número de veces que un alumno ha enviado ejercicios para su corrección. Para este indicador, se contabilizan todos los intentos.
15. *a_try*: número de veces que un alumno ha comenzado a realizar un ejercicio sin finalizarlo o aprobarlo. Para esta variable, no hay valor máximo, ya que un ejercicio se puede comenzar sin acabar un número indefinido de veces.

16. *a_complete*: número de veces que un alumno completa y aprueba un ejercicio por primera vez. Para esta variable, el rango es de 0 al número de ejercicios.
17. *a_CFA*: indica el número de veces que un alumno completa correctamente al 100 % un ejercicio en el primer intento (CFA). Para esta variable, el rango es de 0 al número de ejercicios.
18. *a_review*: número de veces que un alumno revisa un ejercicio que ha completado correctamente con anterioridad. Para esta variable, no hay valor máximo, ya que un ejercicio se puede revisar un número indefinido de veces.
19. *perc_acomplete*: porcentaje de ejercicios completados y aprobados, respecto al total de ejercicios.
20. *perc_areview*: porcentaje de ejercicios revisados una vez completados, respecto al total de ejercicios.

C. Variables sobre actividad en la plataforma

El tercer bloque de variables se refiere a la actividad del alumno en el sentido de cuántas veces o de qué manera el alumno accede a la plataforma educativa. Este tipo de variables también son frecuentes en la literatura y las variables propuestas para el análisis han sido utilizadas en trabajos previos de la literatura (p. ej., [23], [150], [226], [227], [228]). En particular, la lista siguiente muestra las variables utilizadas dentro de esta categoría.

1. *ndias*: número de días diferentes que el estudiante ha accedido a la plataforma educativa.
2. *perc_dias*: porcentaje de días que el estudiante ha accedido a la plataforma educativa, respecto a la duración total del curso o al periodo concreto para el que se calcule la variable. Por ejemplo, en un curso de cuatro semanas, si se añaden datos en el análisis temporal, la primera semana sería el número de días accedidos entre siete, la segunda semana entre 14, y al final de curso entre 28.
3. *avg_con*: número medio de días consecutivos que el estudiante accede a la plataforma educativa. Por ejemplo, si el alumno accede tres días seguidos, más tarde cinco días seguidos, y finalmente otros siete días seguidos, el valor de esta variable sería $(3 + 5 + 7)/3 = 5$ días.
4. *racha_acc*: racha más larga de días consecutivos que el alumno accede a la plataforma educativa. En el ejemplo de la variable anterior, la racha más larga sería de siete días.
5. *tiempo_invertido*: tiempo total que el alumno ha estado realizando actividades en la plataforma educativa.
6. *num_ses*: número de sesiones en las que el alumno ha trabajado en la plataforma educativa. Una sesión se define como el periodo de tiempo en que el alumno tiene una actividad continuada en la plataforma educativa y no se producen periodos de actividad superiores a 45 minutos.

7. *perc_pc*: porcentaje de los accesos del alumno que se han realizado desde un ordenador personal, y no desde otros dispositivos como móviles, tablets, etc.
8. *perc_wk*: porcentaje de los accesos del alumno que se realizan durante el fin de semana.
9. *perc_noche*: porcentaje de los accesos del alumno que se realizan durante la noche. Esta variable solo tiene sentido para aquellos cursos que se desarrollan en un área específica, y no para un MOOC, donde los alumnos pueden estar en diferentes zonas horarias. Como criterio, se ha tomado que la noche comprende desde las 20:00h hasta las 9:00 del día siguiente.

D. Variables sobre la actividad en el foro del curso

El siguiente bloque de variables recoge información sobre las interacciones de los estudiantes con el foro del curso. El foro del curso tiene una gran relevancia en los MOOCs, ya que es el medio preferido de comunicación [229] y donde ocurre la mayor parte de las interacciones entre estudiantes [230], lo que permite un análisis social. Los mensajes del foro, además pueden ser útiles para el análisis, ya que proporcionan una información bastante heterogénea, dado que permiten obtener información tanto sobre la participación del estudiante en el foro (p. ej., número de mensajes que publica), como del contenido de los mensajes. Por ejemplo, el análisis semántico del texto puede dar información sobre qué contenidos del curso presentan más dificultades o sobre el estado anímico del estudiante. Para facilitar el análisis de este tipo de variables, tanto en modelos predictivos, como en el análisis que pueda hacer un profesor sobre cómo interactúan sus estudiantes, en un trabajo previo definimos las tres dimensiones 3S, que son las tres categorías de datos que pueden analizarse en el foro [231]. A continuación, se presentan las tres categorías.

- **Social**: esta dimensión analiza las interacciones que un alumno realiza con otros en el foro (tanto en cantidad como en calidad), sin tener en cuenta el contenido de las mismas. Algunos aspectos considerados en esta dimensión (como los tomados para la predicción por Klüsener y Fortenbacher [129]) son: número de mensajes publicados, número de respuestas y votos positivos a los mensajes, y tiempo de respuesta entre el primer mensaje del hilo y la primera respuesta. Entre ellos, por ejemplo, el número de mensajes publicados sería un indicador de la cantidad y el número de votos positivos de la calidad.
- **Sentimientos**: esta dimensión analiza el estado anímico de los estudiantes, dado que estos pueden afectar al aprendizaje [232]. Para ello, se utilizan técnicas de minería de texto que permitan determinar la polaridad de los mensajes, es decir, si los mensajes son positivos, negativos o neutros.
- **Habilidades** (del inglés *Skills*, que empieza por S): esta dimensión analiza qué habilidades del curso son las que los alumnos tratan en los mensajes del foro (p. ej., sobre qué habilidades debaten o preguntan). El motivo de analizar estas habilidades es que es muy probable que las habilidades que más aparezcan en los mensajes sean aquellas que más dificultades generen a los estudiantes, dado que el uso principal del foro es hacer preguntas [233] (p. ej., cuando los estudiantes se atascan al realizar las tareas [234]).

Estas tres dimensiones pueden servir para analizar el foro de un curso, junto con un cuarto elemento, el tiempo, para ver cuándo se efectúan las interacciones. El análisis de estas interacciones puede dar una gran información al profesor sobre cómo interactúan sus estudiantes. Por ejemplo, a partir de estas dimensiones, se definió la metodología 3S [231], que a través de seis pasos, (1) selección de los intervalos de tiempo, (2-4), el análisis de las tres componentes para cada intervalo y la obtención de (5) conclusiones para cada intervalo y (6) globales, permiten analizar la interacción social. Esta metodología requiere la intervención del profesor para analizar cada uno de los pasos. Para el caso de la predicción, se adoptan las dimensiones 3S y las variables utilizadas en el trabajo previo [231] (que son variables comunes en la literatura y aparecen en otros trabajos, como [129], [195], [225], [235]), aunque éstas deben obtenerse para cada alumno (y no de forma agregada) para ser utilizadas en los modelos predictivos. Este hecho conlleva que la tercera dimensión (habilidad) no pueda obtenerse fácilmente para la predicción, dado que conocer las habilidades o conceptos que contiene un curso requiere un metadato del mismo. Por ello, para el análisis predictivo, se toman las variables de las dos primeras dimensiones (social y sentimientos), que pueden obtenerse a partir de las interacciones del foro y de minería de textos para la obtención de los sentimientos. Teniendo esto en cuenta, a continuación se muestra la lista de variables de foro utilizadas para la predicción (separadas en social y sentimientos).

1. *participations (social)*: indica el número de mensajes que el alumno ha publicado en el foro.
2. *com_threads (social)*: indica el número de hilos (lo que se conoce en Open edX como *CommentThread*) que el estudiante ha comenzado en el foro.
3. *comments (social)*: indica al número de respuestas a hilos (lo que se conoce en Open edX como *Comment*) que el alumno ha realizado.
4. *votos_rcv (social)*: indica el número de votos positivos que el alumno ha recibido en sus mensajes, procedentes de otros usuarios.
5. *votos_emt (social)*: indica el número de votos positivos que un alumno ha emitido hacia mensajes publicados por otros usuarios.
6. *endorsed (social)*: indica el número de mensajes (en particular respuestas) marcados por el instructor o el usuario que originó el hilo por ser mensajes de alto valor o relevancia.
7. *avg_longitud (social)*: indica el número de medio de caracteres de los mensajes publicados por el alumno.
8. *percentil (social)*: indica el percentil en el que se encuentra el alumno tras ordenar los estudiantes de acuerdo al número de mensajes publicados.
9. *sentimiento(sentimientos)*: indica la positividad media de los mensajes del foro. Tiene un valor entre 0 y 1, siendo 0 el valor más negativo (todos los mensajes del alumno son negativos) y 1 el más positivo (todos los mensajes del alumno son positivos). Este indicador se ha obtenido utilizando RF, con un conjunto de mensajes etiquetados manualmente, y tres indicadores: (1)

la longitud del mensaje, (2) la diferencia de días entre la fecha del mensaje y el inicio del curso y (3) una variable denominada *orientacion*. De entre estas variables, la última es la más importante y da una indicación similar entre 0 y 1 de la positividad del mensaje, basada en la positividad de las palabras del mensaje, de acuerdo con diccionarios de palabras positivas y negativas, compilados por Hu y Liu [236], y considerando posibles negaciones y el significado de emoticonos.

En particular, el algoritmo para calcular la orientación está detallado con pseudocódigo en un trabajo previo [237]. En este algoritmo, se parte de una orientación igual a 0 y se itera con cada palabra del mensaje (tras tokenizar el mensaje). En primer lugar, se comprueba si esa palabra es un emoticono, ya que estos normalmente expresan directamente la polaridad del mensaje (si es positivo o negativo) [238]. En dicho caso, la orientación del mensaje es directamente la del emoticono (p. ej., una cara sonriente :), implica que el mensaje es positivo). Si la palabra no es un emoticono, se busca en los diccionarios de palabras positivas y negativas. Si se encuentra, se aumenta o disminuye el valor de orientación en una unidad dependiendo del caso. Para determinar el caso, se mira si ha habido alguna palabra de negación en las últimas palabras (en este caso, se consideraron las últimas palabras, como en Hu y Liu [236]). Si no hay una palabra de negación cercana, las palabras positivas suman a la orientación una unidad y las negativas restas, y si no al contrario. Una vez aplicado este proceso a todas las palabras, se normaliza y ajusta el valor de orientación para que esté en el rango de 0 a 1, teniendo en cuenta el número de palabras que incrementaron o disminuyeron el valor. Por ejemplo, si hay tres palabras positivas y dos negativas (sin contar negaciones), la orientación sería $1+1+1-1-1=1$, que al normalizarse teniendo en cuenta las cinco palabras sería $1/5 = 0.2$. Siguiendo este procedimiento, se obtuvo un AUC superior a 0.7 para la clasificación de los mensajes, lo cual es aceptable, dado que esta clasificación tiene una componente subjetiva importante.

E. Variables sobre la autorregulación del aprendizaje

El siguiente bloque de variables incluye variables sobre SRL. Este grupo de variables es bastante relevante, ya que capacidad de SRL puede tener un impacto en el éxito de los estudiantes [96]. Para este bloque, se utilizan diferentes patrones relacionados con el SRL a partir de las trazas de los estudiantes en la plataforma educativa (en este caso, solo se han aplicado estas variables a MOOCs de Coursera). En particular, se consideran los siguientes patrones como variables independientes, procedentes de un trabajo previo de la literatura [224].

1. *only_vlecture*: patrón identificado cuando los alumnos solo interaccionan con vídeos.
2. *atry_to_vlecture*: patrón identificado cuando los alumnos abren la prueba de evaluación primero y luego buscan la información en los vídeos.
3. *explore*: patrón identificado cuando los alumnos inspeccionan los contenidos de forma superficial sin intención de completarlos.

4. *only_assessment*: patrón identificado cuando los alumnos solo interactúan con las pruebas de evaluación.
5. *vlcomplete_to_atry*: patrón identificado cuando los alumnos interactúan primero con vídeos y luego comienzan una prueba de evaluación (pero no la completan o aprueban).
6. *vlecture_to_acomplete*: patrón identificado cuando los alumnos interactúan con vídeos y luego comienzan una prueba de evaluación (y la completan satisfactoriamente, con nota aprobada).
7. *complex*: variable que incluye otros patrones no incluidos en los seis anteriores.

F. Variables de alto nivel sobre comportamientos del alumno

El siguiente bloque de variables incluye variables de más alto nivel, basadas en la transformación de otras variables respecto a ejercicios. Estas variables se han categorizado aparte ya que aportan información sobre comportamientos de los estudiantes o aspectos de su personalidad. En particular, se han seleccionado tres indicadores. A continuación se muestra la forma de calcular cada uno de ellos.

1. Eficiencia (*eficiencia*): esta variable indica cómo de rápido los estudiantes son capaces de conseguir los resultados. El cálculo de eficiencia utilizado, tomado de [114], es una modificación de una definición previa presentada en la literatura [47], aunque en dicha definición se considera el tiempo invertido o los intentos, pero no en relación con los resultados obtenidos, que son utilizados normalmente para medir la eficacia. En este caso, la eficiencia mide exactamente cuántos puntos consiguen los estudiantes por unidad de tiempo y se mide en puntos/minuto. La Fórmula (1) indica cómo se calcula exactamente el valor de la eficiencia.

$$Eficiencia = \frac{\sum_{i=1}^{n^{\circ} \text{ ejercicios_intentados}} \text{mejor_nota}_{ejercicio_i}}{\sum_{i=1}^{n^{\circ} \text{ ejercicios_intentados}} \text{tiempo_invertido}_{ejercicio_i}} \quad (1)$$

Esta fórmula suma las mejores notas de cada ejercicio intentado por el estudiante (considerando todos los intentos realizados) y divide el resultado por el tiempo total invertido en todos los ejercicios (suma del tiempo invertido en cada ejercicio). De este modo, la eficiencia mide si los estudiantes pueden incrementar su conocimiento de forma rápida en tiempo.

2. Constancia (*constancia*): esta variable indica en qué medida los estudiantes dedican un tiempo similar a realizar ejercicios en la plataforma cada día. La relevancia de esta variable ya fue indicada previamente en la literatura, por ejemplo en [116], y una definición fue presentada en [47]. Para los análisis de esta tesis, la constancia se toma específicamente como la inversa de la desviación típica del tiempo invertido diario realizando las actividades en la plataforma, tal y como se presentó en un trabajo previo [114]. Esta medida es similar a la propuesta en [47], aunque en ese trabajo se utilizaba la varianza en vez de la desviación típica, y en [114] se prefiere la desviación típica para que el resultado sea medible en unidades de tiempo (en lugar de unidades de

tiempo al cuadrado). Para el cálculo, se han utilizado únicamente los días en los que se ha interactuado dada las condiciones del contexto (en el contexto en el que se utiliza los estudiantes no necesariamente interactúan diariamente o semanalmente), aunque en otro contexto podrían utilizarse todos los días. Según esta definición, un alumno es más constante cuando su desviación del tiempo invertido es menor, por lo que para facilitar el entendimiento de la variable, se utiliza la inversa de la constancia ($constancia^{-1}$), que es la desviación típica y está medida en unidades de tiempo (p. ej., min en lugar de min^{-1} , que tendría la constancia). La Fórmula (2) indica el cálculo específico para obtener el valor de la constancia (en particular, su inverso).

$$Constancia^{-1} = SD[tiempo_invertido] = \sqrt{\frac{\sum_{i=1}^{dias_activo} (tiempo_invertido_{dia_i} - tiempo_invertido_{dia})^2}{dias_activo - 1}} \quad (2)$$

3. Persistencia (*persistencia*): esta variable, definida en un trabajo previo [80], indica la habilidad del estudiante de continuar realizando los ejercicios una vez resueltos de forma incorrecta. Para medir esta variable, se tiene en cuenta las interacciones de los alumnos con los ejercicios y, en particular, la secuencia de intentos de los ejercicios. Por ejemplo, una secuencia en la que el alumno lo hace mal (0) y no lo vuelve a intentar sería la de un alumno no persistente. Una secuencia (0-1), hacer el ejercicio mal una vez y luego bien, demostraría persistencia del alumno, aunque no tanta como una secuencia en el que el alumno resuelve bien el ejercicio tras cinco intentos fallidos (0-0-0-0-0-1). Sin embargo, en ambos casos, el estudiante tendría la persistencia máxima, ya que consigue resolver el ejercicio. Del mismo modo en una secuencia de dejar el ejercicio tras fallarlo cuatro veces (0-0-0-0), el alumno no es del todo persistente porque no acierta el ejercicio, pero es más persistente que si lo deja tras fallarlo la primera vez. Teniendo esto en cuenta, se propone la Fórmula (3) para calcular la persistencia a nivel de estudiante. Es decir, esta fórmula calcula la persistencia de un estudiante entre 0 y 1 a partir de todas las interacciones de sus ejercicios y no una persistencia para cada ejercicio.

$$Persistencia = \frac{\sum_{i=1}^n \min(intentos_i - 1, stop)}{\sum_{i=1}^n \min(intentos_i - 1, stop) + penalty \cdot (1 - nota_i)} \quad (3)$$

Los elementos que intervienen en esta fórmula son los siguientes.

- n : indica el número de ejercicios que el alumno ha intentado.
- i : representa un ejercicio en particular que el estudiante ha intentado.
- $intentos_i$: indica el número total de veces que el estudiante ha intentado el ejercicio i . Esta fórmula tiene el término $intentos_i - 1$ para excluir el primer intento, ya que no se muestra persistencia nunca en el primer intento (ya que el alumno no tiene que decidir si continuar o no tras las dificultades).
- $nota_i$: valor binario que indica si el estudiante consiguió la respuesta correcta del ejercicio ($nota=1$, que es la máxima nota) o no ($nota=0$). Para los ejercicios que admitan notas parciales, (p. ej., ejercicios de casillas de verificación), la nota es 1 si el ejercicio está totalmente correcto y 0 en caso contrario.

- *penalty*: variable para penalizar cuando los estudiantes no alcanzan la nota correcta del ejercicio. De este modo, cuando el ejercicio es correcto, el numerador no varía, pero el denominador se incrementa para disminuir el valor global de la persistencia. Cuando el ejercicio es incorrecto, como $\text{nota}=0$, no hay penalización. El valor de penalización se puede ajustar dependiendo del contexto, aunque conviene que no sea ni muy pequeño para que los ejercicios donde el alumno no sea persistente tengan una penalización suficiente, ni muy grande para que sea posible compensar dichos ejercicios.
- *stop*: representa el máximo número de intentos que pueden sumarse por cada ejercicio. Se usa para evitar que un ejercicio con muchos intentos tenga un alto peso en la fórmula que pueda hacer que la persistencia sea alta incluso cuando el estudiante casi nunca intenta los ejercicios que resuelve de forma incorrecta.

Para ilustrar el uso de esa fórmula, la Tabla 3.1 presenta los valores de la persistencia acumulada para un ejemplo en el que un alumno realiza cuatro ejercicios secuencialmente (ID1, ID2, ID3 e ID4) y tiene una secuencia de intentos para cada ejercicio (y consecuentemente una nota y un número de intentos). Teniendo esto en cuenta, la persistencia se actualiza después de cada ejercicio, de modo que la persistencia en la primera fila es la obtenida cuando se ha realizado el primer ejercicio, la de la segunda fila cuando se han realizado los dos primeros y así sucesivamente.

TABLA 3.1
EJEMPLO DE CÁLCULO DE LA PERSISTENCIA

ID	Secuencia	Intentos	Nota	Persistencia (acumulada)
1	0	1	0	$0/4 = 0$
2	0 1	2	1	$(0+1)/(4+1) = 1/5 = 0.2$
3	0 0 0 0 0 1	6	1	$(0+1+5)/(4+1+5) = 6/10 = 0.6$
4	0 0 0 0	4	0	$(0+1+5+3)/(4+1+5+7) = 9/17 = 0.53$

G. Variables sobre la tipología del curso

El último bloque de variables son sobre las características del curso. Hasta ahora, todas las variables están centradas a nivel de alumno. Sin embargo, en caso de querer realizar un modelo predictivo a partir de datos de diferentes cursos, puede ser relevante incluir variables sobre las características del curso (como en [132]), de modo que el modelo tenga en cuenta que las variables del alumno puedan ser diferentes dependiendo del contexto. En este caso, para un mismo curso, estas variables son fijas para todos los alumnos, pero varían de un curso a otro. Este tipo de variables se utilizan para ver cómo afectan a la predicción de múltiples SPOCs. A continuación, se indica la lista de variables sobre la tipología del curso utilizadas.

1. *area*: indica el área temática del curso, de acuerdo con las categorías establecidas por Wu et al. [239]: (1) humanidades, (2) ciencias sociales, (3) ciencias naturales, (4) ciencias formales, y (5) profesiones y ciencias aplicadas.

2. *sincrono*: variable binaria que indica si la apertura de materiales sigue el esquema de un modo de instrucción síncrono (1) o asíncrono (0). No obstante, al aplicarse esta variable en SPOCs, hay que tener en cuenta que el curso en sí puede ser síncrono en el sentido que el profesor va impartiendo un temario cada semana a todos los estudiantes, aunque los contenidos pueden estar abiertos desde el principio.
3. *nejercicios*: indica el número de ejercicios que tiene el curso.
4. *nvideos*: indica el número de vídeos que tiene el curso.
5. *avg_duracion*: indica la duración media de los vídeos (en segundos).
6. *num_personal*: indica el número de personal (principalmente profesores) que están a cargo del curso.
7. *ingles*: variable binaria que indica si el curso se imparte en inglés (1) o en español (0).

3.3.2. Presentación de las variables dependientes

Los modelos predictivos, aparte de tener las variables independientes, vistas en la sección anterior, necesitan definir una variable dependiente, que es la variable que se va a predecir. En la literatura, las variables más comunes en la predicción son las relacionadas con el abandono y con el rendimiento académico [16]. El posible motivo es que las tasas de fracaso académico y abandono pueden ser muy altas, tanto a nivel de curso como a nivel global de titulación [15]. Teniendo en cuenta la importancia de las variables mencionadas, en esta tesis se propone realizar la predicción de tres variables, que se enumeran a continuación.

1. Notas. Es una variable continua (normalmente en rango 0-1) que indica el rendimiento del alumno en algún elemento calificable. El elemento calificable a analizar dependerá del contexto y se especificará para cada caso en el capítulo 5, aunque los posibles elementos incluyen: nota media de los ejercicios del curso, nota de una prueba de evaluación, nota final del curso (como suma ponderada de pruebas sumativas) y nota de un examen final.
2. Éxito. Es una variable binaria que indica si el alumno ha conseguido superar un curso o prueba oficial, o alguna parte de los mismos (p. ej., superar el examen final de un curso).
3. Abandono (conocido por su término en inglés, *dropout*). Es una variable binaria que indica si el alumno ha dejado de trabajar en el curso y no lo ha completado. Esta variable difiere de la anterior (éxito), aunque en algunos casos se mezclen (p. ej., [240]), en que puede haber un alumno que realice todas las tareas del curso (completa el curso), pero las realice incorrectamente y no sea exitoso. Por tanto, esta variable se centra en que el alumno tenga actividad en el curso y realice las tareas y no en el rendimiento de las mismas. Para definir específicamente el abandono, puede haber varias posibilidades, tal y como se argumentó en [241]. En este trabajo, se considerarán principalmente dos aproximaciones: (1) abandono relacionado

con la falta de actividad y (2) abandono relacionado con la completitud del curso. En la primera aproximación, se considera que un alumno ha abandonado cuando presenta un periodo considerable de inactividad en el curso, es decir, el alumno no accede a la plataforma educativa en un periodo suficientemente grande. En la segunda aproximación, un estudiante abandona cuando no completa la mayor parte de los ejercicios o contenidos evaluables del curso, sin importar en el momento en el que se completen. Esta aproximación difiere de la primera en que esta definición se centra en completar actividades (independientemente del periodo), mientras que la primera se centra en acceder al curso (independientemente de si se hacen las actividades o no).

Estas tres variables han sido tratadas en trabajos de la literatura (ver ejemplos de cada una de ellas en la sección 2.4.2). No obstante, es posible analizar en más detalle los factores que afectan a la predicción, ya que puede haber diferencias debido al contexto. De hecho, la propia definición debe adaptarse al contexto. Por ejemplo, la tasa para superar un curso (tener éxito) puede diferir de un curso a otro. Del mismo modo, el abandono puede variar de un curso a otro, y en unos puede tener más sentido utilizar una aproximación relacionada con actividad y en otras con completitud. E incluso tomando la misma aproximación, puede haber diferencias, ya que, por ejemplo, el porcentaje necesario para considerarse un curso completo puede variar. Por ejemplo, puede haber un curso que requiera realizar todas las actividades (en cuyo caso un alumno abandona si no realiza el 100 % de las mismas) y otro en el cual se puede aprobar realizando un porcentaje menor, por lo que se pueda considerar la completitud con un porcentaje menor de actividades completadas. Debido a estas diferencias, la definición específica de cada variable dependiente dependerá del contexto (como se argumentará en el capítulo 5) y en cada contexto puede haber diferentes factores que afecten a la predicción.

3.4. Métodos analíticos y herramientas

Para el desarrollo de los análisis presentados en esta tesis doctoral, se han utilizado métodos descriptivos y métodos predictivos. El núcleo de la tesis se basa en el uso de los modelos predictivos para predecir las variables dependientes mostrada en la subsección anterior, aunque en algunos casos también se utiliza analítica descriptiva para conocer valores agregados de alguna variable o analizar la relación entre variables, de modo que se pueda comprender mejor cómo estas variables pueden estar relacionadas para la predicción.

Para realizar estos análisis, se han utilizado dos herramientas (lenguajes de programación), siendo todas ellas *software* libre: R y Python (utilizando la distribución Anaconda). Para la transformación de los datos, se han utilizado diferentes librerías, dependiendo del formato de origen de los datos. Los datos proporcionados, dependiendo del contexto, estaban almacenados en una base de datos relacional (MySQL) o en ficheros de texto con información en formatos Comma-Separated Values (CSV) o JavaScript Object Notation (JSON). Para el caso de Python, se ha trabajado con formatos CSV y JSON, por lo que se han utilizado las librerías `csv` y `json`. Adicionalmente, para el análisis de datos, se han utilizado las librerías `numpy` y `pandas`. En cuanto a R, se ha trabajado con MySQL y con ficheros en formato CSV, para lo cual se ha utilizado la librería `RMySQL` (las

funcionalidades para `CSV` vienen en la librería `utils`, que aparece por defecto en R). Para la limpieza y procesamiento de los datos, se ha utilizado la librería `dplyr`.

Respecto a las técnicas de analítica descriptiva, se han utilizado, aparte de estadísticos descriptivos básicos (p. ej., media, mediana, desviación típica, etc.), correlaciones (utilizando el coeficiente de correlación de Pearson), técnicas de comparación de grupos (tales como la prueba t de Student, t-test, o el test de Mann-Whitney), minería de texto (para los mensajes del foro, usando la librería `nlTK` de Python) y técnicas de agrupamiento o clústering (en particular clústering jerárquico). La mayoría de estos análisis se han realizado utilizando la librería `stat` de R.

En cuanto a los modelos predictivos, se han utilizado las librerías `sklearn` en Python y `caret` en R. Los algoritmos de ML (aprendizaje supervisado) que se han utilizado durante en los análisis predictivos son los siguientes: (1) regresión (RG), ya sea regresión lineal o un modelo de regresión generalizado (`Generalized Linear Model (GLM)`), como regresión logística para clasificación, (2) `Random Forest (RF)`, (3) `Support Vector Machine (SVM)` y (4) `Decision Tree (DT)`. Estos algoritmos se han seleccionado al ser los más comunes, de acuerdo a varios análisis sistemáticos de la literatura sobre la predicción ([16], [188]). Para la aplicación de los algoritmos, en todos los casos se ha utilizado una validación cruzada (`Cross-Validation (CV)`) 10-fold para entrenar los modelos y seleccionar los hiperparámetros.

En cuanto a las métricas para evaluar los modelos predictivos, se han utilizado dos, dependiendo de si el problema a predecir es de regresión (predicción de una variable continua) o de clasificación (predicción de una variable discreta, binaria en este caso). Para la predicción de variables continuas, se utiliza como métrica la raíz del error cuadrático medio (`RMSE`), como sugiere Pelánek [202], que prefiere esta métrica respecto a otras como el error absoluto medio (`MAE`), ya que esta métrica está sesgada hacia el resultado mayoritario. Para los problemas de clasificación, la métrica utilizada es el `AUC`. Esta métrica es ampliamente utilizada y generalmente apropiada para problemas de clasificación que involucran comportamientos de estudiantes [202]. Además, evita algunos problemas que tienen otras métricas cuando se tienen conjuntos de datos no balanceados [204] (por ejemplo, la exactitud puede ser muy alta incluso cuando el modelo no es bueno si los datos están balanceados). Teniendo esto en cuenta, la Tabla 3.2 resume las métricas utilizadas para cada variable dependiente que se utiliza en la tesis doctoral.

TABLA 3.2
RESUMEN DE LAS MÉTRICAS UTILIZADAS EN LOS ANÁLISIS PREDICTIVOS

Variable dependiente	AUC	RMSE
Nota (media, parcial o final)		X
Éxito	X	
Abandono	X	

A la hora de interpretar estas métricas, hay que tener en cuenta que el `AUC` tiene un rango de 0 a 1, pero el `RMSE` tiene el mismo rango que la variable que se predice. Por ejemplo, si se predice una nota de 0 a 10, el `RMSE` puede ir de 0 a 10. Para facilitar la comprensión de los valores a lo largo de la tesis y poderlos comparar mejor, en todos los casos en los que se use `RMSE`, se escala el resultado de 0 a 1 para poder proporcionar todos los valores en el mismo rango.

Por último, cabe mencionar que para visualizar los resultados obtenidos en los diferentes experimentos (ya sean diagramas de líneas, de dispersión, de caja, etc.) se han utilizado los sistemas gráficos `base` y `ggplot2` de R.

4

Necesidad de la predicción en educación

Contenido

4.1. Estado de adopción de la analítica del aprendizaje y la predicción	53
4.2. Análisis de la necesidad del uso de la predicción	56
4.2.1. Opinión de los líderes institucionales y expertos	57
4.2.2. Opinión de los profesores	58
4.2.3. Opinión de los estudiantes	59
4.2.4. Conclusiones globales	60

Esta tesis doctoral trata sobre la predicción de variables (relacionadas con el abandono y rendimiento) de los estudiantes, con el objetivo de que los modelos diseñados sean de utilidad en los entornos educativos y se puedan extraer conclusiones para su diseño. Sin embargo, para que los modelos predictivos sean efectivos, es importante que los interesados reconozcan su importancia y estén dispuestos a usarlos. Por ello, este capítulo trata de analizar y justificar la necesidad de realizar modelos predictivos en el contexto de la educación superior, a través de la realimentación de diferentes actores, en concreto de líderes institucionales, expertos, profesores y estudiantes. Además, se presenta el estado de adopción de LA en instituciones de educación superior en Europa para poner las bases de partida sobre qué hay ya implementado y puesto en práctica, y conectarlo con la predicción.

4.1. Estado de adopción de la analítica del aprendizaje y la predicción

A lo largo de la tesis, se ha ido comentando sobre el uso de LA y la predicción como un caso particular de LA. Entre los objetivos de la predicción, está la mejora de los procesos educativos y de la enseñanza. A pesar de ello, en algunos casos no se ha podido mostrar una evidencia del impacto positivo de LA [242], [243], y

la investigación se centra en desarrollar sistemas (p. ej., modelos predictivos) que podrían tener un posible potencial para producir dicho impacto. Entre los motivos se encuentran que el estado de adopción de LA no ha llegado a un punto suficiente de madurez. A nivel global, desde un punto bibliométrico, Estados Unidos es el país con mayor investigación sobre LA, seguido de España y Reino Unido [244], aunque la mayoría de estudios son a pequeña escala y con carácter de investigación, con pocos ejemplos de análisis en un ámbito mayor (p. ej., de adopción global en instituciones). Entre dichos ejemplos, Colvin et al. [245] analizaron el caso australiano y concluyeron que el foco de LA ha sido el análisis del abandono y el entendimiento de los factores que afectan a los resultados del estudiante. Por otro lado, Arroway et al. [246] analizaron el caso estadounidense y vieron que la mayor parte de las instituciones usan LA para monitorizar el progreso de los estudiantes, pero que en pocos casos se han aplicado modelos predictivos. Esto es coherente con el modelo de sofisticación de LA, propuesto por Siemens et al. [247], por el cual para llegar a los modelos predictivos hay que pasar por otras etapas previas de darse cuenta del potencial de LA, experimentar con los datos y realizar analítica descriptiva y visualizaciones.

Para profundizar en el estado de adopción de LA y conocer mejor las expectativas de los interesados, se realizaron encuestas, entrevistas y grupos focales, dentro del contexto europeo. Estas entrevistas se realizaron dentro del marco del proyecto *Supporting Higher Education to Integrate Learning Analytics (SHEILA)*¹. Para la realización de dichas entrevistas colaboraron cuatro instituciones de educación superior: The University of Edinburgh, Tallinn University, Open University of the Netherlands y UC3M. En particular, se realizó una encuesta a líderes institucionales, otra a profesores y otra a estudiantes. Adicionalmente, se llevaron a cabo 64 entrevistas en 51 instituciones de educación superior² en 16 países europeos a líderes institucionales o expertos en LA, 18 grupos focales a estudiantes de cuatro instituciones³ y 16 grupos focales a profesores de cuatro instituciones de educación superior⁴ [248]. De entre estas entrevistas y grupos focales, el autor de esta tesis contribuyó en parte de la transcripción, codificación y análisis cualitativo de las mismas. En esta tesis, como parte de la justificación del nivel de adopción, se toman resultados del proyecto SHEILA, en los que colaboró el autor, y el autor especialmente contribuye en la relación de los mismos con la predicción, que es el ámbito de estudio de esta tesis. Teniendo esto en cuenta, a continuación se muestran los resultados globales en torno a la adopción de LA, que permitan hacer una valoración sobre el estado actual de uso de la predicción a nivel institucional en el marco europeo.

Para analizar la adopción institucional, se tomaron las entrevistas a líderes institucionales y expertos. Los resultados muestran que a fecha de las entrevistas (entre agosto 2016 y febrero 2017), 21 de las 51 instituciones tenían proyectos de LA con un apoyo centralizado (ver Fig. 4.1). De entre ellos, solo nueve se habían implementado a nivel institucional, siete se habían realizado en un ámbito parcial (p. ej., dentro de un grupo de investigación determinado o mediante pilotos) y cinco estaban en una fase exploratoria y de limpieza de datos. Por otro lado, 18 instituciones estaban en una fase preparatoria para poder llevar a cabo proyectos

¹<https://sheilaproject.eu/>

²https://www.it.uc3m.es/pedmume/files/Institutional%20interview_questions2.pdf

³https://www.it.uc3m.es/pedmume/files/Student%20FG_questions2.pdf

⁴https://www.it.uc3m.es/pedmume/files/Staff%20FG_questions2.pdf

de LA y 12 no tenían ningún plan concreto para adoptar LA en el momento de la entrevista. Esto implica que en el momento de las entrevistas, muy pocas instituciones tenían incorporado el uso de LA de modo global y una gran parte (59%) no tenían nada implementado y simplemente estaban en preparación o no tenían siquiera planes futuros respecto a la LA. Esto muestra que el estado de adopción todavía está en una fase inicial y teniendo en cuenta el modelo de sofisticación [247], según el cual la fase de predicción es la última de ellas a nivel de adopción institucional, es posible que muy pocas instituciones hayan llegado a usar los modelos predictivos, lo que puede suponer una oportunidad para trabajar en ello, de cara a que las instituciones puedan utilizar las lecciones aprendidas cuando llegue el momento de implementación.

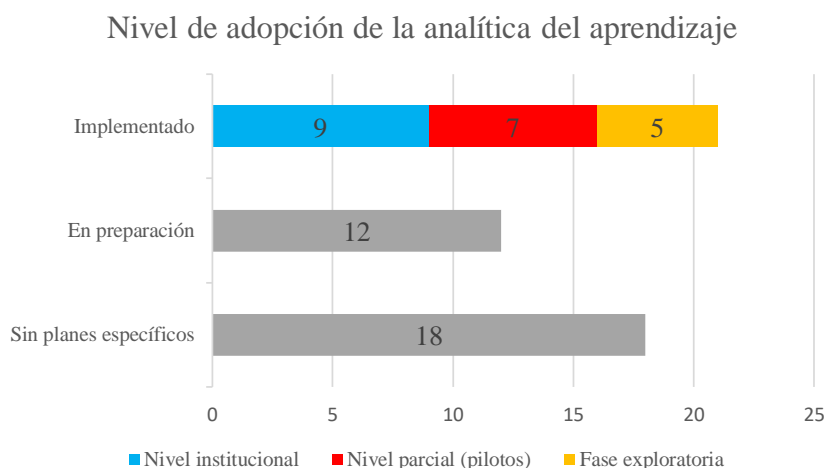


Fig. 4.1. Nivel de adopción de analítica del aprendizaje en Europa, tomado de [248]

A pesar de que el nivel de adopción es bajo, el resultado de las entrevistas muestra que existe interés en el uso de LA. Para el caso de los líderes institucionales y expertos, se analizaron los resultados utilizando NVivo y un esquema de codificación con 99 códigos⁵, que sirvió para analizar los datos de forma sistemática. Los resultados confirman, por un lado, que dado el bajo nivel de adopción, el éxito principal está en la experiencia adquirida, por lo que todavía hace falta más trabajo en LA [249]. Sin embargo, en cuanto a la percepción se observa que LA debe usarse para mejorar la enseñanza y la gestión institucional, por lo que los usuarios principales son los profesores y el personal de apoyo, aunque los resultados muestran una falta de involucramiento de los estudiantes y uso de LA para mejorar las capacidades de autorregulación de los estudiantes.

En este contexto, la predicción tiene sentido, ya que se priorizaría que el profesor tenga que recibir la información sobre los estudiantes y ser proactivo con los estudiantes. Además, de acuerdo con los grupos focales a estudiantes [248], estos particularmente quieren que LA les proporcione retroalimentación y contenido personalizado, a la vez que valoran las conversaciones con profesores y tutores, por lo que una actitud proactiva de los profesores supondría también en este beneficio.

De igual forma, el análisis de las encuestas y entrevistas a los profesores mostró que los profesores quieren que LA se utilice para proporcionar retroalimentación a los estudiantes y reconocen su potencial para mejorar la enseñanza y el aprendizaje, con

⁵http://bit.ly/interview_coding

una mayor adaptación a las necesidades de los estudiantes [248], [250]. Sin embargo, no se ven en la obligación de actuar. Uno de los profesores, por ejemplo, reconoció que es posible predecir alumnos en riesgo, pero el problema es quién debe actuar ante ello. Esto reafirma que todavía quedan bastantes preguntas por responder hasta la implantación efectiva de los modelos predictivos a nivel institucional.

Para ayudar a responder estas preguntas y facilitar la adopción de LA, lo que se han realizado son algunos marcos. Por ejemplo, el marco SHEILA [251], [252] define seis pasos (mapear el contexto político, identificar interesados, identificar cambios de comportamiento deseados, desarrollar la estrategia de compromiso, analizar la capacidad interna para efectuar cambios y establecer monitorización y marcos de aprendizaje) para ayudar a las instituciones de educación superior a desarrollar políticas que ayuden a la adopción de LA. Del mismo modo, Prieto et al. [253] desarrollaron el marco *Orchestrating Learning Analytics (OrLA)* para guiar y apoyar las conversaciones entre los interesados para la adopción de innovaciones sobre LA, y se han desarrollado otros marcos que ayuden a la adopción en algunas regiones, como Australia [245] o Latinoamérica [254].

Estos marcos tratan de ayudar en el camino de la adopción de LA, en el que todavía queda mucho camino por recorrer. Para ello, hace falta solventar los principales problemas, que ahora mismo se pueden resumir en (1) falta de compromiso de los interesados y de reconocimiento de los beneficios de LA [255], que puede venir acentuado por la falta de base pedagógica en las implementaciones de LA [256], (2) falta de recursos, incluyendo infraestructura, recursos humanos y financieros [257], (3) limitaciones en los datos, tanto en la recolección, como en la calidad de los datos y capacidad para su interpretación [251], y (4) problemas éticos y de privacidad sobre el uso de los datos [258].

Por ello, y para concluir esta sección, se puede decir que LA está en un estado inicial en cuanto a adopción y queda mucho por recorrer. Es por ello, que actualmente los modelos predictivos se están usando principalmente a nivel de curso y a través de pilotos iniciados por investigadores relacionados con LA. Esto implica que es necesario todavía bastante trabajo para conseguir consolidar estos modelos predictivos en el ámbito educativo. Por ello, el estado actual de adopción puede generar una oportunidad de investigación para ver cómo poder desarrollar esos modelos, para que sean lo más efectivos posibles y puedan generar un impacto en la educación. No obstante, es importante que dentro de los objetivos de LA de los interesados esté el uso de dichos modelos.

4.2. Análisis de la necesidad del uso de la predicción

En la actualidad, la mayoría de las implementaciones de LA están realizadas a pequeña escala, y existen pocos casos de uso modelos predictivos en la práctica. De hecho, gran parte de los modelos se han realizado en el ámbito de la investigación y están limitados para su uso en la práctica (p. ej., no están diseñados para realizar una predicción temprana o no se ha analizado cómo realizar intervenciones personalizadas [24]). Sin embargo, estos modelos pueden tener un potencial para mejorar el aprendizaje y pueden dar retroalimentación a los profesores y/o estudiantes, que son aspectos vistos en las expectativas de los interesados del apartado anterior. Por ello, los modelos predictivos pueden ser relevantes en el ámbito educativo y su aplicación a gran escala puede ser muy provechosa [259]. De

cara a analizar esta relevancia, y analizar si los diferentes actores del aprendizaje ven positivo y necesario el uso de estos modelos (y cuáles son sus necesidades específicas), en esta sección se realiza un análisis específico sobre la opinión de los interesados respecto a la predicción.

El análisis, tratado en [260], se basa en un análisis cualitativo basado en entrevistas con líderes institucionales y expertos, y grupos focales de profesores y estudiantes españoles. En particular, se utilizan 18 entrevistas a 18 líderes institucionales y expertos de LA de 11 instituciones españolas, cinco grupos focales a estudiantes de la UC3M (23 estudiantes en total) y cuatro grupos focales a profesores (16 profesores en total) de la misma universidad. Los grupos focales de profesores y estudiantes tienen participantes de diferentes disciplinas, incluyendo (1) ingeniería, (2) humanidades y (3) ciencias sociales. La duración media de las entrevistas fue de en torno a 25 minutos, mientras que los grupos focales tuvieron una duración entre 40 y 60 minutos. Las entrevistas se llevaron a cabo entre 2016 y 2017, mientras que todos los grupos focales se realizaron en 2017. Estas entrevistas y grupos focales son un subconjunto de las comentadas en la subsección anterior, de las cual se tenía acceso a la transcripción, y de las que se ha analizado explícitamente el contenido relacionado con la predicción. Esto puede ser una limitación, así como el tamaño de la muestra y que el área geográfica sea limitada. Sin embargo, permite obtener una visión general sobre las expectativas en cuanto a la predicción de diferentes interesados en el contexto español, lo cual es relevante para tenerlo en cuenta en el diseño de los modelos predictivos. En cuanto a la metodología de análisis para esta sección, se han utilizando métodos de análisis jerárquicos dentro de la investigación cualitativa. En particular, primero se realizó un resumen de los aspectos relacionados con la predicción para cada institución o grupo focal, y luego se agruparon respuestas similares. A continuación, se detallan las conclusiones para cada uno de los interesados.

4.2.1. Opinión de los líderes institucionales y expertos

Respecto a los líderes institucionales y expertos, los resultados de la entrevista mostraron que en ninguna de las 11 instituciones españolas analizadas existían herramientas relacionadas con la predicción, aunque en 8 de las 11 entrevistas hubo menciones explícitas sobre el tema de usar predicciones. En particular, las instituciones que tenían un mayor avance en LA y tenían proyectos centralizados de analítica eran más propensos a reflexionar sobre el uso de la predicción, lo que es coherente con el modelo de sofisticación de LA comentado anteriormente [247]. En cuanto a la utilidad de los modelos predictivos, los líderes institucionales y expertos principalmente querían los modelos para poder detectar alumnos en riesgo de abandono. Por ejemplo, en la institución U2 se mencionó que tenían un proyecto (en fase inicial) donde querían conocer más información sobre el trabajo de los estudiantes, dado que tenían poca información sobre las titulaciones y tenían la necesidad de “*predecir abandonos, conocer los indicadores de abandono, o los indicadores de éxito de una titulación*”. En particular se mencionó que:

“En principio creemos que es muy importante disponer de un mecanismo que te dote de inteligencia institucional, que permita por un lado disponer en tiempo real, que actualmente no la tenemos, que te ayude en la toma de decisiones, también intervención temprana si un alumno está en riesgo de abandono o si un profesor no

hace el seguimiento que debería hacer de una asignatura.” - Institución U2.

Aparte del abandono, el bajo rendimiento de los estudiantes parece ser otro punto importante de interés y se mencionaron otras variables a predecir relacionadas con comportamientos de los estudiantes e indicadores de éxito de la titulación. Por ejemplo, en una institución se mencionó:

“Finalmente creemos que la analítica del aprendizaje es crucial para la universidad y para los profesores y nos permite aplicar nuevos modelos para detectar casi en tiempo real problemas en el contenido y en el comportamiento de los estudiantes.” - Institución U1.

Por tanto, la visión de los líderes institucionales sobre la predicción se centra principalmente en detectar alumnos en riesgo para buscar la mejora de indicadores como el éxito de los estudiantes o el abandono, que puede verse tanto a nivel de curso (que los alumnos no abandonen los cursos) o a nivel de titulación.

4.2.2. Opinión de los profesores

En cuanto a los profesores, reconocen la importancia de la predicción y en particular la de predecir estudiantes en riesgo de abandono y de suspenso. Sin embargo, dado que el número de estudiantes en un grupo puede ser muy numeroso (p. ej., grupos de 200 estudiantes), creen que no es posible proporcionar un apoyo personalizado a todos ellos. Por ello, alguna idea podría ser centrarse en los estudiantes más críticos y tomar acciones con ellos. Por ejemplo, un profesor indicó:

“A lo mejor no tienes que hacer una observación de los 200, sino tener una herramienta que te permita detectar, pero detectar a los 10 más críticos. Pues hay un 35 % de los alumnos que no tienen conocimientos suficientes de [la asignatura]. Pues eso te puede dar pie a que tú digas, bueno voy a hacer una clase específica de [la asignatura] o voy a dar una tutoría grupal especial sobre este problema que lo vais a necesitar porque si no, no vais a ser capaz de seguir la asignatura.” - Grupo 1.

Otra preocupación de los profesores era el alto nivel de intervencionismo. Por ejemplo, un profesor indicó que mientras que era importante ayudar a los estudiantes, proporcionar demasiada información podía ser contraproducente.

“Yo creo que estamos ya ante un nivel de intervencionismo excesivo. [...]. Lo de realizar el intervencionismo ese me asusta un poco, porque un excesivo intervencionismo a veces conduce al abandono también del estudiante.” - Grupo 4.

Sin embargo, a pesar de que haya dudas sobre cuál debe ser el nivel de intervención o cómo gestionar los grupos de muchos estudiantes, hay un consenso de que la predicción es una herramienta para mejorar el aprendizaje. Por ejemplo, un profesor comparó la educación con el fútbol y mencionó que si los equipos de fútbol utilizan inteligencia artificial para predecir los goles y mejorar el rendimiento de los jugadores, los contextos educativos deben tomar una aproximación similar.

Me decía el entrenador del [equipo X] con chavales: Este al final va a tener este porcentaje, este va a mejorar el no sé cuántos, este al final... [...]. No, en mayo hará esto, en junio hará no sé cuántos. Entonces cuando ves cómo... claro, parten de datos ¿eh? Y entonces saben cómo llegar a mejorar. [...]. Entonces todos esos datos son importantes, incluso para un curso, para ver cómo se va evolucionando y tal. - Grupo 2.

En cuanto a las variables a predecir, aunque en la mayoría de los casos se trató el abandono y el riesgo de suspenso, algunos profesores mencionaron que no debería enfocarse solo a eso, sino a cualquier comportamiento inusual para poder reaccionar a tiempo. Del mismo modo, otro tema a debatir fue la sobre la necesidad de tener tutores personales para ayudar a los estudiantes, dado que algunos profesores indicaron que éstos podrían ser los encargados de ver los resultados de las predicciones para alertar a los estudiantes y ofrecer apoyo. Además, los modelos predictivos incluso podrían decidir qué estudiantes necesitan un tutor que les apoye. En esta línea del tutor personal, por ejemplo, un profesor indicó:

“Una idea fantástica y que funciona, y que funciona muy bien si se pudiera expandir a tutores de grupo o tutores de un número de alumnos. Probablemente a ese tutor sí que le vendría bien tener esa información y poder alertar o dar un aviso de un comportamiento anómalo a un alumno.” - Grupo 4.

Como visión general, la conclusión que se extrae de los profesores es que ven positivo el uso de los modelos predictivos, dado que pueden ayudar a detectar alumnos en riesgo o que tengan comportamientos anómalos. Sin embargo, tienen algunas dudas en cuanto al nivel de intervencionismo y cómo ofrecer el apoyo en grupos numerosos.

4.2.3. Opinión de los estudiantes

En el caso de los estudiantes, aunque inicialmente, en general, no conocen qué es LA, reconocen su importancia si puede resolver sus problemas. Y, en cuanto a la predicción, el uso principal que les gustaría sería para tener sistemas que les predigan si van a suspender o no, de modo que el sistema les alerte y puedan reaccionar a tiempo. Sin embargo, los alumnos ven ciertas dificultades y problemas a estos sistemas. Por un lado, algunos estudiantes tenían ciertas dudas sobre si los datos recogidos sobre las plataformas educativas pueden ser útiles para realizar las predicciones. Un estudiante, por ejemplo, comentó:

“Creo que los datos que se dispongan van a afectar mucho en la previsión de dicho riesgo. No creo que con los datos de acceso a Aula Global [Moodle] se pueda estimar ese riesgo.” - Grupo 2.

Por otro lado, algunos estudiantes eran reacios a proporcionar los resultados a los profesores, ya que esto podía hacer que los profesores generasen algunos estereotipos basados en los datos. Sin embargo, no hubo consenso respecto a este punto en los grupos focales, ya que para otros estudiantes era importante que los profesores tuvieran acceso a las predicciones, de modo que pudieran utilizarlas para ayudar a los estudiantes. Algunas reflexiones respecto a esto son:

[Sobre la opción de “alertar lo antes posible al personal docente, si está en riesgo de suspender un módulo (un tema), o si pudiera mejorar su aprendizaje”]. “El profesor ya lo sabrá sin necesidad de una alerta. Esa opción puede crear el riesgo de predisponer el suspenso al estudiante.” - Grupo 4.

“La cosa es que se detecte que los estudiantes van mal antes del “cataclismo” de los exámenes. Que el profesor sepa si puede mejorar y lo que está pasando en su clase. [...] Es conveniente que los problemas se detecten antes de los exámenes para poder mejorar la situación antes de que suspendan los estudiantes.” - Grupo 5.

A pesar de este debate, otro problema que apreciaron los estudiantes es que podía haber casos en el que los profesores pudieran recibir la información, pero no saber cómo intervenir. Otros estudiantes lo que criticaron fue que en muchas ocasiones se trata a los estudiantes como si el conocimiento de todos los estudiantes fuese el mismo. Como resultado, algunos estudiantes propusieron utilizar LA para detectar las debilidades de los estudiantes, de modo que los profesores pueden utilizar esa información para proporcionar recursos adicionales y apoyo a los estudiantes. Por ejemplo, un estudiante dijo:

“Las clases se dan como si todos los estudiantes fueran iguales, lo cual no es cierto, y si algún estudiante no es igual que el resto se le etiqueta como un estudiante “que no vale”. Me gustaría que, según los puntos débiles de los estudiantes, se otorgaran diferentes recursos o herramientas para apoyar en los puntos débiles.” - Grupo 3.

Por tanto, la visión de los estudiantes es que los modelos predictivos pueden ser de ayuda para identificar los estudiantes en riesgo y las dificultades de los estudiantes, aunque hay dudas sobre cómo los resultados deben ser utilizados para que en realidad tengan un efecto positivo en el aprendizaje.

4.2.4. Conclusiones globales

En vista de los resultados de las entrevistas y de los grupos de enfoque, se puede concluir que todos los interesados reconocen que el uso de los modelos predictivos puede ser útil para proporcionar apoyo a los profesores y alumnos, y que el interés principal de estos modelos es común para los diferentes interesados. En particular, el mayor interés está en predecir el abandono y los estudiantes en riesgo. Esto, además es común con lo observado en los trabajos de investigación sobre predicción [16] y sirve para justificar las variables dependientes utilizadas en los análisis de predicción de esta tesis. Como se vio en la subsección 3.3.2, las variables utilizadas a predecir son el éxito, el abandono y la nota de los estudiantes. Por tanto, las variables elegidas concuerdan con las necesidades de los interesados.

Otro aspecto común de todos los interesados fue que todos reconocieron la importancia de la intervención en tiempo real o cuando sea posible. Esto conduce a una necesidad de analizar cuándo es el mejor momento para empezar a predecir, de manera que se puedan tener predicciones tempranas. Del mismo modo, en todos los grupos se trató la forma de realizar las intervenciones. Sin embargo, este aspecto no está claro y existen diferencias entre los interesados en varios aspectos (p. ej., quién debe recibir la información y mediante qué modos de comunicación). No obstante,

lo que queda claro en esta etapa es que los interesados son positivos hacia el uso de herramientas de LA con funciones de predictivas y que el uso de las herramientas puede suponer una mejora en la educación.

Actualmente el estado de adopción de estas herramientas es bajo, como se demostró en el análisis del estado de adopción (sección 4.1), y debe haber trabajo para ver cómo incorporar estas herramientas y cómo deben mostrarse a los interesados.

5

Descripción de los escenarios

Contenido

5.1. Descripción general de los escenarios	63
5.2. Escenario 1. MOOC de programación en Java (UC3M)	64
5.3. Escenario 2. MOOC de programación en Java (HKUST)	66
5.4. Escenario 3. MOOCs en Coursera (PUC)	68
5.5. Escenario 4. SPOCs de apoyo en pruebas de acceso (KU Leuven)	71
5.6. Escenario 5. SPOCs de apoyo en cursos universitarios (UC3M)	74
5.7. Escenario 6. Plataforma para alumnos de primaria y secundaria	76
5.8. Similitudes y diferencias entre los diferentes escenarios	79

Para realizar los modelos predictivos se necesitan datos de estudiantes, que realicen cursos en plataformas educativas y de los cuales se disponga de traza a partir sus interacciones. En esta tesis doctoral, se utilizan diferentes escenarios que vienen dados por cursos con diferentes contextos y metodologías docentes para poder entender mejor cómo pueden afectar los modelos predictivos según el contexto. En este capítulo, se presentan los diferentes escenarios en los que se han llevado a cabo los análisis. El capítulo tiene una primera sección para dar una visión global de los escenarios utilizados y después seis secciones para explicar cada uno de ellos. Por último, hay una sección para indicar las similitudes y diferencias de todos los escenarios, para tener una visión global de todos ellos.

5.1. Descripción general de los escenarios

En este trabajo, con el objetivo de poder abordar diferentes contextos educativos, se han considerado seis escenarios de aprendizaje. Estos escenarios son lo suficientemente diferentes para abarcar diferentes tipos de cursos (presenciales, semipresenciales y a distancia), con diferentes temáticas y diferentes perfiles de

alumnos, así como diferentes variables disponibles. En unos casos, los escenarios están compuestos por un curso y, en otros, por varios y, de forma general, se han dividido por el tipo de curso y la fuente de la que proceden. En esta sección, se proporciona una descripción general de los cursos y en las secciones siguientes, se explica la metodología docente y los detalles de cada escenario. A continuación, se indica una descripción general sobre cada escenario de aprendizaje.

- Escenario 1 (MOOC de programación en Java, desarrollado por UC3M). Se trata de un curso online (MOOC), con miles de estudiantes con diferentes tipos de perfiles. Es un curso síncrono de cinco semanas.
- Escenario 2 (MOOC de programación en Java, desarrollado por HKUST). Al igual que el caso anterior, es un curso online (MOOC), con miles de estudiantes y un modo de instrucción síncrono. Difiere del anterior en su duración, ya que este curso tiene 10 semanas, y consecuentemente en parte de los contenidos, dada la duración (aunque existe un solape de contenidos importante).
- Escenario 3 (MOOCs en Coursera, desarrollados por PUC). Se trata de tres cursos online (MOOCs), albergados en la plataforma Coursera. Dos cursos son sobre ciencias sociales y otro sobre electrónica. Estos cursos se han agrupado en un escenario al analizarse conjuntamente y tienen en común que todos están albergados en Coursera y tienen un modo de instrucción asíncrono.
- Escenario 4 (SPOCs de apoyo en pruebas de acceso, desarrollados por KU Leuven). Se trata de dos cursos semipresenciales (uno de física y otro de química) para ayudar a los estudiantes de último curso de educación secundaria a preparar las pruebas de acceso a la universidad en estudios de medicina y odontología en Flandes (Bélgica).
- Escenario 5 (SPOCs de apoyo en cursos universitarios, desarrollados por UC3M). Se trata de cursos online utilizados como apoyo en las clases presenciales de la UC3M. Esos cursos comprenden diversas áreas temáticas impartidas por la universidad: ciencias sociales, ciencias formales e ingeniería. Su duración es equivalente al de los cursos presenciales (en torno a 16 semanas).
- Escenario 6 (plataforma de alumnos de primaria y secundaria). Se trata de cursos online que sirven de apoyo a las clases presenciales en diferentes asignaturas de primaria y secundaria en diferentes centros educativos. Su duración comprende un curso académico completo (de septiembre a junio).

Tras dar una visión general de los escenarios en los que se centra esta tesis, a continuación se proporciona una descripción detallada de cada uno de ellos (secciones 5.2 a 5.7).

5.2. Escenario 1. MOOC de programación en Java (UC3M)

El escenario 1 es un MOOC (xMOOC) sobre programación llamado “*Introduction to Programming with Java - Part 1: Starting to Program in Java*” (en español, “*Introducción a la Programación en Java - Parte 1: Empezando a Programar en Java*”). Este escenario fue tratado en [261] y en un Trabajo Fin de Máster (TFM)

previo del autor [262]. Los resultados de dicho TFM se toman como base para poder ponerlos en contexto y compararlos con otros trabajos en esta tesis, de modo que sirvan para obtener nuevas conclusiones globales en base a varios escenarios educativos.

El MOOC tratado en este escenario está albergado en edX y fue el primero de una trilogía de cursos desarrollados por la UC3M para aprender a programar en Java sin conocimientos previos. El curso está organizado en cinco semanas (i), entre el 28 de abril y el 30 de junio de 2015, siguiendo un modo de instrucción síncrono. En estas semanas, los alumnos deben ver vídeos y realizar actividades antes de enviar las tareas de evaluación sumativas semanales. Entre esas pruebas sumativas, hay una prueba tipo test con preguntas cerradas de respuesta múltiple semanalmente (T_i). Cada una de estas pruebas representa un 15% de la nota final. Además, hay dos pruebas sumativas de programar y entregar código (P_i). Estas pruebas se sitúan en las semanas 3 y 5 del curso y representan un 10% y un 15% de la nota, respectivamente. Estas tareas son evaluadas mediante evaluación entre pares, siguiendo una rúbrica de evaluación común proporcionada a los alumnos.

Respecto a los contenidos del curso, aunque estén relacionados, existe cierta independencia entre ellos. Algunos de ellos requieren mayor pensamiento computacional sobre cómo desarrollar algoritmos, mientras que otros requieren una mayor capacidad de abstracción. En cuanto a los datos demográficos de los alumnos matriculados, la media de edad es de 30 años y el 70% de los alumnos tiene un título universitario. La mayor parte de los alumnos proceden de Estados Unidos (25%), seguido de la India (16%) y de España (3%) [263]. El hecho de que el MOOC esté en inglés puede contribuir a una mayor internacionalización.

En este contexto, las variables dependientes son la nota de las cinco pruebas semanales de preguntas de respuesta múltiple (T_i), la nota de las dos pruebas de programación (P_i) y la nota final del curso (NF) [261]. Además, se realiza un análisis de la predicción del éxito, de modo que el éxito se considera cuando un alumno aprueba el curso, es decir, su nota final es superior a 60%, que es el porcentaje mínimo para aprobar. Por otro lado, las variables independientes utilizadas son:

- Variables sobre la actividad en el foro: `participations`, `com_threads`, `comments`, `votos_rcv`, `votos_emt`, `endorsed`, `avg_longitud`, `percentil` y `sentimiento`.
- Variable sobre interacciones con ejercicios: `nota_media`, `perc_intentados`, `perc_ej_abiertos`, `avg_intentos`, `CFA` y `perc_correctos_total`.
- Variables sobre interacciones con vídeos: `perc_vabiertos`.
- Notas previas: Son las notas de las pruebas sumativas previas (pruebas de respuesta múltiple o programación, T_i o P_i) cuando estén disponibles (p. ej., la semana 1 no tiene notas previas, pero en la semana 2 está disponible la nota de la prueba de la semana 1).

Cabe mencionar que todas estas variables están obtenidas a partir de las exportaciones de la base de datos del MOOC, es decir, de los ficheros contenidos en el fichero comprimido `{org}-{date}.zip`, proporcionado por edX (más detalles en la Sección 3.2). Finalmente, es importante destacar que en este MOOC hay 95 555 alumnos registrados. Sin embargo, la mayoría no llegan a realizar ninguna actividad. Por ello, es necesario establecer un criterio de selección de la muestra que evite

sesgo en los resultados por los alumnos que no interactúan. En este caso, como este escenario tiene bastante información sobre los foros y se quiere analizar el efecto de estas variables, se han incluido solo aquellos alumnos que han contribuido al menos una vez en el foro (publicando un mensaje). Este criterio tiene sentido, ya que, por ejemplo, Anderson et al. [264] concluyeron que el 90 % de los usuarios activos en el MOOC también lo eran en el foro. Además, los estudiantes que se desmatricularon del curso fueron excluidos, ya que no hay nota de ellos y no se dispone de las variables dependientes. Teniendo en cuenta estos criterios, se considera un total de 4358 alumnos en este escenario.

5.3. Escenario 2. MOOC de programación en Java (HKUST)

El escenario 2, tratado en [263], es un MOOC (xMOOC) sobre programación titulado “*Introduction to Computing with Java*” (en español, “*Introducción a la Computación en Java*”), que fue desarrollado por HKUST en lengua inglesa. Este curso, albergado en edX, tiene una duración de 10 semanas, entre el 20 de junio y el 31 de agosto de 2014, y sigue un modo de instrucción síncrono. El foco del curso está en el laboratorio (a diferencia del escenario 1, que también trataba la programación en Java, pero daba un mayor peso a los conceptos teóricos), por lo que evalúa si los alumnos consiguen las habilidades para poner en práctica los conceptos. En este MOOC, los laboratorios consistían en tareas de programación (que involucraban normalmente 4-5 ejercicios de corrección automática). La evaluación consistía en seis laboratorios (L_i) como los mencionados (de los que se cogían las cinco mejores notas, que representaban un 20 % de la nota), un proyecto (valía un 20 % y tenía una parte de evaluación automática, PA y otra con revisión por pares, PR) y un examen final (EF), que representaba un 40 % de la nota final (NF). Los seis laboratorios se realizaban en las semanas 1, 2, 3, 5, 7 y 8; el proyecto se realizaba entre las semanas 5 (apertura) y 9 (entrega), y el examen final se realizaba en semana 10 (última semana). Adicionalmente, el MOOC contenía 42 ejercicios cortos, que incluían tanto conceptos, mediante preguntas respuesta múltiple, como habilidades, mediante pequeñas tareas de realizar código en Java. Estos ejercicios tenían como objetivo preparar a los alumnos para el resto de pruebas de evaluación y eran evaluados automáticamente, pero proporcionaban el 20 % restante de la nota final.

En cuanto a los datos demográficos, eran muy similares a los del escenario 1, tanto en edad, nivel de estudios y área geográfica. La principal diferencia era que en cuanto a área geográfica, el tercer país con más estudiantes era Reino Unido (3 %) en vez de España (el país en el que se diseñó el MOOC del escenario 1). Los dos países con más estudiantes eran Estados Unidos y la India, al igual que en el escenario 1. En este caso, a pesar de que el MOOC estaba diseñado en Hong Kong, el número de estudiantes chinos era menor, a diferencia del escenario 1 donde muchos participantes eran españoles. Un posible motivo es que los vídeos estaban subidos en la plataforma YouTube, por lo que muchos alumnos no podían acceder a ellos desde China continental, lo cual es un problema ya identificado previamente en la literatura [265]. Finalmente en cuanto a la metodología del curso, la Tabla 5.1 muestra un resumen de las características fundamentales de este escenario, en comparación con el anterior [263].

En cuanto a las variables dependientes utilizadas en este escenario, de forma similar al escenario 1, se predicen las notas de las pruebas evaluables del MOOC

TABLA 5.1
RESUMEN Y COMPARATIVAS DE LOS CURSOS EN LOS ESCENARIOS 1 Y 2

Característica	MOOC UC3M	MOOC HKUST
Duración	5 semanas	10 semanas
Aproximación	Orientado a conceptos	Orientación a laboratorios
Modo de instrucción	Síncrono	Síncrono
Número ejercicios cortos*	135	42
Número de vídeos	69	122
Alumnos matriculados	95 555	47 178
Alumnos que han visto algún vídeo	24 055	19 903
Alumnos que han intentado ejercicios	16 317	11 708
Número de alumnos aprobados	1507	1299
Porcentaje requerido para aprobar	60 %	50 %
Tiempo entre apertura y cierre de un entregable	2 semanas (excepto en la semana 1 donde hay 3 semanas)	9 días (excepto en el proyecto, que hay 37 días)

* Los ejercicios cortos se refieren a los ejercicios formativos en el escenario 1

(incluido el examen final). En particular se predice la nota de los seis laboratorios (L_i), la componente de corrección automática del proyecto (PA), la componente de revisión por pares del proyecto (PR), el examen final (EF) y la nota final del curso (NF). Estas notas se encuentran en escala 0-1, tal y como aparecen en el informe de notas. Además, se predice el éxito académico del estudiante (si aprueba o suspende) en el examen final y en la nota final del curso. Para este caso, se utiliza una variable binaria, cuyo valor es 1 si la nota del estudiante está por encima del 50% mínimo requerido para aprobar y 0 en caso contrario.

Respecto a las variables independientes, se han obtenido a partir de las exportaciones de la base de datos y del registro de eventos. En particular, el conjunto de variables independientes utilizadas es el siguiente:

- Variables sobre la actividad en el foro: `participations`, `com_threads`, `comments`, `votos_rcv`, `votos_emt`, `endorsed`, `avg_longitud`, `percentil` y `sentimiento`.
- Variables sobre interacciones con ejercicios: `nota_media`, `perc_intentados`, `perc_ej_abiertos`, `avg_intentos`, `CFA` y `perc_correctos_total`.
- Variables sobre interacciones con vídeos: `perc_vabiertos`, `perc_vtotal`, `perc_compl`, `avg_rep` y `avg_pausas`.
- Variables sobre actividad en la plataforma: `ndias`, `avg_con`, `racha_acc`, `perc_pc` y `perc_wk`.
- Notas previas: Son las notas de las pruebas sumativas previas (pruebas de respuesta múltiple o programación) cuando estén disponibles (p. ej., la semana 1 no tiene notas previas, pero en la semana 2 está disponible la nota de la prueba de la semana 1).

Finalmente, cabe destacar que el criterio de selección de la muestra utilizado es el mismo que el utilizado en el escenario 1, de modo que los escenarios sean fácilmente comparables y se pueden obtener conclusiones acerca de las diferencias debidas al contexto. En este caso, siguiendo dicho criterio (incluir alumnos que hayan contribuido al menos una vez en el foro), se tiene un conjunto de 2168 alumnos.

5.4. Escenario 3. MOOCs en Coursera (PUC)

El escenario 3, tratado en [91], [266], está compuesto por tres MOOCs, desarrollados por PUC en la plataforma Coursera. El primero de los MOOCs era un xMOOC sobre electrónica titulado “*Electrones en Acción*”. Este curso estaba organizado en cuatro módulos y contenía 83 vídeos y 16 ejercicios. Este curso estaba desarrollado en español y tenía un modo de instrucción asíncrono, a diferencia de los escenarios anteriores. Dado que el curso era asíncrono, se eligió un periodo para la recolección de datos y análisis, que fue entre abril y diciembre de 2015. En este MOOC, había un total de 25 706 alumnos matriculados, aunque para el análisis solo se consideraron 2035 estudiantes, que son aquellos que contestaron un cuestionario online sobre sus estrategias de autorregulación. Los datos fueron proporcionados ya filtrados por el equipo de PUC, que realizaron este filtrado previamente para focalizarse en el análisis de SRL en otros trabajos como [224].

Para la validación de resultados de este escenario, se utilizaron el segundo y tercer MOOC de este escenario. Estos MOOCs fueron también xMOOCs desarrollados por PUC en Coursera. Ambos eran sobre ciencias sociales, y sus nombres eran “*Aula Constructivista*” y “*Gestión de Organizaciones Efectivas*”. Estos cursos, al igual que “*Electrones en Acción*”, se impartieron en español y utilizaron un modo de instrucción asíncrono. El hecho de que fueran impartidos en español pudo restringir los usuarios potenciales, pero también hizo que la mayoría de estudiantes fueran latinoamericanos, lo que hace que este análisis contribuya en los estudios de LA en Latinoamérica. En cuanto a la estructura de los cursos, el MOOC “*Aula Constructivista*” tenía nueve módulos y 11 pruebas de evaluación sumativas (tres de ellas se evaluaban mediante revisión por pares), mientras que “*Gestión de Organizaciones Efectivas*” tenía siete módulos y seis tareas sumativas. El periodo de recolección de datos en estos cursos era el mismo que para “*Electrones en Acción*” y el número de estudiantes tras el filtrado (considerando solo los estudiantes que rellenaron el cuestionario) era 337 y 526 (sin filtrar 18 653 y 10 576) para ambos MOOCs respectivamente.

Teniendo en cuenta estos cursos, para el escenario 3 se propusieron diferentes análisis. En primer lugar, para el primero de los MOOCs, se realizó un primer análisis exploratorio para predecir la nota final del curso (variable continua) y para predecir el éxito de los estudiantes [266]. En este caso, el éxito se definió de dos formas diferentes:

- Definición 1. Obtener la nota mínima para aprobar el curso (80 %) independientemente de los contenidos que se hayan trabajado en el curso.
- Definición 2. Obtener la nota mínima para aprobar el curso (80 %), habiendo visualizado al menos el 50 % de los vídeos.

A continuación, se realizó un análisis de la predicción del abandono de los estudiantes, que fue extendido a los tres cursos [91]. Para este análisis, el primer

problema fue cómo definir abandono, ya que en un MOOC un estudiante puede estar inactivo durante un tiempo y luego volver, y esto particularmente puede ocurrir en un MOOC asíncrono ya que no han un calendario definido para completar las actividades. Por ello, se realizó un análisis de cuántos estudiantes vuelven a interactuar después de un periodo de inactividad. Para este análisis, se calculó el periodo de actividad como el máximo número de días inactivos entre dos interacciones y se calcularon los percentiles sobre el periodo de inactividad (ver Tabla 5.2).

TABLA 5.2
PERCENTILES DEL PERIODO INACTIVO DE LOS ALUMNOS EN EL MOOC

Percentil	25	50	60	70	80	90	94	95	96	97	100
Número días	0	3	5	8	10	16	21	22	26	29	221

Los resultados muestran que, por ejemplo, la mitad de los alumnos presentan periodos de inactividad de más de tres días, pero luego vuelven a interactuar con el curso. Sin embargo, para la definición de abandono es importante seleccionar un periodo a partir del cual los alumnos vuelvan a interactuar con muy poca probabilidad, de modo que se pueda considerar ese periodo de inactividad como sinónimo de abandono. A la vista de los datos, se consideró un periodo de inactividad de cuatro semanas (28 días), ya que más del 96 % de los alumnos que están inactivos durante más de 28 días no vuelven a interactuar en el curso.

Aparte de este criterio, es importante considerar el periodo desde la última interacción hasta el final del periodo de recolección de datos, ya que un alumno puede haber interactuado los primeros días, pero luego no volver. Entonces, para el análisis, el periodo de inactividad se considera como el máximo entre el periodo más largo entre dos interacciones y el periodo desde la última interacción hasta el final del periodo de recogida de datos. Si ese periodo de inactividad es superior a 28 días, se considera abandono. Sin embargo, se tienen en cuenta dos consideraciones adicionales:

- Un alumno no se considera como que ha abandonado si ha completado al menos el 80 % de las pruebas de evaluación, ya que en este caso se podría decir que ha completado el curso.
- Si un alumno empieza hacia el final del periodo de recolección de datos o avanza despacio en el curso, es posible que no se sepa si el alumno va abandonar o no. Por ello, el alumno se descarta si no tiene un periodo de inactividad inferior a 28 días y el alumno no ha completado el 80 % de las evaluaciones.

Tomando todas estas consideraciones, que están resumidas en la Fig. 5.1, se consideran finalmente 926 estudiantes para el curso “Electrones en Acción”, 337 para “Aula Constructivista” y 526 para “Gestión de Organizaciones Efectivas”.

Respecto a las variables independientes, se utilizaron diferentes tipos de variables, incluyendo las tradicionales sobre interacciones con ejercicios, vídeos y actividad en la plataforma. Además, este escenario tiene como especial que también se recogieron variables sobre la autorregulación del alumno (a partir de patrones de eventos en la plataforma), de modo que puede analizarse el efecto del SRL en los modelos predictivos. En particular, la siguiente lista de variables fue utilizada:

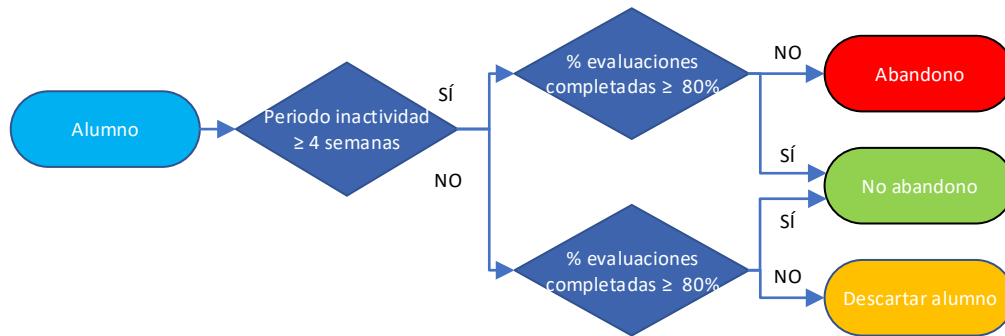


Fig. 5.1. Diagrama de flujo con las reglas para determinar el abandono

- Variables relacionadas con patrones de SRL: `only_vlecture`, `atry_to_vlecture`, `explore`, `only_assessment`, `vlcomplete_to_atry`, `vlecture_to_acomplete` y `complex`.
- Variables sobre actividad en la plataforma: `ndias`, `tiempo_invertido` (en minutos), y `num_ses`.
- Variables sobre interacciones con vídeos: `vl_complete`, `vl_begin`, `vl_review`, `perc_vabiertos`, `perc_compl` y `perc_vlreview`.
- Variables sobre interacciones con ejercicios: `a_try`, `a_complete`, `a_review`, `perc_intentados`, `perc_acomplete` y `perc_areview`.

Asimismo, un aspecto importante de este escenario respecto a los MOOCs anteriores es que el modo de instrucción es asíncrono. Este hecho tiene un impacto sobre el modo de calcular las variables dependientes e independientes en cada instante de tiempo (para el análisis temporal). El motivo es que en un momento determinado de tiempo (p. ej., semana X del calendario) cada estudiante puede llevar un progreso diferente en el curso, según cuando haya comenzado el curso y el ritmo que lleve. Por ello, la solución típica de considerar un periodo de tiempo (típicamente la semana) e ir añadiendo datos semana a semana para ver cómo evoluciona el poder predictivo con el paso del tiempo (usada, por ejemplo, en [267]) no es válida. Para el caso de MOOCs asíncronos, Vitiello et al. [268] normalizaron el tiempo en función del tiempo total activo en el curso, de manera que comparaban alumnos que llevasen el mismo porcentaje de tiempo activo en el curso. Un problema de utilizar el tiempo activo es que los porcentajes no pueden ser calculados hasta que se termina el curso, por lo que no podrían aplicarse los modelos sobre un estudiante que esté cursando, lo que limita su aplicabilidad. Como alternativa, se podría usar el porcentaje completado del curso, aunque en este escenario se detectó una baja varianza entre las variables cuando el porcentaje completado de los estudiantes era similar. Teniendo en cuenta esto, la solución propuesta para el análisis de MOOCs asíncronos fue la de normalizar el periodo de tiempo.

Esta normalización consiste en calcular el periodo desde la primera interacción de cada alumno hasta el final del periodo seleccionado (p. ej., si se considera una semana, el periodo incluye los siete primeros días desde la primera interacción para cada alumno, pudiendo estar esta primera interacción en diferentes fechas para cada uno). Siguiendo esta normalización, se puede realizar el análisis temporal por

semanas, como en MOOCs síncronos, pero teniendo en cuenta el periodo temporal en el que cada uno realiza el curso. La Fig. 5.2 representa esta normalización. En este caso, el análisis de la primera semana incluye el primer bloque (semana) de cada alumno, el análisis de las dos primeras semanas, incluye los dos primeros bloques, y así sucesivamente.

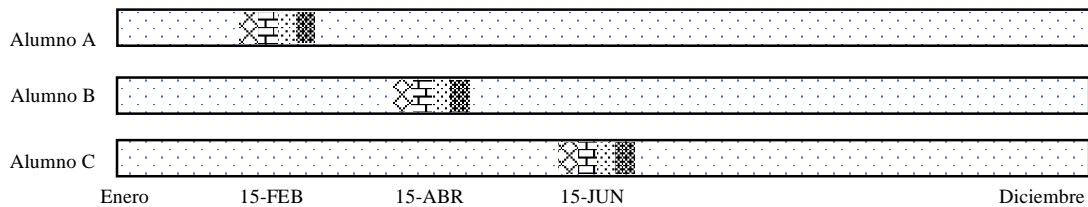


Fig. 5.2. Ejemplo de las cuatro primeras semanas (bloques) del curso para tres alumnos que comienzan en diferentes periodos

Un posible problema es que esta aproximación no tiene en cuenta la velocidad a la que los estudiantes avanzan en el curso. Sin embargo, en este escenario, los datos mostraron que los alumnos completaban los módulos en el periodo sugerido (si había cuatro módulos, el periodo sugerido eran cuatro semanas), aunque pudieran ir más rápido en ciertos momentos. En concreto, el 48 % de las interacciones ocurrían en la primera semana, el 69 % en las dos primeras y el 94 % en las cuatro primeras, lo que hace que la aproximación mencionada tenga sentido para este escenario. No obstante, es importante señalar que para otro escenario podría ser diferente si el ritmo de avance tuviera más variaciones.

5.5. Escenario 4. SPOCs de apoyo en pruebas de acceso (KU Leuven)

El escenario 4, tratado en [29], [269], incluye datos de SPOCs albergados en edX Edge y desarrollados en un proyecto conjunto entre la Facultad de Ciencias y la Facultad de Medicina de KU Leuven. Estos SPOCs eran parte de un programa semipresencial de apoyo en la preparación a las pruebas de acceso a la universidad en medicina y odontología en Flandes. Esta prueba consiste en una parte científica con pruebas de química, física, matemáticas y biología, y una parte de habilidades de comunicación y procesamiento de la información. Esta prueba es requisito indispensable para acceder ser admitido en los programas de medicina u odontología en Flandes, aunque la tasa de aprobados de esta prueba es muy baja y fluctúa entre el 20 %. Por ello, los estudiantes deben realizar una preparación intensa para la prueba. En esta prueba además, el perfil académico del estudiante parece que tiene una alta importancia, y los estudiantes que han cursado itinerarios que incluyan ciencias y/o matemáticas tienen una ventaja [270].

Teniendo todo esto en cuenta, los estudiantes a los que están dirigidos los SPOCs son estudiantes en el último curso de educación secundaria que quieren acceder a estudios de medicina u odontología en Flandes y pagan una tasa para registrarse al programa de apoyo semipresencial para estas pruebas. Los SPOCs funcionan de forma síncrona y los módulos del curso se van abriendo de forma gradual cada quincena (entre septiembre y mayo de cada curso académico). Además, se realizan

sesiones interactivas presenciales utilizando una metodología de clase invertida con el propósito de estimular a los alumnos a que vayan trabajando poco a poco durante el curso. No obstante, en la práctica, la mayoría de estudiantes se apuntaban tarde al programa y estudiaban a su propio ritmo, como si fuera un SPOC asíncrono. Respecto a la asistencia a las clases, no era obligatoria dado que uno de los objetivos era que los contenidos se pudieran trabajar en cualquier momento y lugar, lo que daba flexibilidad al programa. En definitiva, el objetivo del curso era preparar a los alumnos para la prueba de acceso, por lo que cada uno debía usar el contenido de la forma más provechosa posible de cara a la prueba real, que es independiente del SPOC.

Para el análisis de este escenario, se tomó como referencia un SPOC sobre química, y luego se extendió a otros SPOCs, principalmente para analizar la generalización, que se abordará en el capítulo 8. El primer SPOC a analizar se desarrolló en el curso 2016/2017 (septiembre 2016 a agosto 2017, ambos incluidos) y tenía 11 módulos, con 66 vídeos y 121 ejercicios. Este curso se impartía en flamenco (lengua de Flandes) y cubría los contenidos necesarios para la prueba de Química de las pruebas de acceso. Aparte del SPOC, se realizaron tres sesiones presenciales interactivas durante el curso, en las cuales se resolvieron ejercicios. El número de estudiantes matriculados fue 1062, aunque solo 680 completaron al menos un ejercicio y solo 750 realizaron interacciones con vídeos.

A continuación, se consideraron otros tres SPOCs para analizar la generalización de los modelos predictivos. Estos otros cursos eran un SPOC de Física del curso académico 2016/2017 y los SPOCs de Física y de Química del curso académico 2017/2018. La Tabla 5.3 presenta un resumen de los participantes de los cursos. En esta tabla, es importante destacar que los estudiantes no se matriculaban en un SPOC en particular, sino en el programa de preparación. Por tanto, los alumnos del SPOC de Física y de Química del mismo curso académico son los mismos. De este modo, con estos datos, se tiene información de dos cohortes, la de los alumnos del 2016/2017 y la de los alumnos del 2017/2018.

TABLA 5.3
RESUMEN DE LOS PARTICIPANTES EN LOS SPOCS

Curso	Año	Nº alumnos	Alumnos que vieron al menos 1 vídeo	Alumnos que intentaron al menos 1 ejercicio
Química	2016/2017	1062	750	680
Física			730	606
Química	2017/2018	1131	939	834
Física			856	767

En cuanto al SPOC de Física del curso 2016/2017, se diferencia del de Química en que tiene una componente mayor de comprensión de conceptos. Sin embargo, el número de alumnos que interaccionan en uno y otro son similares. En cuanto a los dos SPOCs (Física y Química) del curso académico 2017/2018, el contenido es muy similar a los del 2016/2017, aunque cambia el cronograma de las sesiones presenciales.

En el curso 2017/2018, los alumnos se matricularon antes dado que al haberse realizado el curso con anterioridad, los estudiantes tenían mayor conocimiento sobre

la existencia del programa, y consecuentemente la actividad en el curso fue mayor. La Tabla 5.3 muestra que hubo tanto una mayor actividad en vídeos y ejercicios como un mayor número de matriculados. En el curso 2017/2018, se organizaron nueve sesiones presenciales y hubo una asistencia media de 351 estudiantes cada sesión. Un posible motivo del incremento de interés también fue el cambio de formato en la prueba de acceso. Durante el curso 2016/2017 había una primera prueba en julio y había una segunda oportunidad para los que suspendían en agosto. Además, todos los estudiantes que conseguían superar la prueba eran admitidos. En el curso 2017/2018, se eliminó la recuperación de agosto, y los estudiantes solo tenían un intento en julio. Además, se fijó un número máximo de admisiones, por lo que ya solo los mejores estudiantes que aprobasen las dos partes (ciencias y habilidades de comunicación e información) eran admitidos (en lugar de todos los aprobados).

Teniendo en cuenta la información del contexto mencionada, la variable dependiente es el éxito académico, que en este caso consiste en aprobar la parte de ciencias de la prueba de acceso. Dado que la prueba de acceso es una prueba oficial, independiente del SPOC, no se tiene información de la misma a partir de los eventos del curso. Por ello, los resultados que se conocen son los obtenidos a través de un cuestionario, que se enviaba a todos los alumnos y en el cual los alumnos autorreportaban su nota de la parte de ciencias del examen (que era la que se trataba en los SPOC). Sin embargo, no todos los alumnos rellenaban este cuestionario, lo que cual es una limitación importante del estudio. Teniendo esto en cuenta, el criterio de selección de la muestra utilizado en este estudio fue el de coger todos aquellos estudiante que habían accedido a la plataforma al menos una vez y que habían completado el cuestionario (para conocer la variable dependiente). Una implicación del criterio de tener al menos un acceso es que el número de estudiantes varía al realizar el análisis temporal, ya que, según este criterio, los alumnos se van incorporando a los modelos una vez que se hayan inscrito al curso. No obstante, la limitación principal es la de conocer la variable dependiente. Con estos criterios, el número de alumnos considerados en el análisis es de 114 estudiantes para el curso 2016/2017 y 116 para el 2017/2018. Para este número se tiene presente que un mismo alumno solo rellena una vez el cuestionario, por lo que si lo rellena, afecta al curso de Física y al de Química, lo que hace que no haya variaciones entre las dos materias.

Respecto a las variables independientes, se obtuvieron a través del registro de eventos de edX Edge, con los eventos mencionados en la Sección 3.2. En particular, la lista de variables utilizadas es la siguiente:

- Variables sobre interacciones con vídeos: `perc_vabiertos`, `perc_vtotal`, `perc_compl`, `avg_rep` y `avg_pausas`.
- Variables sobre interacciones con ejercicios: `perc_intentados`, `avg_intentos`, `media_intentados_pr`, `perc_correctos`, `a_CFA`, `racha_ej` y `nshow`.
- Variables sobre actividad en la plataforma: `racha_acc`, `ndias`, `avg_con`, `perc_pc`, `perc_wk` y `perc_noche`.

Aparte de estas variables, se analizaron otras variables obtenidas fuera del SPOC. Entre ellas, se analizó la formación académica. En cuanto a esta variable, se clasificó a los estudiantes en tradicionales y no tradicionales [269]. Los estudiantes tradicionales eran aquellos que tenían un itinerario de secundaria con asignaturas de ciencias

y matemáticas, mientras que los no tradicionales eran aquellos que preparaban la prueba de acceso habiendo realizado otros itinerarios (p. ej., asignaturas relacionadas con humanidades). Se separó entre los estudiantes tradicionales que aprobaban ($n = 92$) y los que no aprobaban ($n = 22$), y los no tradicionales que aprobaban ($n = 6$) y no aprobaban ($n = 10$) del SPOC de Química del 2016/2017. Al realizar test de Mann-Whitney entre los estudiantes tradicionales y no tradicionales, se encontraron diferencias estadísticamente significativas ($p - \text{valor} < 0.05$). Los estudiantes no tradicionales eran más activos porque les faltaban conocimientos previos, pero en muchos casos suspendían incluso trabajando más que estudiantes tradicionales que aprobaban. Dado que los comportamientos eran bastante diferentes, el análisis se realiza por cada grupo. Dado que hay muy pocos estudiantes no tradicionales, de momento no se han realizado análisis predictivos, aunque se espera que se puedan realizar en ediciones futuras, cuando haya más estudiantes. Teniendo esto en cuenta, los 114 y 116 estudiantes incluidos en el análisis en cada cohorte son aquellos estudiantes tradicionales que contestaron la encuesta y reportaron sus resultados, no únicamente los que contestaron la encuesta.

Por otro lado, se analizó el efecto de la asistencia a las sesiones presenciales y la fecha de registro en los SPOCs del curso 2017/2018 (en el 2016/2017 no se llevó registro de la asistencia), aunque de cada sesión los estudiantes se registraban previamente a la sesión diciendo si iban a ir, pero luego no se controlaba la asistencia real. Sin embargo, estas variables no aportaban valor a los modelos [29] y la diferencia entre los estudiantes que aprobaban y no en términos de estas variables no era significativa (el p-valor era 0.10 para la asistencia y 0.31 para la fecha de registro). Un posible motivo es que los estudiantes que se registraban tarde no eran necesariamente peores y que como los estudiantes también cubrían los contenidos en el instituto, podían tener los apuntes y obtener buenas notas incluso registrándose tarde. En cuanto a la asistencia, dado que los SPOCs tenían estudiantes de todo Flandes, podía haber alumnos que vivieran lejos del lugar donde se realizaban las sesiones presenciales y decidieran no ir. Por ello, aunque potencialmente se podían incluir variables obtenidas fuera de la plataforma educativa, finalmente el conjunto de variables independientes estaba basado exclusivamente en las interacciones con los SPOCs.

5.6. Escenario 5. SPOCs de apoyo en cursos universitarios (UC3M)

El escenario 5, que fue tratado en [80], [108], consiste en SPOCs ofrecidos por la UC3M y albergados en una instancia local de Open edX. Estos cursos sirven como apoyo a las clases presenciales (ofrecidas de forma síncrona) de las asignaturas incluidas en las titulaciones que ofrece la universidad y tienen tres posibles modelos de uso: (1) SPOCs necesarios para aprobar el curso o con un peso significativo en la nota final, (2) SPOCs que son parte del curso y a menudo se combinan con la metodología de clase invertida, aunque no cuentan para la evaluación sumativa del curso, y (3) SPOCs que son únicamente de apoyo al curso, pero no obligatorios. En total, en este escenario, se dispone de datos de 17 SPOCs, que incluyen diferentes áreas temáticas de las ofrecidas en la universidad (ciencias sociales, ciencias formales e ingeniería). Sin embargo, el tipo de uso del SPOC y las características de cada

curso (p. ej., contenidos, estructura, etc.) son desconocidos para el propósito del análisis.

Sobre el contenido, lo que se conoce es que tienen vídeos y ejercicios, y que los ejercicios pueden ser de muchos tipos, incluyendo ejercicios de respuesta múltiple, de casillas de verificación, con desplegados, de respuesta numérica y de respuesta textual (p. ej., escribir una palabra). Todos los ejercicios permiten corrección automática y proporcionan una nota en un rango continuo de 0 a 100 %. Sin embargo, tras un análisis exploratorio, se vio que en el 95 % de los casos, los ejercicios están evaluados de forma binaria (bien o mal), ya que la mayoría de los ejercicios solo admiten valores binarios (p. ej., los ejercicios de respuesta múltiple solo pueden ser estar bien o mal, mientras que los de casillas de verificación pueden tener notas parciales dependiendo del número de aciertos). Para este análisis, la tipología de cada ejercicio no se conoce y es una limitación del escenario.

Para el análisis de los datos, se ha utilizado el registro de eventos, que sigue el formato de Open edX, aunque parcialmente preprocesado y filtrado por el equipo de Servicio de Informática y Comunicaciones de la UC3M. Los eventos utilizados son los mencionados en la Sección 3.2, exceptuando *problem_show* y *stop_video* que no están en la versión preprocesada. No obstante, la duración de los vídeos sí estaba almacenada, por lo que sin el evento de *stop_video*, se conocía la duración de los vídeos. En cuanto al criterio de selección de la muestra, se han tomado todos los alumnos disponibles, pero considerando el par “*estudiante-curso*”. Es decir, las interacciones de los estudiantes que están en más de un curso se consideran de forma independiente. De este modo, en los 17 SPOCs hay 1534 estudiantes, aunque para el análisis hay 1625 pares de “*estudiante-curso*”. No obstante, en algunos casos en los que se analice la predicción a mitad de curso, se usarán 16 SPOCs, ya que uno de ellos se utiliza solo en la segunda mitad del periodo lectivo y no hay trazas en las ocho primeras semanas.

Respecto a las variables dependientes, se ha considerado el éxito y el abandono. En cuanto al éxito, se considera que un alumno ha sido exitoso cuando la nota media de todas las actividades del SPOCs es igual o superior a 5.0 sobre 10 (nota necesaria para aprobar en España). Al incluir en la nota media todas las actividades del curso, las actividades no intentadas cuentan como 0. Sobre la variable de éxito, es importante hacer notar que tener/no tener éxito no implica aprobar o suspender la asignatura. Esta variable se calcula en base a las interacciones del SPOCs, que es lo único que se conoce, y no en función de la nota obtenida en la asignatura. Por lo tanto, puede haber alumnos que no hayan trabajado en el SPOCs (y se consideren como no exitosos), pero terminen aprobando la asignatura (sobre todo, si el uso del SPOCs no es obligatorio. No obstante, este análisis da una información relevante a los profesores sobre el uso del SPOCs, que dependiendo de la asignatura, puede correlar o no con lo que ocurra en la nota final de la misma.

En cuanto al abandono, se consideraron dos posibles definiciones:

- Definición relacionada con la actividad. Según esta definición, un estudiante abandona el curso cuando no interactúa con el curso durante más de dos semanas consecutivas. Dado que puede haber periodos donde los estudiantes no tengan que interactuar necesariamente (p. ej., en festivos y/o en periodos en los que el profesor se centra en actividades que no están en el SPOC), las semanas en las que menos de un 10 % de los estudiantes interactúan no se consideran para el cálculo del abandono.

- Definición relacionada con la completitud. Según esta definición, un estudiante abandona cuando no completa (entendiendo completar como realizar la actividad y enviarla, independientemente del resultado) la mayor parte de las actividades del curso. En particular, se considera que un estudiante completa el curso (y no abandona) cuando completa el 75 % de las actividades, aunque este umbral puede ser ajustado dependiendo del contexto. En particular, se ha elegido este umbral, ya que es usado por otras plataformas conocidas, como MiríadaX¹, que requiere completar el 75 % de los módulos (aunque requiere también aprobarlos, lo cual no se contempla aquí) para obtener el certificado de participación. Esta definición difiere de la anterior en que se centra en que los estudiantes completen las actividades (independientemente del momento), mientras que la anterior se centra en acceder al SPOC (sin importar si se realizan las actividades o no).

Para predecir las variables dependientes mencionadas, se utilizan diferentes variables independientes. Entre estas variables, aparte de incluir las variables típicas sobre interacciones con vídeos y ejercicios, y actividad en la plataforma, se incluyen nuevas variables. En particular, se analiza el efecto de la persistencia (obtenida a través de eventos) y de las variables sobre la tipología del curso. Estas últimas variables tienen como objetivo que los modelos sean capaces de adaptarse a cada contexto y se analizarán en el capítulo siguiente. La lista completa de variables independientes utilizadas en este contexto es la siguiente:

- Variables sobre interacciones con vídeos: `perc_vabiertos`, `perc_vtotal`, `perc_compl`, `avg_rep` y `avg_pausas`.
- Variables sobre interacciones con ejercicios: `perc_intentados`, `avg_intentos`, `media_intentados`, `media_intentados_pr`, `perc_correctos`, `CFA`, y `racha_ej`.
- Variables sobre actividad en la plataforma: `racha_acc`, `perc_dias` y `avg_con`.
- Variables de alto nivel sobre comportamientos de alumnos: `persistencia`.
- Variables sobre la tipología del curso: `area`, `sincrono`, `nejercicios`, `nvideos`, `avg_duracion`, `num_personal` e `ingles`.

5.7. Escenario 6. Plataforma para alumnos de primaria y secundaria

El último de los escenarios considerados en esta tesis doctoral (escenario 6), tratado en [114], incluye datos de estudiantes de primaria y secundaria (lo que se conoce como estudiantes K-12) de centros españoles. Estos alumnos se conectan a una plataforma y acceden, a través de servicios web, a actividades que les sirven de apoyo a las clases presenciales. Entre los contenidos ofrecidos, hay también algunos recursos explicativos, aunque la traza que se genera es exclusivamente sobre el resultado de los ejercicios realizados. Por ello, este contexto, trata el análisis de un contexto, donde a diferencia de los anteriores, solo hay traza de ejercicios, lo que asemeja este contexto a un ITS en cuanto a información disponible. Teniendo esto en cuenta, se recogieron datos del curso académico 2017-2018 (de septiembre 2017 a

¹<https://miriadax.net/es/faq?faqid=8635212>

junio 2018). Estos datos incluían los registros de 10 171 alumnos, que pertenecían a 480 cursos de diversos centros educativos de primaria y secundaria. En este caso, el término curso se refiere a un grupo de estudiantes que cursan la misma asignatura (p. ej., Ciencias Sociales) en el mismo nivel (p. ej., 4º de primaria) y centro educativo. Por tanto, el término curso no se refiere al nivel del alumno, que es comúnmente conocido como curso. De este modo, un ejemplo de curso sería “Ciencias Sociales de 4º de educación secundaria del centro X”.

La metodología de cada curso concreto no se ha proporcionado, pero sí se conoce la información sobre los ejercicios que realizan en la plataforma educativa. Sin embargo, dichos ejercicios no son necesariamente sumativos y pueden tener un nivel de importancia diferente, dependiendo del profesor. Por ejemplo, puede haber un profesor que pida a sus estudiantes que realicen todas las actividades cada semana, mientras que puede haber otro profesor que quiera centrarse solo en determinadas actividades para reforzar conceptos concretos. El hecho de desconocer la metodología del curso puede afectar al tratamiento de ciertas variables en los modelos predictivos (p. ej., la nota media de todos los ejercicios pierde parte del sentido si el profesor solo pide realizar un subconjunto). Asimismo, el hecho de que el resultado de estos ejercicios no tenga que ver con la nota de la asignatura hace que lo que se pueda predecir es el desempeño de los alumnos en la plataforma. De este modo, al igual que en el escenario anterior, puede haber un estudiante que no haga nunca las actividades de la plataforma y/u obtenga un mal rendimiento, pero luego pueda aprobar los exámenes. Sin embargo, si un estudiante realiza muy pocas actividades en comparación con sus compañeros del curso, puede ser un problema (en este caso, aunque el profesor no mande muchas actividades, el alumno no está realizando las tareas que se le piden). Por ello, el análisis de la predicción en este escenario es interesante y puede aportar información valiosa al profesor sobre lo que el alumno va a hacer en la plataforma educativa. Además, probablemente, cuanto más uso se dé a la plataforma, se podrá obtener una información más útil sobre qué estudiantes pueden tener dificultades en el curso y sobre cuáles debe intervenir el profesor para ayudarles.

Respecto a los datos, se han recogido a partir de la información de la traza de cada actividad, que tiene formato Marsupial, como se mencionó en la sección 3.2. En este caso, se han incluido todos los estudiantes que tenían traza en la plataforma, por lo que únicamente quedan excluidos aquellos potenciales estudiantes que no realizaron ninguna actividad en la plataforma. Sobre estos últimos estudiantes, al no interactuar en la plataforma, no se recoge traza de ellos y por tanto, no se tiene constancia siquiera de su presencia en el curso, por lo que son excluidos.

En cuanto a las variables dependientes, en este escenario se considera el abandono y el éxito académico. Para el caso del abandono, en este caso no se puede definir un criterio basado en la actividad, ya que al no conocer la información sobre la metodología, no se sabe cuándo el estudiante debería interactuar con la plataforma. Por ejemplo, puede haber cursos donde el profesor no use la plataforma para ciertos módulos, lo que puede hacer que los alumnos no presenten actividad durante cierto tiempo.

Por ello, el abandono se define en base a las actividades completadas. La diferencia de este contexto es que no se puede establecer un porcentaje mínimo de actividades completadas, como por ejemplo, se realizó en el escenario de los SPOCs de la UC3M. El motivo también reside en la metodología. Si un profesor utiliza la

plataforma únicamente en una parte del curso sería normal que los estudiantes no tengan un alto porcentaje de ejercicios completados. Por ello, se propone el uso de una variable para mitigar este efecto: el número de ejercicios vistos. Esta variable indica el número de ejercicios que los estudiantes deberían ver (e intentar) en un curso. Para el cálculo de esta variable, hay que definir un umbral que indique el porcentaje de estudiantes que deben intentar un ejercicio para que el ejercicio sea incluido (se considere que los alumnos deberían hacer ese ejercicio). Los valores extremos para este umbral serían considerar un ejercicio como visto si lo realiza al menos un estudiante, o si lo realizan todos. Sin embargo, estas opciones pueden no ser las más apropiadas. Por un lado, es altamente improbable que todos los alumnos realicen los ejercicios, y por otro, puede haber estudiantes que quieran explorar otros ejercicios. En este último caso, si se considera un ejercicio como visto si lo realiza un estudiante, todos los ejercicios que explora ese estudiante deberían ser realizados por todos. Por ello, se ha definido un criterio de que un ejercicio se considera como visto si lo intentan al menos el 15 % de los estudiantes. Este umbral es considerado para incluir todos los ejercicios posibles y excluir solo aquellos ejercicios que se realizan por estudiantes que quieren explorar el curso (considerando que si al menos el 15 % realiza el ejercicio, no es porque un alumno lo realice por explorar el curso).

Teniendo esto en cuenta, se considera que un estudiante ha abandonado cuando no intenta al menos el 75 % de los ejercicios que intentan al menos el 15 % de los alumnos. En esta definición, se toma el umbral del 75 %, siguiendo el mismo criterio que en el anterior escenario. Además, este umbral es razonable porque es normal que no todos los ejercicios se hagan por el alumno. Por ejemplo, puede haber ejercicios que sean opcionales y dado que esta información es desconocida, el umbral del 75 % da cierto margen por esos posibles ejercicios.

Respecto a la variable dependiente del éxito académico, se ha medido considerando la media de todos los ejercicios. Para este cálculo, también se ha considerado el número de ejercicios vistos, por lo que el éxito en realidad se calcula como el producto de la nota media de los ejercicios intentados por el porcentaje de ejercicios intentados (teniendo en cuenta que el total de ejercicios solo incluye los ejercicios vistos por al menos el 15 % de los estudiantes).

Para realizar los análisis, se han considerado diferentes variables independientes, aunque todas ellas obtenidas a partir de los eventos relacionados con los ejercicios. Algunas de estas variables se han normalizado respecto al número de ejercicios vistos. Por ejemplo, el porcentaje de ejercicios intentados en vez de referirse al total de ejercicios de la plataforma se refiere al número de ejercicios vistos, o el número de ejercicios enviados (`nacesos`) se ha dividido entre el número de ejercicios vistos para que la medida quede normalizada en distintos cursos (si en un curso se hacen más ejercicios habrá más ejercicios enviados, pero al normalizar los ratios serán más similares). Aquellas variables que se han normalizado con los ejercicios vistos han sido marcadas con (*). Teniendo esto en cuenta, la lista de variables independientes utilizadas es la siguiente.

- Variables sobre actividad en la plataforma: `ndias`.
- Variables sobre interacciones con ejercicios: `nacesos` (*), `tiempo_medio` (en min.), `perc_intentados` (*), `avg_intentos`, `nota_media` (*), `media_primer` (*), `perc_correctos_total` (*), `CFA` (*).

- Variables de alto nivel sobre comportamientos de alumnos: *eficiencia* y *constancia*⁻¹.
- Otros: estado de abandono (*estado_dr*). Esta variable, utilizada en este escenario, indica si un estudiante tiene la condición de abandono (llevar menos del 75 % de ejercicios intentados respecto a los ejercicios vistos) en un instante determinado X, anterior al final del curso, que es cuando se mide la variable dependiente. Esta variable permite determinar si el hecho de llevar menos del 75 % de los ejercicios en un periodo implica que el alumno no los vaya a tener hechos al final.

5.8. Similitudes y diferencias entre los diferentes escenarios

A lo largo del capítulo, se han ido presentando los diferentes escenarios que se utilizan en esta tesis. Para finalizar este capítulo, se realiza un resumen de las características de todos ellos, así como de las similitudes y diferencias. En cuanto a las similitudes, todos los escenarios comparten que hacen uso de alguna plataforma educativa online y por tanto, todos los datos utilizados son a partir de traza de las plataformas, y no de datos obtenidos por parte del profesor durante las clases. También, todos los escenarios comparten que hay información sobre ejercicios que los alumnos realizan en la plataforma.

Sin embargo, hay diferencias significativas entre los escenarios. Por un lado, tenemos el tipo de plataforma utilizada. En unos casos, tenemos MOOCs, que son cursos con muchos estudiantes y que pueden proceder de cualquier parte del mundo. En otros, tenemos SPOCs, que tienen un ámbito más reducido, aunque hay diferencias importantes entre ellos. Mientras que los de KU Leuven tienen un ámbito de todo Flandes y están dirigidos a estudiantes de último curso de educación secundaria, los de la UC3M están dirigidos a sus propios alumnos. Además, la metodología utilizada en cada uno de ellos es diferente. Y por último, existen los datos de alumnos de primaria y secundaria (alumnos K-12), que por tanto, están dirigidos a un perfil de estudiantes muy diferente. Aparte del tipo de plataforma utilizada, existen diferencias en la metodología docente y en la temática y duración de los cursos. Por ejemplo, hay unos cursos con modo de instrucción síncrono y otros con modo asíncrono, hay cursos que pueden ser simplemente de apoyo a clases presenciales, etc. Por lo que puede haber diferencias en la predicción también por la metodología. Todos estos hechos hacen que cada escenario sea particular y relevante para su análisis.

Además, y de especial relevancia, están las diferencias en cuanto a los datos disponibles, ya que no todos los escenarios permiten recoger el mismo tipo de información. Por ejemplo, solo se dispone de datos respecto al foro en los MOOCs de edX, ya que por ejemplo, el uso de los foros en los SPOCs no es tan habitual, ya que en muchos casos hay también interacción presencial, y en los casos analizados, no había ese tipo de interacciones. Asimismo, la plataforma para alumnos de primaria y secundaria solo recoge información sobre ejercicios, ya que es un entorno más cercano a los ITSS y por tanto, se dispone de menos tipos de información en recursos para predecir, aunque la información que se produce en los ejercicios puede ser más rica que en otros escenarios. También, hay variables que aunque potencialmente se pudieran obtener en varios contextos, se han probado únicamente en ciertos escenarios para comenzar a analizar su efecto. Por ejemplo, las variables de SRL

solo están disponibles en los MOOCs de Coursera, en los cuales se quería analizar este efecto en particular, por su relevancia, dado que los cursos eran asíncronos. Sin embargo, este tipo de variables se podrían incorporar en el futuro en otros cursos.

Del mismo modo, la definición de las variables dependientes varía según el contexto. Aunque en todos los escenarios se tratan variables similares, la definición debe adaptarse en cada caso. Por ejemplo, el éxito en un MOOC es conseguir aprobarlo, mientras que en el SPOC de KU Leuven el objetivo no era tener muy buenos resultados en el SPOC, sino aprobar la prueba de acceso a la universidad. Incluso dentro de los propios MOOC, la tasa para aprobar es diferente entre los cursos. También el abandono es una variable que depende mucho del contexto y el usar una definición relacionada con inactividad o con completitud también depende del caso.

Por todos estos motivos, los escenarios presentados en la tesis son bastante diferentes atendiendo a realidades educativas típicas y los modelos predictivos deben desarrollarse en cada uno de ellos para entender cómo puede variar la predicción de un contexto a otro y qué conclusiones globales se pueden extraer. La Tabla 5.4 muestra un resumen de todas las características de todo los escenarios descritos en este capítulo de una forma comparativa.

TABLA 5.4
RESUMEN DE LOS ESCENARIOS ANALIZADOS EN LA TESIS

Dato	Escenario 1	Escenario 2	Escenario 3	Escenario 4	Escenario 5	Escenario 6
Tipo curso	MOOC	MOOC	MOOC	SPOC	SPOC	Online
Plataforma	edX	edX	Coursera	edX Edge	Open edX	Propietaria
Institución	UC3M	HKUST	PUC	KU Leuven	UC3M	Omitido
Ámbito	Mundial	Mundial	Mundial	Flandes	UC3M	España
Idioma	Inglés	Inglés	Español	Flamenco	Español / Inglés	Varios posibles
Alumnos	95 555	47 178	54 935	2193	1534	10 171
Nº cursos	1	1	3	4	17	480
Modo instrucción	Síncrono	Síncrono	Asíncrono	Síncrono	Síncrono	Síncrono
Duración	5 semanas	10 semanas	4-9 semanas	1 curso (10 - 11 meses)	16 semanas	1 curso (10 meses)
Tipos de variables independientes						
Vídeos	X	X	X	X	X	
Ejercicios	X	X	X	X	X	X
Actividad		X	X	X	X	X
Foro	X	X				
SRL			X			
Comportamientos					X	X
Tipología curso					X	
Variables dependientes						
Notas	X	X	X			X
Éxito	X	X	X	X	X	
Abandono			X		X	X

6

Efecto de las variables predictoras

Contenido

6.1. Variables sobre interacciones con vídeos, ejercicios y actividad en la plataforma	82
6.1.1. Análisis de los escenarios 1 y 2	83
6.1.2. Análisis del escenario 3	85
6.1.3. Análisis del escenario 4	86
6.1.4. Análisis del escenario 5	87
6.1.5. Análisis del escenario 6	89
6.2. Notas previas de la evaluación sumativa del curso	90
6.3. Variables de la actividad del alumno en el foro	92
6.4. Variables sobre la autorregulación del aprendizaje	94
6.5. Variables sobre comportamientos del alumno	97
6.5.1. Efecto de la persistencia en la predicción	98
6.5.2. Efecto de la eficiencia y constancia en la predicción	100
6.6. Tipo de metodología de recogida de datos	102
6.7. Lecciones aprendidas	105

Uno de los factores principales que afectan a los modelos predictivos es el conjunto de variables predictoras utilizadas en los modelos. Hay ciertas variables como la nota de los alumnos en ejercicios que ya han sido tratadas ampliamente en el estado del arte como variables predictoras, pero de las que se pueden extraer más conclusiones, y hay otras menos exploradas o que no se han tratado antes en la literatura. Sobre estas variables, es relevante conocer si pueden mejorar los modelos predictivos al usar otras variables y qué poder predictivo pueden conseguir por sí mismas, ya que en algunos casos, puede haber otras variables que no estén disponibles por diversos motivos. Este capítulo se centra en analizar el poder predictivo y el efecto de distintos tipos de variables en los modelos. El capítulo se divide en siete secciones para analizar diferentes tipos de variables predictoras.

Las cinco primeras secciones concretamente analizan el efecto de las (1) variables sobre interacciones con vídeos, ejercicios y actividad en la plataforma, (2) notas previas de la evaluación sumativa, (3) variables de la actividad en el foro, (4) variables sobre SRL, y (5) variables sobre comportamientos del alumno, tales como la persistencia, eficiencia y constancia. La sexta sección analiza dos estrategias de recogida de datos en un MOOC para la realización de la predicción, la de recoger todos los datos desde el inicio del curso hasta la semana actual, o la de utilizar únicamente los datos recientes. Finalmente, la séptima sección muestra las lecciones aprendidas respecto a las variables predictoras teniendo en cuenta las enseñanzas de las diferentes experiencias.

6.1. Variables sobre interacciones con vídeos, ejercicios y actividad en la plataforma

La primera sección analiza el efecto de las variables relacionadas con las interacciones con vídeos, ejercicios y actividad en la plataforma, combinadas todas ellas a la vez. Este tipo de variables se han utilizado ampliamente en la literatura (p. ej., [28], [128]) y en la mayoría de los casos, se ha concluido que estas variables pueden proporcionar un buen poder predictivo (p. ej., [28], [128]). Por ello, el objetivo de esta sección, es corroborar mediante nuevos escenarios que se cumple que la combinación de este tipo de variables ofrece buenas predicciones, y proporcionar unos primeros modelos globales. Estos modelos podrán servir posteriormente para comparar con otros modelos que incluyen otros tipos de variables concretas que se analizan (variables de foro, SRL, etc.) y/o para tomarse como base para otros análisis posteriores, como el análisis temporal o la generalización.

Para este análisis, se consideran los seis escenarios de la tesis, ya que en todos se han podido obtener variables sobre la actividad en la plataforma y/o interacción con vídeos y ejercicios. No obstante, cada escenario tiene sus particularidades y en cada uno las variables a predecir pueden variar. Además, las variables predictoras concretas también pueden variar. Las variables independientes se han seleccionado, teniendo en cuenta variables con alto poder predictivo del estado del arte (p. ej., [30], [47]) y trabajos previos [261], [262]. Sin embargo, existen diferencias porque los datos disponibles eran diferentes en cada escenario, lo que condicionaba las posibles variables, y porque teniendo en cuenta el estado del arte, se han realizado algunas variaciones para comparar algunos aspectos específicos y corroborar que las variables utilizadas ofrecen un alto poder predictivo, aun siendo ligeramente distintas.

La sección se estructura en varias subsecciones para abordar cada uno de los escenarios. Para guiar el desarrollo de la sección, se describen los elementos comunes y diferencias entre los análisis realizados en los distintos escenarios para los modelos de esta sección.

- Escenarios 1 y 2: presentan la predicción de notas parciales en MOOCs, y para ello utilizan las mismas variables y metodología.
- Escenario 3: presenta la predicción de abandono en un MOOC. Las variables son diferentes a los casos anteriores al utilizarse otra plataforma. Al ser un curso asíncrono, en lugar de realizar predicciones parciales teniendo en cuenta los plazos de cada tarea, se considera el abandono a partir de los datos del curso.

- Escenario 4: presenta la predicción del éxito en una prueba de acceso (externa a la plataforma), por lo que la predicción es diferente a los casos anteriores porque lo que se mide es sobre algo externo, a diferencia de los MOOCs, en los que todas las interacciones ocurren dentro de la plataforma. En este caso, se analiza el poder predictivo justo antes de la prueba para ver cuál es el poder predictivo de todas las interacciones del curso. Algunas variables utilizadas coinciden con el escenario 1 y 2, al proceder los datos también de una plataforma Open edX, pero hay diferencias debido al contexto y a los datos disponibles en cada escenario.
- Escenario 5: presenta la predicción del éxito y abandono en SPOCs. Dado que las variables dependientes no se refieren a actividades externas (prueba externa o pruebas sumativas diferenciadas), sino que a las actividades del curso (sin ninguna diferenciación), se predice a mitad de curso, ya que a partir de las variables a final de curso se pueden derivar las variables dependientes. Las variables utilizadas son similares a las del escenario 4.
- Escenario 6: presenta la predicción del abandono y notas, siguiendo una aproximación similar al escenario 5, en cuanto a que las variables se miden respecto a las actividades de la plataforma, por lo cual se predice a mitad de curso para ver el efecto en un instante anterior. Sin embargo, los datos son diferentes al escenario 5 y por tanto, las variables utilizadas.

Aparte de las características mencionadas, es importante notar que dependiendo de la variable dependiente utilizada, la métrica utilizada es el AUC o RMSE (como se justificó en la sección 3.4). De este modo, el poder predictivo de la predicción de la nota se mide en RMSE, y el del abandono y éxito en AUC. Teniendo en cuenta todo lo comentado anteriormente, en las siguientes subsecciones se describen los resultados obtenidos para cada escenario con modelos de predicción que incluyen variables sobre interacciones con vídeos, ejercicios y actividad en la plataforma.

6.1.1. Análisis de los escenarios 1 y 2

En primer lugar, se considera el escenario 1 (MOOC de la UC3M), analizado en [261] y en un TFM previo [262]. En este TFM se presentó el modelo que se presenta en esta sección, aunque aquí se pone dicho modelo en contexto con otros escenarios de la tesis y se compara. Con el escenario 1, el objetivo es predecir la nota de las siete evaluaciones, de las cuales cinco eran tipo test (numeradas como T_i , siendo i el número de la prueba, que corresponde con la semana del curso) y dos eran tareas de programación (P_3 y P_5 , ya que estas tareas se encuentran en las semanas 3 y 5). Además, se predice la nota final del curso (NF). Para la predicción, se consideran únicamente los datos de las interacciones pertenecientes a las semanas del curso en las que se realiza la prueba (y eliminando las interacciones que ocurran después de la fecha límite, si las hubiera). Para el análisis, se ha utilizado un modelo, que incluye las variables de vídeos y ejercicios (ver variables concretas en la sección 5.2). Los resultados de este modelo se presentan en la Tabla 6.1.

Una observación inicial es que las primeras pruebas de evaluación son más difíciles de predecir [262], como ocurre en otros trabajos (p. ej., [27], [153]). Los resultados (dados como RMSE) muestran que el poder predictivo para predecir el primer test es de 0.26, mientras que es de 0.16 en el último (tomando el mejor valor, que se

TABLA 6.1
 PREDICCIÓN DE NOTAS CON VARIABLES DE EJERCICIOS Y VÍDEOS EN ESCENARIO 1

Algoritmo	T_1	T_2	T_3	T_4	T_5	P_3	P_5	NF
RG	0.27	0.22	0.20	0.18	0.16	0.25	0.20	0.14
SVM	0.27	0.22	0.21	0.19	0.17	0.25	0.21	0.15
DT	0.34	0.28	0.26	0.22	0.18	0.31	0.27	0.16
RF	0.26	0.21	0.20	0.19	0.16	0.25	0.21	0.14

Nota: El mejor poder predictivo (medido en **RMSE**) de cada columna se marca en negrita

obtiene con **RF** en ambos casos). Dado que únicamente se dispone de datos de la semana en la que está la prueba calificable, este resultado puede implicar que los alumnos interactúan en las primeras semanas del curso no están necesariamente comprometidos con las pruebas sumativas. Sin embargo, aquellos que interactúan en las últimas semanas tienen unos comportamientos más predictivos, y por ello puede mejorar el poder predictivo.

En cuanto a los resultados obtenidos en la predicción, mejoran a partir de la semana 3, ya que se consigue tener un **RMSE** no superior a 0.20, que indica que aunque no se pueda predecir la nota exacta, al menos se podría acertar en el rango de la misma. Para poner el valor del **RMSE** en contexto, en otros trabajos, como [27] se alcanzó un **RMSE** de entre 0.14 y 0.15 en las mejores tareas, y Pérez-Sanagustín et al. [271] consiguió un mejor **RMSE** de 0.18 al final de curso cuando predijeron la nota final. Por ello, en comparación con otros trabajos, los resultados al predecir la nota final o los T_4 y T_5 son buenos, pero el poder predictivo en general es mejorable. En este caso, hay que tener en cuenta que prácticamente se utilizan solo variables sobre ejercicios, ya que la única variable sobre vídeos disponible es `perc_vabiertos`, que es menos representativa al indicar solo los vídeos que se abren y no los que realmente se completan. Respecto a los algoritmos, los mejores son **RF** y **RG**. En las primeras semanas, **RF** obtiene mejores resultados, pero luego **RG** obtiene los mejores valores de **RMSE**. No obstante, las diferencias entre un algoritmo y otro son pequeñas.

En el escenario 2 (MOOC de **HKUST**, tratado en [263]), se desarrolló un modelo utilizando las mismas variables que en el escenario 1 (variables relacionadas con ejercicios y `perc_vabiertos`). En este caso, se predijeron las notas de los seis laboratorios (L_i), situados en las semanas S_i del curso, la nota de la componente del proyecto de corrección automática (PA), la parte del proyecto evaluada mediante revisión por pares (PR) y la nota del examen final (EF). Los resultados de los modelos predictivos con las variables mencionadas para el escenario 2 se muestran en la Tabla 6.2.

A la vista de los resultados de la tabla, se pueden obtener unas conclusiones similares a las obtenidas en el escenario 1. Es decir, se observa cómo el poder predictivo mejora significativamente a lo largo de las pruebas y el poder predictivo en las últimas es mejor. Este hecho puede ser debido, al igual que se comentó en el escenario 1, por el hecho de que al inicio del curso, muchos estudiantes exploran al curso pero no tienen intención de completarlo. Esto se corrobora al comprobar que, mientras que 13 005 y 10 826 alumnos realizan ejercicios en la semana 1 en los escenarios 1 y 2, respectivamente, esta cifra baja a 4719 y 5162 en la semana 2, y

TABLA 6.2
 PREDICCIÓN DE NOTAS CON VARIABLES DE EJERCICIOS Y VÍDEOS EN ESCENARIO 2

Algoritmo	L_1 W_1	L_2 W_2	L_3 W_3	L_4 W_5	L_5 W_7	L_6 W_8	PA W_9	PR W_9	EF W_{10}
RG	0.33	0.33	0.28	0.30	0.26	0.28	0.26	0.25	0.17
SVM	0.36	0.36	0.29	0.33	0.27	0.34	0.29	0.26	0.19
DT	0.38	0.41	0.33	0.36	0.32	0.31	0.28	0.27	0.18
RF	0.34	0.36	0.30	0.33	0.28	0.30	0.27	0.26	0.18

Nota: El mejor poder predictivo (medido en **RMSE**) de cada columna se marca en negrita

baja a 1609 y 1443 en la última semana del curso. Este hecho, además, es bastante común en MOOCs de la literatura (p. ej., [272]).

Respecto al poder predictivo obtenido en el escenario 2, no es muy alto salvo en el caso del examen final. Un posible motivo puede ser por el tipo de preguntas. En el MOOC de HKUST las evaluaciones son mayoritariamente de laboratorio (orientación a laboratorios), mientras que el MOOC de la UC3M está más orientado a conceptos.

También, en cuanto a los algoritmos predictivos, los resultados muestran que el algoritmo que mejor predice es RG, aunque la diferencia con otros algoritmos es pequeña. Esto muestra que es posible obtener buenos resultados con un algoritmo simple como RG y que las variables relacionadas con ejercicios pueden tener una alta relación lineal con la nota. Es decir, que si un alumno tiene buenos resultados en los ejercicios formativos, es probable que tenga buenos resultados en los sumativos.

A la vista de los dos escenarios anteriores, se ha comprobado que para la predicción de las notas parciales, las variables sobre interacción con ejercicios pueden ser muy buenas predictoras y suficientes para predecir (como en trabajos previos, como [27], [28]), si bien es posible que para algunas pruebas iniciales, la cantidad de datos sea limitada y el poder predictivo obtenido sea peor.

6.1.2. Análisis del escenario 3

Para continuar el análisis de las variables sobre interacción con vídeos, ejercicios y actividad en la plataforma, se considera el escenario 3 (MOOCs de Coursera), tratado en [91]. En este caso, al incluir cursos asíncronos en los que cada alumno puede llevar su ritmo y en los que está permitido repetir las pruebas indefinidamente si no se superan, ya no es tan interesante saber la nota de cada prueba en un intervalo de tiempo como en los escenarios anteriores, sino conocer si el alumno al final va a completar todas las pruebas. Por ello, en este caso, se considera el análisis de predicción del abandono. Para este análisis, se han utilizado los datos de Coursera, y en particular, se han utilizado variables sobre la actividad del alumno en la plataforma, y las interacciones con vídeos y ejercicios (ver variables concretas en cada caso en la sección 5.4). Teniendo en cuenta estos datos, se han realizado modelos predictivos usando todos los datos de cada categoría por separado, para ver qué poder predictivo pueden alcanzar los diferentes grupos de variables. Los resultados se muestran en la Tabla 6.3.

A la vista de los resultados, las variables que ofrecen un mejor poder predictivo son las variables relacionadas con los ejercicios en los tres MOOCs (lo cual es coherente con otros trabajos, como [23], [28]). Este hecho es razonable, ya que los

TABLA 6.3
 PREDICCIÓN DE ABANDONO EN EL ESCENARIO 3

Variables	Electrones en acción				Aula Constructivista				Gestión de Organizaciones Efectivas			
	RG	SVM	DT	RF	RG	SVM	DT	RF	RG	SVM	DT	RF
Actividad	0.94	0.89	0.86	0.92	0.93	0.96	0.92	0.95	0.94	0.94	0.87	0.94
Vídeos	0.93	0.92	0.91	0.93	0.97	0.97	0.92	0.97	0.94	0.97	0.92	0.97
Ejercicios	0.96	0.96	0.96	0.96	0.98	0.98	0.97	0.99	0.97	0.97	0.97	0.97

Nota: El mejor poder predictivo (medido en AUC) de cada curso se marca en negrita

modelos están realizados al final del periodo de recogida de datos y aquellos alumnos que tengan un porcentaje bajo de ejercicios realizados necesariamente deben haber abandonado. Sin embargo, este tipo de variables son muy relevantes al realizar el análisis temporal, ya que pueden ver, por ejemplo, si el hecho de que un alumno haga pocos ejercicios al inicio tendrá o no relevancia sobre lo que ocurra el final. De hecho, pueden alcanzar muy buen poder predictivo desde el inicio (p. ej., un AUC de 0.82 desde la primera semana en “Electrones en Acción”). En cuanto a las variables de actividad y vídeos, éstas tienen también un alto poder predictivo sobre el abandono en los tres MOOCs. Esto quiere decir que puede existir una alta relación entre quién ve los vídeos y quién realiza las pruebas de evaluación (como se observó en [124]), y entre quién accede al curso y quién realiza las pruebas (como se concluyó en [133]).

6.1.3. Análisis del escenario 4

Hasta ahora, los resultados obtenidos son en escenarios de MOOCs, en los que toda la actividad del alumno ocurre dentro del curso. El escenario 4 es diferente (SPOCs de KU Leuven, tratado en [29], [269]), ya que es un contexto de SPOCs y lo que se mide es si el alumno aprobará la prueba de acceso a la universidad, que es externa al curso. Este hecho puede afectar a la predicción, ya que además, el alumno no solo aprende los contenidos en el SPOC, sino en las clases de secundaria, lo que puede afectar en la predicción.

Debido a estos condicionantes por el contexto, en el escenario 4, se han realizado inicialmente las predicciones a final de curso, para ver si toda la información del SPOC tiene un alto poder predictivo en la prueba externa. Los datos utilizados en este caso son sobre vídeos, ejercicios y actividad en la plataforma (ver variables concretas en la sección 5.5), utilizados de forma conjunta en un solo modelo. Para comenzar el análisis, se realizó un modelo (por cada algoritmo) con los datos del SPOC de Química del 2016/2017. Los resultados obtenidos de dicho modelo por cada algoritmo (medidos en AUC) son: 0.87 para RF, 0.85 para SVM, 0.80 para DT y 0.77 para RG. Por tanto, el RF es el algoritmo que mejor funciona, lo cual es consistente con los escenarios previos. Sin embargo, la RG ofrece un poder predictivo moderado y es el peor de todos, al contrario de lo que ocurría en los escenarios previos. Esto implica que la relación entre las interacciones en el SPOCs y el éxito en la prueba de acceso tiene una menor linealidad.

De cara a validar los resultados con otros SPOCs, se han desarrollado modelos predictivos en los otros tres cursos utilizando RF y al final de curso. Además, se han generado modelos con la combinación de las interacciones de los SPOCs de Física y Química del mismo cohorte, ya que los alumnos son los mismos en ambos y la

variable dependiente también. Los resultados de estos modelos y se muestran en la Tabla 6.4.

TABLA 6.4
PODER PREDICTIVO EN LOS CURSOS DEL ESCENARIO 4

Curso	AUC	Curso	AUC
Química 2016/2017	0.87	Química 2017/2018	0.77
Física 2016/2017	0.87	Física 2017/2018	0.86
Física y Química 2016/2017	0.88	Física y Química 2017/2018	0.84

Los resultados muestran que es posible obtener un poder predictivo bueno (AUC superior a 0.8) en todos los casos, salvo en el SPOC de Química del 2017/2018, cuyo AUC es 0.77. Esto significa que las variables de ejercicios, vídeos y actividad en la plataforma tienen un buen poder predictivo en conjunto, pero es peor que en los escenarios previos. Uno de los motivos principales puede ser debido al hecho de que en este escenario, a diferencia de los anteriores, la variable dependiente se mide respecto a una prueba externa. Aunque los contenidos de la prueba se asemejan a los del SPOC, al ser una prueba única y externa, puede haber varios factores que condicionen el resultado, teniendo en cuenta que la prueba suele ser bastante dura y la tasa de éxito baja [270] (en torno al 20%). Además, en este escenario, al ser el SPOC un apoyo al contenido visto en secundaria, el alumnado puede hacer un uso diferente. No obstante, los resultados muestran un alto poder predictivo a partir de las interacciones del SPOC, lo que demuestra que la actividad con el SPOC puede ser muy relevante para tener éxito en las pruebas de acceso.

En este escenario, una limitación fue que la muestra de cada curso era limitada, ya que no se disponía del resultado del examen de todos los estudiantes. Sin embargo, puede haber escenarios en los que lo único que se dispone son los datos de la plataforma (sin ningún dato de resultados externos), por lo que el análisis se debe realizar únicamente en base a la misma. Este es el caso de los escenarios 5 y 6, que se aborda en las siguientes subsecciones.

6.1.4. Análisis del escenario 5

El siguiente escenario analizar es el escenario 5, que es el de los SPOCs de la UC3M. En este escenario, una particularidad es que, a diferencia del escenario 4, y al igual que ocurre con el escenario 6 (que se aborda en la siguiente subsección), no se sabe la nota de los alumnos obtenida en los cursos, sino únicamente lo obtenido en la plataforma. Por ello, en estos casos, la variable a predecir es sobre lo que ocurre en la plataforma. Dado que al final de curso ya se dispone de todos los datos de la plataforma y la variable dependiente podría conocerse, en estos casos, la predicción se realiza a mitad de curso (aunque se realizará un análisis mayor en el tiempo en el capítulo siguiente), al ser un momento del curso donde la anticipación tiene sentido y donde se esperaría que el poder predictivo fuera adecuado.

Teniendo esto en cuenta, en el escenario 5, las variables dependientes son el abandono y el éxito académico. En concreto, el abandono se definió según la actividad y la completitud de las actividades (ver sección 5.6). En este escenario, se han realizado unos modelos utilizando las variables de vídeos, ejercicios y actividad por separado y agrupándolas (ver variables concretas en la sección 5.6). Para el análisis se han considerado los 16 SPOCs que se impartieron durante el primer

cuatrimestre del curso 2018/2019 y se han elaborado modelos globales con todos los cursos a mitad de curso (semana 8 de 16). Los resultados obtenidos en los modelos, para los cuatro algoritmos considerados, se muestran en la Tabla 6.5.

TABLA 6.5
RESULTADOS DE LA PREDICCIÓN EN EL ESCENARIO 5

Variables	Abandono (actividad)				Abandono (completitud)				Éxito			
	RG	SVM	DT	RF	RG	SVM	DT	RF	RG	SVM	DT	RF
Actividad	0.74	0.73	0.71	0.72	0.74	0.70	0.72	0.71	0.69	0.66	0.65	0.68
Vídeos	0.71	0.71	0.66	0.71	0.69	0.69	0.63	0.69	0.65	0.67	0.64	0.70
Ejercicios	0.73	0.73	0.70	0.70	0.91	0.91	0.89	0.93	0.90	0.89	0.86	0.93
Todas	0.77	0.77	0.69	0.77	0.92	0.91	0.87	0.92	0.91	0.91	0.86	0.93

Nota: El mejor poder predictivo (medido en *AUC*) de cada columna se marca en negrita

Los resultados muestran que en cuanto al abandono relacionado con la actividad, el mejor tipo de variables son las relacionadas con la actividad, lo cual tiene sentido ya que la definición de abandono está relacionada con la actividad/inactividad de los alumnos en el curso. No obstante, el poder predictivo es muy similar al obtenido con otros tipos de variables, lo que implica que tanto las variables de vídeos como ejercicios pueden ser buenos predictores. El mejor modelo es el que combina todas las variables, por lo que todos los tipos pueden aportar al modelo. En cuanto a los algoritmos, se observa que hay también muy poca diferencia, al igual que lo ocurrido en otros escenarios, aunque el *DT* ofrece un peor poder predictivo. Otra observación es que el poder predictivo a mitad de curso solo es moderado (entre 0.7 y 0.8), por lo tanto las predicciones sobre el abandono relacionado con la actividad no son muy precisas en este momento. En el capítulo siguiente se analizará la evolución temporal para ver el mejor momento para empezar a predecir.

Respecto al abandono relacionado con la completitud, se observa que las variables más relevantes son las relacionadas con los ejercicios, y las relacionadas con actividad y vídeos tienen peor poder predictivo. Para la predicción de este tipo de abandono, las variables de ejercicios tienen mayor relación porque lo que se mide es que el alumno complete un porcentaje de ejercicios, no que se conecte frecuentemente a la plataforma (como en el caso anterior). Sin embargo, la principal diferencia es que en el caso anterior, los tres tipos de variables tenían un poder predictivo similar mientras que ahora las variables de ejercicios tienen un poder predictivo bastante superior. De hecho, las variables de ejercicios son suficientes para predecir (lo cual es bastante común en entornos tipo *ITS*, como [128]) ya que el resto de variables no mejoran el poder predictivo al que se obtiene con las variables de ejercicios. En cuanto al *AUC* obtenido, es excelente con las variables de ejercicios, pero moderado o pobre con el resto de variables. Por lo tanto, en este caso, se podría usar únicamente variables de ejercicios. Respecto a los algoritmos, *RF* es el que obtiene el mejor poder predictivo, aunque con poca diferencia al resto. Otra observación respecto al *AUC* es que los valores son mucho mejores que los del abandono relacionado con la actividad. Esto puede significar que en los *SPOCs* puede no ser necesaria una conexión frecuente dependiendo del contexto mientras que se hagan las actividades. En caso de que un *SPOC* esté diseñado para una parte del curso, puede haber muchos alumnos que hayan abandonado según la primera definición, pero hayan hecho las actividades y no hayan abandonado según la segunda. Por ello, el primer tipo de abandono debería ser interpretado por cada profesor según el contexto, es decir, que si un

profesor utiliza el SPOC para una parte del curso, podría no tener en cuenta esta predicción.

En tercer lugar, se muestran los resultados de la predicción del éxito académico. En este caso, las conclusiones son muy similares a las del abandono relacionado con completitud, ya que las variables sobre ejercicios son las que ofrecen un mejor poder predictivo y son suficientes para predecir. En este caso, el éxito también está más ligado a lo que se realiza en los ejercicios por lo que también es coherente. Sin embargo, el poder predictivo de la actividad y los vídeos es pobre o moderado, lo que implica que estas variables pueden no ser efectivas en todos los contextos. Para el caso de los ejercicios, el poder predictivo es excelente por lo que se pueden obtener predicciones precisas a mitad de curso. En cuanto a los algoritmos, RF es el algoritmo que proporciona mejor resultados, por lo que teniendo en cuenta las otras variables dependientes, este algoritmo sería el mejor en este contexto, aunque con una diferencia pequeña respecto a los demás. Como conclusión final respecto a las variables, las variables de ejercicios nuevamente vuelven a mostrarse unas variables con un alto poder predictivo. Las variables de vídeos y ejercicios tienen un menor poder predictivo y su efecto puede venir determinado dependiendo de la aplicación (p. ej., tipo de variable a predecir) y el contexto.

6.1.5. Análisis del escenario 6

Por último, se analiza el escenario 6 (plataforma para alumnos de primaria y secundaria). En este contexto, como se explicó en el capítulo 5, el objetivo es predecir el abandono y la nota de final de los estudiantes en la plataforma. Este contexto es similar al escenario 5, aunque la plataforma es diferente, así como el nivel educativo de los estudiantes. La principal diferencia de esto es que las variables que se pueden obtener son diferentes. En este caso, es un escenario en el que solo se dispone de información de ejercicios, por lo que se asemeja más a un ITS. Adicionalmente, se dispone de la variable de actividad `ndias`, que se ha obtenido a partir de los días que hay interacción con ejercicios. Dado que solo hay interacciones con ejercicios, en este análisis se han separado dos tipos de variables sobre ejercicios. El primer grupo trata las variables relacionadas con la participación en los ejercicios, es decir, su foco está en ver el número de actividades que se hacen, o el tiempo o número de días que se emplean. En este grupo se incluyen las siguientes variables: (1) `ndias`, (2) `nacesos`, (3) `tiempo_medio`, (4) `perc_intentados` y (5) `estado_dr`. El segundo grupo incluye las variables relacionadas con el rendimiento del alumno en los ejercicios, es decir, cómo de bien lo hace y cuántos intentos necesita. En este grupo, se incluyen las siguientes variables: (1) `nota_media`, (2) `media_primer`, (3) `avg_intentos`, (4) `perc_correctos_total` y (5) `CFA`. Teniendo en cuenta estos dos grupos de variables, se han realizado los modelos predictivos con los cuatro algoritmos para predecir las dos variables dependientes teniendo en cuenta los alumnos del curso 2017/2018. Los modelos que se presentan en esta sección están elaborados a mitad de curso (fecha 31 de enero de 2018), de modo que el modelo se desarrolle con cierta antelación para poder tomar medidas. Los resultados de estos modelos se presentan en la Tabla 6.6.

Los resultados muestran que a mitad de curso pueden obtenerse unas buenas predicciones, tanto del abandono como de la nota, con los dos conjuntos de variables (participación y rendimiento). El AUC es en todos los casos bueno (superior a 0.8) e incluso excelente para el caso en el que se utilizan todas las variables con RF (el AUC es 0.90 en dicho caso). En cuanto a la predicción de la nota, el RMSE también

TABLA 6.6
 PREDICCIÓN DE ABANDONO Y NOTAS EN EL ESCENARIO 6

Variables	Predicción de abandono				Predicción de la nota			
	RG	SVM	DT	RF	RG	SVM	DT	RF
Participación	0.89	0.88	0.85	0.89	0.16	0.16	0.18	0.16
Rendimiento	0.86	0.83	0.81	0.85	0.13	0.14	0.15	0.13
Ambas	0.89	0.87	0.83	0.90	0.13	0.14	0.15	0.13

El poder predictivo se mide en **AUC** para la predicción de abandono

El poder predictivo se mide en **RMSE** para la predicción de la nota

es bastante bueno, teniendo en cuenta el obtenido en otros trabajos (p. ej., Pérez-Sanagustín et al. [271] alcanzaron un mejor **RMSE** de 0.18, Elbadrawy et al. [27] de 0.15, y Ding et al. [153] de hasta 0.12) y que este poder predictivo se obtiene a mitad de curso y puede ser mejorado con el paso del tiempo. En cuanto a los algoritmos, al igual que en otros escenarios, se observa que las diferencias entre ellos son bajas, si bien el **RF** es el que consigue el mejor poder predictivo en ambos casos (aunque empatado con la **RG** para la predicción de la nota).

El aspecto que tiene una mayor relevancia es el conjunto de variables utilizadas. Para la predicción del abandono, las variables que ofrecen mejor poder predictivo son aquellas relacionadas con la participación. Esto es lógico, ya que el abandono se mide en función de que el estudiante complete las actividades y no de la nota que tenga. Sin embargo, para la predicción de la nota, las mejores variables son las relacionadas con el rendimiento y además, se observa que los modelos no mejoran cuando se añaden las variables sobre participación. Esto implica que dependiendo de la definición de la variable dependiente, unas variables pueden tener mayor o menor poder predictivo. De este modo, las variables sobre participación están más relacionadas con el abandono y las de rendimiento con las notas. No obstante, se observa que el rendimiento de las variables de rendimiento para el abandono y las de participación para la predicción de notas también es buena (aunque peor). Por tanto, se puede decir que las variables relacionadas con las interacciones con ejercicios son muy buenas predictoras del abandono y las notas, y pueden ser utilizadas por su alto poder predictivo, si bien unas variables concretas pueden tener mayor poder predictivo dependiendo de la definición concreta de la variable dependiente.

Para concluir, en esta sección se han expuesto diferentes experiencias en las que se han utilizado variables relacionadas con la interacción con vídeos, ejercicios, y actividad en la plataforma. A pesar de que cada escenario tenía sus propias particularidades, se ha observado que las variables sobre ejercicios son muy buenas predictoras y en muchos casos son suficientes para predecir, lo cual coincide con otros trabajos (p. ej., [23], [28]). Además, las variables de vídeos y de actividad pueden tener un buen poder predictivo por sí solas, o pueden servir para mejorar los modelos predictivos al añadirse a otras variables, por lo que este tipo de variables deben tenerse en cuenta, como ya sugirieron otros trabajos (p. ej., [124], [133]).

6.2. Notas previas de la evaluación sumativa del curso

En la sección anterior, se ha analizado el efecto de las variables sobre las interacciones con vídeos, ejercicios y la actividad en la plataforma. Sin embargo, al analizar las interacciones con ejercicios, no se consideraba la nota individual de cada

prueba sumativa pasada. Esta sección trata de analizar cómo las notas sumativas previas afectan en la predicción.

Para analizar este efecto, se han considerado los escenarios 1 y 2. Estos escenarios fueron tratados en [261], [263], y los resultados del escenario 1 fueron presentados en un TFM previo [262], y en [261], y sirven para poder comparar los resultados con otro escenario (escenario 2, tratado en [263]). En estos escenarios, se predice la nota de cada una de las pruebas sumativas y existe una distinción clara entre parte formativa y sumativa. En otros escenarios, como el 4 y 6, todas las actividades son formativas, y en el 5 se desconoce la separación entre formativa y sumativa. En el escenario 3, todas las notas son sumativas, por lo que estas notas ya se consideraron para el abandono en la sección anterior, y no se consideran en esta sección, donde el foco es ver qué aportan las notas sumativas respecto a usar solo notas formativas.

Teniendo en cuenta estas consideraciones, para los escenarios 1 y 2, se han generado modelos en los que, aparte de incluir las variables sobre ejercicios formativos y el porcentaje de vídeos abiertos (variables consideradas en los modelos presentados en la sección anterior, ver Tablas 6.1 y 6.2), se consideran las notas previas de las pruebas sumativas, una vez que están disponibles. Por ejemplo, la nota del test de la primera semana se utiliza en los modelos a partir de la segunda semana, la nota del test de la segunda semana se utiliza para la predicción a partir de la tercera semana, y así sucesivamente. Las Tablas 6.7 y 6.8 muestran los resultados de incluir las notas previas, tanto para el escenario 1 (Tabla 6.7) como para el escenario 2 (Tabla 6.8). En dichas tablas, se compara el modelo sin notas sumativas previas (modelo A, que es el usado en la sección anterior) con el modelo que utiliza las notas sumativas previas (modelo B). Es de destacar que el modelo B de la Tabla 6.7 no incluye la predicción de la nota final, ya que la nota final es la media ponderada de las notas previas, por lo que si se incluyese, se tendrían todos los datos para calcular la nota final.

TABLA 6.7
PREDICCIÓN DE NOTAS CON NOTAS PREVIAS EN ESCENARIO 1

Modelo	Algoritmo	T_1	T_2	T_3	T_4	T_5	P_3	P_5	NF
A	RG	0.27	0.22	0.20	0.18	0.16	0.25	0.20	0.14
A	SVM	0.27	0.22	0.21	0.19	0.17	0.25	0.21	0.15
A	DT	0.34	0.28	0.26	0.22	0.18	0.31	0.27	0.16
A	RF	0.26	0.21	0.20	0.19	0.16	0.25	0.21	0.14
B	RG	0.27	0.21	0.18	0.15	0.13	0.24	0.19	-
B	SVM	0.27	0.20	0.18	0.15	0.13	0.24	0.20	-
B	DT	0.34	0.26	0.23	0.20	0.17	0.32	0.26	-
B	RF	0.26	0.20	0.18	0.15	0.13	0.24	0.19	-

Nota: El mejor poder predictivo (medido en RMSE) de cada columna se marca en negrita

Modelo A: sin notas previas. Modelo B: con notas previas

En vista de los resultados obtenidos introduciendo variables con las notas sumativas previas (modelo B), y comparando con los obtenidos cuando no se utilizaban las notas sumativas previas (modelo A), se puede ver que en los dos escenarios existe una mejora del poder predictivo al introducir estas notas. Esto sugiere que el resultado en las pruebas sumativas es un buen predictor del rendimiento futuro y puede mejorar los modelos. Este resultado es coherente con

TABLA 6.8
 PREDICCIÓN DE NOTAS CON NOTAS PREVIAS EN ESCENARIO 2

Modelo	Algoritmo	L_1	L_2	L_3	L_4	L_5	L_6	PA	PR	EF
		W_1	W_2	W_3	W_5	W_7	W_8	W_9	W_9	W_{10}
A	RG	0.33	0.33	0.28	0.30	0.26	0.28	0.26	0.25	0.17
A	SVM	0.36	0.36	0.29	0.33	0.27	0.34	0.29	0.26	0.19
A	DT	0.38	0.41	0.33	0.36	0.32	0.31	0.28	0.27	0.18
A	RF	0.34	0.36	0.30	0.33	0.28	0.30	0.27	0.26	0.18
B	RG	0.33	0.31	0.26	0.27	0.24	0.26	0.25	0.24	0.16
B	SVM	0.36	0.35	0.27	0.27	0.24	0.28	0.27	0.25	0.17
B	DT	0.38	0.38	0.31	0.30	0.29	0.30	0.30	0.28	0.22
B	RF	0.34	0.33	0.27	0.27	0.24	0.27	0.26	0.25	0.17

Nota: El mejor poder predictivo (medido en RMSE) de cada columna se marca en negrita

Modelo A: sin notas previas. Modelo B: con notas previas

otros trabajos (p. ej., [125], [144]), que ya usaron notas previas para refinar los modelos a lo largo del tiempo [125], o concluyeron que las notas sumativas previas eran las mejores variables en la predicción [144]. En cuanto a los algoritmos, los resultados también indican que la diferencia de poder predictivo entre ellos (al menos entre los mejores algoritmos) es baja, por lo que el hecho de introducir variables tiene mayor efecto que el cambiar de un algoritmo a otro.

6.3. Variables de la actividad del alumno en el foro

Las secciones anteriores se centraron únicamente en las interacciones de vídeos, ejercicios (incluyendo notas previas) y actividad en la plataforma. Sin embargo, uno de los componentes importantes de los MOOCs son las interacciones en el foro, que permiten la comunicación entre usuarios del MOOC. Esta sección trata de analizar el poder predictivo de las variables del foro. Para ello, se utilizan el escenario 1 (los resultados de los análisis de las variables del foro en este escenario 1 fueron presentados en un TFM previo [262], y aquí sirven para como base para comparar con el escenario 2) y el escenario 2 (tratado en [263]), que son los escenarios de MOOCs donde se dispone de ese tipo de interacciones. Para otros escenarios, como los LMSs, ITSs o SPOCs, este tipo de variables solo tiene sentido si el foro se usa, lo cual no ocurre con tanta frecuencia. Además, en algunos casos, esta información ni siquiera está disponible. Es por ello que los otros escenarios no se han tenido en cuenta para esta sección.

Para el análisis del foro, se han desarrollado tres modelos adicionales a los ya mostrados en el escenario 1 y 2. El primer modelo (llamado modelo C) incluye únicamente las variables del foro, y los modelos D y E son los equivalentes a los A y B mostrados anteriormente, pero incluyendo las variables del foro. Los resultados para el escenario 1 se muestran en la Tabla 6.9. En este caso, al igual que en el modelo B, se elimina la predicción de la nota final en los modelos que incluyen las notas que constituyen la nota final.

Los resultados muestran que las variables relacionadas con el foro tienen un bajo poder predictivo. En base al modelo C, que solo incluye variables de foro, el mejor resultado se obtiene en la tarea de programación de la semana 5, en la cual el RMSE es 0.24, lo cual es peor que los resultados presentados en los modelos A y B. Para

TABLA 6.9
 PREDICCIÓN DE NOTAS CON VARIABLES DE FORO EN ESCENARIO 1

Modelo	Algoritmo	T_1	T_2	T_3	T_4	T_5	P_3	P_5	NF
C	RG	0.41	0.36	0.33	0.31	0.27	0.34	0.24	0.26
C	SVM	0.46	0.40	0.35	0.33	0.28	0.34	0.26	0.28
C	DT	0.46	0.40	0.35	0.33	0.30	0.36	0.28	0.27
C	RF	0.42	0.37	0.33	0.31	0.27	0.34	0.25	0.25
D	RG	0.26	0.22	0.20	0.18	0.16	0.25	0.20	0.14
D	SVM	0.26	0.22	0.21	0.19	0.17	0.26	0.21	0.15
D	DT	0.34	0.28	0.26	0.23	0.19	0.32	0.28	0.17
D	RF	0.25	0.21	0.20	0.19	0.16	0.25	0.20	0.14
E	RG	0.26	0.21	0.18	0.15	0.13	0.24	0.19	-
E	SVM	0.26	0.20	0.18	0.15	0.14	0.25	0.20	-
E	DT	0.34	0.26	0.23	0.20	0.17	0.32	0.26	-
E	RF	0.25	0.20	0.18	0.15	0.13	0.24	0.19	-

Nota: El mejor poder predictivo (medido en **RMSE**) de cada columna se marca en negrita
 Modelo C: solo foro. Modelo D: ejercicios, vídeos y foro. Modelo E: modelo D más notas previas

corroborar el bajo poder predictivo, se han calculado dos modelos base. El primero consiste en asignar valores aleatorios a cada nota. Este modelo produce un **RMSE** de entre 0.49 y 0.53, lo cual es peor que el poder predictivo proporcionado por las variables del foro. El segundo modelo base consiste en asignar la nota media a todos los estudiantes. En este caso, el poder predictivo está entre 0.34-0.40 (en **RMSE**), por lo que las variables del foro (modelo C) no mejoran significativamente este modelo.

Respecto a los modelos D y E, tras comparar los resultados con los modelos A y B (sus equivalentes sin las variables del foro), los resultados muestran que el poder predictivo prácticamente no cambia entre los modelos A y D, y B y E (no hay mejoría tras incluir las variables del foro). Además, un análisis estadístico entre los modelos (t-test pareado muestra que la diferencia entre los modelos no es estadísticamente significativa, con p -valor = 0.98). Por tanto, en base al escenario 1, se puede concluir que la actividad en el foro no está relacionada con las notas de los estudiantes.

Este hecho puede estar relacionado con algunos artículos en los que se concluye que los buenos estudiantes tienden a ser individualistas (p. ej., [273], [274]). Si ocurre este caso, podría suceder que los alumnos no contribuyan en el foro a pesar de que entiendan los contenidos y aprueben las evaluaciones sumativas. En el escenario 1 en particular, solo 754 (de los 1507 que aprobaron el curso) contribuyeron con algún mensaje en el foro. Por tanto, el 49.97% de los usuarios aprobaron el curso sin escribir mensajes y, por otro lado, hubo 3604 usuarios que escribieron mensajes, pero no aprobaron. Esto sirve para justificar que las variables del foro no tienen un alto poder predictivo en la predicción.

Para validar estos resultados en otro contexto, se utiliza el escenario 2. En este escenario, siguiendo la misma metodología que en el escenario 1, se han calculado los modelos C, D y E, que incluyen variables sobre las interacciones en el foro. Los resultados de estos modelos se presentan en la Tabla 6.10.

Los resultados de los modelos predictivos utilizando las variables de foro muestran unas conclusiones similares a las del escenario 1, ya que los modelos

TABLA 6.10
 PREDICCIÓN DE NOTAS CON VARIABLES DE FORO EN ESCENARIO 2

Modelo	Algoritmo	L_1 W_1	L_2 W_2	L_3 W_3	L_4 W_5	L_5 W_7	L_6 W_8	PA W_9	PR W_9	EF W_{10}
C	RG	0.40	0.45	0.47	0.45	0.43	0.38	0.33	0.37	0.38
C	SVM	0.45	0.56	0.57	0.53	0.48	0.42	0.35	0.41	0.45
C	DT	0.47	0.53	0.56	0.50	0.46	0.42	0.35	0.38	0.39
C	RF	0.42	0.47	0.50	0.46	0.43	0.39	0.33	0.37	0.38
D	RG	0.32	0.33	0.28	0.30	0.26	0.29	0.26	0.24	0.17
D	SVM	0.33	0.36	0.29	0.34	0.27	0.33	0.29	0.26	0.19
D	DT	0.39	0.45	0.38	0.39	0.33	0.34	0.31	0.30	0.19
D	RF	0.32	0.34	0.29	0.33	0.28	0.30	0.27	0.26	0.18
E	RG	0.32	0.31	0.26	0.26	0.24	0.26	0.25	0.24	0.16
E	SVM	0.33	0.34	0.27	0.27	0.24	0.29	0.27	0.25	0.17
E	DT	0.39	0.40	0.35	0.33	0.30	0.31	0.32	0.30	0.22
E	RF	0.32	0.32	0.27	0.27	0.24	0.27	0.26	0.25	0.17

Nota: El mejor poder predictivo (medido en RMSE) de cada columna se marca en negrita
 Modelo C: solo foro. Modelo D: ejercicios, vídeos y foro. Modelo E: modelo D más notas previas

no mejoran su poder predictivo al introducir las variables del foro. En el caso del escenario 2, el uso del foro era más intensivo y aunque en ambos MOOC se fomentaba el uso del foro, en el escenario 2 había actividades específicas formativas para que los alumnos resolvieran en el foro. Estas actividades tenían una alta participación (p. ej., una actividad tuvo 826 respuestas) y el foro funcionaba muy bien, con un 94% de hilos con respuesta y un tiempo media de respuesta de 6 horas. Sin embargo, las variables del foro no mejoraban el poder predictivo. Por ello, y en vista de los resultados de los dos escenarios, la conclusión extraída es que la actividad en el foro no influye en los resultados del alumno. Esto coincide con trabajos previos, en los que las variables del foro eran las que peor poder predictivo obtenían al predecir quién iba a obtener el certificado [131], o tenían poco efecto en la nota final del curso [275]. Sin embargo, difiere de otros trabajos (p. ej., [129]) en el que el foro tenía más relevancia. Aunque en este contexto, el uso del foro era muy intenso. En un MOOC analizado de matemáticas, los resultados de los ejercicios tenían que subirse a través del foro, y en otro MOOC sobre narrativa, los ejercicios se debatían en el foro. Por ello, aunque en general se concluye que las variables de foro tienen un bajo poder predictivo y se corrobora con otros trabajos, esto podría variar en un contexto muy intensivo en los foros.

6.4. Variables sobre la autorregulación del aprendizaje

En los MOOCs, la autorregulación del aprendizaje (SRL) puede ser un elemento crítico para el éxito [224], y para lograr completar satisfactoriamente MOOCs los alumnos necesitan habilidades para establecer sus objetivos, planificarse o monitorizar su aprendizaje [276]. Dada la importancia del SRL, esta sección trata de analizar si las variables relacionadas con el SRL tienen un efecto en la predicción. Para realizar este análisis, se considera únicamente el escenario 3, ya que es el único en el que se dispone de variables sobre SRL, al ser éstas proporcionadas por el equipo de PUC.

Para abordar el análisis, en primer lugar se llevó a cabo un estudio exploratorio utilizando regresión y el curso “Electrones en Acción” (con 2035 alumnos) [266]. En particular, se realizaron modelos para predecir la nota y el éxito (definido como obtener una nota superior a 80 sobre 100, y como obtener una nota superior a 80, habiendo visto al menos el 50 % de los vídeos). Para la elaboración de estos modelos, se consideraron tres clústeres de alumnos, que fueron identificados por el equipo de PUC en un trabajo previo [224], en base a patrones de SRL. Estos tres grupos de estudiantes son los siguientes [266]:

- Alumnos de muestreo (clúster 1). Son alumnos que generalmente acceden al curso para “muestrear” algunos materiales y después dejan el curso, por lo tienen muy poca actividad. En este grupo hay 1530 alumnos, de los que solo 7 completan el curso.
- Alumnos completos (clúster 2). Son los alumnos que siguen el itinerario sugerido por el profesor y trabajan tanto en vídeos como en las evaluaciones. En particular, estos alumnos siguen el patrón lógico de ver primero los vídeos y después intentar las pruebas de evaluación. En este grupo solo hay 85 alumnos, de los que 30 consiguen completar el curso.
- Alumnos estratégicos (clúster 3). Son alumnos que no tienen interés en completar todos los contenidos como los alumnos completos y se centran en aquello necesario para superar las evaluaciones, de ahí el nombre de estratégicos. Esta orientación estratégica hace que visualicen menos vídeos y se centren más en las evaluaciones. En este grupo hay 420 alumnos, de los cuales 143 logran completar el curso.

Teniendo en cuenta estos clústeres, se utilizó regresión con variables sobre patrones de SRL y variables sobre actividad (ver variables concretas en la sección 5.4) para predecir las variables dependientes mencionadas. Además, en este caso, los modelos presentados contenían otras variables demográficas y sobre intenciones [266], pero se excluyen de este análisis para poner el foco en el SRL, que es el objeto de esta sección.

En primer lugar, se analizó el porcentaje de variabilidad explicada por los modelos para predecir la nota final (R^2). En este caso se obtuvo un R^2 ajustado de 0.80 para los estudiantes completos, 0.72 para los estudiantes estratégicos y 0.86 para el conjunto con todos los alumnos (con p – valor < 0.001 en todos los casos). A la vista de estos resultados, se observa que las variables explican una alta variabilidad del modelo (aunque en menor grado para los alumnos estratégicos, que siguen un itinerario menos convencional). Teniendo en cuenta que estos modelos están contruidos con patrones de SRL (y tras comprobar que estos patrones tenían coeficientes significativos en la regresión, y con valores superiores a otros tipos de variables), una primera hipótesis es que los patrones de SRL pueden tener un efecto importante para la predicción.

Como segundo experimento, se dividió el conjunto de datos en entrenamiento y test (80 % para entrenamiento y 20 % para test) y se llevó a cabo una validación cruzada (CV) 10-fold con el conjunto de entrenamiento. Con estos modelos se evaluó el poder predictivo mediante el RMSE para la nota (variable continua) y el AUC para el éxito (variables discretas). Los resultados se muestran en la Tabla 6.11. Estos resultados muestran que el poder predictivo es mayor para el grupo con todos

los alumnos (AUC superior a 0.9). Esto es normal ya que se incluyen también los alumnos de muestreo, cuya nota es más fácil de predecir, ya que no hacen actividad en la plataforma y por tanto se predice que suspenden (por eso para estos alumnos tiene menos sentido realizar un modelo separado). Para los alumnos completos, se aprecian algunas diferencias entre el conjunto de entrenamiento y test. El motivo es que hay muy pocos alumnos completos y las limitaciones de los datos pueden suponer un problema al realizar los modelos. No obstante, el AUC para CV es de 0.84, que es bueno. Para el grupo donde las predicciones son peores (en particular en la predicción de la nota final) es para los alumnos estratégicos. Un posible motivo es que estos alumnos tienen comportamientos más anómalos, ya que en muchas ocasiones no siguen parte de los materiales y se centran únicamente en lo necesario para las evaluaciones. No obstante, los valores del AUC obtenidos son bastante buenos, y por encima de 0.9 en la mayoría de los casos, y los valores de RMSE son también buenos, en comparación con otros trabajos (p. ej., [27], [271]). Teniendo en cuenta que los coeficientes de la regresión eran mayores para los patrones de SRL que para otros tipos de variables (salvo las de actividad, que tenían coeficientes similares), la conclusión de este análisis inicial es que las variables sobre SRL pueden ser buenos predictores.

TABLA 6.11
EVALUACIÓN DE PREDICCIÓN DE NOTAS Y ÉXITO EN ESCENARIO 3

Clúster	Datos	Nota	$Nota \geq 80$	$Nota \geq 80$ y
		(continua) RMSE	(binaria) AUC	$perc_compl \geq 0.5$ (binaria) AUC
Todos	CV	0.11	0.98	0.98
	Test	0.12	0.98	0.98
Completos	CV	0.17	0.84	0.84
	Test	0.12	0.92	0.92
Estratégicos	CV	0.17	0.92	0.92
	Test	0.18	0.90	0.92

Para validar esta hipótesis, se analizó la predicción del abandono en el escenario 3, usando los tres MOOCs considerados en este escenario [91]. En primer lugar, se desarrollaron diferentes modelos utilizando las variables de patrones de SRL, junto con la combinación de estas variables con otros tipos de variables (p. ej., de vídeos, ejercicios y actividad). Los resultados de los modelos predictivos, usando las variables de SRL se muestran en la Tabla 6.12.

A la vista de los resultados, los patrones de SRL, por sí mismos, pueden alcanzar un AUC de hasta 0.96 en dos de los MOOC y 0.97 en “Aula Constructivista”, por lo que estas variables pueden obtener un muy buen poder predictivo por sí mismas. Esto apoya la idea de que las habilidades de SRL son muy importantes para tener éxito (y no abandonar) en los MOOCs.

Además, los resultados muestran que otros modelos mejoran su poder predictivo al añadir los patrones de SRL. Por ejemplo, el poder predictivo de las variables de actividad aumenta ligeramente en todos los algoritmos (con una mejora en AUC de entre 0.02 y 0.05) al incluir los patrones de SRL. Esta misma mejora se aprecia, aunque en menor medida y únicamente en “Electrones en Acción” para el caso de las variables de vídeo. Sin embargo, las variables de SRL no mejoran el modelo cuando

TABLA 6.12
 PREDICCIÓN CON VARIABLES DE SRL EN ESCENARIO 3

Variables	Electrones en acción				Aula Constructivista				Gestión de Organizaciones Efectivas			
	RG	SVM	DT	RF	RG	SVM	DT	RF	RG	SVM	DT	RF
Patrones SRL	0.96	0.94	0.85	0.96	0.97	0.94	0.88	0.95	0.94	0.96	0.75	0.95
Actividad	0.94	0.89	0.86	0.92	0.93	0.96	0.92	0.95	0.94	0.94	0.87	0.94
Actividad y patrones SRL	0.96	0.94	0.89	0.96	0.96	0.94	0.92	0.97	0.96	0.96	0.87	0.96
Vídeos	0.93	0.92	0.91	0.93	0.97	0.97	0.92	0.97	0.94	0.97	0.92	0.97
Vídeos y patrones SRL	0.95	0.94	0.92	0.96	0.97	0.93	0.92	0.97	0.95	0.97	0.92	0.97
Ejercicios	0.96	0.96	0.96	0.96	0.98	0.98	0.97	0.99	0.97	0.97	0.97	0.97
Ejercicios y patrones SRL	0.96	0.95	0.96	0.97	0.98	0.96	0.97	0.98	0.95	0.98	0.97	0.97
Todas las variables	0.96	0.93	0.96	0.96	0.89	0.95	0.97	0.99	0.95	0.94	0.97	0.99

Nota: El poder predictivo se mide en AUC. Los mejores modelos para cada curso se marcan en negrita.

están presentes las variables de ejercicios. Un posible motivo es que los patrones de SRL tienen cierta relación con las interacciones con los ejercicios (y en cierta medida con los vídeos) y por eso aquí no se observa una mejora del poder predictivo. No obstante, aunque las variables de ejercicios tengan un alto poder predictivo (y sean las mejores), los patrones de SRL alcanzan un muy alto poder predictivo por sí solos, lo que hace que estas variables tengan un interés para la predicción. Por ejemplo, pueden tener un gran interés cuando las variables sobre ejercicios no estén disponibles por su alto poder predictivo.

Teniendo en cuenta todo lo anterior, se concluye que los patrones de SRL pueden alcanzar un alto poder predictivo por sí mismos, lo que implica que estas variables tienen una relación con el éxito académico y las notas (como ya se había visto en trabajos previos, como [105], [107]), y pueden utilizarse en los modelos predictivos. Cuando hay otro tipo de variables, como las de ejercicios, puede que los patrones de SRL no añadan mucho valor (al menos en el escenario analizado), posiblemente por la relación con dichas variables, pero si las variables sobre ejercicios no están disponibles, los patrones de SRL pueden dar un valor importante al modelo por su alto poder predictivo, y su uso es recomendable.

6.5. Variables sobre comportamientos del alumno

En las plataformas educativas, aparte de obtener variables sobre la propia interacción con la plataforma, tales como porcentajes de recursos, vídeos o ejercicios realizados o el rendimiento, es posible obtener comportamientos de más alto nivel sobre el alumno, que puedan estar asociados con la personalidad del estudiante. Aunque la obtención de estas variables puede ser muy interesante para desarrollar sistemas adaptativos y personalizados a cada estudiante, y para proporcionar visualizaciones que ayuden a los profesores (como demostró, por ejemplo, Molenaar y Knoop-van Campen [277]), no se conoce qué poder predictivo pueden tener muchas de estas variables. Esta sección analiza el poder predictivo de variables de comportamientos en la predicción. Para ello se analizan tres variables sobre comportamientos: la persistencia, la eficiencia y la constancia del alumno. Estas variables se analizan en escenarios concretos, si bien podrían extenderse a otros escenarios.

6.5.1. Efecto de la persistencia en la predicción

En primer lugar, se analiza el efecto de la persistencia en los modelos predictivos. Para ello, se utiliza el escenario 5 (SPOCs de la UC3M), tratado en [108]. En este escenario, las variables dependientes consideradas son (1) el abandono relacionado con la actividad, (2) el abandono relacionado con la completitud y (3) el éxito, definido como obtener una nota igual o superior a 5.0 sobre 10. Para el análisis de la persistencia, se ha seleccionado un subconjunto de variables independientes, de modo que no se pierda el foco en la persistencia, pero la lista sea lo suficientemente heterogénea (tenga al menos alguna variable de vídeos, ejercicios y actividad en la plataforma). En concreto, las variables utilizadas son: `perc_dias`, `perc_compl`, `perc_intentados`, `media_intentados`, y `persistencia`. Con estas variables, independientes, se han generado inicialmente tres modelos: el primero incluye todas las variables excepto la persistencia (modelo de interacciones, INT), el segundo solo incluye la persistencia (modelo PER) y el último incluye todas las interacciones (TOD).

Para el caso de la persistencia, un problema que hay es que la persistencia no está definida para todos los estudiantes. El motivo es que algunos estudiantes solo han realizado actividades que se han resuelto con dos intentos como máximo (y fueron filtradas, ya que no sirven para medir la persistencia) o han realizado todas las actividades de forma correcta, por lo que no han demostrado persistencia. Para dichos casos, se ha utilizado imputación de la media con una variable *dummy* [278], que es una técnica de imputación bastante conocida y usada en muchos artículos de investigación (p. ej., [278], [279]). Esta imputación consiste en tener una variable *dummy* que contiene 0 o 1 dependiendo de si la persistencia está disponible para dicho alumno, y una segunda variable con el valor de la persistencia. Para el análisis, se han considerado 13 SPOCs del primer semestre del curso académico 2018/2019, de los cuales se dispone de los datos al completo. Para este primer análisis, se han realizado los modelos tomando los datos hasta la semana 8 (que es justo la mitad del curso, al tener los cursos una duración de 16 semanas). De este modo, el poder predictivo debe ser suficientemente bueno y hay una suficiente antelación. Los resultados de los tres modelos mencionados (INT, PER y TOD) se encuentran en la Tabla 6.13.

TABLA 6.13
PREDICCIÓN UTILIZANDO PERSISTENCIA EN ESCENARIO 5

Datos	Variables	Predicción	RG	SVM	DT	RF
TOD	<code>perc_dias</code> , <code>perc_compl</code> , <code>perc_intentados</code> , <code>media_intentados</code> , y <code>persistencia</code> (incluida la variable <i>dummy</i>)	ABA	0.77	0.76	0.69	0.76
		ABC	0.91	0.91	0.87	0.92
		EXT	0.90	0.89	0.89	0.93
INT	<code>perc_dias</code> , <code>perc_compl</code> , <code>perc_intentados</code> y <code>media_intentados</code>	ABA	0.77	0.76	0.70	0.76
		ABC	0.91	0.90	0.89	0.92
		EXT	0.90	0.89	0.89	0.93
PER	<code>persistencia</code> (incluida la variable <i>dummy</i>)	ABA	0.51	0.54	0.54	0.52
		EXT	0.57	0.64	0.64	0.64

Nota: El poder predictivo se mide en AUC

TOD: Todas las variables, INT: Solo interacciones, PER: Solo persistencia

ABA: Abandono relacionado con actividad, ABC: Abandono relacionado con completitud, EXT: Éxito

Los resultados muestran que es posible obtener buenas predicciones del abandono relacionado con la competitividad y del éxito con tan solo cinco variables y a mitad de curso (el mejor AUC que se obtiene es superior a 0.9). Sin embargo, el poder predictivo que se puede obtener para el abandono relacionado con la actividad es menor, y el mejor AUC que se obtiene en la semana 8 es de 0.77, lo cual es un AUC moderado. Un posible motivo de que el AUC sea peor en este caso puede estar en el contexto. Dado que el contexto de cada curso es desconocido, no se sabe en qué momento cada alumno debe interactuar en el SPOC. Por ejemplo, puede ser que un SPOC esté diseñado únicamente para una parte del curso. En este caso, los estudiantes no tendrían que interactuar todas las semanas y eso causaría que muchos estudiantes fueran considerados como que han abandonado incluso si han seguido el ritmo del curso. Por ello, la definición de abandono debe adaptarse a cada contexto.

Respecto a la persistencia, los resultados muestran que su poder predictivo es bajo. Puede añadir algo de información, ya que el AUC obtenido es superior a 0.5, pero no tiene un alto poder predictivo. Dado que el poder predictivo con una sola variable tampoco se espera que sea alto, otro motivo para ver el bajo poder predictivo es que el poder predictivo de los modelos con y sin persistencia es muy similar (e incluso mejor sin la persistencia en algún caso). Esto implica que la persistencia parece que no es un buen predictor del abandono o éxito de los estudiantes. Sin embargo, es posible que esta variable pueda añadir cierto valor si otras variables no están presentes.

Para analizar este hecho, se han generado varios modelos utilizando la persistencia con otras variables de forma individual. Estos modelos se han generado utilizando RF para analizar la importancia de las variables mediante la disminución media de Gini [280]. La lista de modelos generados se muestra en la Tabla 6.14. Hay un modelo para cada variable individual, añadiendo la persistencia (junto con la variable *dummy*) y un modelo que incluye todas las variables. Teniendo en cuenta estos modelos, la Fig. 6.1 muestra la importancia relativa (en porcentaje) de cada variable al predecir las tres variables dependientes, de modo que la suma de todas las variables suman un 100% para cada modelo.

TABLA 6.14
MODELOS UTILIZADOS EN LA PREDICCIÓN EN EL ESCENARIO 5

Datos	Variables
TOD	<code>perc_dias</code> , <code>perc_compl</code> , <code>perc_intentados</code> , <code>media_intentados</code> , y <code>persistencia</code> (incluida la variable <i>dummy</i>)
DIA	<code>perc_dias</code> y <code>persistencia</code> (incluida la variable <i>dummy</i>)
PEJ	<code>perc_intentados</code> y <code>persistencia</code> (incluida la variable <i>dummy</i>)
VID	<code>perc_compl</code> y <code>persistencia</code> (incluida la variable <i>dummy</i>)
NOT	<code>media_intentados</code> y <code>persistencia</code> (incluida la variable <i>dummy</i>)

Los resultados muestran que cuando los modelos predictivos se desarrollan utilizando la persistencia y otra variable, la persistencia tiene un peso relevante en la importancia, lo que significa que la persistencia puede añadir cierto valor a los modelos predictivos. La única excepción es el caso en el que la persistencia se combina con el porcentaje de ejercicios intentados (modelo PEJ), y en particular cuando se predice el abandono relacionado con la completitud y el éxito. En esos casos, se observa que la persistencia local tiene poca importancia (menos de 13%, incluyendo la variable *dummy*) en el modelo. El porcentaje de ejercicios intentados es

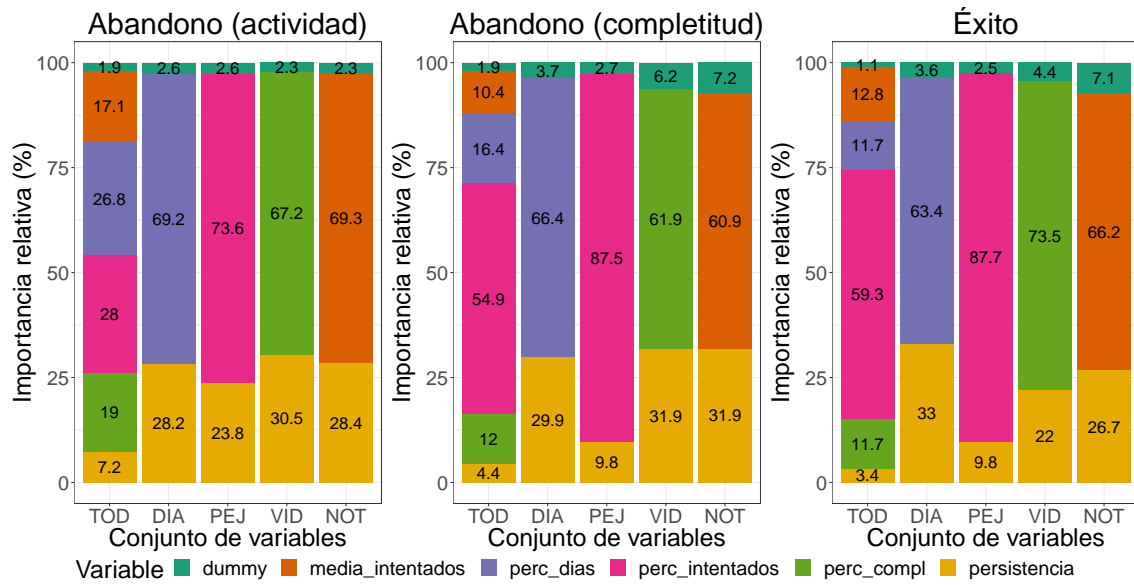


Fig. 6.1. Importancia relativa de las variables en la predicción, utilizando persistencia en escenario 5

el mejor predictor en todos los modelos y tiene una alta importancia en los modelos predictivos. Esto implica que completar las actividades es importante para tener éxito en el curso. Además, dado que estas variables se miden a mitad de curso, el resultado implica que los estudiantes que no interaccionan en las primeras semanas probablemente no lo hagan después.

Sin embargo, cuando no se utiliza el porcentaje de ejercicios intentados (modelos DIA, VID y NOT), puede apreciarse que la persistencia puede añadir valor a los modelos. El valor predictivo de la persistencia no es alto, pero puede mejorar el poder predictivo. Por ejemplo, el poder predictivo de `perc_dias` al predecir el abandono relacionado con la completitud en la semana 8 es de 0.72 y aumenta a 0.76 al añadir la persistencia, y el poder predictivo de `media_intentados` al predecir el éxito en la semana 8 es 0.67 y aumenta a 0.71 al añadir la persistencia). No obstante, la mejora no es muy alta. Por ello, la conclusión es que la persistencia podría ligeramente mejorar el poder predictivo cuando hay pocas variables disponibles, pero cuando hay bastante información, su poder predictivo es prácticamente nulo. Este resultado puede contradecir a otros trabajos en los que se observó una relación positiva entre persistencia y resultados de aprendizaje (p. ej., [112], [281]), aunque en esos trabajos no analizaron la persistencia en los modelos predictivos, y puede haber otros factores debido al contexto. Por ejemplo, si los ejercicios son sumativos, el alumno puede tender más a ser persistente. No obstante, a pesar de que esta variable no aporte mucho valor a los modelos y pueda no merecer la pena incluirla, sí que puede ser utilizada de forma descriptiva, ya que puede aportar valor a los interesados.

6.5.2. Efecto de la eficiencia y constancia en la predicción

Aparte de la persistencia, otros comportamientos de interés son la eficiencia y la constancia de los alumnos. Para analizar estas variables, se utilizaron los datos del escenario 6 (tratado en [114]), que es sobre el que se obtuvieron estas variables. Como análisis inicial, se realizó un análisis de la correlación y de regresión utilizando estas

variables [114]. En este análisis, se vio que la correlación de la eficiencia con la nota final es de $r = 0.12$, mientras que la de la inversa de la constancia (la constancia, tal y como se definió en la sección 3.3.1 es la inversa de la desviación típica del tiempo invertido, y para los análisis se utiliza la inversa de la constancia al estar en unidades de tiempo) es de $r = 0.22$. A la vista de estos resultados, se puede ver que la correlación no es muy elevada en ambos casos. Para analizar mejor este caso, la Fig. 6.2 muestra la relación entre la eficiencia, la inversa de la constancia y la nota final.

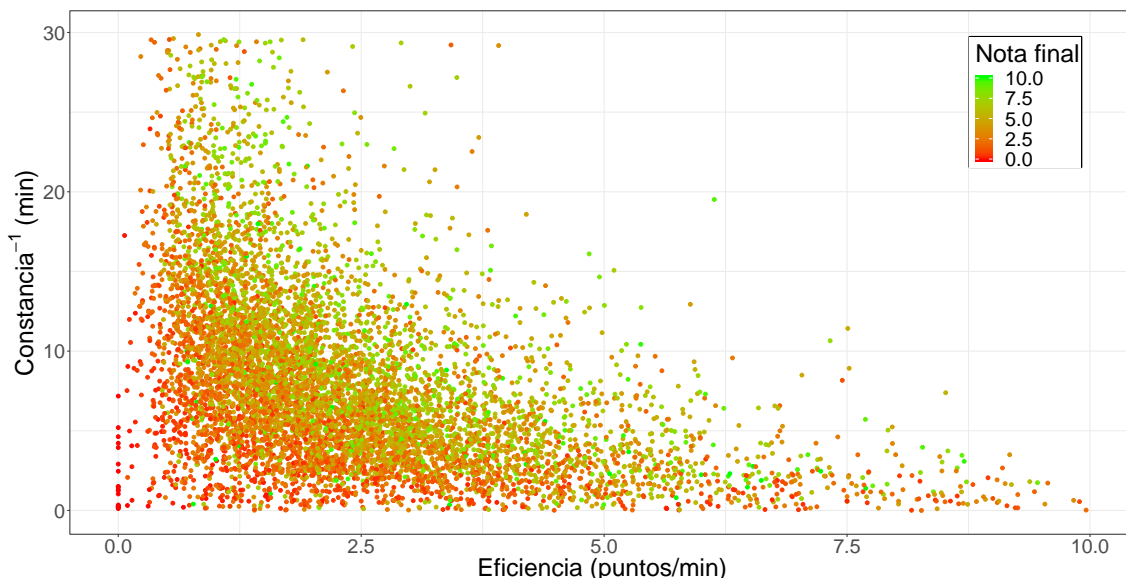


Fig. 6.2. Relación entre la eficiencia, el inverso de la constancia y la nota final

En la figura, se observa que existe cierta relación entre la eficiencia y el inverso de la constancia, de modo que los estudiantes más eficientes tienden a ser más constantes, lo que implica que trabajar de forma más constante puede ser más eficiente. Sin embargo, no se observa ningún patrón con respecto a ninguna de las dos variables en referencia a la nota final. Por lo que se observa, hay alumnos más y menos eficientes y constantes con mayor y menor nota. Teniendo esto en cuenta, se desarrolló un modelo de regresión simple con la eficiencia y la inversa de la constancia para predecir la nota final. Este modelo tuvo un R^2 de 0.11 (p -valor < 0.05), lo que muestra que estas variables explican poca variabilidad de la nota final. Sin embargo, el porcentaje de ejercicios intentados alcanzaba un R^2 de 0.81 (p -valor < 0.05). Por tanto, el análisis inicial muestra que el ser más eficiente o constante en la plataforma no necesariamente implica obtener mejores resultados.

Para profundizar mejor en el poder predictivo de la eficiencia y la constancia, se realizaron modelos predictivos adicionales incluyendo la eficiencia, el inverso de la constancia, y las variables sobre participación y rendimiento (en los ejercicios), que fueron utilizadas en la sección 6.1. Estos modelos se realizaron a mitad de curso, utilizando RF (ya que obtuvo los mejores resultados en los primeros análisis en la sección 6.1), y las variables a predecir son el abandono y la nota final. Los resultados de estos modelos predictivos se presentan en la Tabla 6.15.

Los resultados muestran que el poder predictivo de la eficiencia y el inverso de la constancia son muy bajos al predecir el abandono. El AUC es de 0.51 para la eficiencia y de 0.53 para el inverso de la constancia, lo cual son valores bastante

TABLA 6.15
 PREDICCIÓN UTILIZANDO EFICIENCIA Y CONSTANCIA EN ESCENARIO 6

Datos	VARIABLES	Abandono	Nota
PAR	ndias, nacesos, tiempo_medio, perc_intentados y estado_dr	0.89	0.16
REN	nota_media, media_primer, avg_intentos, perc_correctos_total y CFA	0.85	0.13
EFC	eficiencia	0.51	0.28
CON	constancia ⁻¹	0.53	0.27
COM	eficiencia y constancia ⁻¹	0.56	0.24
PARC	VARIABLES en PAR y COM	0.89	0.14
RENC	VARIABLES en REN y COM	0.87	0.13
PR	VARIABLES en PAR y REN	0.90	0.13
TOD	Todas las variables	0.90	0.13

El poder predictivo se mide en AUC para la predicción de abandono

El poder predictivo se mide en RMSE para la predicción de la nota

malos y muy cercanos a 0.50, que es el AUC teórico que se obtendría al realizar la predicción al azar. Además, el poder predictivo de las variables de participación es el mismo incluyendo o sin incluir la eficiencia y el inverso de la constancia, y hay solo una pequeña mejora para el caso de las variables de rendimiento. En cuanto al modelo que incluye todas las variables, también se observa que no hay ninguna mejora al incluir la eficiencia y la constancia, por lo que estas variables no son buenas variables para predecir el abandono.

En cuanto a la nota final, se observa que la eficiencia y la constancia alcanzan un RMSE de 0.28 y 0.27, respectivamente, y que en combinación alcanzan un RMSE de 0.24. Estos valores no son muy buenos, pero teniendo en cuenta los obtenidos en otros escenarios y para otras variables, no son extremadamente malos. Por ejemplo, las variables de foro obtenían valores de RMSE de entre 0.24 a 0.41 en el escenario 1 (y los valores inferiores a 0.30 eran únicamente en la última semana del curso) y siempre superiores a 0.33 en el escenario 2. Además, para algunas tareas de los escenarios 1 y 2, el mejor RMSE estaba en torno 0.25, y aun no siendo buenos resultados, podían dar una cierta indicación de si el estudiante podía estar en una franja de aprobar, suspender o de estar en borde. Por otro lado, el modelo con variables de participación presenta una pequeña mejora al incluir la eficiencia e inversa de la constancia, pero el modelo con las variables de rendimiento y con todas las variables no mejora en absoluto. Por ello, la conclusión es que si bien la eficiencia y constancia pueden aportar algo de información sobre la predicción de la nota, aporta poco valor a los modelos (tal y como se vio con la correlación). En un escenario con pocas variables, podrían llegar a ser útiles porque aportan algo de información (aunque para esto se deberían realizar más análisis). Sin embargo, el poder predictivo por sí mismas no es elevado y cuando hay otras variables con buen poder predictivo, puede no merecer la pena utilizar estas variables para la predicción, al igual que ocurre con la persistencia.

6.6. Tipo de metodología de recogida de datos

Las secciones anteriores han ido abordando diferentes tipos de variables a predecir, tales como las variables sobre vídeos, ejercicios, foros, SRL, comportamientos,

etc. Sin embargo, hay otro aspecto que también puede afectar a las variables predictoras, que es la metodología de recogida de datos. La metodología de recogida de datos, en este trabajo, se refiere al periodo temporal para el cual se calculan las variables que se van a utilizar para la predicción. Por ejemplo, si se quiere predecir una prueba que ocurre al final de la semana 3, se podrían tomar dos aproximaciones: (1) tomar todos los datos desde el inicio del curso hasta la semana 3 o (2) considerar solo las interacciones dentro de la semana 3. El usar una aproximación u otra no genera indicadores adicionales, pero sí cambia el valor de las variables que se calculen, por lo que esta metodología de recogida de datos tiene un impacto en la predicción.

En esta sección, se trata el efecto de las variables a predecir (en particular para la predicción de notas) cuando se cambia la metodología de recogida de datos. Para el análisis se considera el escenario 1, tratado en [261], y cuyos resultados de análisis de cambio de la metodología de recogida de datos proceden de un TFM previo [262] (que se utiliza para comparar los resultados con el escenario 2), y el escenario 2, tratado en [263]. Se usan los escenarios 1 y 2 al ser los escenarios sobre los que se predicen notas parciales, que son aquellos en los que el tipo de metodología tiene más sentido de analizar, al ser aquellos en los que la variable dependiente puede estar referida a una única parte del curso (p. ej., resultado de la prueba de la semana X). En otros escenarios, como el 5 o el 6, en los que se predice el resultado de aprendizaje de todos los ejercicios al final, teniendo en cuenta todo el curso, es más habitual que se consideren todos los datos, porque si no, se perdería parte de la información. Sin embargo, cuando lo que se predice es en base a un módulo concreto, sí tiene más sentido ver qué datos previos se utilizan (si solo los de ese módulo o todos los disponibles).

Teniendo esto en cuenta, para el escenario 1 y 2 se consideran dos modos de recogida de datos para predecir la nota de las pruebas sumativas, que son las variables dependientes en este caso. El primer modo, el no acumulativo toma los datos de la semana en la que se realiza la prueba y el segundo modo, el acumulativo, recoge los datos desde el inicio del curso hasta la semana en la que se realiza la prueba. En realidad, el periodo es algo mayor, ya que se considera como “semana” el periodo desde que se abre el contenido hasta la entrega y aunque el curso lleve un ritmo semanal, en el escenario 1 se abren los contenidos con una semana de antelación (por lo que normalmente hay dos semanas para las tareas) y en el escenario 2 se abren los contenidos con dos días de antelación (por lo que los periodos serían de nueve días). Teniendo esto en cuenta, se han elaborado los modelos predictivos para los dos modos utilizando las variables de vídeos y ejercicios (las utilizadas en el modelo A, mostrado anteriormente). Las variables de foro se han excluido por su bajo poder predictivo y las notas previas se han excluido en este análisis ya que su valor no varía dependiendo de la metodología de recogida de datos (y así el número de variables es constante para cada modelo). Con estas consideraciones, la Tabla 6.16 muestra los resultados para los dos modos de recogida de datos en el escenario 1.

Los resultados muestran que el RMSE es más bajo en el modo no acumulativo para todas las pruebas (exceptuando el test 1 en el que los datos son los mismos para los dos modos). Esto implica que obtener las variables a partir de más datos no siempre es mejor y, en este caso, es mejor utilizar únicamente las interacciones de la semana actual. Este resultado es coherente con el trabajo de Xing et al. [177], en el que observan que se obtienen mejores resultados concatenando variables de

TABLA 6.16
 PREDICCIÓN SEGÚN EL MODO DE RECOGIDA DE DATOS EN ESCENARIO 1

Modo	Algoritmo	T_1	T_2	T_3	T_4	T_5	P_3	P_5	NF
No acumulativo	RG	0.27	0.22	0.20	0.18	0.16	0.25	0.20	0.14
	SVM	0.27	0.22	0.21	0.19	0.17	0.25	0.21	0.15
	DT	0.34	0.28	0.26	0.22	0.18	0.31	0.27	0.16
	RF	0.26	0.21	0.20	0.19	0.16	0.25	0.21	0.14
Acumulativo	RG	0.27	0.25	0.24	0.23	0.21	0.27	0.22	0.15
	SVM	0.27	0.23	0.22	0.22	0.19	0.27	0.22	0.14
	DT	0.34	0.31	0.30	0.29	0.25	0.36	0.30	0.17
	RF	0.26	0.23	0.22	0.21	0.18	0.26	0.22	0.13

Nota: El mejor poder predictivo (medido en $RMSE$) de cada columna se marca en negrita

diferentes semanas que calculando variables acumuladas, aunque en este caso no se concatenan las variables, sino que solo se calculan las variables para una semana en concreto en el modo no acumulativo.

A pesar de obtener estos resultados al predecir notas intermedias, el resultado varía según el algoritmo cuando se predice la nota final, y los valores son muy similares en ambos modos. Un posible motivo es que como la nota final es la media ponderada de todas las pruebas, entonces es mejor utilizar todos los datos (los correspondientes a cada prueba) para realizar la predicción. No obstante, las últimas interacciones son mejores en general. La posible explicación es que como los alumnos empiezan haciendo las actividades pero luego abandonan, en las últimas interacciones ya solo están las de los alumnos que permanecen en el curso.

Otra observación es que la diferencia entre los modos aumenta según pasan las semanas. El motivo puede ser que en las primeras semanas, los datos previos son más recientes y tienen más relación con la semana actual. En términos de contenido, los resultados pueden implicar que hay cierta independencia entre los contenidos por lo que un alumno puede dominar unas partes mejor que otras y las interacciones de otras partes pueden no ser relevantes. En el escenario 1 puede tener sentido esta afirmación ya que hay una diferencia importante entre las tres primeras semanas (que se centran más en el desarrollo de algoritmos) y las dos últimas (que son sobre orientación a objetos y requieren mayor abstracción). Sin embargo, el contenido es incremental (así como la dificultad) y los resultados pueden ser diferentes en otros contextos. Para terminar con el escenario 1, en cuanto a los algoritmos, lo que muestran los resultados es que **RF** y **RG** son los mejores algoritmos, como sucedía en los modelos desarrollados anteriormente.

Para el escenario 2 se realizó un análisis similar para analizar las diferencias entre los dos modos, utilizando como referencia el modelo A (variables de ejercicios y `perc_vabiertos`). Los resultados de este análisis se presentan en la Tabla 6.17. En este caso, se observa que el modo no acumulativo es mejor en la mayoría de los casos, aunque la diferencia con el modo acumulativo es pequeña. De hecho en tres pruebas se obtiene el mismo poder predictivo y para el laboratorio 4 se obtiene mejor poder predictivo con el modo acumulativo. Esto implica que en este escenario el modo de recogida de datos es menos significativo. Un posible motivo para favorecer al modo acumulativo en este caso es que hay muy pocos ejercicios cortos en el escenario 2 (solo 3-4 cada semana). Por tanto, el modo no acumulativo se basa en las interacciones de solo 3-4 ejercicios, lo que puede hacer que sea menos representativo y por ello el

modo acumulativo tenga un poder predictivo más similar. En este caso, es también de notar que es mejor el modo no acumulativo para el caso del examen final. Aunque el examen final contenga contenidos de todas las unidades, como está situado al final del curso, también puede tener una mayor influencia de lo que se hace en los días previos. Por ello, también presentar mejores predicciones con el modo no acumulativo. Esto es diferente de la nota final, ya que la nota final no se obtiene en una fecha, sino que es la media ponderada de lo que ocurre a lo largo del curso, por lo que para la nota final, sí tendría más sentido que funcionara mejor el modo acumulativo, como ocurría en el escenario 1.

TABLA 6.17
PREDICCIÓN SEGÚN EL MODO DE RECOGIDA DE DATOS EN ESCENARIO 2

Modo	Algoritmo	L_1	L_2	L_3	L_4	L_5	L_6	PA	PR	EF
		W_1	W_2	W_3	W_5	W_7	W_8	W_9	W_9	W_{10}
No acumulativo	RG	0.33	0.33	0.28	0.30	0.26	0.28	0.26	0.25	0.17
	SVM	0.36	0.36	0.29	0.33	0.27	0.34	0.29	0.26	0.19
	DT	0.38	0.41	0.33	0.36	0.32	0.31	0.28	0.27	0.18
	RF	0.34	0.36	0.30	0.33	0.28	0.30	0.27	0.26	0.18
Acumulativo	RG	0.33	0.34	0.32	0.30	0.30	0.31	0.30	0.27	0.23
	SVM	0.36	0.35	0.31	0.29	0.26	0.28	0.29	0.26	0.20
	DT	0.38	0.46	0.42	0.40	0.36	0.38	0.38	0.35	0.26
	RF	0.34	0.36	0.31	0.29	0.26	0.28	0.28	0.25	0.19

Nota: El mejor poder predictivo (medido en $RMSE$) de cada columna se marca en negrita

Para finalizar, en vista de los dos escenarios, la conclusión que puede extraerse es que dependiendo del contexto, la metodología de recogida de datos puede tener un mayor o menor impacto. No obstante, el modo no acumulativo puede tener mejor poder predictivo para predecir una prueba concreta al focalizarse en interacciones relacionadas con esa prueba (sobre todo cuando haya independencia entre los temas). Sin embargo, cuando se predice una nota global (p. ej., nota final), puede ser mejor utilizar el modo acumulativo, ya que incluiría todas las interacciones y cada una de ellas tendría su impacto en la nota.

6.7. Lecciones aprendidas

Para completar este capítulo, esta sección realiza una síntesis de los resultados y establece una lista de lecciones aprendidas en torno a las variables, considerando todos los escenarios analizados. En particular, los principales resultados obtenidos en cuanto a las variables predictoras son los siguientes:

1. Las variables sobre interacciones con ejercicios (tanto formativas como sumativas) obtienen un alto poder predictivo. En todos los escenarios analizados, este tipo de variables ofrecen un buen poder predictivo. Este resultado corrobora lo obtenido en trabajos previos (p. ej., [28], [128]) en el que las variables sobre ejercicios eran muy buenas predictoras. Además, las notas de las pruebas sumativas previas siempre mejoran el poder predictivo, respecto a solo utilizar notas formativas. En los escenarios 1 y 2 se comprobó que los modelos para predecir las pruebas sumativas mejoraban al incluir las notas sumativas previas, lo que es coherente, por ejemplo, con [144], que concluyó que las notas sumativa previas eran las mejores variables para la predicción.

2. Dentro de las variables de ejercicios, las variables sobre la participación (es decir, qué ejercicios hace el alumno) pueden ser más relevantes para el abandono, mientras que las de rendimiento pueden ser más relevantes para predecir las notas. No obstante, ambos tipos de variables pueden ser buenas predictoras para ambas cosas.
3. Las variables sobre vídeos y actividad en el curso normalmente ofrecen un buen poder predictivo por sí solas, aunque en algunos casos podría no ser suficiente (como en el escenario 5 al predecir el éxito académico). No obstante, en otros casos, como el escenario 3, sí pueden obtener un muy alto poder predictivo por sí mismas o estar entre las mejores predictoras (como en el escenarios 4). En otros trabajos (p. ej., [126], [133]) también se han obtenido buenos resultados con estas variables. Por ello, se recomienda utilizarlas si están disponibles.
4. La actividad en el foro de los MOOCs no influye en la predicción de los resultados académicos. Tanto en el escenario 1 como en el 2, las variables de foro no mejoraron los modelos predictivos para predecir la nota. Aunque hay cursos con uso muy intensivo de foro donde se han obtenido buenos resultados (p. ej., [129]), en otros trabajos se corrobora el bajo poder predictivo de estas variables (p. ej., [131], [275]). Por ello, este tipo de variables pueden no ser útiles para predecir esta variable y no merecería la pena recogerlas en este caso.
5. Las variables de SRL tienen un alto poder predictivo en la predicción del abandono en MOOCs, lo que apoya la idea de que la autorregulación puede ser importante para tener éxito en los MOOCs ([96], [99]) y en especial aquellos asíncronos, que son en los que se analiza esta variable. Estas variables pueden añadir menor valor cuando hay otras variables de ejercicios (probablemente por su relación), pero alcanzar un alto poder predictivo por sí solas. Por ello, podría merecer la pena su uso cuando estén disponibles.
6. Las variables de comportamientos, tales como la persistencia, eficiencia y constancia aportan poco valor a los modelos predictivos. Podrían añadir algo de valor y alguna mejora al modelo, pero la mejora no es significativa. Por ello, se podría prescindir su cálculo para el uso en la predicción del abandono o el rendimiento académico. No obstante, en algún caso, con la ausencia de otras variables relevantes en el modelo, podrían llegar a ser útiles y como parte del trabajo futuro se podrían analizar más los posibles casos. También, estas variables aportan una componente alternativa al rendimiento y añaden un valor descriptivo, por lo que su información en sí misma puede ser útil para un profesor. Asimismo, puede haber otro tipo de comportamientos que podrían mejorar las predicciones. Por ejemplo, Wei et al. [282] identificaron que variables derivadas de las trayectorias del ratón al realizar los ejercicios (p. ej., tiempo en pensar la pregunta) podrían mejorar el poder predictivo. Por ello, más trabajo se debería realizar en explorar nuevas variables.
7. Aunque puede depender del contexto, al predecir pruebas parciales, la mejor aproximación puede ser coger datos de las interacciones relacionadas con los contenidos de esa prueba (modo no acumulativo) en vez de coger los datos desde el inicio del curso (modo acumulativo). Este resultado es coherente con

Xing et al. [177], que obtenía mejores resultados concatenando variables de distintas semanas que utilizando variables acumuladas. Un posible motivo para que el modo no acumulativo sea mejor es que en este modo, las interacciones están más relacionadas con la prueba a predecir (sobre todo si hay independencia en los temas) y se filtran las interacciones de los alumnos que interaccionaron al inicio y después fueron abandonando el curso. No obstante, si hay pocos ejercicios relacionados con la prueba a predecir, una aproximación acumulativa podría funcionar mejor, por lo que se debe tener en cuenta el contexto educativo.

Aparte de los resultados respecto a las variables, otros resultados relevantes obtenidos en este capítulo son los siguientes:

1. Las pruebas situadas al inicio del curso son más difíciles de predecir. Dado que al inicio del curso mucha gente interactúa pero no está comprometido con el curso, las predicciones de las primeras pruebas pueden obtener peor poder predictivo. Esta observación coincide con otros trabajos, como [27], [153].
2. Los algoritmos de predicción tienen menos impacto que los datos utilizados en los modelos. En los análisis, se ha observado que RF y RG son los mejores modelos de entre los cuatro utilizados, pero las diferencias entre un modelo y otro son generalmente pequeñas. Este resultado es coherente con la revisión de Gardner y Brooks [188], en la que se indica que las variables tienen mayor impacto con los algoritmos y no se concluye que haya algún algoritmo consistentemente mejor que el resto.
3. La definición de las variables debe adaptarse a cada contexto. Por ejemplo, en el escenario 5 se observa que la definición de abandono relacionada con la actividad puede no ser apropiada en algunos casos si el curso no está definido para usarse durante todo el semestre.

Para concluir, pueden extraerse múltiples indicadores para los modelos predictivos. Sin embargo, de estos indicadores, los que normalmente tienen mejor poder predictivo son los de ejercicios, por los que la recomendación es incluir este tipo de variables siempre que sea posible. También es conveniente añadir las notas de las pruebas sumativas que ya se hayan realizado y se conozca la nota. Además, si se pueden añadir otras variables, las que más pueden aportar son las relacionadas con la actividad, los vídeos y el SRL. Asimismo, se debe tener en cuenta qué datos se toman para la predicción ya que para pruebas parciales, los datos relacionados con esa prueba pueden ser más relevante. No obstante, es importante analizar el contexto particular en cada caso para ver si existe algún factor que deba incluirse en el modelo o que modifique la definición de los indicadores existentes debido al contexto.

7

Análisis temporal

Contenido

7.1. Evolución temporal del poder predictivo en cursos síncronos	110
7.2. Evolución temporal del poder predictivo en cursos asíncronos	122
7.3. Lecciones aprendidas	124

Uno de los objetivos fundamentales de desarrollar modelos predictivos es que puedan ser utilizados durante la ejecución de los cursos para poder detectar qué estudiantes están en riesgo o tienen dificultades de forma temprana [267], de manera que puedan tomarse medidas que mejoren la situación. Sin embargo, un factor relevante para que esas medidas puedan tener un efecto positivo en el aprendizaje es el instante de tiempo en que se puede predecir con suficiente precisión. Si las predicciones se realizan con suficiente antelación, puede ser posible realizar intervenciones que produzcan una mejora. En cambio, si las predicciones se realizan en el último momento, puede no haber tiempo suficiente para tomar las medidas. No obstante, en los primeros momentos de un curso habrá menos cantidad de interacciones de los alumnos y será más difícil realizar una predicción precisa. Por este motivo, es importante analizar cómo evoluciona el poder predictivo a lo largo del tiempo y ver a partir de qué momento se pueden obtener unas predicciones lo suficientemente buenas para ser utilizadas en los modelos. Este capítulo aborda este tema y realiza un análisis temporal de la predicción considerando varios escenarios de aprendizaje. El capítulo se divide en tres secciones. La primera de ellas se centra en analizar la evolución del poder predictivo en escenarios de aprendizaje síncronos a lo largo del curso y determinar cuál es el mejor momento para empezar a predecir, teniendo en cuenta el balance entre anticipación y poder predictivo. La segunda sección se centra en el análisis temporal en cursos asíncronos, en el cual cambia el eje temporal al poder matricularse y acceder al curso en cualquier momento. Finalmente, la tercera sección muestra las lecciones aprendidas en este capítulo.

7.1. Evolución temporal del poder predictivo en cursos síncronos

En esta sección, se analiza la evolución del poder predictivo a lo largo del tiempo en los diferentes escenarios. Este problema ha sido abordado en trabajos previos de la literatura (p. ej., [23], [24]), aunque generalmente se ha trabajado en contextos de un solo curso / tipo de curso. En esta tesis, se extiende el análisis con nuevos escenarios heterogéneos, que abarcan diferentes tipos de cursos y niveles educativos, incluyendo enseñanza primaria, secundaria, preparación a la universidad, enseñanza universitaria y enseñanza a distancia (MOOCs). El análisis de estos escenarios sirve para conocer mejor en diferentes contextos cuál es el momento a partir del cual las predicciones pueden ser lo suficientemente precisas, y para en base al análisis de diferentes escenarios, extraer conclusiones globales sobre la evolución temporal de los modelos predictivos.

Para el análisis de esta sección, se consideran cursos síncronos, de modo que se pueda determinar a partir de qué momento en la línea temporal del curso se obtiene un buen poder predictivo y cuál es el mejor momento para empezar a predecir. Por cursos síncronos se entiende a aquellos cursos que siguen el ritmo marcado por el profesor y en el que todos los alumnos deben interactuar con las mismas actividades en el mismo tiempo [78].

En este caso, todos los escenarios son síncronos excepto el escenario 3, que se aborda en la sección 7.2. No obstante, a pesar de que el resto pueden considerarse síncronos, presentan ciertas diferencias en el contexto.

- Los escenarios 1 y 2 son más similares, ya que en ellos se abren los contenidos periódicamente y hay entregas semanales, que quedan cerradas una vez que se vencen los plazos. Dado que estos dos escenarios son los únicos que tienen entregas parciales, de los considerados en esta sección, únicamente en estos escenarios se analiza la predicción de notas parciales.
- En el escenario 4, en cambio, no hay entregas obligatorias. Dado que el curso sirve para preparar una prueba de acceso, los alumnos pueden interactuar libremente con los ejercicios, por lo que algunos podrían usar los cursos como si fueran asíncronos. Sin embargo, para guiar a los estudiantes a que sigan un ritmo, los instructores van liberando los contenidos periódicamente y organizan sesiones presenciales, de modo que los estudiantes deberían idealmente seguir el ritmo marcado por el profesor. En este caso, la predicción se centra en el éxito en la prueba de acceso.
- En el escenario 5, los SPOCs sirven de apoyo a los cursos universitarios, aunque cada profesor puede darle un uso diferente. Se pueden considerar síncronos porque los cursos universitarios tienen un temario que se imparte de forma semanal, por lo que los estudiantes deberían realizar las actividades del SPOC según se cubran los contenidos. No obstante, dependiendo del curso, el momento en el que los alumnos realizan las actividades puede variar. Puede haber cursos en los que el profesor exija entregar algo periódicamente, lo que puede obligar a los alumnos a interactuar periódicamente. Pero también puede haber cursos en los que se suban los materiales y el alumno se organice en cubrirlos, de manera que al final tenga los conocimientos necesarios para el

examen o prueba de evaluación. En este caso, la predicción se centra en el éxito y el abandono en los cursos.

- En el escenario 6, todas las actividades están disponibles desde el principio, pero el maestro o profesor típicamente manda a sus estudiantes realizar las actividades según va cubriendo los temas a lo largo del curso. Por tanto, se asume que todos los estudiantes llevan el mismo ritmo. En este caso, la predicción se centra en el abandono y la nota final obtenida en la plataforma.

Debido a estas diferencias entre escenarios, el tratamiento de cada escenario es diferente. En los dos primeros, se toma la semana como la unidad de tiempo y se añaden datos semanalmente para predecir las diferentes pruebas parciales. En el escenario 4, se toman como referencia diferentes puntos claves del curso, en los que se espera que los alumnos hayan dedicado mayor tiempo al curso. En los escenarios 5 y 6, dado que no hay ninguna referencia temporal de pruebas o hitos importantes del curso, se toma como referencia la semana y quincena, respectivamente, para el análisis temporal.

Además, para este análisis, hay que tener en cuenta que el poder predictivo se mide en **AUC** para unas variables y **RMSE** para otras (tal y como se justificó en la sección 3.4). Para aquellos casos donde se utiliza el **AUC**, se toma el criterio para empezar a predecir cuando hay un valor de al menos 0.8 (que es un **AUC** bueno, como se introdujo en la sección 2.4.4), aunque se evalúa el momento a partir del cual el **AUC** es excelente (superior a 0.9). Únicamente para los casos en los que el poder predictivo es peor, se considera un **AUC** moderado (entre 0.7 y 0.8). Para la predicción de notas, en la que se usa el **RMSE**, no se establecen umbrales concretos, aunque se toma como referencia que Pérez-Sanagustín et al. [271] alcanzaron un mejor **RMSE** de 0.18, Elbadrawy et al. [27] de 0.15, y Ding et al. [153] de hasta 0.12. Por tanto, se espera que el valor del **RMSE** esté en la línea de dichos trabajos para que pueda proporcionarse una estimación de la nota razonable. Con estas consideraciones, a continuación se presenta el detalle de cada uno de los escenarios.

Para empezar el análisis, se considera la predicción del éxito en el escenario 1, que fue tratado en [261], y en un TFM previo [262] (cuyos resultados se presentan para poder ponerlos en contexto y compararlos con otros escenarios a lo largo del capítulo). En este caso, se considera el éxito como aprobar el MOOC. Para analizar el éxito, se desarrollaron modelos predictivos con los cuatro algoritmos utilizados, añadiendo los datos semana a semana y utilizando el modelo que incluye las variables de ejercicios y `perc_vabiertos` como variable sobre vídeos (modelo A, de los considerados en el capítulo 6). Los resultados de estos modelos se presentan en la Fig. 7.1.

Los resultados muestran que el mejor modelo se obtiene con el algoritmo SVM, con un **AUC** de 0.97. Este resultado es muy bueno a final de curso y acorde con otros trabajos. Por ejemplo, Ruipérez-Valiente et al. [23] analizó la evolución del poder predictivo en la predicción de certificados en un MOOC y aunque obtuvo un **AUC** de casi 1.0 al final de curso, el **AUC** era de 0.95 tras la quinta semana, que es la última en este curso. Aunque la duración podría ser relevante según el contexto, en los resultados de la evolución temporal de este escenario 1, se aprecia que se puede obtener un **AUC** de 0.83 desde la semana 1 y de 0.95 a partir de la semana 3. Esto implica que se pueden obtener unas predicciones buenas desde el inicio del MOOC. En este caso, teniendo en cuenta que el **AUC** es bueno (superior a 0.8) en

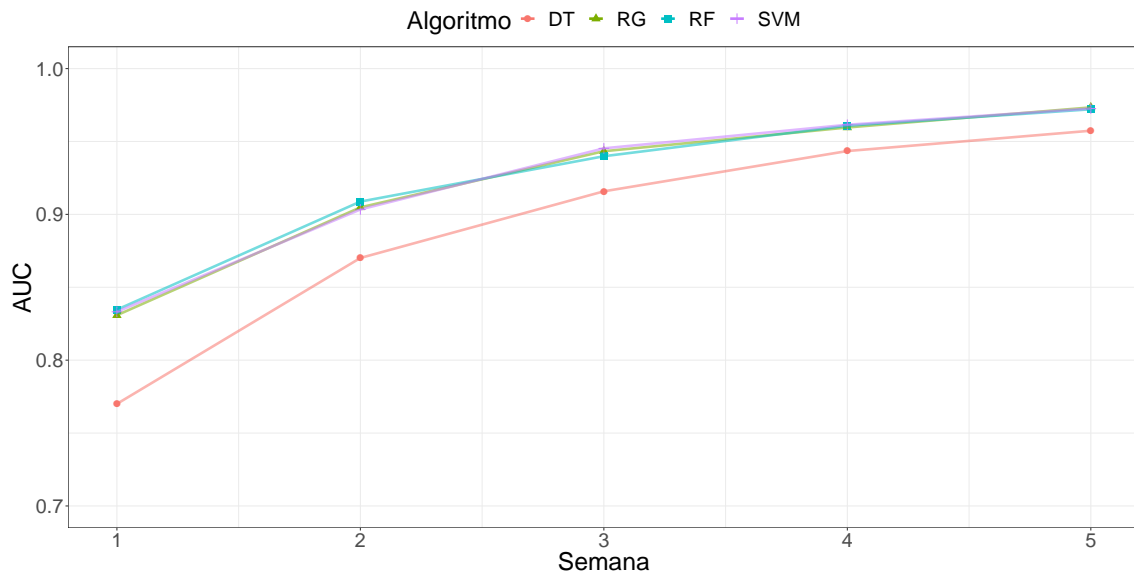


Fig. 7.1. Evaluación de los resultados de la predicción del éxito a lo largo del tiempo en el escenario 1

la primera semana, se podrían utilizar los modelos predictivos desde el inicio (tras una semana) para realizar intervenciones tempranas que pudieran mejorar el éxito de los alumnos.

En el escenario 1, aparte del éxito, la variable dependiente que se ha utilizado en el capítulo anterior es la nota de las pruebas que se van realizando a lo largo del MOOC. En el capítulo anterior, se utilizaban los datos de las interacciones de los alumnos desde el inicio o desde que se abría la prueba hasta la fecha de finalización de dicha prueba. El periodo desde que se abría la prueba hasta la finalización normalmente era de 15 días (salvo en la primera semana que eran tres semanas). Sin embargo, el problema es que al tomar los datos hasta la fecha máxima para realizar la prueba, los modelos no se desarrollan con suficiente antelación como para que puedan tener un efecto en los estudiantes. Por ello, en esta sección se analiza el poder predictivo que se obtiene al utilizar datos de los módulos o semanas previas. Además, se analiza cómo es el efecto de utilizar un modo acumulativo (datos desde el inicio hasta cierta semana previa) o solo de una semana anterior (modo no acumulativo). La Fig. 7.2 muestra la evolución del RMSE al predecir cada una de las pruebas del curso utilizando datos de semanas previas (en el eje X) en modo acumulativo y no acumulativo. Para estos modelos se utiliza RF, por su alto poder predictivo y el modelo con las variables de ejercicios y vídeos (llamado modelo A en el capítulo anterior). En la Fig. 7.2, nótese que después de realizar la prueba, no tiene sentido la predicción. Por ello, las líneas terminan en la semana en la que se realiza la prueba (p. ej., la predicción de T_3 se realiza hasta la semana 3).

En esta figura, se observa que al predecir los test (T_i), la mejor opción es usar los datos de los módulos previos sin considerar todos los datos desde el principio. Esto podría suponer que los contenidos inmediatamente previos tienen una mayor relación y son más útiles para predecir las notas. Sin embargo, al predecir las tareas de programación (P_i), se obtiene un mejor poder predictivo usando todos los datos disponibles aunque hay un punto donde los resultados mejoran utilizando solo los datos de la última semana. Un posible motivo para que los modelos sean mejor solo

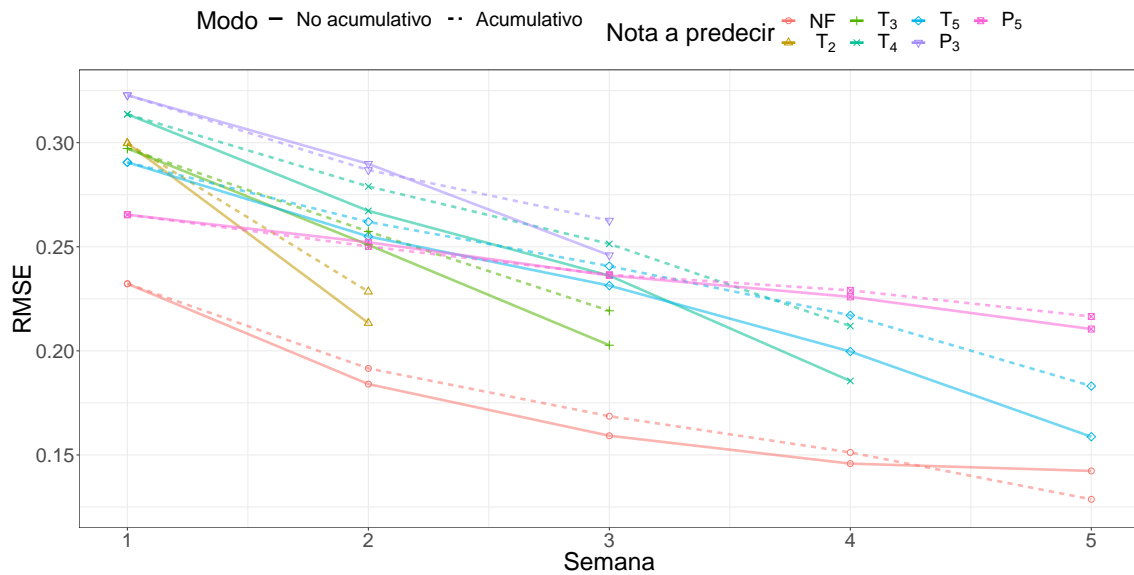


Fig. 7.2. Evolución temporal de la predicción de notas usando datos de módulos previos en el escenario 1

con las últimas interacciones es que puede haber muchos alumnos al inicio, pero luego abandonan el curso, y las interacciones de esos alumnos pueden afectar en las predicciones. Como conclusión de este análisis es que los datos de los módulos previos pueden ser usados para conseguir una mayor anticipación. El poder predictivo será peor que si se utilizan los datos de las interacciones del módulo sobre el que se va a predecir la nota, pero se conseguirá una mayor anticipación, que será fundamental para conseguir un impacto en el aprendizaje. Por ello, dada la evolución del poder predictivo, la recomendación en este caso sería utilizar las interacciones en el módulo anterior para construir el modelo desde el inicio del módulo y realizar intervenciones lo antes posible.

Para continuar el análisis, se consideró el escenario 2, tratado en [263]. En primer lugar, se analizó la evolución temporal de la predicción de notas para cada una de las pruebas intermedias de dicho MOOC, siguiendo la misma metodología aplicada en el escenario 1. Por tanto, se utilizó el modelo con variables de ejercicios y vídeos y RF, y se predijeron las notas utilizando los datos de los módulos previos de forma acumulativa y no acumulativa. Los resultados se muestran en la Fig. 7.3.

Los resultados muestran que a diferencia del escenario 1, no está claro si es mejor utilizar un modo acumulativo o no acumulativo. El modo acumulativo es mejor para algunas pruebas pero no para otras, y en algunos casos (como en el examen final, *EF*) el mejor modo varía lo largo del tiempo. Además, la figura muestra que es difícil obtener unas predicciones lo suficientemente precisas hasta el último momento en el escenario 2. Una posible diferencia con respecto al escenario 1 es que la desviación típica de los resultados de los estudiantes en las diferentes tareas es mayor que en el escenario 1 (0.29 frente a 0.14), lo que puede hacer que sea más difícil predecir en el escenario 2. Este hecho también puede estar relacionado con la anticipación. Los resultados muestran que el poder predictivo mejora semana a semana, pero como el poder predictivo no es muy alto hasta el último momento (teniendo en cuenta que otros trabajos, como [153], [271] se han obtenido valores entre 0.12-0.18), no es posible realizar predicciones muy tempranas. En este caso, teniendo en cuenta el

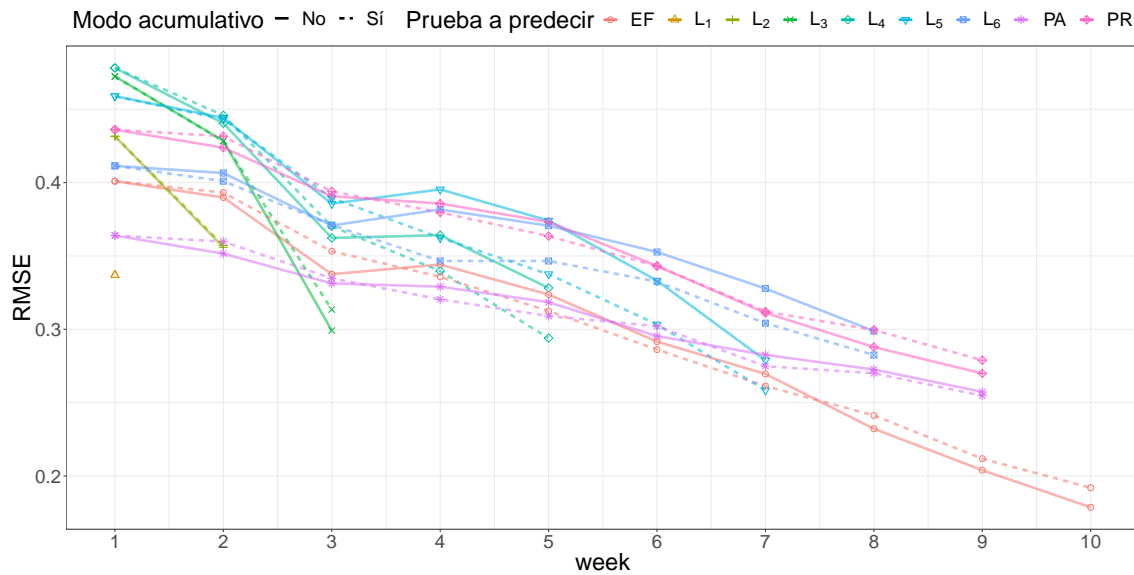


Fig. 7.3. Evolución temporal de la predicción de notas usando datos de módulos previos en el escenario 2

balance entre poder predictivo y anticipación, una buena opción podría ser predecir con una semana de antelación, de modo que se obtenga el mejor poder predictivo y haya cierta anticipación. De este modo, a partir de la segunda semana podrían utilizarse los modelos y realizar intervenciones, aunque cada modelo particular se implementaría utilizando datos de la semana anterior (ya sea en modo acumulativo o no acumulativo, según el mejor modo en cada caso).

Los modelos presentados hasta ahora sirven para predecir cada una de las pruebas del curso, pero en el escenario 2, a diferencia del escenario 1, está el problema de que las variables de ejercicios cuentan directamente para de la nota final. Por ello, para predecir la nota final se ha utilizado un modelo con otras variables. En particular, se ha utilizado un modelo incluyendo todas las variables disponibles en el escenario 2, a excepción de las variables de ejercicios y notas previas. Es decir, se incluyen variables de vídeos, foro y actividad de la plataforma (de las cuales, algunas no estaban disponibles en el escenario 1). Teniendo en cuenta estos modelos, la Fig. 7.4 muestra la evolución temporal del poder predictivo al predecir la nota final del MOOC del escenario 2. Dado que en el escenario 1 también se analiza el éxito académico y este depende de la nota final, la Fig. 7.4 también muestra la predicción del éxito académico (resultado de aprobar/suspender), tal y como se presentó en el escenario 1, en la Fig. 7.1.

Los resultados muestran, en cuanto al poder predictivo, que el mejor RMSE al predecir la nota final se obtiene al final de curso con RF y es de 0.18 (al igual que en [271]), lo que implica que aunque haya errores en la predicción, se puede dar una estimación aceptable. No obstante, por ejemplo, el RMSE solo baja de 0.20 a partir de la semana 8 (y usando el modo no acumulativo), por lo que no se puede tener una predicción de la nota precisa con demasiada antelación. Sin embargo, los resultados de la predicción del éxito sí que indican que es posible obtener una predicción temprana del éxito (aunque no tan temprana como en el escenario 1 u otros trabajos como [23]). En particular, es posible obtener un AUC bueno, superior a 0.8, a partir de la semana 4 (40% del curso) y un AUC excelente, superior a

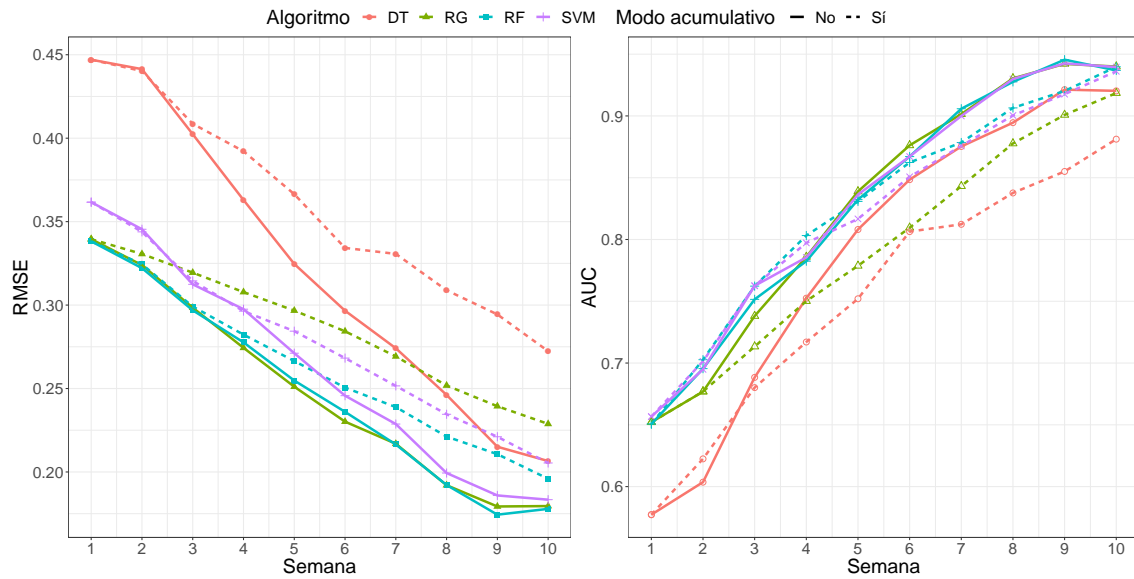


Fig. 7.4. Evolución temporal de la predicción de la nota final en el escenario 2

0.9, a partir de la semana 7 (70%) del curso. Por tanto, en este caso se podrían proporcionar predicciones del éxito a partir de, por ejemplo la semana 4, y en las últimas semanas complementarlas con la predicción de la nota (que es una variable más difícil de predecir). En cuanto a los algoritmos, en este caso, el poder predictivo de todos ellos (salvo DT) es muy similar. Respecto al modo de recogida de datos, se observa que el modo no acumulativo funciona en general mejor, si bien al final de curso no hay diferencia con algunos algoritmos (RF y SVM), lo que indica que cualquiera de las dos aproximaciones podría ser válida.

En relación con el escenario 1, se observa que el poder predictivo de la nota final es peor en el escenario 2, que en el 1. Esto puede ser debido a que el poder predictivo de las notas parciales también es peor, como se ha comentado previamente. También, en cuanto al éxito, es necesario más tiempo para conseguir predicciones precisas. Por ejemplo, se necesita un 40% del curso para llegar a un AUC de 0.8 en el escenario 2, pero solo un 20% en el escenario 1. Esto implica que aunque los escenarios sean similares en cuanto a temática puede haber diferencias debido al contexto.

Para comprender mejor cómo puede variar el análisis temporal en otros contextos, se continúa el análisis con el escenario 4, analizado en [29], [269] (el escenario 3 se analiza de forma separada, en la sección 7.2 al ser un curso asíncrono). Este escenario, a diferencia de los anteriores es de SPOCs y la predicción se no se realiza por cada prueba intermedia, ya que solo hay una prueba externa, y se realiza al final. En particular, el objetivo en este caso es predecir el éxito en la parte de ciencias de la prueba de acceso a la universidad para medicina y odontología. Para el análisis, se tomó el SPOC de Química del curso 2016/2017 y se realizó el análisis temporal. En cuanto a los periodos a analizar, se seleccionaron siete periodos (T_i) correspondientes a siete puntos claves en el programa de aprendizaje semipresencial. T_1 , T_2 y T_4 se corresponden con las sesiones presenciales interactivas, que fueron organizadas para realizar problemas o tratar temas específicos del SPOC. En T_3 , se organizaron clases sobre temas que no formaban parte del SPOC de Química, pero que eran fundamentales para el examen. La primera sesión de la prueba de acceso fue en T_5 , y la recuperación fue organizada en T_6 (había dos sesiones para dar una segunda

oportunidad a los que suspendían la primera prueba). Finalmente T_7 incluye todas las interacciones del SPOC. Teniendo en cuenta dichas fechas, se desarrollaron los modelos predictivos tomando los datos desde el inicio del curso (7 de septiembre de 2016) hasta cada T_i . Los resultados de los modelos predictivos en los diferentes periodos se muestran en la Tabla 7.1.

TABLA 7.1
EVOLUCIÓN TEMPORAL DE LA PREDICCIÓN EN EL ESCENARIO 4

Periodo	T_1	T_2	T_3	T_4	T_5	T_6	T_7
Fecha final del periodo	22/10	14/01	07/04	06/05	05/07	30/08	
Nº estudiantes incluidos	33	79	104	107	113	114	114
% de actividad incluido	2.2 %	13.4 %	31.0 %	42.6 %	77.6 %	99.8 %	100 %
RF	0.46	0.45	0.70	0.78	0.84	0.87	0.87
RG	0.59	0.71	0.72	0.73	0.74	0.77	0.77
SVM	0.55	0.51	0.72	0.73	0.84	0.85	0.85
DT	0.50	0.50	0.70	0.71	0.78	0.80	0.80

Nota: El mejor poder predictivo (medido en AUC) de cada columna se marca en negrita

Los resultados muestran que al inicio del curso el poder predictivo es pobre. Considerando un umbral de 0.8 para considerar que el AUC es bueno, no se alcanza un poder predictivo bueno hasta T_5 , que es la fecha de la primera sesión del examen (y se corresponde con el 90 % de la duración del curso, contando hasta desde el inicio en septiembre hasta el final de agosto, cuando es la segunda sesión del examen). Un posible motivo es que hay muy baja actividad al inicio del SPOC (57.4 % de las interacciones ocurren después de T_4). Este factor es de interés en este escenario en concreto, ya que mientras que en los escenarios previos hay unas pruebas semanales y debe haber actividad desde el inicio, en este contexto puede haber mayor variabilidad de la interacción al inicio, ya que es un curso de apoyo y cada alumno puede inscribirse en un momento diferente del curso. Este factor también se diferencia de los escenarios 5 y 6, porque aunque el uso de los cursos pueda variar dependiendo del profesor en dichos escenarios, todos los alumnos (salvo casos excepcionales) deberían estar matriculados desde el inicio del curso y seguir el ritmo, lo cual no ocurre aquí.

Dado que en este escenario no se alcanza un AUC bueno hasta el final, se analiza el caso de obtener un AUC moderado (al menos 0.7). Tomando este criterio, se pueden considerar las predicciones a partir de T_3 (que se corresponden con el 58 % del periodo total del curso), e incluyen el 31 % de las interacciones del SPOC. Este valor es bastante superior al 13.4 % incluido en T_2 , donde todavía muchos alumnos no se han matriculado, lo que hace que los modelos se calculen con pocos casos, puedan ser menos robustos y alcancen peor poder predictivo. Por otro lado, el bajo nivel de actividad podría indicar que los alumnos no utilizan el SPOC de forma síncrona como inicialmente se había planeado. Y este hecho podría afectar a los modelos predictivos ya que la actividad de los estudiantes no es uniforme a lo largo del curso. Este hecho también invita a la reflexión sobre el diseño del SPOC, ya que éste puede tener un impacto relevante en el éxito y/o abandono [283]. Dentro del diseño semipresencial, sería bueno enfatizar la relevancia de las sesiones presenciales para asegurar que los alumnos interactúan con el SPOC desde el inicio y que asisten a todas las sesiones presenciales.

En términos de algoritmos, el mejor algoritmo es el **RF** a partir de T_4 , y este algoritmo es el que alcanza un mejor **AUC** de 0.87. Aunque las diferencias no son muy grandes en algunos periodos, este algoritmo parece ser el más consistente cuando hay suficientes datos para predecir. Sin embargo, en T_3 se alcanzan mejores resultados con **RG** o **SVM**. Por ello, y teniendo en cuenta todo lo comentado anteriormente y el balance entre anticipación y poder predictivo, una recomendación para este escenario sería predecir a partir de T_3 (usando **RG** o **SVM** para ese periodo) o T_4 (usando **RF** para ese periodo y los sucesivos). De este modo, se conseguiría un poder predictivo razonable, y las predicciones se realizarían 2-3 meses antes de la prueba de acceso, por lo que podría haber tiempo suficiente para realizar posibles intervenciones con los estudiantes.

En este escenario, a pesar de que el trabajo de los estudiantes no era uniforme a lo largo del año, la predicción estaba centrada en una prueba concreta, cuyo resultado se conocía para el entrenamiento. En los dos últimos escenarios, se toman cursos de diferentes niveles educativos, pero no se conoce el resultado que obtuvieron en los cursos. Por tanto, el análisis se realiza en base a sus indicadores en la plataforma.

En cuanto al escenario 5 (**SPOCs** de la **UC3M**), las variables dependientes son el abandono y el éxito en las actividades del **SPOC**. El abandono se midió de dos formas, según la actividad de los estudiantes, de modo que un estudiante abandona si no accede a la plataforma durante más de dos semanas consecutivas. La segunda aproximación es la relacionada con completar un mínimo de actividades (en este caso 75 %). Para predecir estas variables, se han utilizado las variables de actividad, ejercicios y vídeos, y un modelo con todas las variables para cada uno de los cuatro algoritmos propuestos en el capítulo 3. Para el análisis temporal, se debe tener en cuenta que los **SPOCs** tienen una duración de 16 semanas y que se utiliza un modelo global, por lo que el **AUC** es el **AUC** global respecto a todos los cursos. Teniendo en cuenta estas consideraciones, los resultados, se muestran en la Fig. 7.5.

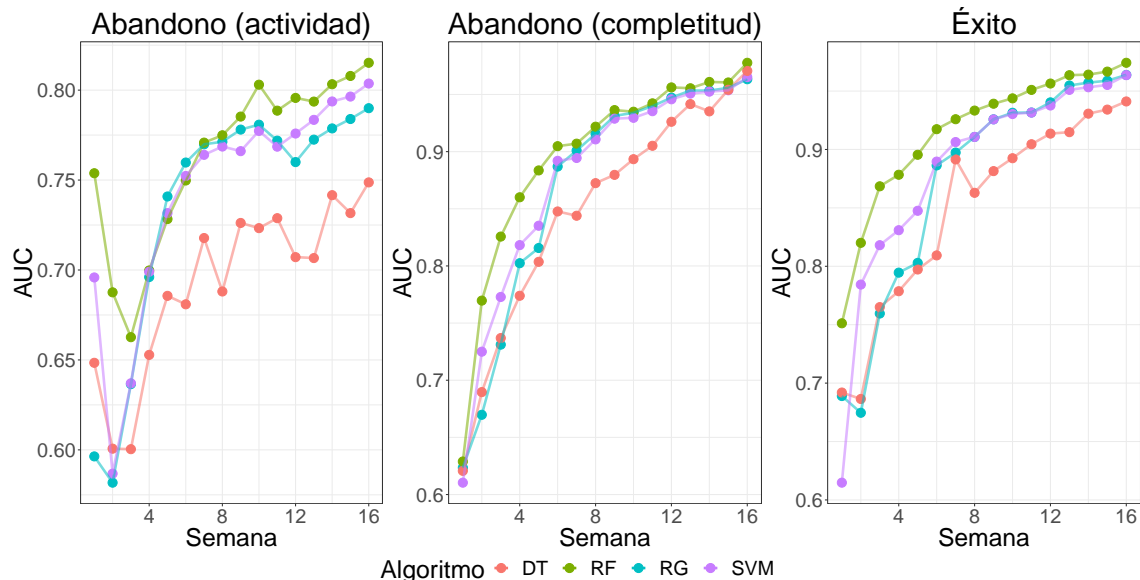


Fig. 7.5. Evolución temporal de la predicción en el escenario 5

Los resultados muestran que para el caso del abandono relacionado con la actividad, se observa que solo se consigue un **AUC** bueno (superior a 0.8) en el final del curso. Por tanto, los modelos para predecir la actividad no permiten una

buena anticipación y si se usan, deberían usarse a partir de al menos la mitad de curso, donde los AUC se acercan más a 0.8. Como mucho, los modelos se podrían utilizar a partir de la semana 4 (25 %) del curso, donde el AUC es moderado. No obstante, si la información es complementaria, los modelos podrían centrarse más en el abandono relacionado con la completitud, ya que también muestra una idea del seguimiento del curso y los resultados son bastante más precisos. En cuanto a esta última definición de abandono, se observa que es posible tener un AUC superior a 0.8 a partir de la semana 3 y superior a 0.9 a partir de la semana 6. Por ello, es posible obtener unas predicciones tempranas y antes de llegar a la mitad del curso, se podrían presentar unas predicciones bastante precisas. En este caso, se podrían proporcionar las predicciones desde la semana 3, de modo que el profesor pueda realizar algún tipo de intervención con aquellos estudiantes con mayor riesgo. De forma similar, es posible obtener buenas predicciones tempranas en cuanto al éxito académico. En este caso, es posible obtener un AUC superior a 0.8 a partir de la semana 2 y de 0.9 a partir de la semana 6. En cuanto a porcentaje, teniendo en cuenta la semana 3, donde se superaría el AUC de 0.8 para tanto el abandono (relacionado con completitud) como el éxito, se necesitaría un 19 % del curso. Para superar un AUC de 0.9 se necesitaría un 38 % del curso. Por tanto, en ambos casos se podrían obtener predicciones bastante tempranas que permitan realizar intervenciones, lo cual es positivo. No obstante, una limitación es que se utiliza un modelo global y el AUC que se muestra es global a todos los cursos. Sobre esta limitación se volverá a tratar en el capítulo siguiente (sección 8.2).

Por último, se analiza el caso del caso del escenario 6 (plataforma educativa para estudiantes de primaria y secundaria). Para este caso, las variables dependientes son el abandono y la nota final. Para el análisis, se han considerado variables obtenidas a partir de interacciones con los ejercicios. En particular, se ha utilizado el grupo de variables que incluye tanto las variables de participación como las de rendimiento, ya que ambos grupos de variables tienen un alto poder predictivo, de acuerdo con lo visto en el capítulo anterior. Teniendo esto en cuenta, las Fig. 7.6 y 7.7 muestran la evolución temporal de la predicción del abandono y la nota, respectivamente para el escenario 6.

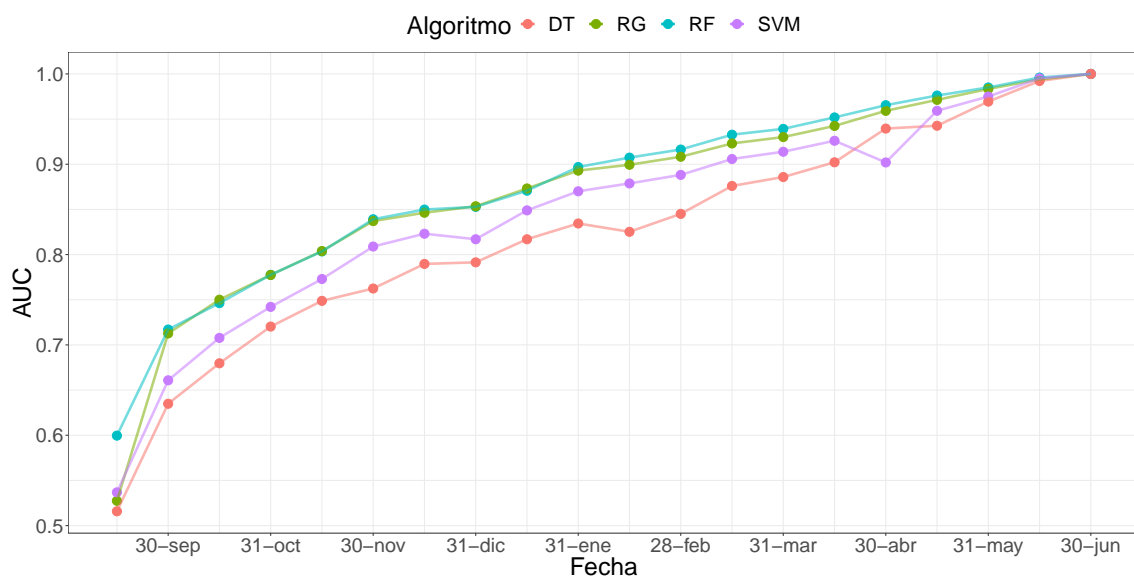


Fig. 7.6. Evolución temporal de la predicción del abandono en el escenario 6

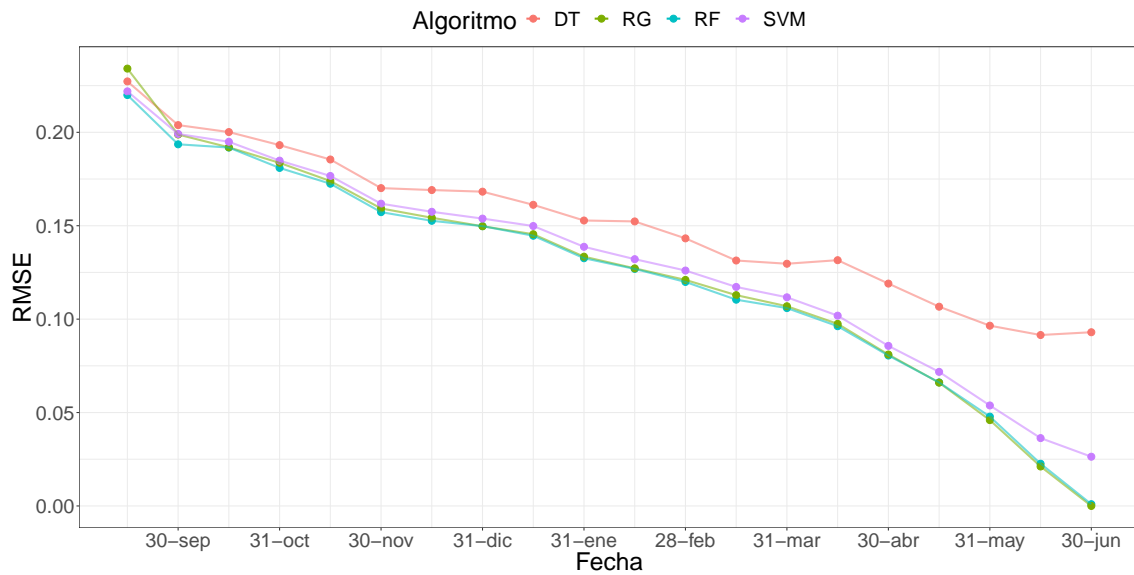


Fig. 7.7. Evolución temporal de la predicción de la nota en el escenario 6

Los resultados muestran que, en cuanto al abandono, es posible obtener un **AUC** bueno (superior a 0.8) desde mediados de noviembre y un **AUC** excelente (superior a 0.9) desde mediados de febrero. En cuanto a la predicción de la nota, es posible obtener un **RMSE** inferior a 0.2 a partir del primer mes, y para los periodos citados en la predicción del abandono, es posible obtener un **RMSE** de 0.17 a mediados de noviembre y de 0.13 a mediados de febrero. Estos valores están en la línea de otros trabajos que reportan **RMSE** entre 0.12-0.18 (p. ej., [153], [271]). Por tanto, el periodo para empezar a predecir la nota puede ser similar al del abandono en este caso, lo cual sería más favorable que en el escenario 2, en el que la predicción de la nota se proporciona más tarde. En este caso, considerando el balance entre poder predictivo y anticipación, una recomendación podría ser empezar a mostrar las predicciones desde mediados de noviembre (a los 2.5 meses del inicio del curso), ya que el poder predictivo es razonablemente bueno y se podrían realizar intervenciones, lo cual indica que es posible realizar predicciones tempranas. Teniendo en cuenta que la primera evaluación del curso se realiza a mediados de diciembre (en el sistema educativo español, hay tres evaluaciones en primaria y secundaria, que ocurren normalmente a mediados de diciembre, justo antes de Semana Santa, y a final de curso), estas predicciones pueden servir incluso para tomar medidas que mejoren los resultados en la primera evaluación. Y a partir de la primera evaluación, los profesores pueden tener una retroalimentación en base a los resultados obtenidos, y pueden seguir recibiendo predicciones más precisas para realizar el seguimiento de los estudiantes.

La limitación principal que deben tener en cuenta los profesores es que las predicciones se realizan en base a la actividad en la plataforma y no en base a la actividad presencial. Por ello, puede ocurrir que un alumno obtenga muy buenas notas en las evaluaciones presenciales, pero la predicción de su nota y los indicadores asociados muestren malos resultados. En ese caso, podría pasar que el alumno comprenda los conceptos, pero no trabaje en la plataforma. A partir de la primera evaluación, el profesor podría tratar también esos casos y ver por qué el alumno no trabaja en la plataforma y analizar cómo podría hacer para que el alumno trabajara más en las actividades propuestas.

En cuanto a los algoritmos, los resultados muestran que **RG** y **RF** son los algoritmos que proporcionan un mejor poder predictivo, con valores de poder predictivo muy similares, tanto en el abandono como en la nota. El algoritmo que obtiene un peor poder predictivo es **DT** y **SVM** obtiene en la mayoría de los casos un valor intermedio entre **RG/RF** y **DT**. No obstante, los resultados muestran que el uso de unas variables u otras puede tener más impacto que el propio algoritmo. Por ello, entre estos cuatro algoritmos, se preferiría el **RF/RF**, aunque se pondría mayor atención en las variables.

Respecto a las variables, una conclusión bastante interesante del capítulo anterior fue que las variables de participación afectan más al abandono y las de rendimiento a la nota. Para profundizar en este aspecto, en este escenario particular, se ha realizado un análisis de la importancia de las variables, utilizando **RF** y la disminución media de Gini [280]. En este caso, como el capítulo trata sobre el análisis temporal, se ha analizado cuál es la evolución de la importancia de las variables a lo largo del tiempo. La Fig. 7.8 y Fig. 7.9 muestran la evolución temporal de la importancia de las variables para la predicción del abandono y de la nota, respectivamente. La importancia se expresa de 0 a 1, de modo que la variable con más importancia tiene el valor 1 y la importancia del resto es relativa a la variable con mayor importancia.

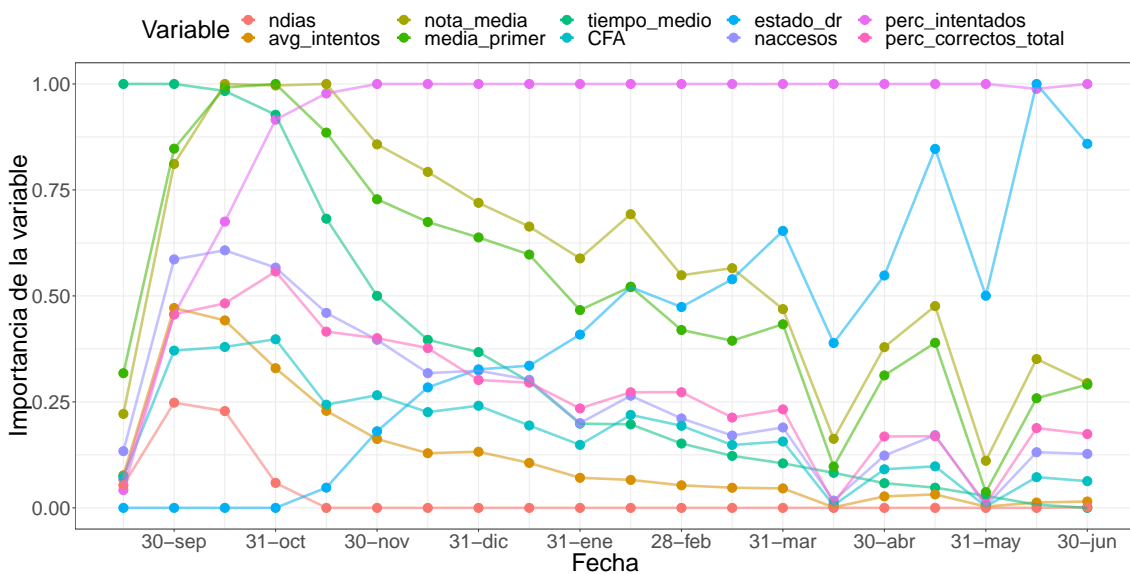


Fig. 7.8. Evolución temporal del poder predictivo de las variables al predecir el abandono en el escenario 6

En cuanto a la predicción del abandono, los resultados muestran que la variable con más importancia es el porcentaje de ejercicios intentados (**perc.intentados**). Esto es razonable, ya que un alumno se considera que abandona si al final de curso no realiza un porcentaje mínimo de ejercicios. Sin embargo, los resultados muestran que a partir de final de noviembre (tres meses del inicio del curso), el porcentaje de ejercicios intentados es muy relevante y el alumno que no lleve un porcentaje alto respecto al que debería llevar en ese momento puede tener alta probabilidad de abandonar. En cambio, el poder predictivo de **perc.intentados** al inicio no es muy alto. Esto implica que si un alumno no hace las actividades al inicio, no tiene por qué abandonar. Del mismo modo, si un alumno tiene un 1.00 (valor máximo de la variable) al inicio, significa que ha hecho todo lo que se le ha mandado hasta

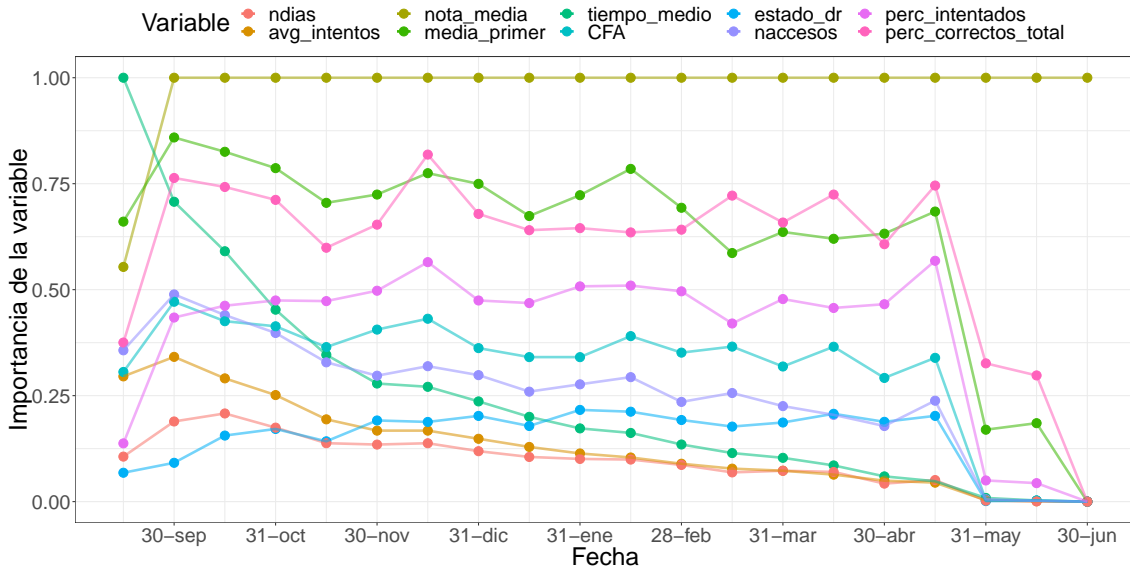


Fig. 7.9. Evolución temporal del poder predictivo de las variables al predecir la nota en el escenario 6

el momento, pero no implica que abandonará después. Al inicio, el tiempo medio por ejercicio es la variable con más relevancia, seguido de la nota media, las cuales pueden dar una indicación de si el alumno hace bien las actividades y si emplea mucho o poco tiempo en cada una. Sin embargo, la nota media pierde importancia según va pasando el curso. Del mismo modo, el estado de abandono (*estado_dr*) va ganando cada vez más importancia. Esto tiene sentido, al igual que el porcentaje de intentados, por la relación con la variable dependiente. Respecto a otras variables de rendimiento, se aprecia que tienen menor relevancia. Lo que llama la atención es que el número de días activos (*ndias*) no tiene impacto en el modelo. Esto puede ser porque en cada curso empleen la plataforma de una manera diferente, por lo que sea más relevante los porcentajes de actividad relativos a lo que se ha visto en la clase, que los valores absolutos sobre los días activos.

Respecto a la predicción de la nota, se observa que la nota media (*nota_media*) en un instante previo es la variable con mayor importancia. Esto es razonable, ya que la variable a predecir es la nota media con todos los intentos, por lo que esta variable indica el estado de la nota en un estado previo. De este modo, la nota al principio tiene bastante incertidumbre porque se han hecho pocos ejercicios, pero a medida que pasa el curso va quedando más definida y es por ello que al final el *RMSE* es 0. A pesar de que esta variable tiene relación con la variable dependiente al final, no habría problema, ya que en cada momento se utiliza la información conocida, por lo que toda la información sobre la nota hasta una fecha dada se utiliza para conocer una fecha futura, lo que hace que esta variable sea implementable. Tras la nota media, las variables con mayor importancia son la media del primer intento (*media_primer*) y el porcentaje de ejercicios correctos (*perc_correctos_total*), que son todas variables relacionadas con el rendimiento. A continuación la variable más importante es el porcentaje de intentados, por lo que la participación también tiene cierta importancia (de ahí que los modelos también ofrezcan un alto poder predictivo, aunque peor). Esto es razonable, ya que si un alumno no intenta los ejercicios, difícilmente puede tener una nota elevada. Finalmente, las variables con menor

relevancia son el número medio de intentos (`avg_intentos` y `ndias`). Esto implica que en este escenario, el intentar muchas o pocas veces los ejercicios no influye en la nota, al igual que los días activos, como se comentó anteriormente. Sobre este escenario, la conclusión fundamental en cuanto a las variables es que para ver el abandono hay que centrarse más en lo que el alumno hace y no en cómo lo hace, al contrario que para la nota, que se centra más en el rendimiento.

Para finalizar esta sección, en relación con los seis escenarios, se ha visto que es posible obtener un elevado poder predictivo con suficiente antelación, lo que puede servir para poder llevar a cabo intervenciones. En otros trabajos similares de la literatura (p. ej., [153], [267]) también se han desarrollado modelos y se ha visto que es posible lograr un grado de anticipación, lo que corrobora la utilidad de los modelos predictivos desde el punto temporal. En este trabajo, además, se han analizado varios contextos. En la sección 7.3 se analizan de forma global todos ellos, junto con los modelos para cursos asíncronos (que se abordan en la siguiente sección), para extraer conclusiones globales.

7.2. Evolución temporal del poder predictivo en cursos asíncronos

En la sección previa, el análisis predictivo se ha realizado añadiendo datos de cada semana del curso y analizando cómo evoluciona el poder predictivo cuando los indicadores se calculan con interacciones de más semanas. Esto tiene sentido cuando se espera que los alumnos trabajen a la vez en el curso de manera síncrona. Sin embargo, en el escenario 3, este procedimiento no puede ser aplicado ya que en un momento determinado del tiempo (fecha X), cada alumno puede avanzar en una parte distinta del curso al ser un MOOC asíncrono. Por ello, esta sección tiene como objetivo analizar el caso concreto del análisis temporal para cursos asíncronos.

Para este caso, el eje temporal se transforma y en lugar de tener un eje con unas fechas fijas de calendario, se adapta para cada alumno según su matriculación (según lo comentado en la sección 5.5). De este modo, la semana 1 del curso es la primera semana desde que cada uno comienza el curso, en lugar de una semana de calendario. La segunda semana es la segunda semana de cada uno en el curso y así sucesivamente.

Teniendo esto en cuenta, se han desarrollado modelos para predecir el abandono en el escenario 3 (tratado en [91]). En particular, se ha considerado que cada módulo del curso equivale a una semana y se ha desarrollado un modelo por cada semana del curso (al haber cuatro módulos, son cuatro semanas), más una semana adicional para contemplar posibles retrasos, y un modelo extra considerando todos los datos del estudiante. Estos modelos se han diseñado utilizando los cuatro algoritmos utilizados a largo de la tesis. Como variables, se han utilizado las variables sobre ejercicios, ya que eran las que obtenían un mejor poder predictivo, y los patrones de SRL, que también eran muy buenos predictores (como se analizó en el capítulo 6. En primer lugar, se ha tomado como referencia el curso “Electrones en Acción” y luego se ha extendido el análisis para los otros dos MOOCs. Los resultados de la evolución temporal del poder predictivo para el curso mencionado se muestran en la Fig. 7.10.

Los resultados muestran que es posible predecir con un AUC bueno (entre 0.8 y 0.9) desde la primera semana y con un AUC excelente (superior a 0.9) a partir de la segunda semana. Esto implica que es posible predecir el abandono en este

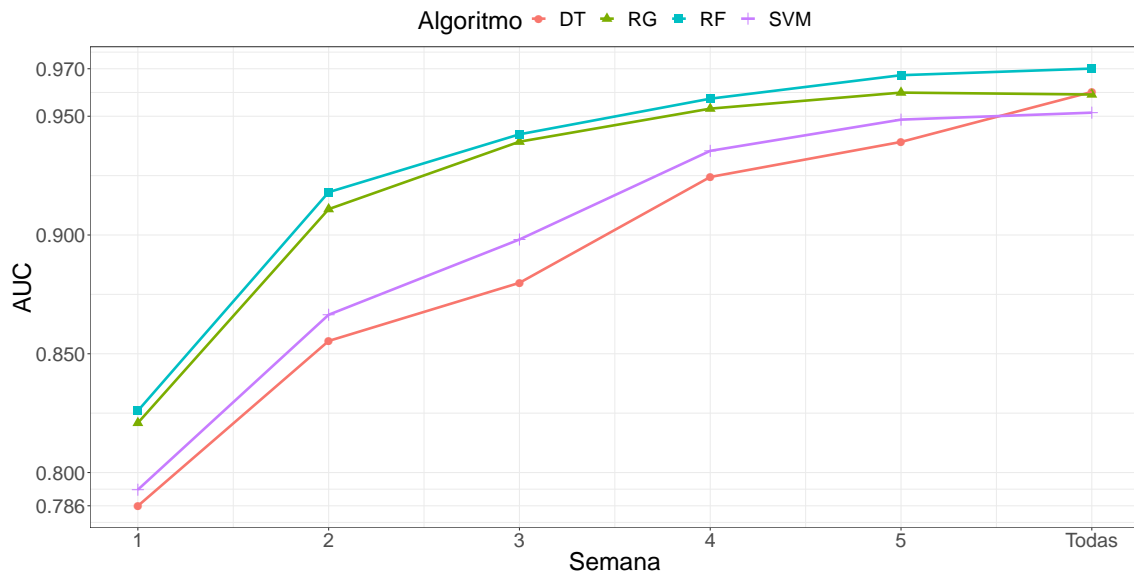


Fig. 7.10. Evolución temporal de la predicción del abandono en en el MOOC “Electrones en Acción”

curso con muy pocos días desde la primera interacción. Además, los primeros días parecen muy significativos, ya que pueden alcanzar un alto poder predictivo. Este resultado coincide con el resultado de Jiang et al. [284], que consiguieron predecir quién iba a obtener el certificado del MOOC con datos de únicamente la primera semana. En esa contribución, el MOOC era síncrono, pero los resultados obtenidos aquí muestran que también es posible obtener predicciones tempranas en cursos asíncronos. Además, este resultado es coherente con el análisis temporal de Vitiello et al. [268] en MOOCs asíncronos, en el que concluyeron que las interacciones de la primera semana eran significativas. Respecto a los algoritmos, la figura muestra que RF es algoritmo que mejor funciona entre los cuatro utilizados en el análisis. Sin embargo, la diferencia con RG es pequeña, particularmente en las primeras semanas. Sin embargo, SVM y DT obtienen un poder predictivo, si bien los resultados obtenidos con estos algoritmos también son bastante buenos.

Para validar los resultados, se ha repetido el análisis con los otros dos MOOCs de Coursera. Este análisis es bastante relevante porque además, la duración de cada MOOC es diferente (el número de módulos es distinto), teniendo nueve módulos el curso “Aula Constructivista” y siete el de “Gestión de Organizaciones Efectivas”. Dado que el MOOC más largo tiene nueve módulos, se han desarrollado módulos de las diez primeras semanas y un modelo con todas las interacciones. Fig. 7.11 muestra los resultados de la evolución temporal del poder predictivo en los tres MOOCs.

Una observación inicial es que el poder predictivo se estabiliza en “Electrones en Acción” tras la semana 6. El motivo es que el 99% de las interacciones ocurren dentro de dicho periodo, y por tanto no hay información nueva tras la semana 6. Esto también implica que los alumnos que no abandonan procuran seguir el calendario sugerido de un módulo por semana (aunque en algunos casos pueden emplear un par de semanas más de lo esperado). Un efecto similar con la cantidad de interacciones ocurre alrededor de la semana 10 del curso “Gestión de Organizaciones Efectivas”, lo cual es tres semanas más de la duración teórica. Para la semana 10, ocurren el 99.6% de las interacciones. Sin embargo, para el curso de “Aula Constructivista”,

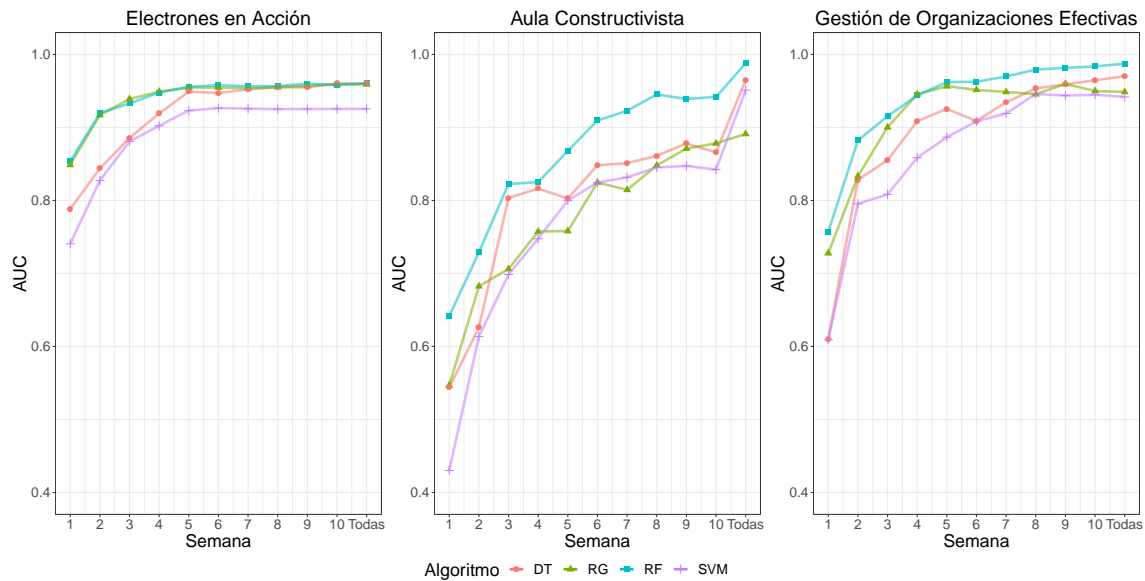


Fig. 7.11. Evolución temporal de la predicción del abandono en los tres MOOCs del escenario 3

que es el más largo, solo el 89.7% de las interacciones han ocurrido a las 10 semanas, probablemente por la mayor duración.

En cuanto al poder predictivo, es posible alcanzar un **AUC** de al menos 0.8 a partir de las semanas 3 y 2 para “Aula Constructivista” y “Gestión de Organizaciones Efectivas”, respectivamente. Por otro lado, el umbral para conseguir un **AUC** excelente (al menos 0.9) se alcanza en las semanas 6 y 3, respectivamente. Teniendo en cuenta la duración de los tres cursos, estos resultados muestran que se necesita entre un 25 y un 33% de la duración teórica del **MOOC** para alcanzar un **AUC** de 0.8 y entre un 43 y 67% de la duración teórica del **MOOC** para alcanzar un **AUC** de 0.9. Esto implica que es posible obtener modelos con una alta precisión y con suficiente antelación en los **MOOC** asíncronos. Por ello, la recomendación para los cursos asíncronos sería ir descargando los datos de los estudiantes, calcular el periodo desde que se registraron y empezar a proporcionar predicciones del abandono a los alumnos que lleven al menos un número de semanas equivalente al 25-33% de la duración teórica del curso. De esta manera, se podrían realizar intervenciones con estos estudiantes a partir de las predicciones para intentar mejorar el abandono, como propusieron Xing et al. [24].

7.3. Lecciones aprendidas

A lo largo de este capítulo, se han presentado diferentes análisis temporales en varios cursos síncronos y uno asíncrono. Para concluir este capítulo, se presentan las conclusiones principales.

Respecto a la predicción de la nota final de curso, se ha visto que en algunos casos (escenarios 1 y 2) puede requerirse más tiempo para obtener predicciones precisas que para el éxito, por lo que la precisión de la nota debería ser acompañada de la del éxito para mayor anticipación. Incluso en algunos casos, el nivel de precisión podría no ser muy alto, como en algunos casos del escenario 2. Esto implica que puede tardarse más en conseguir una alta predicción de la nota, al ser una variable

más difícil de predecir (es más difícil predecir un valor de nota concreto, que un resultado binario). No obstante, en el escenario 6, se podía lograr un nivel similar de anticipación entre el abandono y la nota, por lo que este hecho podría variar dependiendo del contexto. En la literatura, la mayoría de trabajos abordan solo una de las dos variables dependientes (p. ej., [23], [271]). Por ello, se deberían realizar más trabajos con varias variables para analizar este caso.

En relación con las notas intermedias, el análisis temporal también confirma que el modo no acumulativo es consistentemente mejor en el escenario 1, pero no en el escenario 2, en el que varía dependiendo de la prueba. Un posible motivo es que el escenario 2 contiene muy pocas pruebas por semana (solo hay 3-4 ejercicios por semana), lo que puede hacer que las variables sean menos significativas en el modo no acumulativo y por ello tener un peor poder predictivo. Por ello, a la hora de seleccionar un modo u otro debe tenerse en cuenta el diseño del curso y la metodología, ya que también una independiencia entre los temas puede favorecer al modo no acumulativo.

En cuanto a las predicciones del éxito y el abandono, sí que se ha visto que es posible obtener una alta precisión temprana (lo que es coherente con otros trabajos, como [23], [24]), de modo que el modelo puede ser utilizado en las primeras etapas para detectar los alumnos en riesgo y poder intervenir con ellos. Aunque el momento concreto a partir del cual se obtenga una alta precisión puede variar dependiendo del contexto, se pueden extraer algunas conclusiones en función de todos los contextos analizados. Por ello, la Tabla 7.2 muestra un resumen de en qué momento se superan ciertos umbrales del *AUC* en los modelos para la predicción del éxito y el abandono en los diferentes escenarios analizados.

TABLA 7.2
RESULTADOS AGREGADOS DEL ANÁLISIS TEMPORAL

Variable a predecir	Escenario (número)	<i>AUC</i> \geq 0.7		<i>AUC</i> \geq 0.8		<i>AUC</i> \geq 0.9	
		Periodo	%	Periodo	%	Periodo	%
Éxito	MOOC (1)	Semana 1/5	20%	Semana 1/5	20%	Semana 2/5	40%
Éxito	MOOC (2)	Semana 2/10	20%	Semana 4/10	40%	Semana 7/10	70%
Abandono	MOOC 1 (3)	Semana 1/4	25%	Semana 1/4	25%	Semana 2/4	50%
Abandono	MOOC 2 (3)	Semana 2/9	22%	Semana 3/9	33%	Semana 6/9	67%
Abandono	MOOC 3 (3)	Semana 1/7	14%	Semana 2/7	29%	Semana 3/7	43%
Éxito	SPOC (4)	Mes 7/10	58%	Mes 9/10	90%	-	-
Abandono 1	SPOCs (5)	Semana 4/16	25%	Semana 14/16	88%	-	-
Abandono 2	SPOCs (5)	Semana 2/16	13%	Semana 3/16	19%	Semana 6/16	38%
Éxito	SPOCs (5)	Semana 1/16	6%	Semana 2/16	13%	Semana 6/16	38%
Abandono	K-12 (6)	Mes 1/10	10%	Mes 3.5/10	35%	Mes 5.5/10	55%

Abandono 1 se refiere a la definición relacionada con actividad y abandono 2 a la de completitud

En el escenario 3, los MOOCs 1, 2 y 3 son “Electrones en Acción”,

“Aula Constructivista” y “Gestión de Organizaciones Efectivas”, respectivamente

Los resultados muestran que para el caso de los MOOCs (escenarios 1-3), se alcanza un *AUC* superior a 0.7 (moderado) entre el 14-25% del curso, un *AUC* superior a 0.8 (bueno) entre el 20-40% del curso y un *AUC* excelente entre el 40-70% del curso. Estos umbrales incluyen tanto cursos síncronos como asíncronos en edX y Coursera, y las temáticas de los mismos son variadas. En cuanto a los umbrales, no se observa ningún patrón dependiendo de si se predice abandono o éxito, o se cambian las características del curso. Teniendo en cuenta estos umbrales, aunque

no se pueda asegurar que vayan a generalizar para todos los contextos, se podría tomar como referencia el periodo del 20-40 % del curso como periodo para empezar a realizar las predicciones. Si se realizan antes, el poder predictivo podría no ser suficiente y se podrían obtener resultados poco precisos y si se espera más, se podrá obtener mejores resultados (incluso AUCs excelentes), pero habrá menos tiempo de anticipación. No obstante, el mejor momento exacto para empezar a predecir en cada curso puede variar, ya que puede depender de lo que interaccionan los alumnos en el curso y de la metodología del curso.

En cuanto a los SPOCs y la plataforma para alumnos de primaria y secundaria (alumnos K-12), se observa que en todos los casos excepto el escenario 4 y el abandono según la actividad se pueden obtener unas predicciones tempranas y se podría obtener un AUC superior a 0.8 entre el 13-35 % del curso, lo cual es un poco antes que en los MOOCs, pero no muy diferente. En estos casos, hay que tener en cuenta que la variable a predecir en un instante previo es conocida, por lo que es normal que cuando se haya cubierto una parte significativa del curso la predicción sea alta. Por ello, aquí se pueden obtener unas predicciones bastante tempranas. El caso del abandono del escenario 5 puede ser algo diferente porque la no actividad puede depender del contexto, lo que puede hacer que en algunos casos sea menos representativo.

Un caso bastante interesante es el escenario 4. En este caso, no se obtiene una precisión aceptable hasta final de curso. El motivo principal está en el contexto y la variable que se predice. Mientras que en los otros contextos de SPOCs la variable a predecir se obtiene dentro del SPOC y la información parcial te va idea sobre la variable a predecir, en este caso es totalmente independiente. Aquí, lo que se predice es la nota de una prueba oficial externa, que es independiente al SPOC. Además, el uso del SPOC no es obligatorio, sino que es de apoyo a lo que se ve en las clases del último curso de secundaria. Por ello, puede haber menor relación entre el uso del SPOC y el resultado del examen, lo que hace que se obtenga un peor predictivo. También, el hecho de que no hay mucha interacción al inicio puede contribuir a una menor anticipación. Por todos estos factores, es muy importante tener en cuenta el contexto del curso y sus posibles variaciones a lo largo del tiempo dependiendo del curso.

No obstante, los resultados de esta sección indican que para contextos de MOOCs o SPOCs similares a los analizados en esta tesis, podría ser sensato aplicar los modelos predictivos a partir del 13-40 % del curso. Esto implicaría que habría que dejar un periodo al inicio para que se generaran suficientes datos para poder conseguir unas predicciones con una precisión aceptable. También otra implicación sería que se podrían empezar a proporcionar las predicciones antes de la mitad del curso, lo cual daría suficiente margen para poder llevar a cabo intervenciones. Sin embargo, hay que tener presente que el tipo de curso y la metodología pueden tener un papel muy relevante, que puede hacer que estos valores fluctúen.

Por otro lado, respecto al poder predictivo de las variables, en los análisis se ha visto que los mejores predictores se pueden mantener a lo largo del tiempo, aunque puede haber variables que al inicio sean más representativas y luego pierdan poder predictivo, y viceversa. Por ejemplo, en el escenario 6, el tiempo medio por ejercicio es la variable más relevante para el abandono y el éxito al inicio y luego pierde poder predictivo. En dicho escenario, el porcentaje de ejercicios intentados, más relacionado con la participación, es la variable más relevante para el abandono a

partir del final del mes 3 (de 10 meses) hasta el final. Lo mismo ocurre con la nota media (relacionada con el rendimiento) para el éxito desde el final del mes 1. Por lo que al inicio puede haber mayores variaciones con las variables que deben tenerse en cuenta. Esto se corrobora con un trabajo previo [23], en el que se concluye que el progreso en los problemas de un MOOC es la variable más relevante para predecir el certificado del curso, aunque esta variable no es tan relevante al inicio, donde hay mayores variaciones con las variables, y en este caso, el número de sesiones era la variable más relevante al inicio.

Para concluir, en este capítulo se ha visto cómo el tiempo tiene un efecto muy importante en la predicción (como se ha visto en otros trabajos, como [23], [24]), y se ha podido comprobar cómo a medida que pasa el curso puede mejorar el poder predictivo (e incluso puede cambiar la importancia de alguna variable). Sin embargo, a medida que pasa el tiempo, se disminuye el poder de reacción. Por ello, es importante encontrar el punto óptimo entre el poder predictivo y la anticipación (tiempo a partir del que se muestran las predicciones). En este caso, se ha visto que hay que dejar un tiempo al inicio para que haya datos suficientes y los modelos sean precisos. Sin embargo, una vez que el poder predictivo sea aceptable (lo cual puede ser en torno al 13-40 % del curso, aunque puede depender del contexto) se deberían mostrar las predicciones para poder tener el mayor impacto posible con las mismas.

8

Generalización de los modelos predictivos

Contenido

8.1. Transferencia de modelos	130
8.2. Agregado de cursos	133
8.3. Reutilización y adaptación de los modelos	140
8.4. Lecciones aprendidas	144

A lo largo de la tesis, se ha analizado el efecto de las diferentes variables para realizar los modelos predictivos en diferentes escenarios y se ha realizado un análisis temporal para ver si es posible conseguir una anticipación en los modelos y cuál es el mejor momento para empezar a predecir en diferentes escenarios. Pero además, los modelos utilizados se han diseñado a posteriori, una vez que el curso ha terminado. Entonces, las conclusiones obtenidas son válidas para el curso finalizado, pero aunque se indique que se puede obtener una predicción temprana, ya no se podría realizar ninguna intervención. Una utilidad fundamental es que estos modelos puedan ser utilizados en tiempo real y puedan suponer una mejora en el aprendizaje, para lo que es necesario que los resultados sean generalizables a otros cursos, de modo que si se utiliza un modelo diseñado, se obtenga el mismo o similar poder predictivo para otro curso a partir del mismo modelo, y se pueda obtener una predicción temprana, como se indicó en el análisis temporal. Aunque este caso sería el deseable, en la realidad no es posible obtener un modelo que generalice para todos los cursos [285]. Por ello, es relevante analizar bajo qué condiciones es posible que un modelo pueda ser utilizado en otro curso y qué poder predictivo se puede obtener. Este capítulo aporta en relación con este reto y presenta los resultados en escenarios específicos de dos tipos de aproximaciones para la generalización: la transferencia de modelos y el agregado de cursos. Estas aproximaciones se tratan en las dos primeras secciones. Después, se discute sobre otra aproximación para conseguir el funcionamiento de los modelos predictivos, que es la reutilización y adaptación de los modelos. Por último, se presentan las lecciones aprendidas en base a las tres aproximaciones señaladas.

8.1. Transferencia de modelos

Para el análisis de la generalización, la primera aproximación que se considera es la de transferencia de los modelos. Esto es, entrenar el modelo predictivo con un curso, para usarlo con un curso diferente, ya sea el mismo curso de la materia pero en ediciones diferentes o bien en cursos de materias diferentes. En esta sección, se va a analizar qué poder predictivo puede obtenerse cuando se transfiere un modelo. Para ello, se toman el escenario 3 y el escenario 4.

En el escenario 3, tratado en [91], se realizó un experimento inicial para analizar la transferencia. Este experimento consistió en entrenar un modelo con todos los datos de un MOOC y después utilizarlo con los alumnos de los otros dos MOOCs considerados (usando todas las combinaciones de transferencia posibles). La Tabla 8.1 muestra los resultados de la transferencia de los modelos.

TABLA 8.1
TRANSFERENCIA DE MODELOS EN CURSOS DEL ESCENARIO 3

Curso para entrenar	Electrones en Acción (EEA)		Aula Constructivista (AC)		Gestión de Organizaciones Efectivas (GOE)	
	AC	GOE	EEA	GOE	EEA	AC
Curso para predecir						
DT	0.50	0.50	0.94	0.50	0.89	0.96
RG	0.95	0.54	0.28	0.75	0.96	0.96
RF	0.99	0.97	0.97	0.96	0.97	0.98
SVM	0.95	0.96	0.91	0.94	0.89	0.92

Nota: El mejor poder predictivo se mide en AUC

Los resultados muestran que es posible obtener un AUC muy alto (superior a 0.96) al utilizar datos de un MOOC diferente al utilizado para entrenar el modelo. Esto significa que puede ser posible reutilizar el modelo a pesar de la duración y del área temática del curso. Sin embargo, los resultados muestran que DT y RG no son siempre consistentes, y proporcionan modelos que no son transferibles cuando se entrena con “Electrones en Acción” y “Aula Constructivista”. En este caso, RF es el mejor algoritmo, ya que es el más consistente al transferir a otros MOOCs y también obtiene el mejor predictivo cuando se prueba con el mismo curso con el que se entrenó el modelo (como se comprobó en capítulos anteriores).

A pesar de que los modelos predictivos muestran que en este escenario es posible transferir los modelos, este hecho puede cambiar dependiendo del contexto. En este caso, los tres cursos eran MOOCs asíncronos impartidos en español y con un sistema de evaluación muy similar, aunque la duración y el área temática era diferente. Si el tipo de curso o metodología fuera más diferente, podría haber más diferencias en cuanto a la generalización. Por ello, es importante analizar cada caso individual y el contexto para ver si pudiera ser posible la transferibilidad a otro curso. No obstante, este primer resultado es positivo, ya que muestra que se podría transferir un modelo predictivo en un contexto de MOOC, en el que las tasas de abandono pueden ser muy elevadas.

Para profundizar más en la generalización, se ha considerado el escenario 4, tratado en [29]. En este escenario se dispone de cuatro SPOC: dos de Física y dos de Química, de los cuales uno de cada materia son del curso 2016/2017 y los otros del 2017/2018. En este escenario se analiza la generalización cuando se entrena con un curso y se predice en otro de un cohorte (curso académico) diferente y el mismo o

diferente curso. Un ejemplo de esto es entrenar con los datos del SPOC de Química del curso 2016/2017 y predecir con los datos del SPOCs de Física o de Química del curso 2017/2018. Además, se puede analizar la generalización usando cursos del mismo cohorte (Física y Química del mismo año) o combinando cursos del mismo cohorte. Es decir, se puede crear un modelo combinando los dos cursos del mismo cohorte y transfiriendo el modelo al cohorte siguiente. Teniendo en cuenta los posibles experimentos para la generalización, se propone la siguiente lista de experimentos para el análisis en este escenario (usando RF y todos los datos disponibles).

- Transferencia dentro del mismo cohorte: consiste en entrenar un modelo utilizando datos de un SPOC y predecir utilizando datos de otro SPOC perteneciente al mismo cohorte.
- Transferencia dentro del mismo curso, pero en diferente cohorte: consiste en entrenar un modelo utilizando datos de un SPOC y predecir utilizando datos del mismo SPOC pero en un cohorte (en una edición) diferente.
- Transferencia en un diferente curso y cohorte: consiste en entrenar un modelo utilizando datos de un SPOC y predecir utilizando dato de otro curso diferente en un cohorte diferente.
- Transferencia usando la combinación de SPOCs en diferentes cohortes: consiste en entrenar un modelo a partir de las interacciones de los dos SPOCs (Física y Química) del mismo cohorte, y predecir utilizando las interacciones de los dos SPOCs en un cohorte diferente.

El primer experimento consiste en predecir utilizando datos de estudiantes del mismo cohorte (mismo curso académico). Dado que para curso académico, existen datos de los cursos de Física y Química y los alumnos son los mismos en los dos cursos, se entrena utilizando datos de un curso y se predice utilizando las interacciones del otro. Los resultados de este experimento se encuentran en fila (a) de la Tabla 8.2 y reflejan que hay una buena transferibilidad dado que el poder predictivo es similar e incluso algo mejor que el poder predictivo obtenido al entrenar y predecir con los datos del mismo curso. Esto significa que el modelo puede adaptarse a los datos nuevos y puede transferirse a otro SPOC (de una temática diferente) durante el mismo curso académico (y cohorte). Una posible explicación es que, como los estudiantes son los mismos, su comportamiento podría ser similar en los diferentes cursos. Otro factor a favor puede ser que lo que se predice no es el resultado del examen de Física o Química por separado, sino el resultado (aprobado/suspense) de la parte de ciencias de la prueba de acceso, que contiene los resultados relacionados con ambos cursos.

El segundo experimento es más común en la literatura y consiste en entrenar el modelo con datos de una edición del curso y predecir con datos de una edición diferente (p. ej., una edición futura del curso). En este caso, se utilizan datos de cursos del 2016/2017 para predecir en cursos del 2017/2018. Esta aproximación ha sido también realizada en la literatura (p. ej., [30]) dado que es la aplicación común de la predicción a posteriori, es decir, un uso común es entrenar con un curso pasado y usar ese modelo durante el desarrollo (p. ej., cada semana o cada mes) de una edición futura del curso. Los resultados de este análisis (fila (b) de la Tabla 8.2) muestran que el poder predictivo, tanto al transferir desde el SPOC de

TABLA 8.2
 RESULTADOS DE LA TRANSFERIBILIDAD DE LOS MODELOS EN EL ESCENARIO 4

Experimento	Curso utilizado para entrenar	Curso utilizado para predecir	AUC
(a) Dentro del mismo cohorte	Química 16/17	Física 16/17	0.88
	Física 16/17	Química 16/17	0.89
	Química 17/18	Física 17/18	0.87
	Física 17/18	Química 17/18	0.83
(b) Dentro del mismo curso, pero en diferente cohorte	Química 16/17	Química 17/18	0.79
	Física 16/17	Física 17/18	0.79
(c) Utilizando un curso y cohorte diferente	Química 16/17	Física 17/18	0.75
	Física 16/17	Química 17/18	0.63
(d) Utilizando la combinación de SPOCs en diferentes cohortes	Física y Química 16/17	Física y Química 17/18	0.74

Física como el de Química del curso 2016/2017 al curso 2017/2018, es de 0.79. Este valor está justo en el límite entre un AUC moderado y bueno. Esto implica que es posible usar el modelo predictivo en una edición futura del curso, aunque se pierde poder predictivo. Otra observación es que cambiar el cohorte tiene un mayor impacto negativo que el hecho de cambiar de curso (dentro del mismo curso académico), ya que no se pierde poder predictivo al cambiar de curso dentro del mismo académico, pero sí cuando se cambia de alumnos (aunque sea el mismo curso). Aunque hay muchos factores que puedan cambiar de un curso a otro, este resultado sugiere que cambiar de cohorte tiene un efecto significativo en la generalización, aunque dado el poder predictivo obtenido, es posible usar el modelo entrenado para predecir en una edición futura del SPOC.

El tercer experimento consiste en predecir utilizando datos de un curso y cohorte diferente a los datos utilizados para el entrenamiento, de manera que tanto los alumnos como el curso sean diferentes. A pesar de que en la mayoría de los casos, es posible que se quiera entrenar y predecir con el mismo curso, este experimento es bastante relevante para entender la generalización. También puede ser práctico porque podría darse el caso de que se cambien los cursos en los años siguientes o que solo se tenga un modelo entrenado de un curso específico y quiera utilizarse para otros cursos y estudiantes. Los resultados obtenidos en este caso (fila (c) de la Tabla 8.2) muestran que el poder predictivo es peor que en los experimentos previos. Al transferir del SPOC de Química 16/17 a Física 17/18, el poder predictivo es moderado (0.75), pero cuando se transfiere de Física 16/17 a Química 17/18, el poder predictivo es pobre (0.63). Esto significa que transferir los modelos a diferentes cursos y estudiantes puede ser más difícil ya que cambian más factores, lo que limita la generalización de los modelos.

El cuarto y último experimento consiste en combinar las variables de los dos SPOCs (Física y Química) del mismo cohorte. En la literatura, los modelos se desarrollan utilizando normalmente datos de un curso (p. ej., [153], [154]) (o si se utilizan varios cursos, se usan varios cursos con diferentes estudiantes, como en [126], [141]). Sin embargo, dado que los SPOCs tienen los mismos estudiantes, es posible desarrollar modelos con la combinación de los indicadores de ambos SPOCs. A pesar de que el poder predictivo de la combinación no difiere mucho del poder predictivo

del mejor de los dos SPOCs (como se vio en el capítulo 6) al analizar y predecir en el mismo SPOC o combinación, es interesante analizar el poder predictivo al transferir la combinación de los dos cursos al siguiente cohorte. Los resultados (fila (d) de la Tabla 8.2 muestran que el AUC es 0.74 cuando se transfiere la combinación del curso académico 2016/2017 al 2017/2018. Esto es similar al mejor AUC obtenido en el tercer experimento pero peor que el AUC obtenido cuando se utilizan datos de un solo curso para transferir a la edición siguiente (segundo experimento). Esto implica que la combinación de SPOCs no es mejor y que es preferible utilizar modelos entrenados con un solo curso para la generalización. No obstante, esta aproximación podría ser útil en algunos casos (su poder predictivo es moderado) en los que haya pocos datos de cada curso individual, de modo que la combinación tenga una muestra suficiente para generar los modelos que no la tengan los cursos individuales.

Como conclusión de este escenario, los resultados muestran que es posible transferir los modelos predictivos a otros cursos (como también se vio en el escenario 3), pero el poder predictivo depende de cómo cambia el contexto. Los mejores resultados se obtuvieron cuando se utilizaba el mismo cohorte en otros cursos, pero los resultados también eran aceptables cuando se utilizaba el mismo curso pero en una edición posterior (con diferente cohorte). Este último resultado es coherente con otros trabajos como [211], [212], aunque también debe tenerse en cuenta, como reportaba [30], que al transferir a otra edición de un curso va a existir una pérdida de poder predictivo.

8.2. Agregado de cursos

Una segunda aproximación para intentar conseguir la generalización consiste en el agregado de cursos. Esta aproximación genera un modelo global que incluye datos de muchos cursos, de manera que la posible variabilidad por el contexto del curso quede recogida en el modelo por la presencia de datos variados. Esta aproximación fue utilizada anteriormente por Kicilcec y Halawa [216]. En dicho trabajo, se entrenó un modelo global con los datos de 20 MOOCs para predecir el abandono, y se obtuvo un AUC de 0.92 al predecir con datos de otros 20 MOOCs, lo cual es un AUC excelente. Aunque en muchos casos puede interesar crear modelos con un curso y transferir (como se ha realizado en la sección anterior), en otros puede haber limitaciones que hagan que esta aproximación sea más interesante. Por ejemplo, cuando hay muchos cursos, éstos se podrían caracterizar (para diferenciarlos) con variables relevantes para la predicción, y puede tener sentido un modelo global. Esto puede tener también relevancia en el caso de que dentro de esos cursos haya algunos que son nuevos, por lo que no haya ningún curso como referencia para transferir de ediciones anteriores. También podría ser interesante para cursos donde el número de casos sea bajo, por lo que los modelos predictivos no sean muy significativos. Es el caso de la enseñanza K-12 (primaria y secundaria), donde los grupos suelen tener 20-30 alumnos, lo que haría que el número de cursos para entrenar fuera muy elevado y con muy pocos casos.

Entre los escenarios tratados en la tesis, hay dos en los que ajusta mejor sentido esta aproximación porque se dispone de un número elevado de datos de diferentes cursos: los escenarios 5 y 6. En el escenario 5, se consideran los SPOCs de la UC3M. En este escenario, hay un número considerable de cursos (que va en aumento) para poder elaborar un modelo global. Además, puede ocurrir que el número de alumnos

en algún curso puede ser bajo. Por ello, es de interés poder generar un modelo con los datos de todos los cursos y poder usarlo para ofrecer información a los profesores de todos los SPOCs acerca de los estudiantes en riesgo de sus cursos. Del mismo modo el escenario 6 es un escenario de K-12, en el cual actualmente se trabaja con 480 cursos. Estos cursos pueden ser muy variados y el número de alumnos en cada uno puede no ser muy alto como para generar un modelo propio. Por ello, en el escenario 6, puede ser muy interesante ver si se puede generar un modelo global para utilizarlo con un grado de precisión aceptable en todos los cursos.

En los dos capítulos anteriores, se analizó el poder predictivo en los escenarios 5 y 6, y como conclusión, se vio que era posible obtener un alto poder predictivo (si bien el de abandono relacionado con actividad del escenario 5 era peor). Este poder predictivo es el del modelo global, por lo que se corroboraría lo obtenido por Kizilcec y Halawa [216] en los dos escenarios. Sin embargo, un posible problema de esta aproximación es que aunque el poder predictivo sea elevado de forma global, puede haber cursos en los que el modelo no funcione. Esto es posible ya que el modelo podría no capturar bien todo tipo de contexto. Para analizar este hecho, en primer lugar, en esta sección, se analiza el poder predictivo de cursos concretos.

En primer lugar, se considera el escenario 5. En este escenario, hay 16 cursos, por lo que se propone analizar el poder predictivo en cada curso específico. Para ello, se ha tomado la semana 8 y el algoritmo RF, ya que el poder predictivo es suficientemente bueno en este punto del curso y el algoritmo era el más consistente en este escenario. Con este algoritmo y punto temporal, se han elaborado 16 modelos globales. En cada modelo global, se han utilizado los datos de 15 cursos para entrenar y los datos del curso restante para evaluar el poder predictivo. El curso se ha ido rotando para que se pueda evaluar el poder predictivo de cada curso por separado con modelos globales. La Tabla 8.3 muestra el poder predictivo del modelo entrenado con los 15 cursos y el obtenido al predecir con el curso restante para las tres variables dependientes. Al realizar el análisis, se ha observado que hay algunos cursos en los que la tasa de abandono o de fracaso es del 100%, probablemente porque esos SPOCs estén diseñados para apoyo del curso más que como material obligatorio. En concreto hay un curso según la definición de abandono con actividad, cinco según la de completitud y cuatro según el éxito. Los cuatro cursos donde ningún alumno consigue una media superior a 5 sobre 10 coinciden con los del abandono según la completitud, pero el obtenido según la definición de abandono según la actividad es diferente. Dado que en estos cursos no es posible calcular el AUC al haber solo una clase posible en los datos, las filas correspondientes aparecen con un guión.

Los resultados de la tabla muestran que a pesar de que se obtiene un AUC global que puede calificarse de bueno en general, el AUC de cada curso puede variar. En el conjunto de entrenamiento, como se varía solo un curso, los resultados son muy similares en todos los casos, pero a la hora de probar con un curso concreto puede haber diferencias, en unos casos positivas y en otros casos negativas. Para el caso del abandono relacionado con el abandono, se observa que el AUC global está siempre entre 0.76-0.79. Sin embargo hay 6 cursos (40%) donde el AUC baja en más de 0.1 respecto al AUC global y 3 cursos (20%) en los que hay una mejora de más de 0.1. Esto muestra que el modelo puede tener una alta variación dependiendo del curso y por tanto, no sería conveniente poner este modelo en la práctica.

Respecto al abandono relacionado con completitud, el AUC del modelo global está siempre entre 0.92-0.94. Para los cursos concretos, de los 11 cursos donde se

TABLA 8.3
 RESULTADOS DE LA PREDICCIÓN EN CURSOS CONCRETOS EN EL ESCENARIO 5

Curso	Abandono (actividad)		Abandono (completitud)		Éxito	
	AUC_GLB	AUC_CUR	AUC_GLB	AUC_CUR	AUC_GLB	AUC_CUR
Curso 1	0.78	-	0.92	1.00	0.94	1.00
Curso 2	0.76	0.74	0.94	0.71	0.95	0.75
Curso 3	0.78	0.63	0.92	-	0.93	-
Curso 4	0.78	1.00	0.92	0.86	0.93	1.00
Curso 5	0.77	0.71	0.92	-	0.93	-
Curso 6	0.77	0.82	0.93	0.93	0.93	0.98
Curso 7	0.78	0.62	0.92	-	0.93	-
Curso 8	0.77	1.00	0.93	-	0.93	-
Curso 9	0.78	0.68	0.92	0.90	0.93	0.93
Curso 10	0.79	0.68	0.92	0.95	0.93	0.96
Curso 11	0.78	0.56	0.92	0.90	0.93	0.93
Curso 12	0.78	0.70	0.93	1.00	0.94	1.00
Curso 13	0.78	0.57	0.93	0.87	0.94	0.92
Curso 14	0.75	0.92	0.92	0.95	0.93	0.99
Curso 15	0.76	0.40	0.92	-	0.94	0.54
Curso 16	0.79	0.64	0.94	0.80	0.95	0.85

AUC_GLB: AUC del modelo global. AUC_CUR: AUC del curso concreto

puede calcular el AUC, en 2 disminuye el AUC en más de 0.1 (18%) y solo en 1 de ellos el AUC baja de 0.8. Para el éxito académico, las conclusiones son similares y solo 2 cursos tienen una bajada de AUC superior a 0.1 (17%), aunque en los dos casos el AUC obtenido es inferior a 0.8. Por ello, respecto a estos dos modelos, se aprecia que son más potentes que el de abandono relacionado con actividad y podrían funcionar en la mayoría de contextos. Por ello, la recomendación sería utilizar este modelo, aunque excluyendo los cursos más problemáticos y pilotar cómo funciona en la práctica. También sería recomendable analizar diferentes cohortes del mismo curso para ver si el comportamiento es similar. En caso afirmativo, el modelo podría utilizarse. No obstante, se debe tener en cuenta que el modelo no va a funcionar bien para todos los contextos (como sugieron Gašević et al. [285]) y debería centrarse en aquellos donde sí se obtenga un buen poder predictivo.

Para corroborar el hecho de que un modelo global puede tener errores en cursos concretos, se utilizó el escenario 6. En este escenario, dado que el número de cursos es muy elevado (480), lo que se hizo fue seleccionar aleatoriamente un 80% de los cursos para entrenar el modelo y el 20% de los cursos restantes para evaluar el poder predictivo de cursos concretos. El análisis se realizó a mitad de curso (datos hasta el 31 de enero de 2018) y usando RF al ser el algoritmo que mejores resultados proporcionaba en dicho contexto. La Tabla 8.4 muestra los cuartiles del poder predictivo obtenido en los diferentes cursos, junto con el poder predictivo obtenido en el modelo global, tanto para la predicción del abandono como de la nota. En particular, para el conjunto de cursos de pruebas (test), se muestra el poder predictivo del conjunto y los cuartiles del poder predictivo de los cursos concretos. Además, se muestra el porcentaje de cursos que tienen un AUC mayor/menor a 0.1 sobre el de entrenamiento y un RMSE mayor/menor a 0.05 sobre el de entrenamiento.

TABLA 8.4
RESULTADOS DE LA PREDICCIÓN EN CURSOS CONCRETOS EN EL ESCENARIO 6

Predicción	Cuartil 1	Mediana	Cuartil 3	Debajo	Encima	Entrenamiento	Test
Abandono	0.84	0.94	1.00	18 %	32 %	0.90	0.89
Nota	0.09	0.11	0.17	15 %	20 %	0.13	0.13

El abandono se mide en **AUC** y la nota en **RMSE**

Encima implica una subida del **AUC** superior a 0.1 o del **RMSE** superior a 0.05 respecto al entrenamiento

Debajo implica una bajada del **AUC** superior a 0.1 o del **RMSE** superior a 0.05 respecto al entrenamiento

A la vista de los resultados, una primera observación es que el poder predictivo, tanto para el abandono como para la nota, es muy similar en los conjuntos de entrenamiento y test. Esto significa que el modelo de forma global puede funcionar bien sin problemas de sobreajuste aparentes. Sin embargo, al analizar el poder predictivo, se puede ver que hay cursos donde el modelo puede funcionar algo mejor y otros en los que el modelo no funciona. En cuanto a los cuartiles, se observa que no hay una diferencia muy grande desde el valor del modelo global (en test) a los cuartiles 1 y 3, por lo que al menos para la mitad de los cursos puede funcionar. Sin embargo, se observa que en cuanto a la predicción de la nota, un 18 % de los cursos tiene un **AUC** menor de 0.8 (más de 0.1 respecto al de entrenamiento, que sería el **AUC** teórico que se sabría al diseñar el modelo). También, hay un 32 % de cursos donde el modelo funciona mejor, aunque esto sería menos problemático. Del mismo modo, se observa que hay un 15 % de cursos donde el **RMSE** es superior a 0.18 (más de 0.05 respecto al de entrenamiento). Aunque este **RMSE** todavía sería aceptable, se aprecia que hay un porcentaje que no sigue lo esperado. Por tanto, en este escenario, hay entre un 15-18 % de cursos donde los resultados son considerablemente peores que los esperados según el entrenamiento y, en muchos de ellos, podría ser mejor no ofrecer las predicciones. Este resultado es similar al obtenido en el escenario 5 para el abandono relacionado con la completitud y el éxito, lo que implica que los modelos globales pueden funcionar para muchos casos, pero hay un porcentaje no despreciable de cursos donde pueden fallar. La implicación es que esta aproximación podría usarse pero para ello sería importante identificar los cursos a los que no se ajuste para descartarlos. Para estos cursos, se podría analizar si se puede desarrollar un modelo específico u otro con otro conjunto de cursos. Para la identificación de estos cursos, más trabajo hace falta para identificar las características de estos cursos y si el poder predictivo de estos cursos se mantiene en distintos cohortes en este tipo de modelos. También se puede analizar cómo poder mejorar estos modelos globales.

Como propuesta para mejorar los modelos globales, una idea consiste en utilizar variables sobre el contexto del curso. Estas variables serían comunes para todos los estudiantes del mismo curso y servirían para que el modelo pudiera distinguir los alumnos en función de algunas características del curso. Por ejemplo, se podría considerar el número de vídeos o ejercicios que tiene el curso, la duración media de los vídeos, el modo de instrucción (síncrono/asíncrono), el área temática, el número de profesores o el idioma (p. ej., si es inglés o español), entre otros.

De los dos escenarios considerados en esta sección, en el escenario 6 no se puede añadir este tipo de variables ya que se desconoce en principio el contexto de cada curso concreto (aunque se podría obtener en un futuro). Sin embargo, en el escenario 5 sí se conoce algunos datos sobre los **SPOCs**. Por tanto, es posible analizar si este tipo de variables pueden añadir poder predictivo a los modelos. Para realizar este

análisis se han considerado las variables sobre la tipología del curso citadas en el párrafo anterior, que son las que se describieron en el capítulo 5. Teniendo en cuenta estas variables, se ha realizado el análisis temporal con las tres variables dependientes del escenario 5 y RF. En particular, se han elaborado tres modelos. Un modelo contiene las variables de actividad, vídeo y ejercicios (denominado “alumno”, al ser variables referidas a la actividad de cada alumno), otro modelo para las variables de curso y otro mezclando ambas. Los resultados de estos modelos se muestran en la Fig. 8.1.

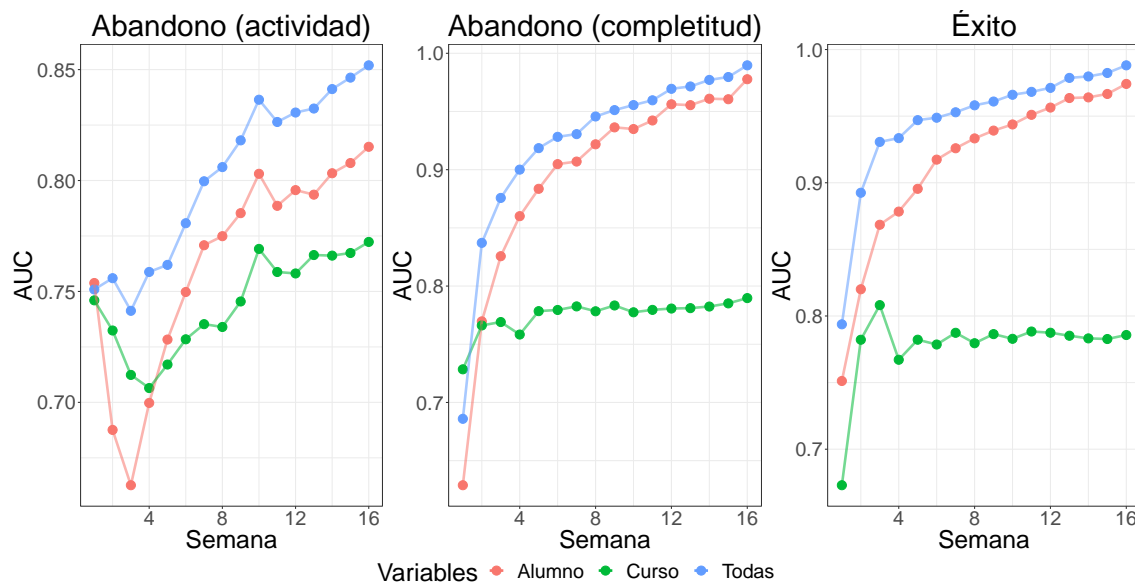


Fig. 8.1. Predicción utilizando variables de curso en el escenario 5

A la vista de los resultados, se observa que en los tres casos se produce una mejora al introducir las variables de curso. Al realizar la media de la mejora del AUC a lo largo de las semanas para los tres casos, se observa que para el abandono relacionado con actividad se produce una mejora media de 0.04 en el AUC al introducir las variables de curso, mientras que la mejora es de 0.03 para las otras dos variables dependientes. Esto implica que las variables de curso pueden proporcionar una mejora moderada del poder predictivo. Además, se observa que esta mejora normalmente es más elevada al inicio del curso, cuando todavía se dispone de menos datos sobre los alumnos. En dicho periodo, se ha observado una mejora en el AUC de hasta 0.08 en el abandono relacionado con la actividad y de 0.07 en las otras dos variables, lo cual es considerable. Otra observación es que el AUC utilizando únicamente las variables de curso es moderado, lo que indica que el diseño del curso puede tener un efecto importante en el abandono o éxito académico.

Para profundizar más en las variables, se calculó la importancia de las variables, usando la disminución media de Gini [280]. Para el caso del abandono relacionado con la actividad, la variable con mayor importancia es el porcentaje de ejercicios intentados (`perc_intentados`), seguido del número medio de repeticiones en vídeos (`avg_rep`) y el porcentaje de días que se accede a la plataforma (`perc_dias`). En cuanto a las variables de curso, no aparece ninguna entre las primeras, aunque la importancia relativa es similar en varias variables. La variable de curso con mayor efecto es la longitud media de los videos (`avg_duracion`), seguida del número de ejercicios del curso (`nejercicios`). Para el caso del abandono relacionado con la

completitud, las variables con mayor importancia son el porcentaje de ejercicios intentados, seguida del porcentaje de ejercicios correctos en primer intento (*CFA*) y el porcentaje de visualizado de vídeos (*perc_vtotal*). Sin embargo, en cuarto lugar aparte la longitud media de los vídeos (*avg_duracion*), por lo que esta variable tiene una alta importancia. El resto de variables tienen menor importancia, aunque de entre el resto, la que más importancia tiene es el número de ejercicios (*nejercicios*). Para el éxito académico, de nuevo el porcentaje de ejercicios intentados (*perc_intentados*) es la variable más relevante, lo que indica que para poder completar el curso o tener éxito lo más importante es que se hagan los ejercicios. La segunda variable más importante es *CFA*, pero la tercera es la longitud media de los vídeos (*avg_duracion*), lo que implica que las variables de curso pueden tener un gran efecto. En este caso, en 4ª y 6ª posición aparecen el número de ejercicios y de vídeos del curso (*nejercicios* y *nvideos*), lo que demuestra que las variables de curso tienen una alta influencia cuando los modelos se construyen a partir de datos de múltiples cursos. No obstante, en vista de los resultados, las variables referidas al alumno tienen mayor importancia que las del curso, por lo que aunque el curso pueda tener una influencia importante, el comportamiento del alumno es lo más relevante.

Dada la importancia de algunas variables sobre el curso, se ha realizado un análisis de dichas variables para ver qué valores tienen para los distintos grupos. En particular, se han realizado diagramas de caja con la duración media de los vídeos para las tres variables dependientes, el número de ejercicios y vídeos para el éxito, y el número de ejercicios para el abandono relacionado con completitud. Se han elegido estas variables al ser las que tenían mayor importancia en cada contexto. Los diagramas de caja se muestran en la Fig. 8.2.

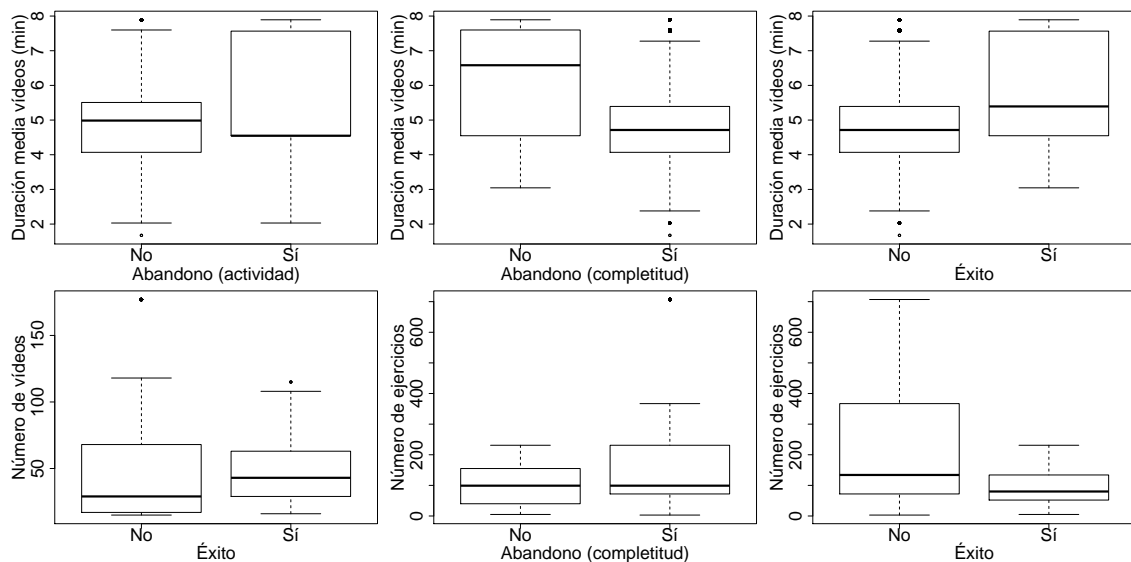


Fig. 8.2. Diagramas de caja con la relación entre las variables de curso y las variables dependientes

En estos diagramas se observa en primer lugar la duración media de los vídeos. Para el abandono relacionado con la actividad, se muestra que la mediana de la duración de los vídeos en los cursos donde más alumnos abandonan es menor, aunque hay muchos alumnos que abandonan en cursos con duraciones de vídeo altas. Para el caso del abandono relacionado con la completitud y el éxito, que parecen más

significativos por el poder predictivo, se observa que hay un menor abandono (y mayor éxito) cuando la duración de los vídeos es mayor. La mediana del tiempo medio de los vídeos que toman los alumnos que abandonan/fracasan es de entorno a 5 minutos, mientras que es superior para los que completan/aprueban. En concreto, la mediana es cercana a los 7 minutos para los alumnos que completan y de en torno a 5.5 minutos para los que aprueban. Esto podría indicar que una duración entre 5.5-6.5 minutos por vídeo podría ser adecuada para que los vídeos sean lo suficientemente completos y no desmotiven. Esto es coherente con el resultado de Guo et al. [286], que indicó que los vídeos deberían tener una duración máxima de seis minutos, ya que a partir de ahí se perdía la atención, e indica que una duración en torno a seis minutos podría ser adecuada. En cuanto al número de vídeos, el resultado muestra que los cursos con mayor número de vídeos tienen un mayor número de éxito, al contrario que los ejercicios. Esto podría indicar que habiendo más vídeos y siendo completos, los estudiantes podrían estar más motivados a usar el SPOC, pero si hay demasiados ejercicios, los alumnos podrían realizar solo un subconjunto y no llegar a los umbrales de completar el curso o éxito en el SPOC. No obstante, esto podría no ser necesariamente negativo, ya que el alumno podría adquirir los conocimientos y aprobar la asignatura, que es algo que no se evalúa dentro del SPOC. También, estos resultados se presentan con un subconjunto de cursos, pero podrían cambiar en otro contexto, aunque pueden aportar ideas para el diseño de los cursos.

Tras ver cómo pueden afectar algunas variables sobre la tipología del curso, el último paso es ver si esta mejora también se produce en los cursos individuales y no únicamente en el modelo global. Para ello, se ha vuelto a calcular el AUC para cada curso concreto cuando se entrena con los cursos restantes, pero ahora introduciendo las variables de curso. Sin embargo, en este caso se ha observado que la mejora no es demasiado significativa. Para el abandono relacionado con la actividad, la mejora es prácticamente nula en media. Hay algunos cursos (6) donde el AUC se mejora en media 0.04, lo que va en la línea de lo mostrado en el AUC global, pero en otros empeora, particularmente en cursos con número de casos muy bajo. Para el abandono relacionado con la completitud, el AUC por cada curso concreto mejora en media 0.02 y solo empeora en dos cursos. Para el caso del éxito, la mejora es prácticamente nula. Hay una mejora del AUC de media 0.03 para los que presentan mejora, pero hay otros cursos donde empeora, también los que tienen un número bajo de casos. Esto implica que incluso utilizando variables de curso, el problema de la generalización no se resuelve. Para algunos casos, este tipo de variables pueden mejorar el modelo y en global se producía una mejora, pero dado que el contexto de cada curso es diferente, puede variar.

Por ello, teniendo en cuenta todo lo visto en este apartado, la conclusión respecto a la aproximación del agregado de cursos es que hay que utilizarla con precaución. Es posible obtener un buen poder predictivo, pero eso no significa que vaya a funcionar bien en todos los cursos. Según los análisis, el poder predictivo de los cursos individuales puede ser considerablemente peor que el poder predictivo global en alrededor del 15-18% de los cursos, por lo que este hecho hay que tenerlo en cuenta. Los resultados también demuestran que el uso de variables de cursos pueden mejorar de forma moderada el poder predictivo de forma global, aunque eso no implica que mejore el poder predictivo en todos los cursos. En algunos sí puede haber una mejora, pero en otros no. Por ello, habría que ver en cada escenario si con la información semántica de los cursos se presenta una mejora y en qué cursos.

También habría que pensar si la inclusión de nuevas variables no consideradas pueden suponer una mejora en el poder predictivo.

Dado que no es posible obtener un modelo que generalice a todos los cursos, si se usa esta aproximación, debe analizarse bien en qué tipos de cursos va a poder funcionar, y el modelo debería utilizarse solo para aquellos cursos en los cuales puede proporcionar predicciones precisas. De este modo, el poder predictivo global debería tomarse únicamente como referencia y cuando se desarrollen estos modelos, debería analizarse el de los cursos concretos para ver dónde se puede aplicar. En los cursos donde funcione, esta aproximación podría ser válida y podría aportar información valiosa a los profesores. En los que no habría que ver si se puede elaborar algún modelo particular o si se puede utilizar la transferencia de algún otro modelo.

8.3. Reutilización y adaptación de los modelos

En las secciones anteriores, se ha abordado el problema de la generalización desde dos aproximaciones. Aunque en algunos casos se ha demostrado que un modelo puede ser transferido a otro curso o que en algunos casos puede utilizarse un modelo global, hay muchos casos en los que el contexto del curso puede hacer que el modelo no sea generalizable, como discutieron Gašević et al. [285]. Como también se argumentó en un trabajo previo [269], puede haber muchos factores que afecten al curso. Estos factores pueden ser el modo de instrucción, la obligatoriedad de realizar más o menos tareas (como ocurre en los SPOCs del escenario 5) o la influencia de variables externas. Por ejemplo, en el escenario 4, probablemente se podrían mejorar los modelos si se conociera información sobre los resultados del alumno en el último curso de secundaria (fuera de la plataforma).

Dado que existen muchos factores del contexto que pueden afectar a los modelos predictivos, una tercera aproximación, que se debe utilizar en muchas ocasiones, es la de reutilizar y adaptar los modelos. Esto quiere decir que aunque no se pueda usar un modelo generado para otro curso o un modelo global, es posible reutilizar parte del desarrollo y realizar una adaptación del modelo implementado para el nuevo curso. El grado de reutilización y adaptación también dependerá del contexto particular. Para argumentar mejor este tema, se han definido cuatro etapas para el desarrollo de los modelos predictivos. Estas etapas son similares a las mencionadas en la metodología de la tesis (capítulo 3), aunque no centradas en la componente investigadora, sino en la implementación del modelo. Las cuatro etapas son [287]:

- Procesado de los datos en crudo y limpieza de datos, incluyendo el filtrado de los alumnos.
- Obtención de las variables independientes.
- Obtención de la(s) variable(s) dependiente(s).
- Entrenamiento del modelo predictivo a partir de las variables obtenidas.

Teniendo en cuenta que la elaboración de todo modelo predictivo debe seguir esas etapas, se puede ver que es posible reutilizar parte del trabajo de unas a otras. Para ilustrar esto se proponen tres casos, de mayor a menor reutilización. El primer caso consiste en realizar una adaptación para otro curso de la misma plataforma, el

segundo para otro curso en otra plataforma, y el tercero para otros casos de predicción en el ámbito educativo.

Caso 1. Reutilización y adaptación en la misma plataforma

El primer caso es el de realizar los modelos para datos de la misma plataforma. Un ejemplo sería intentar reutilizar los modelos entre los escenarios 1, 2, 4 y 5, en las que en todos ellos se utiliza la plataforma Open edX. A la vista de estos escenarios, a primera instancia podría parecer que algunos escenarios son muy diferentes. Por ejemplo, el MOOC de HKUST del escenario 2 puede parecer muy diferente de los SPOCs para preparar la prueba de acceso en Flandes. Por ello, de partida, es sensato pensar que el modelo para predecir la nota final del MOOC no puede ser utilizado para la prueba de acceso del escenario 4. Sin embargo, ambos cursos comparten que los datos proceden de una plataforma Open edX, por lo que el formato de los datos es el mismo. Por tanto, el código para procesar los datos en crudo y limpiar los datos puede ser común para ambos cursos. Del mismo modo, muchas de las variables pueden ser comunes a ambos cursos (p. ej., nota media, porcentaje de vídeos vistos, etc.). Dado que el formato de los datos es el mismo, es posible reutilizar un código para el cálculo de dichas variables. También, una vez obtenidas las variables, el código del entrenamiento puede ser reutilizado. Por ello, ante dos escenarios bastante diferentes, es posible reutilizar gran parte del procesado. En este caso, habría que ver si se añade/elimina alguna variable o algún filtrado, pero el cuerpo general del procesado sería equivalente. También habría que ver si cambia la definición de algunas variables. Por ejemplo, los criterios de abandono pueden cambiar según el contexto o del éxito académico. Por ejemplo, la tasa mínima para aprobar es diferente en los escenarios 1 y 2, por lo que el indicador del éxito cambiaría. Sin embargo, para otros casos, el análisis podría ser muy similar. Para el caso, por ejemplo, de los cuatro SPOCs del escenario 4, dado que en todos ellos se utilizan las mismas variables y filtrado, podría ejecutarse el mismo código con los datos de los cuatro cursos. Por tanto, en este caso, se aprecia que sin un esfuerzo significativo, es fácil hacer una reutilización del modelo y se podría construir un modelo fácilmente para cada curso, lo que podría mejorar los resultados.

Caso 2. Reutilización y adaptación en otra plataforma

Para el caso en el que los cursos estén en distintas plataformas (como el caso de los MOOCs de edX y Coursera), el proceso puede resultar algo más complejo que en el caso anterior, pero también es posible reutilizar ciertos aspectos. La parte del procesado y limpieza de datos será diferente, dado que el formato del que proceden los datos es diferente. Sin embargo, los indicadores utilizados pueden ser similares, por lo que aunque haya que adaptar el contexto, la definición de los mismos se podría reutilizar en parte, y parte de la algoritmia. De hecho, si se consigue transformar los datos a un formato interoperable, el cálculo de los indicadores, podría no variar. No obstante, puede ser que haya algunos indicadores que en una plataforma no tengan sentido, y en otra aporten algún tipo de información particular que pueda enriquecer los modelos. En ese caso, podría merecer la pena añadir nuevos indicadores. Del mismo modo, puede ser que la definición de las variables dependientes tenga que cambiar dependiendo de las políticas de la plataforma. Por ejemplo, puede haber

una plataforma donde sea necesario superar todas las pruebas sumativas para obtener el certificado (como Coursera en la actualidad) u otra que solo requiera alcanzar un porcentaje mínimo para aprobar, lo que llevaría adaptar algún indicador. Sin embargo, en general, el proceso de diseño de las variables se podría extrapolar y adaptar la implementación para la nueva plataforma. En cuanto a los algoritmos, se podría reutilizar bastante ya que en ambos casos se tendría una entrada con las variables para realizar el entrenamiento. Por ello, la adaptación, aunque conllevaría un tiempo para contemplar las nuevas características de la plataforma, sería posible, con la ventaja de que si en una misma institución se utiliza la misma plataforma, una vez que se haya hecho la primera adaptación, el proceso para generar los modelos en los nuevos cursos sería muy sencillo (caso 1).

Caso 3. Reutilización y adaptación para otros casos en el ámbito educativo

Por último, en el ámbito educativo, puede haber casos en los que la predicción sea más diferente por el tipo de datos o por la variable a predecir. En dichos casos, no será posible reutilizar algunas partes del código, pero sí se puede reutilizar el conocimiento adquirido y la metodología para el diseño. También se podrán reutilizar parte de los algoritmos, ya que serán independientes de las entradas proporcionadas.

Como ejemplo de este caso se podría enmarcar la predicción académica. En la predicción académica, el objetivo es llevar la predicción a nivel de una titulación completa en lugar de en un curso específico. Por ejemplo, Luo y Pardos [288] predijeron si un estudiante iba a completar el grado a tiempo (en cuatro años), Tekin [289] predijo la nota media al final de la carrera y Chen et al. [290] si un alumno iba a abandonar la carrera. En este caso, el tipo de información normalmente viene del registro académico, almacenado en el sistema de información estudiantil (del inglés, *Student Information System (SIS)*), que es el que guarda registro de todas las notas del estudiante.

Para ilustrar este ejemplo, se comenta la experiencia llevada a cabo con la *Universidad Austral de Chile (UACH)*, dentro del contexto del proyecto LALA¹. En este caso, tratado en [291], el objetivo es predecir qué estudiantes van a abandonar la carrera de Ciencias de la Computación. Este caso es similar a los escenarios tratados en las plataformas, aunque ahora cambia el nivel en el que se trata. En las plataformas, las principales aproximaciones para el abandono son que el alumno lleve un tiempo sin acceder a la plataforma o que no complete un porcentaje de vídeos/ejercicios. Esta aproximación se podría extrapolar a este caso y aquí se podría definir como estar un periodo sin matricularse en ningún curso (actividad) o como no completar la totalidad de los cursos necesarios para obtener el título (completitud).

Para el ejemplo propuesto, se dispone de los datos del registro académico entre 2011 y 2017. Dado que la carrera dura 11 semestres (5.5 años), hay pocos datos de alumnos que puedan haber completado la titulación. Sin embargo, se conoce que los estudiantes normalmente tardan mucho en completar su carrera, y en especial sus dos primeros años (llamado Bachillerato) [292]. Además, se sabe que, de acuerdo con el histórico, los que aprueban el Bachillerato tienen una probabilidad muy alta de terminar. Por ello, el criterio para el abandono se adapta a este contexto siguiendo la definición de completitud y se considera que un estudiante termina cuando completa el Bachillerato. La predicción está más enfocada, por tanto, en

¹<https://www.lalaproject.org>

terminar el Bachillerato, pero puede ser un indicador relevante dada la alta deserción temprana.

Teniendo esto en cuenta, para el proceso de análisis, es necesario adaptar la variable dependiente a este contexto, como se ha comentado y procesar los datos del SIS, que se hace de forma similar a una plataforma. En cuanto a las variables independientes utilizadas, en este caso se deben definir unas variables nuevas, aunque pueden adaptarse de las variable sobre ejercicios (rendimiento) ya definidas. En particular, en este escenario se usaron las siguientes variables.

- Nota media: en este caso, en lugar de ser la media de las actividades del curso, es la nota media de los cursos matriculados.
- Ratio de cursos aprobados: indica la relación entre los cursos aprobados y los cursos matriculados. Esta variable tiene una idea similar a la del porcentaje de ejercicios correctos.
- Ratio de repeticiones: esta variable tiene una idea similar al número de intentos, pero extrapolado en ejercicios. Representa la relación entre los cursos que se están cursando y el número de intentos totales en esos cursos.
- Ratio de cursos anulados: esta variable sería propia de este contexto e indica el número de cursos anulados entre los matriculados.
- Ratio de asignaturas cursadas: indica la relación entre las asignaturas cursadas y las matriculadas (que incluyen también las convalidadas, reconocidas, anuladas o transferidas). Esta variable sería particular de este contexto.

Como se observa, en este caso, las variables son diferentes, pero la experiencia adquirida en otros sistemas contribuye a la adaptación, En cuanto a los algoritmos, al igual que en el caso 2, dado que lo que requieren es una entrada de variables para hacer un cálculo, sí que se puede reusar parte, aunque se entrene con unos datos diferentes.

No obstante, en este escenario de UACH, se propusieron tres aproximaciones para realizar los modelos predictivos. La primera aproximación consistió en calcular las variables usando los datos de todos los semestres cursados del alumno y la segunda en calcular las variables por cada semestre y juntar las de todos los semestres cursados (si hay 5 variables en la lista anterior y 3 semestres, habría 15 variables). La tercera aproximación consistió en un modelo de configuraciones (conjunto de asignaturas concretas que se toman a la vez). Con este modelo, se calculaba un indicador basado en el rendimiento del alumno en un conjunto de asignaturas, de modo que el modelo tuviera en cuenta no el rendimiento en las asignaturas de un semestre sino en asignaturas concretas. La ventaja de este modelo es que también aportaba un poder descriptivo al poder identificar qué conjuntos de asignaturas podrían generar más dificultades.

En definitiva, esta experiencia muestra un caso diferente de los escenarios que se han tratado en la tesis, ya que el foco está a nivel de titulación en vez de curso y las variables, tanto dependientes como independientes son diferentes. Sin embargo, algunos elementos como los algoritmos y la experiencia en el diseño de las variables pueden utilizarse para extrapolar a este tipo de contextos. Evidentemente, en este caso habrá que hacer un esfuerzo bastante mayor para poner en funcionamiento los modelos predictivos, pero hay ciertos aspectos que pueden ser muy similares.

Así, puede haber muchos escenarios en los que por su contexto particular no sea posible llevar a cabo la transferencia o un modelo global. En esos casos, lo mejor será realizar una adaptación específica al contexto, reutilizando todo lo que sea posible. Para realizar esa adaptación pueden seguirse unos pasos comunes y en función de si el contexto comparte ciertas características o no (p. ej., plataforma), la adaptación será más o menos sencilla. Sin embargo, en cualquier caso, la experiencia del desarrollo de un modelo servirá para la extrapolación a otros contextos, incluso más desconectados. Por ello, la recomendación en este sentido es valorar el nivel de generalización que se puede obtener con otros modelos y considerar realizar una adaptación más profunda si es necesario, ya que no será posible obtener un modelo que generalice para todos los casos.

8.4. Lecciones aprendidas

A lo largo de este capítulo, se han analizado diferentes aproximaciones en referencia a la generalización de los modelos predictivos. En primer lugar, se analizó la transferencia de los cursos, después el agregado de cursos y por último, la reutilización y adaptación. Esta sección presenta las conclusiones finales respecto a estas aproximaciones.

En primer lugar, al realizar la transferencia de los cursos, se observó que había unos casos en los que se podía conseguir transferir un modelo a otro y otros casos en los que había una pérdida significativa de poder predictivo. En particular, se observó que para el caso de los MOOCs (escenario 3), era posible transferir los modelos entre los tres modelos analizados (aunque no con todos los algoritmos). Sin embargo, en el escenario 4, la generalización dependía de las condiciones (curso y cohorte) en las que se realizaba la transferencia del modelo. Un posible motivo puede estar en la metodología y/o contexto del curso. En los MOOCs, a pesar de ser diferentes, tienen una estructura similar y si en cierto punto del curso no han realizado un porcentaje de las tareas, es probable que no las hagan, teniendo en cuenta el alto porcentaje de abandono en los MOOCs [147]. Aunque la capacidad de transferencia varía dependiendo del MOOC concreto y puede ser que en MOOCs diferentes no se pueda lograr, sí que puede ser posible detectar muchos casos de abandono tras ver los alumnos que dejan de interactuar las primeras semanas, lo que puede ayudar a la transferencia. Sin embargo, este patrón típico de los MOOCs es diferente en otros contextos donde cada curso puede tener sus características concretas. Esto puede hacer que la transferencia sea más complicada en el escenario 4. En este escenario, aunque los SPOCs se diseñan con el mismo objetivo, puede haber diferencias entre las materias y el uso que hagan los alumnos (al ser cursos de apoyo). Teniendo en cuenta los distintos experimentos realizados en el escenario 4, se puede pensar que deben cumplirse ciertas condiciones para cumplir la transferencia de los modelos. En particular, del análisis realizado [29], se pueden extraer las siguientes condiciones:

- Las características de los estudiantes deben ser lo más similares posibles. Cuando se generó un modelo con un curso y se transfirió a otro curso con los mismos estudiantes, el poder predictivo era muy alto. Esto implica que si el comportamiento de los estudiantes es similar, es posible transferir los modelos. A pesar de que no siempre es posible utilizar los modelos con los estudiantes del mismo cohorte, los resultados muestran que sería posible por el alto poder

predictivo. Sin embargo, cuando se cambian los cohortes (estudiantes), el poder predictivo es peor pero podría también ser aceptable.

- Los cursos deben ser lo más similares posible. Los resultados mostraron que el poder predictivo se reducía considerablemente cuando se predecía utilizando un modelo entrenado con estudiantes diferentes en un curso diferente. Este hecho puede dificultar la transferencia de los modelos, aunque el poder predictivo podría ser aceptable si el curso es similar. Si el curso es el mismo, pero en una edición diferente, podría ser posible alcanzar un poder predictivo razonable. Por tanto, sería mejor generar un modelo a partir de los datos del mismo curso en otra edición (si está disponible) que utilizar un modelo entrenado con otro curso.

Aparte de estas condiciones, que fueron analizadas en la sección 8.1, hay otras condiciones que tenían en común los cuatro SPOCs del escenario 4 [29]. Estas condiciones podrían contribuir a la generalización, ya que están muy relacionadas con el contexto del curso. Dichas condiciones son:

- La metodología. Los cuatro SPOCs utilizaban enseñanza semipresencial (*blended learning*) y se organizaban de forma similar. Todos los SPOCs se desarrollaban en el mismo periodo (desde septiembre hasta la fecha del examen) cada año, y se desarrollaban sesiones presenciales siguiendo una aproximación similar en todos los cursos. Este hecho puede afectar a la generalización.
- Modo de instrucción. Todos los SPOCs eran síncronos, ya que los materiales se publicaban secuencialmente cada quince días. Los cambios en el modo de instrucción podrían afectar al comportamiento de los estudiantes (p. ej., los estudiantes podrían tener más problemas de autorregulación en contextos asíncronos [266]) y consecuentemente a la generalización.
- Importancia percibida sobre el curso. En este caso, todos los SPOCs servían de material de apoyo para la prueba de acceso, aunque se percibían como una herramienta importante para la preparación de la prueba. Incluso si dos cursos son iguales, puede haber diferencias si la percepción de los mismos es diferente. Por ejemplo, si un profesor no enfatiza la importancia del SPOC y lo considera como material complementario, los resultados podrían variar respecto a otro curso donde se hace un uso más prioritario del SPOC.
- Duración del curso. Todos los SPOCs fueron lanzado al inicio del curso académico y se desarrollaban a lo largo del curso académico en le que se desarrollaba la prueba. Si la duración variase, podría haber diferencias en la predicción.
- Relación entre la variable dependiente entre los cursos. La variable dependiente en este caso es el resultado (aprobar/suspender) la parte de ciencias del examen, lo que incluye tanto Física como Química. Esto podría contribuir a la generalización entre los cursos que forman parte de la parte de la nota de ciencias, que es lo que se predice.

Cuando se cumplan todas esas condiciones, aumentarán las opciones de que los modelos se pueden transferir de un curso a otro. Sin embargo, si solo se cumplen algunas de ellas, el poder predictivo ser lo suficientemente bueno para que el modelo sea usado. Por ejemplo, el análisis mostró que el poder predictivo era aceptable cuando se transfería el modelo al mismo curso en una edición futura. Este resultado coincide con otros artículos que analizaron la generalización (p. ej., [211], [216]), aunque la precisión del modelo puede variar dependiendo del contexto. No obstante, el resultado es positivo en el sentido de que bajo ciertas condiciones es posible que un modelo funcione para otro curso.

En caso de no poder utilizar la primera aproximación, la segunda aproximación es la de utilizar un modelo global. De acuerdo con los resultados del análisis, es posible obtener un alto poder predictivo en estos modelos (tal y como también demostró Kizilcec y Halawa [216]), aunque el problema es que el hecho de obtener un buen poder predictivo global no implica que el poder predictivo sea alto en todos los cursos debido al contexto. No obstante, según los análisis, en los que se obtenía un poder predictivo alto, los modelos solo mostraban un poder predictivo considerablemente peor en un 15-18% de los cursos. Por ello, se podría utilizar esta aproximación, aunque habría que evaluar que el curso para el que se quiera usar no entre dentro de los cursos en los que el modelo no funcione suficientemente bien. Además, en este análisis se ha analizado solo una edición del curso, por lo que sería importante asegurar que el comportamiento en distintas ediciones pueda ser similar (como se vio en la aproximación 1) para ver que si el modelo global funcionaba en una edición, pueda funcionar en la siguiente. Si esto es así, esta aproximación podría ser válida.

No obstante, dado que el contexto puede afectar significativamente a los modelos predictivos, es posible que haya casos en los que el modelo global no sea válido ni se pueda utilizar otro contexto. En esos casos, habrá que ver si es posible reutilizar y adaptar otro modelo para crear un modelo específico para ese curso. El grado de adaptación y complejidad dependerá del contexto y en algunos casos la adaptación puede ser más o menos sencilla, pero de esta manera se podrán solventar algunos problemas del contexto. El problema de esta aproximación es que en algunos casos podría perder utilidad práctica. En cualquier caso, podría ser posible diseñar el modelo con datos pasados y que tenga poder predictivo, pero para que tenga una aplicabilidad se necesitaría que el modelo pudiera ser usado en el futuro. Por ello, esta aproximación tiene sentido cuando se quiere crear un modelo por primera vez en un curso para luego poder transferirlo a otros (usando la aproximación 1). De este modo, en lugar de usar otro curso existente, se adapta el modelo para el curso para su uso posterior. Sin embargo, para que tenga efecto se debe asumir al menos la condición de transferencia dentro de la misma edición, que según los resultados del análisis, debería cumplirse si no cambia significativamente el contexto.

Para evitar este problema, otra posible aproximación sería la de utilizar modelos *in-situ*, que son modelos que utilizan los datos del curso actual para la generación de los modelos de manera que no cambie el contexto. Por ejemplo, Er et al. [293] tenían como objetivo predecir el compromiso en las diferentes actividades de revisión por pares. En su caso, realizaron una vez que pasó la primera actividad realizaron un modelo con las interacciones hasta esa actividad y conociendo la variable dependiente. Después, recogieron las interacciones hasta la segunda actividad y utilizaron el modelo generado anteriormente (que estaba hecho en el mismo curso). Posteriormente, se entrenaba un nuevo modelo a partir de la variable dependiente

de la segunda actividad y se usaba para la tercera y así sucesivamente. Esta aproximación, que puede dar buenos resultados, podría ser válida, por ejemplo, para el escenario 1 y 2 en el que se predicen notas parciales, de modo que el modelo realizado para la prueba X a posteriori sirviera para predecir en tiempo real para la prueba $X+1$. Aunque el análisis de esta aproximación (que forma parte del trabajo futuro) pueda ser relevante, hay una limitación a destacar. La limitación es que estos modelos podrían no ser aplicables cuando no hay mediciones parciales de las variables dependientes, como por ejemplo, el éxito o el abandono, que solo se mide al final.

En definitiva, a lo largo de la sección se han visto diferentes aproximaciones para conseguir la generalización de los modelos predictivos. Una conclusión es que no puede haber un modelo que generalice a todos los casos (como indicó Gašević et al. [285]). Sin embargo, a pesar de estas limitaciones, se podría transferir de un curso a otro bajo ciertas circunstancias o podría funcionar un modelo global. No obstante, en muchos casos puede interesar adaptar un modelo para que sea específico de ese curso o si es posible desarrollar un modelo con los datos del mismo curso. Por ello, se concluye que los modelos predictivos pueden ser desarrollados y usados en otros cursos o ediciones del curso, por lo que los modelos pueden ser aplicables en cursos durante su desarrollo (no solo a posteriori). Sin embargo, no en todos los casos se puede conseguir la generalización y se debe analizar cada contexto para ver cuál es la mejor aproximación, de modo que los modelos puedan ser aplicados garantizando una precisión aceptable para su uso eficaz en los contextos educativos.

9

Efecto de las variables a predecir

Contenido

9.1. Relación y diferencias entre predicción de la nota final y la del examen final	150
9.2. Relación y diferencias entre la predicción del abandono y de la nota final	152
9.3. Selección de las variables dependientes	154
9.4. Lecciones aprendidas	157

En el desarrollo de los modelos predictivos, uno de los componentes fundamentales es la variable dependiente. La medición concreta de esta variable y lo que representa puede tener un efecto importante en el modelo. Por ejemplo, en el abandono, en un curso con tareas sumativas semanales, se podría decir que un alumno ha abandonado si una semana no hace las tareas. Sin embargo, en un curso con tareas mensuales, el no realizar actividad en una semana no necesariamente implica abandono, por lo que es importante adecuar las variables dependientes a cada contexto y una elección inapropiada puede hacer que el profesor interprete mal la información. Además, las variables a predecir pueden tener relación entre ellas y por ejemplo, la nota del examen final puede estar bastante relacionada con la nota del final. También, la predicción de la nota puede estar relacionada con la del abandono. En este último caso, las variables no son necesariamente iguales, ya que un alumno puede no abandonar, pero suspender, pero pueden estar bastante ligadas. Para analizar algunos de estos aspectos, este capítulo analiza cómo varía la predicción y cuál es el efecto de utilizar algunas variables dependientes. La primera sección trata la relación entre la predicción de la nota final y la del examen final, por lo que se analiza cómo cambia el poder predictivo al cambiar de predecir de una a otra y cómo se relacionan. La segunda sección se centra en la predicción del abandono y la nota final, y en poder identificar diferentes perfiles de estudiantes según los valores de estas predicciones. La tercera sección, se centra en justificar la importancia de la selección de las variables dependientes en los escenarios de la tesis. Por último, la cuarta sección muestra las lecciones aprendidas a lo largo de este capítulo.

9.1. Relación y diferencias entre predicción de la nota final y la del examen final

En esta sección, el objetivo principal es comparar el poder predictivo que se obtiene al predecir el examen final de un curso y la nota final del curso (obtenida esta última como una nota ponderada de diferente notas realizadas a lo largo de un curso). En la literatura, estas variables se han tratado normalmente de forma separada (p. ej., [154], [294]), pero también puede ser relevante su comparación para ver cómo afecta la variable dependiente en la predicción. De este modo, el objetivo es ver qué relación hay entre estas notas y cómo cambia el poder predictivo al pasar de predecir una variable u otra. Para analizar este caso se utiliza el escenario 2 (tratado en [263]), ya que en dicho escenario hay un examen final (EF) que representa el 40% de la nota final (NF), por lo que se espera que el examen sea representativo de NF . En primer lugar, se realiza un análisis exploratorio sobre la relación de estas variables para los alumnos del MOOC. La Fig. 9.1 muestra la relación entre ambas notas en el escenario 2. La figura muestra que para aquellos alumnos que aprueban, existe una relación lineal positiva entre EF y NF , (la correlación es 0.95 utilizando todos los puntos y de 0.58 cuando se utilizan únicamente los alumnos que aprueban). Además, hay muchos estudiantes que no realizaron el examen. Casi todos ellos suspendieron el curso y probablemente abandonaron, aunque hay 12 casos de alumnos que alcanzaron el 50% mínimo requerido para aprobar, pero decidieron no realizar el examen. Además, hay muy pocos casos ($n = 48$) de alumnos que suspendieron el examen (habiéndolo hecho), lo que implica que la mayoría de los que hicieron el examen lo aprobaron, y entre los que suspendieron, el 52% aprobaron el curso. Por tanto, se puede decir que ambas notas están altamente relacionadas y casi todos los estudiantes que aprobaron el curso también aprobaron el examen. Esta alta relación se corrobora con otros trabajos, que incluso indican que el resultado de los primeros exámenes de un curso son muy buenos predictores de la nota final (p. ej., [295], [296]).

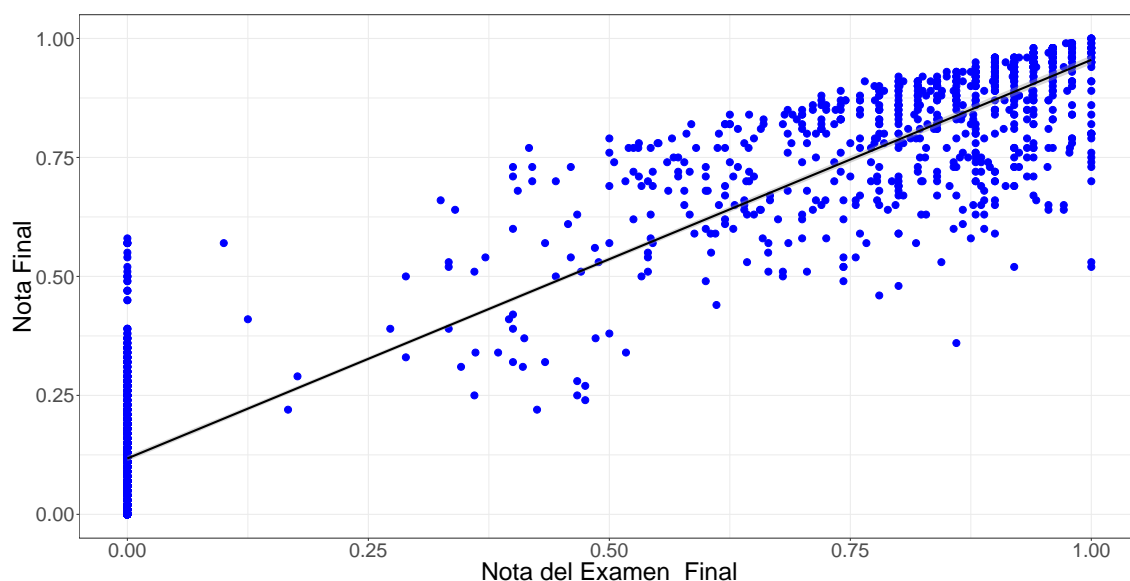


Fig. 9.1. Relación entre la nota del examen final y la nota final

Teniendo en cuenta que ambas notas están muy relacionadas, es interesante analizar las diferencias del poder predictivo al predecir tanto NF como EF . Dado que NF depende de los ejercicios (que cuentan un 20%), los modelos que se utilizan para este análisis contienen las variables sobre la actividad en la plataforma, la actividad en el foro y las interacciones con vídeos (como se usó en el capítulo 7 al predecir NF). De este modo, las predicciones de ambas notas serán comparables al utilizar los mismos indicadores, tomados de los mismos datos. Teniendo esto en cuenta, la evolución del poder predictivo para ambas notas (usando modo no acumulativo) se muestran en la Fig. 9.2. Además, para entender la relación entre estas variables, se ha realizado un modelo adicional para predecir la NF utilizando las predicciones del EF a lo largo del tiempo (un modelo con una sola variable).

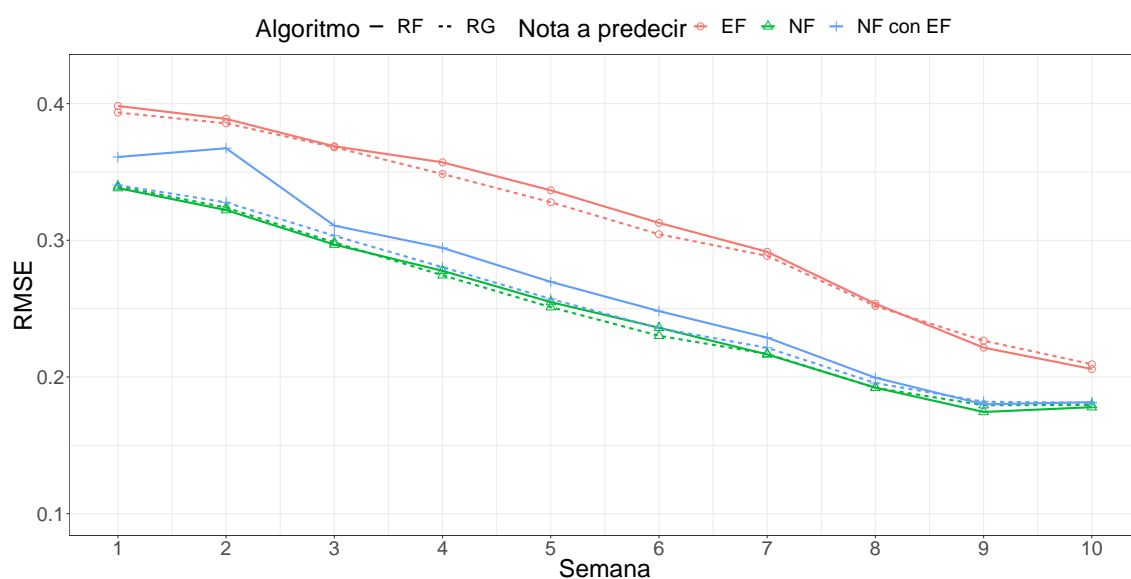


Fig. 9.2. Comparación entre los modelos predictivos para la nota final y el examen final en el escenario 2.

Una primera observación es que el poder predictivo de NF es siempre mejor que la predicción de EF . También, en el caso de EF , se necesita más tiempo para conseguir el mismo grado de precisión que en NF (p. ej., en la semana 8 se alcanza un RMSE de 0.2 para NF y ese mismo RMSE se alcanza en la semana 10 para EF). Esto sugiere que el conocimiento medio a largo plazo es más fácil de predecir que el conocimiento real en un momento determinado. Otra observación es que el poder predictivo obtenido al predecir NF utilizando únicamente las predicciones de EF no es muy diferente al obtenido al predecir NF con las interacciones de los alumnos. A pesar de que las diferencias son mayores al inicio (sobre todo con RF), los resultados son muy similares en las últimas etapas. Esto implica que debido a la alta (y lineal) relación entre ambas notas, como se mostró en la Fig. 9.1, es posible predecir NF a partir de únicamente las predicciones de EF sin perder mucho poder predictivo. El resultado también implica que las variables a predecir pueden estar relacionadas y se debe realizar investigación para analizar la relación entre las variables dependientes [16]. No obstante, el resultado de que NF es más fácil de predecir que EF puede ser relevante y puede ser considerado para la implementación de los sistemas predictivos.

9.2. Relación y diferencias entre la predicción del abandono y de la nota final

A lo largo de la tesis, se ha presentado el análisis de la predicción del abandono y la nota final del curso por separado. En la literatura, también existen muchos trabajos que han abordado estas variables por separado, ya sea tratando una sola en el artículo (p. ej., [133], [153]) o analizando ambas de forma separada (p. ej., [297]). Sin embargo, es interesante analizar de forma conjunta estas variables, ya que aunque las dos variables tienen una amplia relación (el alumno que abandona normalmente obtiene malas notas), puede haber casos en los que estas dos variables no estén relacionadas. Por ejemplo, puede haber alumnos que continúen en el curso hasta el final aunque lo hagan mal y suspendan, por lo que podrían tener una probabilidad de abandono baja y de suspenso alta. Esta sección trata de modelar los posibles perfiles en función de las predicciones del abandono y de la nota.

Para realizar estos perfiles, se ha tomado el escenario 6, en el que se predice tanto el abandono como la nota final de los estudiantes. No obstante, un análisis similar podría realizarse en otros escenarios donde se disponga de estas dos variables. En este escenario 6, se han entrenado dos modelos (para el abandono y la nota) elaborados con todas las variables del escenario 6 hasta el 31 de enero de 2018 (mitad de curso). Se ha escogido esta fecha, ya que el poder predictivo en este momento era bastante elevado, como se mostró en el capítulo 7. Estos modelos se han entrenado utilizando CV 10-fold con el 80% de los casos. Con los modelos entrenados, se ha realizado la predicción con el 20% restante de los casos ($n = 1713$, al haber interactuado 8568 alumnos en la primera mitad de curso). Para el caso de la predicción del abandono, ya que el abandono es una variable binaria, se ha calculado la probabilidad de abandono, de forma que pueda compararse con la nota, que sí es un valor continuo.

Una vez obtenidos los dos valores de las predicciones, se ha procedido a generar los perfiles en función de las mismas. Para ello, se ha utilizado agrupamiento (clustering) jerárquico con distancia euclídea. Esta técnica de agrupamiento es común en e-learning [298] y ha sido utilizada en otros trabajos para agrupar a estudiantes según su comportamiento (p. ej., [299], [300]). En este caso, mediante esta técnica, se han obtenido siete grupos de alumnos en función de sus predicciones del abandono y la nota. La elección de siete perfiles se ha realizado empíricamente para obtener perfiles de la forma más clara y visible. La Fig. 9.3 muestra la relación entre las predicciones del abandono y la nota en el escenario 6. En esta figura, se aprecia que los mejores alumnos son aquellos situados cerca del vértice superior izquierdo, ya que tienen una probabilidad baja de abandono y se predice que van a sacar buena nota. Los peores serían los situados cerca del vértice inferior derecho, ya que tendrían una alta probabilidad de abandonar y de suspender. Teniendo esto en cuenta los perfiles (clústeres) identificados mediante el agrupamiento son los siguientes:

- Clúster C1 ($n = 774$). Son estudiantes con alto riesgo y es en los que el profesor debería centrarse más. Son alumnos con una alta probabilidad de abandono y con una predicción de la nota muy baja. El profesor debería intentar identificar si hay algún problema en la motivación o en el proceso de aprendizaje para ayudar a estos estudiantes.

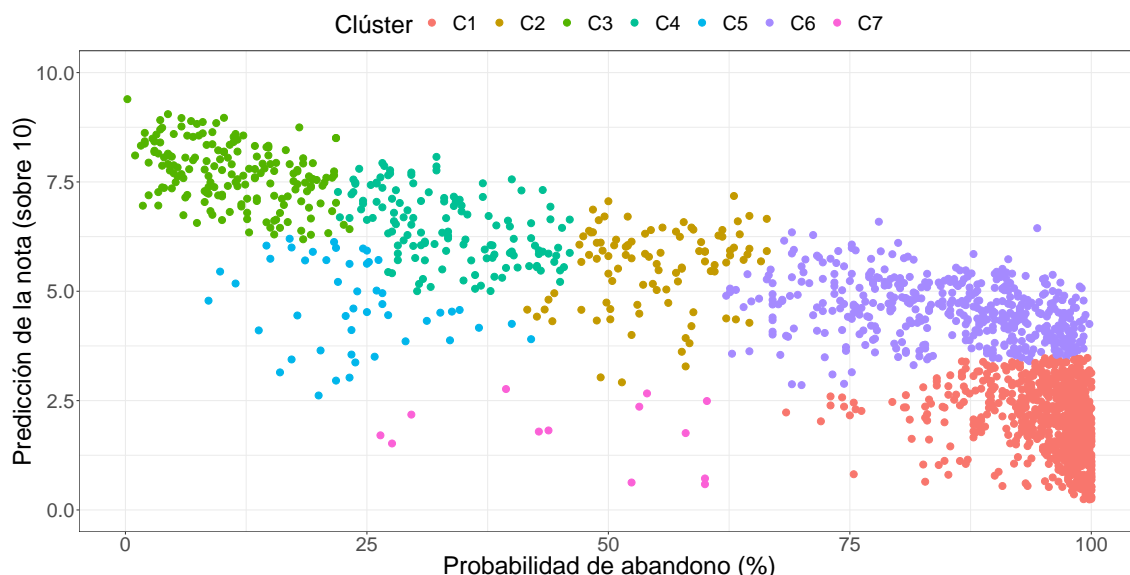


Fig. 9.3. Relación entre la predicción del abandono y de la nota final

- Clúster C2 ($n = 100$). Son estudiantes que están en el borde de aprobar o suspender y abandonar o no. El profesor debería intentar motivarles para que trabajen un poco más y salgan de la zona de riesgo.
- Clúster C3 ($n = 179$). Son los mejores estudiantes. Tienen una probabilidad de abandono baja y la predicción del abandono es siempre superior a 6 sobre 10, siendo en muchos casos por encima de 7 (umbral para la calificación de notable) y en algún caso, aunque menos, por encima de 9 (umbral para la calificación de sobresaliente, que es la máxima en el sistema educativo español para primaria y secundaria).
- Clúster C4 ($n = 135$). Son alumnos en los cuales la predicción de la nota es que van a aprobar, con una nota de suficiente [5-6), bien [6,7) o notable [7,9). La probabilidad de abandono es superior a la del clúster C3, aunque en todos los casos es inferior al 50%. Sobre este clúster habría que fijarse en los alumnos cuya predicción de la nota es inferior para intentar motivarles.
- Clúster C5 ($n = 48$). Son estudiantes que tienen una probabilidad de abandonar relativamente baja, pero la predicción de la nota dice que van a suspender o como mucho obtener un suficiente. Para muchos de los casos que suspenden, la nota que se les predice es un suspenso cercano al aprobado. Por tanto, estos estudiantes pueden ser estudiantes que se esfuerzan y quieren continuar con el curso, pero tienen dificultades en el aprendizaje. El profesor debería tratar de identificar las dificultades para ayudarlos.
- Clúster C6 ($n = 464$). Son estudiantes cuya predicción del abandono es muy alta, pero la predicción de la nota es de obtener un suficiente o un suspenso cercano al aprobado. Dado que es posible tener abandono (no hacer el 75% de las actividades), pero tener una nota superior a 5 si las actividades se hacen bien, es posible que estos alumnos tengan capacidad para hacer las actividades y en muchos casos las hagan bien cuando se ponen, pero les falte motivación para ponerse a hacer los ejercicios. El profesor debería motivar

más a los alumnos para que trabajen en las tareas que se les manda y debería monitorizar si realmente tienen dificultades o no en los contenidos cuando los realizan.

- Clúster C7 ($n = 13$). Son estudiantes con predicción de la nota muy baja (siempre inferior a 3 sobre 10), pero que no tienen una probabilidad de abandono tan alta como los del clúster C1. Estos estudiantes probablemente tengan bastantes problemas en el proceso de aprendizaje que les hagan obtener malos resultados, aunque en muchos casos intenten continuar. El profesor podría tratar de adaptar los contenidos o hacer refuerzo con estos alumnos para intentar mejorar su aprendizaje.

La obtención de estos clústeres demuestra que el abandono y la nota pueden tener una gran relación, pero puede haber diferencias entre ellas, lo que genera diferentes perfiles. Además, estos perfiles pueden ser muy prácticos para la aplicabilidad de los modelos predictivos en los cursos reales, ya que aportarían una información adicional al mero dato de la predicción y pueden ayudar al profesor a entender mejor el estado de los alumnos. De este modo, el profesor puede ver las predicciones y los perfiles de los alumnos, entender el riesgo de cada estudiante e intervenir de alguna forma para intentar mejorar el aprendizaje de los estudiantes en riesgo. Por ejemplo, en base a estos perfiles, podría ser posible la adaptación o personalización de contenidos si fuera necesario, ya sea de forma manual o a través la plataforma, para que se ofrecieran contenidos que se adaptasen a cada estudiante.

9.3. Selección de las variables dependientes

A lo largo de la tesis doctoral, se ha presentado el análisis de seis escenarios educativos, en los cuales se han utilizado diferentes variables dependientes. En algunos casos, estas variables dependientes coincidían de un modo conceptual amplio (p. ej., abandono), aunque no en todos los casos coincidía la forma en la que estaban definidas y su forma de calcularlas. Esta sección justifica la importancia de la selección de las variables dependientes y el motivo por el que una misma variable se define de diversas formas para cada escenario. Este hecho es bastante relevante, como argumentaron Gardner y Brooks [188], ya que la adaptación de las variables a su contexto puede contribuir a que las medidas estén alineadas al contexto y los objetivos del curso [301]. A continuación, se justifica la selección de cada variable en cada escenario, tratando también de ayudar con ello en la selección de las variables dependientes en nuevos contextos.

En primer lugar, se ha analizado la predicción de notas parciales. Mientras que la nota final sí puede tener sentido en todos los escenarios, para el caso de la nota parcial es diferente. En los escenarios 1 y 2, al haber pruebas semanales durante el transcurso de las semanas del curso, sí tiene sentido predecir el resultado de cada prueba para entender cuál va a ser el desempeño del alumno a lo largo de las semanas. En el escenario 3, al ser un contexto similar, también podría tener sentido la predicción de notas intermedias. Sin embargo, dado que en el escenario 3 es posible repetir la prueba un número indefinido de veces y de cara al progreso en el MOOC no es tan relevante el intento en el que se consiga, ha sido más relevante ver cuál va a ser la nota global o si el alumno va a conseguir completar todas las pruebas requeridas o no. Respecto al escenario 4, el SPOC estaba centrado en preparar la

prueba de acceso a la universidad, por lo que en dicho caso, la predicción tiene más sentido que esté centrada en la prueba de acceso, que es el objetivo para el que está dirigido el curso. En cuanto a los escenarios 5 y 6, éstos contenían actividades, pero no había una distinción entre pruebas formativas y sumativas, por lo que no era posible focalizar las predicciones en unas pruebas concretas. Por ello, en estos casos tampoco se realizó una predicción de pruebas intermedias.

Respecto a la predicción de la nota final y el éxito, ambos tipos de predicción pueden ser utilizados indistintamente, dependiendo del grado de precisión que se quiera. Si se quiere mayor precisión, es decir, se quiere un valor numérico en lugar de un valor binario de aprobado/suspense, se puede predecir la nota final, aunque la fiabilidad de la predicción puede ser menor o el nivel de anticipación (como se concluyó en el capítulo 7). Si es suficiente con saber si el estudiante aprueba/suspende, se puede utilizar la predicción del éxito, que proporcionaría la probabilidad de que al alumno complete satisfactoriamente el curso. A lo largo de los escenarios, se han utilizado ambas opciones para ilustrar que ambas tienen sentido.

En cuanto a la variable de abandono, pueden utilizarse diversas definiciones, y la definición utilizada debe ajustarse en cada caso dependiendo del contexto. En la tesis, el abandono se ha analizado en los escenarios 3, 5 y 6. Se podría haber analizado también en el resto de escenarios, aunque en ellos se puso el foco en otras variables por los objetivos concretos de la investigación. Por ejemplo, en el escenario 4, dado que el curso estaba muy centrado en la prueba de acceso, lo más interesante para el profesor era si el alumno aprobaría o no esa prueba, más que otras predicciones dentro del SPOC.

En cuanto a los escenarios en los que se ha analizado el abandono, principalmente se han utilizado dos aproximaciones: (1) analizar el abandono relacionado con la actividad y (2) analizar el abandono relacionado con la completitud. En el escenario 3, el abandono relacionado con la actividad sí tenía sentido porque en base a los datos se identificó que, por ejemplo, más del 96 % de los alumnos que llevaban una inactividad superior a 28 días abandonaban el curso. Sin embargo, esto no aplica a los escenarios 5 y 6, debido al contexto. En el escenario 5, se dispone de SPOCs, pero el uso particular de cada SPOC depende del profesor. Puede haber SPOCs en los que el profesor mande realizar tareas semanalmente, y en ese caso sí tiene sentido una definición de abandono relacionado con la actividad. En cambio, puede haber SPOCs de apoyo a las clases presenciales, en las que el alumno no necesariamente debe acceder todo el tiempo. Por ejemplo, si el SPOC está enfocado en reforzar algunos contenidos vistos en las clases presenciales, un alumno no necesariamente debe cubrir todo lo que está en el SPOC, sino que puede centrarse en los conceptos que tenga más dificultades. También, puede ser que un SPOC esté diseñado solo para una parte del curso. En ese caso, no se espera que haya una actividad continua a lo largo de todo el curso. Por estos motivos, el abandono relacionado con la actividad tiene menor sentido. Como justificación de esto, en los capítulos 6 y 7 se observó que el poder predictivo del abandono relacionado con la actividad era bastante peor en el escenario 5, probablemente porque su definición ajusta peor en dicho escenario (al depender del uso del curso de cada profesor).

Para un contexto en el que la interacción puede no estar centrada en un periodo de tiempo, el abandono puede medirse mejor a partir del porcentaje del curso completado. De este modo, se dice que un alumno completa el curso cuando completa un porcentaje mínimo de actividades. El porcentaje mínimo de actividades es un

parámetro ajustable y podría depender del curso. En este caso, se ha utilizado un umbral del 75 %, al ser un umbral típico y usado en otras plataformas (como MiríadaX¹), pero el valor podría ser refinado en otros contextos. Para el caso de los SPOCs, se ha tomado esta medida y se asume que el alumno debería completar los contenidos del SPOC. Si el SPOC es meramente de apoyo y el profesor no espera que el alumno complete la mayor parte del contenido, entonces el profesor puede interpretar la información del abandono según su caso. Sin embargo, para un caso general, se podría esperar que si un profesor diseña su SPOC es para que la mayor parte la cubran los alumnos.

Esta última asunción, sin embargo, no puede aplicarse para el escenario 6, lo que hace que en este escenario no pueda ser válida la definición de abandono, tal y como se ha mostrado. En el escenario 6, las actividades provienen de una plataforma propietaria externa al centro educativo. De este modo, las actividades son proporcionadas en la plataforma y cada profesor en su curso es libre de usarlas y/o asignarlas a los alumnos como tarea o no. Este hecho hace que no pueda esperarse que el alumno vaya a completar la mayoría de las tareas, ya que el profesor puede estar interesado solo en unas tareas concretas o temas concretos. Este hecho es similar al escenario 5, pero con la diferencia de que en el escenario 5, el profesor sí puede recomendar al alumno hacer todas las tareas, aunque no todas sean de obligatorio cumplimiento, pero aquí solo una parte de las tareas forman parte del curso. Por ello, la aproximación utilizada ha sido la de considerar el abandono según el porcentaje de ejercicios completados pero normalizado a cada curso concreto. Es decir, de cada curso, en lugar de considerar todas las actividades y calcular el ratio entre ejercicios hechos y ejercicios totales, se calcula el porcentaje de alumnos que ha realizado cada ejercicio. Para intentar evitar el caso de que un alumno haga un ejercicio que no se le haya mandado hacer, se considera un umbral mínimo para incluir un ejercicio (en este caso se ha considerado un 15 %). Entonces, el porcentaje completado para mirar el abandono ahora se convierte en el ratio de ejercicios hechos respecto al número de ejercicios que hace al menos un 15 %. De este modo, en cada curso concreto puede haber un uso diferente de las actividades, pero el profesor va a ver el abandono respecto a quién hace o no las actividades que él asigna exclusivamente, lo que le aportará más información.

Como se ha visto en esta sección, la elección de la variable dependiente es bastante relevante. Mientras que la selección de variables independientes es importante para conseguir que el modelo tenga un alto poder predictivo, la selección de variables dependientes puede afectar tanto al poder predictivo como a la interpretación de los resultados. Si la medida utilizada para la variable dependiente no es apropiada, el poder predictivo puede ser peor (como se observó al comparar las dos definiciones de abandono en el escenario 5). Además, si la definición no es apropiada, puede que la interpretación que haga el profesor de los resultados no sea buena. Por ejemplo, si se asume que el estudiante tiene que acceder todas las semanas y no lo hace, se puede alertar de que el estudiante va a abandonar cuando en realidad está siguiendo el curso correctamente. Por ello, en los modelos predictivos se debe reflexionar qué es lo que se quiere medir en cada contexto y elegir las variables dependientes que mejor se adapten al contexto.

¹<https://miriadax.net/es/faq?faqid=8635212>

9.4. Lecciones aprendidas

En este capítulo, se ha analizado el efecto de las variables a predecir y en particular se ha analizado la relación entre la predicción del examen final y la nota final, y entre el abandono y la nota final del estudiante. A partir de los análisis realizados, se ha obtenido el siguiente listado de lecciones aprendidas.

1. Se obtiene mejor poder predictivo al predecir la nota final del curso que la nota del examen final, por lo que es más fácil predecir el conocimiento medio a lo largo de un periodo que el conocimiento en un momento determinado al final del curso. Una posible hipótesis podría ser que las interacciones relacionadas con cada prueba intermedia pueden ser significativas para esa prueba (y por ende, con el agregado de las notas, en la nota final), pero no tanto para el examen final, dado que los alumnos pueden haber olvidado algunos conceptos previos (sobre todo si son conceptos que no se practican varias veces a lo largo del curso [302]). También, se observa que las predicciones de la nota del examen final sirven como predictor de la nota final, por lo que existe una alta relación entre dichas variables. Aunque este resultado se ha probado para un MOOC y podría variar en otro contexto, esta relación se podría considerar en la implementación de los modelos predictivos. Además, si existen diferencias entre las dos notas, puede ser relevante analizar las dos predicciones de forma independiente. Esto puede ser útil, por ejemplo, para identificar si hay alumnos que pudieran aprobar el curso gracias a la evaluación continua, pero vayan a tener un mal rendimiento en la prueba final. Sobre todo, esto puede cobrar importancia en cursos donde no se requiera un mínimo en el final y pudiera haber más discrepancias entre estas dos notas.
2. La predicción de la nota y el abandono, a pesar de estar muy relacionadas, puede variar en algunos alumnos. En particular, es posible identificar diferentes perfiles de estudiantes en función de su probabilidad de abandono y su predicción de la nota. Por ejemplo, es posible identificar alumnos que vayan a seguir haciendo las actividades hasta el final, aunque es muy probable que las hagan mal y no aprueben. También, es posible identificar alumnos a los que quizás les falte motivación y puedan hacer las actividades bien, pero tengan una alta probabilidad de que no las hagan. El estudio de estos perfiles puede ayudar al profesor a entender mejor el comportamiento de sus estudiantes y detectar el riesgo. Esto puede contribuir a una mejora del apoyo adaptativo o personalizado, lo cual puede tener un impacto positivo en el aprendizaje de los estudiantes.
3. La selección de la variable dependiente es relevante y puede afectar a los resultados de los modelos predictivos. En cada caso, es necesario adaptar la variable a los objetivos del curso [301]. De cara a una implementación, se puede reflexionar si el foco debe estar, por ejemplo, en completar las actividades (relacionado con el abandono) o en realizar bien las actividades (nota). Por ejemplo, en un curso meramente de apoyo donde no se requiera realizar todas las actividades, podría haber mayor interés en ver que las actividades que se hagan estén bien. No obstante, como se comentó anteriormente el abandono y la nota pueden ser analizadas de forma conjunta para obtener mayor información del estudiante. Si se quiere predecir una nota, otra decisión

que se puede tomar, en función del grado de la anticipación y la precisión es si se requiere una nota continua o un valor binario de éxito, y si se predice una nota al final o notas parciales. Esta decisión depende del diseño. En un curso con muchas pruebas podría tener sentido mayor detalle y en otro con menos, podría tener sentido analizar el éxito, que podría obtenerse con mayor anticipación. Si el objetivo es el abandono, también es importante tener en cuenta qué significa abandonar el curso. En un contexto, abandonar puede significar no conectarse a realizar actividad en mucho tiempo, pero en otro, si el estudiante se conecta, pero realiza muy pocas actividades y no está comprometido con el curso, también podría considerarse abandono. Por ello, es importante realizar un análisis de cada contexto para poder adaptar las variables dependientes a la realidad específica de cada curso.

Para concluir, en este capítulo se ha visto que las variables dependientes también influyen en el poder predictivo. Esto puede dar a un análisis más detallado del alumno y, al igual que se ha visto que puede haber diferentes perfiles con respecto a la nota y el abandono, se podrían analizar otros perfiles dependiendo de las habilidades en diferentes partes del curso. Además, teniendo en cuenta las diferencias del contexto, puede haber cursos en los que por su diseño cambien las variables dependientes (p. ej., forma en la que se define el abandono) y esto puede afectar al poder predictivo. Por ello, en la implementación de los modelos, es importante considerar cómo se define la variable dependiente y qué elementos pueden afectar a la misma. Adicionalmente, y como recomendación, puede ser bueno analizar si las predicciones pueden relacionarse para dar un valor añadido al profesor sobre la situación del alumno y sus posibles dificultades.

10

Conclusiones / Conclusions

Contenido

10.1. Versión en español	160
10.1.1. Síntesis de resultados	160
10.1.2. Aplicabilidad de los resultados	163
10.1.3. Limitaciones y trabajo futuro	165
10.2. English version	167
10.2.1. Summary of results	167
10.2.2. Applicability of the results	170
10.2.3. Limitations and future work	172

Esta tesis doctoral ha analizado diferentes aspectos relacionados con la predicción en diferentes escenarios. En particular, se ha analizado el efecto de algunas variables predictoras, el efecto de algunas variables a predecir, el análisis temporal de la predicción y el problema de la generalización de los modelos predictivos. Este último capítulo presenta una síntesis de los resultados obtenidos a lo largo de la tesis, la aplicabilidad de los mismos, así como limitaciones y trabajos futuros.

Además, de acuerdo al Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de Doctorado y el Reglamento de la Escuela de Doctorado de la Universidad Carlos III de Madrid, este capítulo aparece también escrito en inglés para cumplir los requisitos para la obtención de la Mención Internacional en el título de Doctor.

This thesis has analyzed several aspects related to prediction in several scenarios. Particularly, the effect of some predictor variables, the effect of some variables to predict, the temporal analysis of prediction, and the generalizability issue of predictive models have been analyzed. This last chapter presents a summary of the results obtained in this thesis, the applicability of them, and the limitations and possible future works.

Moreover, according to Royal Decree 99/2011, 28th January, which regulates Ph.D studies in Spain, and the Regulations of the Doctoral School at Universidad

Carlos III de Madrid, this chapter is provided in English to fulfill the requirements for the “Doctorado Internacional” distinction in the Ph.D. degree certificate.

10.1. Versión en español

10.1.1. Síntesis de resultados

A continuación se presentan los principales resultados respecto a los objetivos presentados en la tesis doctoral.

En torno al primer objetivo sobre el análisis de la necesidad del uso de los sistemas predictivos en el ámbito educativo, se identifica que, a pesar de que el nivel de adopción de los sistemas predictivos en las instituciones es bajo, los diferentes actores del aprendizaje (líderes institucionales, expertos, profesores y estudiantes) reconocen la importancia del uso de la predicción y en particular, están interesados y en particular, están interesados en sistemas que proporcionen información sobre el abandono y los alumnos en riesgo. Esto justifica en parte que los análisis predictivos de esta tesis se centren en variables relacionadas con el abandono y los resultados académicos.

Además, los análisis de las entrevistas y grupos focales también muestran que los diferentes interesados reconocen la importancia de la intervención temprana. Dado que el nivel de anticipación depende en cierta medida del grado de precisión de los modelos, se identifica también la necesidad de analizar a partir de qué momento se pueden realizar predicciones lo suficientemente buenas, de modo que sea posible realizar intervenciones tempranas.

Respecto al segundo objetivo planteado al inicio, relacionado con el análisis del efecto de las variables predictoras, en primer lugar, se comprueba que las variables relacionadas con la interacción con ejercicios tienen un muy buen poder predictivo en todos los escenarios de aprendizaje tratados. Este resultado corrobora lo expuesto en trabajos previos (p. ej., [23], [128]) sobre el alto poder predictivo de estas variables. Otro aspecto que se observa sobre estas variables es que aquellas relacionadas con la participación (qué ejercicios hace el alumno) pueden ser más relevantes para la predicción del abandono, mientras que aquellas más relacionadas con el rendimiento (p. ej., notas de los ejercicios) pueden ser más relevantes para predecir la notas. En relación con las notas, otra conclusión es que las notas sumativas previas mejoran el poder predictivo de los modelos obtenidos utilizando únicamente las notas formativas (como se observó en los escenarios 1 y 2), por lo que es conveniente combinar la parte formativa y sumativa en la elaboración de los modelos.

Otro tipo de variables que pueden ofrecer un buen poder predictivo son las variables sobre la interacción con los vídeos y la actividad con la plataforma, de acuerdo con los escenarios analizados. Este tipo de variables han sido ya exploradas en trabajos previos (p. ej., [126], [133]) y se corrobora su buen poder predictivo. Sin embargo, en cuanto a las variables de actividad en el foro, se observa que no son buenas predictoras y no influyen en la predicción en los dos escenarios considerados para su análisis (como ocurre en otros trabajos, como [131], [275]). Por tanto, puede no merecer la pena usar estas variables.

Un tipo de variables que sí puede obtener un buen predictivo (de acuerdo con los análisis del escenario 3), son las variables de SRL. En los análisis, se observa que las variables de este tipo pueden obtener un poder predictivo muy alto por sí mismas, aunque su aportación puede ser menor cuando otras variables relacionadas están presentes. Por ello, este tipo de variables debe tenerse en

consideración para los modelos predictivos. En cambio, al analizar variables sobre comportamientos, como la persistencia, eficiencia y constancia, se concluye (en base a los dos escenarios analizados) que estas variables aportan poco valor a los modelos y la mejora que proporcionan no es significativa. Únicamente podrían aportar al modelo cuando hay pocas variables disponibles y sobre todo, cuando faltan las variables sobre ejercicios con mayor poder predictivo (aunque como trabajo futuro se podrían realizar más análisis sobre las variables concretas). Por último, en cuanto a las variables predictoras, se observa, en base a dos escenarios, que al predecir pruebas parciales, generalmente se obtienen mejores resultados utilizando solo las interacciones relacionadas con esa prueba y no todas las interacciones del curso.

El siguiente objetivo de la tesis (objetivo 3) es sobre el análisis temporal para determinar cuál es el mejor momento para empezar a predecir. Para el caso de los MOOCs analizados, se observa que es posible obtener un AUC bueno (superior a 0.8) a partir del 20-40 % de la duración del curso. Este resultado es además válido tanto para cursos síncronos como asíncronos, considerando para los asíncronos la fecha de inicio de cada alumno en el MOOC y tomando los indicadores en base a dicha referencia. Para el caso de los SPOCs analizados de apoyo a cursos universitarios y los datos de primaria y secundaria, se obtienen buenas predicciones a partir del 13-35 % del curso. La única excepción es el curso de apoyo a la prueba de acceso. En este caso, dado que lo que se predice es una prueba externa y el SPOC tiene un carácter de apoyo, el poder predictivo es peor. Por ello, una conclusión es que, para contextos similares a los analizados en esta tesis, puede ser sensato tomar como referencia el 13-40 % del curso para empezar a proporcionar las predicciones. No obstante, esto podría variar para algunos contextos específicos. El aspecto positivo es que los valores indican que es posible tener predicciones precisas de forma temprana (como también se reportó en otros trabajos como [23], [24]), incluso antes de la mitad de curso, lo que podría servir para tomar ciertas actuaciones correctivas en el curso.

También, otra de las conclusiones del análisis temporal, en base al escenario 6, es que el poder predictivo de las variables puede variar a lo largo del tiempo y puede haber variables que al inicio sean más representativas y luego pierdan poder predictivo, y viceversa. Por ejemplo, para la predicción del abandono en el escenario 6, el tiempo medio por ejercicio es más relevante al inicio y luego pierde poder predictivo, mientras que el porcentaje de ejercicios intentados es la variable más relevante al final. Por ello, es relevante utilizar diferentes variables que capturen la información del alumno y analizar qué factores son más importantes en cada momento del curso.

Otro de los objetivos de esta tesis (objetivo 4) es analizar el grado de generalización de los modelos predictivos. En relación con este objetivo, se han considerado tres aproximaciones. La primera de ellas consiste en transferir un modelo de un curso a otro. Esta aproximación puede ser válida según ciertas condiciones y según los análisis realizados, el caso más favorable es cuando el curso utilizado para entrenar y el curso donde se predice tienen los mismos alumnos. Para el caso en el que el curso es el mismo, pero los estudiantes distintos (transferir a otra edición del curso), se pierde poder predictivo (como ocurre en [30]) aunque el modelo puede llegar a ser aceptable (lo que es coherente con [211], [212]). El peor caso es cuando los cursos y los estudiantes son diferentes. En este último caso, los modelos pueden no generalizar bien en la mayoría de los casos (aunque se observaron algunas excepciones en los MOOCs analizados). Teniendo en cuenta las características de los cursos donde

se realizaron los análisis, algunos factores importantes para la generalización pueden ser tener los mismos estudiantes (o con comportamientos lo más similares posibles) y tener cursos similares, lo que incluye algunos factores como la metodología, el modo de instrucción, la duración del curso, la importancia percibida del curso (si el curso es de apoyo o es obligatorio), y la relación entre el curso y la variable dependiente que se quiere medir (p. ej., si esa variable se mide en el curso o es externa).

Otra aproximación para analizar la generalización es la de utilizar modelos globales, es decir, modelos que incluyan datos de varios cursos. En los análisis realizados, se observa que es posible obtener un alto poder predictivo con estos modelos a nivel global (como también se observó en [216]), pero eso no implica que el poder predictivo sea alto en todos los cursos individuales. En particular, se identifica que en un 15-18 % de los cursos analizados se pierde un poder predictivo significativo (más de 0.1 en AUC y de 0.05 en RMSE). Mediante el uso de variables sobre la tipología del curso, el poder predictivo global puede mejorar, pero esto no resuelve el problema de la generalización y puede seguir habiendo cursos donde el modelo global no ajusta bien, de acuerdo con los resultados en el escenario 5. No obstante, dado que el modelo puede funcionar en muchos casos, esta aproximación puede ser válida, aunque si se usa, es importante verificar que el modelo funcione para el curso específico en el que se utilice.

Dado que el contexto puede tener un papel muy relevante en la generalización de los modelos, una recomendación es utilizar, en la medida de lo posible, modelos predictivos para cada curso, que es la tercera aproximación considerada. En ese caso, la generalización consiste en reutilizar y adaptar el modelo para cada contexto. Dependiendo de las características del contexto, puede ser más fácil o difícil la adaptación, pero de esta manera, puede ser más fácil conseguir la generalización. En este caso, la idea es generar modelo para cada curso y transferirlo a otras ediciones, salvo que sea posible generar un modelo a partir de las interacciones del mismo curso (por ejemplo, para pruebas parciales, se podría usar un modelo de una parte del curso para predecir en la siguiente), que sería preferible ya que se reducirían problemas relacionados con la metodología del curso.

Otro de los objetivos de la tesis (objetivo 5) es sobre el análisis del efecto de las variables a predecir. Respecto a este objetivo, los análisis también muestran que las variables a predecir pueden tener un gran impacto en los modelos y es importante analizar cómo definir específicamente cada variable en cada contexto y qué variable es más adecuada para cada contexto. Por ejemplo, el abandono puede tener un significado diferente según el escenario y en unos contextos puede ser más adecuado definirlo a partir de un periodo de actividad, mientras que en otros escenarios que no tengan una actividad continua en el tiempo, puede ser mejor utilizar otros criterios relacionados con la parte del curso que se completa. También, se observan diferencias (en base a los análisis del escenario 2) entre la predicción de la nota final y de la nota del examen final, obteniéndose un mejor resultado al predecir la nota del examen final, lo que indica que las interacciones relacionadas con cada prueba intermedia podrían ser significativas para esa prueba y el agregado de la nota final, pero no para la predicción de la nota del examen final. Del mismo modo, se identifican diferencias en los análisis entre la predicción del abandono y de la nota final, lo que permite identificar diferentes perfiles. Por ejemplo, se pueden identificar alumnos con bajo riesgo de abandonar, pero alto riesgo de suspender, lo que indica que son alumnos que siguen el curso pero tienen problemas para entender los conceptos. También,

se pueden identificar alumnos con probabilidad alta de abandono (definido como no realizar un porcentaje de actividades) aunque con una probabilidad moderada de aprobar, lo que indica que pueden ser alumnos que puedan hacer las cosas bien, pero no estén motivados a hacerlas. Estos perfiles pueden ser útiles para identificar las posibles dificultades de los estudiantes y poder proporcionar un seguimiento adaptativo y personalizado.

Por último, de forma transversal a los objetivos marcados en la tesis, se han analizado diversos algoritmos de predicción. Respecto a los algoritmos, la conclusión obtenida a lo largo de los diferentes análisis es que **RF** y **RG** son generalmente los algoritmos que mejores resultados proporcionan en los análisis, y que **DT** suele ofrecer un peor poder predictivo. Sin embargo, las diferencias entre los modelos no son generalmente muy grandes. Por tanto, se observa que las variables utilizadas en los modelos tienen un mayor impacto que los algoritmos, lo que es coherente con [188].

10.1.2. Aplicabilidad de los resultados

Esta sección tiene como objetivo presentar cuál es la aplicabilidad de todos los resultados presentados en la sección anterior.

En primer lugar, estos resultados pueden ser utilizados para la implementación de sistemas predictores. En el caso de las variables predictoras, los resultados sirven para que a la hora de diseñar un sistema se puedan elegir las variables más apropiadas en función del contexto. Por ejemplo, sabiendo que las variables del foro aportan poco poder predictivo, en un futuro sistema, el foco no se pondría en estas variables. Del mismo modo, si el contexto es limitado, se puede saber que hay ciertas variables que pueden contribuir a los modelos en ausencia de otras variables y podrían calcularse en dichos casos. También, conocer el poder predictivo de los diferentes grupos de variables puede ayudar a entender mejor las predicciones. Si al profesor se le dice qué variables son las que tienen mayor efecto en la predicción y se proporciona el valor de dichos indicadores junto con la predicción, el profesor puede entender mejor la información.

Respecto al análisis temporal, los resultados también pueden servir tanto en el diseño de los sistemas como en la interpretación. El que diseñe el sistema de predicción puede utilizar estos resultados para decidir cuándo empezar a mostrar las predicciones y no hacerlo antes para evitar proporcionar resultados que no sean lo suficientemente precisos. También, estos resultados les pueden dar una indicación al profesor de la confiabilidad del sistema, es decir, si el sistema en un momento es lo suficiente preciso o no. De esta manera, un profesor podría ser más cauto en sus actuaciones cuando las precisiones tienen menor fiabilidad que cuando sean más precisas.

Unos resultados muy relevantes para el diseño de los modelos son los relacionados con la generalización. Para entrenar los modelos se usan datos del pasado, pero para poner en marcha un sistema predictivo en cursos durante su ejecución, hay que predecir en cursos diferentes a los que se usan para el entrenamiento. Dado que el contexto puede afectar de forma significativa a los modelos predictivos, estos resultados sirven de guía para ver cuál es la mejor aproximación a elegir cuando se diseñe un modelo predictivo para conseguir que pueda funcionar con datos diferentes a los que se entrena.

Los resultados de las variables a predecir tienen una diferente aplicabilidad dependiendo de los interesados. Por un lado, a los que diseñen los sistemas predictivos les puede ser bastante útil conocer los criterios para seleccionar la variable dependiente, ya que esto tendrá repercusiones en todo el modelo y en la interpretación que puedan dar posteriormente los distintos interesados sobre los resultados. Por otro lado, de cara a los que van a ver las predicciones (profesores, alumnos, gestores, etc.), los resultados muestran que la combinación de diversas variables dependientes puede aportar información adicional a los modelos y sirve para entender mejor a los estudiantes. Por ejemplo, mostrar los resultados de la predicción del abandono combinados con la predicción de la nota puede ayudar a comprender mejor las posibles dificultades del alumno (p. ej., si un alumno tiene muy baja probabilidad de abandono pero alta de suspenso, puede ser que al alumno vaya a intentar hacer las tareas, pero con muchas dificultades en el proceso de aprendizaje).

También, combinando los diferentes resultados, es posible mejorar el diseño y la orquestación de los cursos. Para ello, un profesor podría tener en cuenta la precisión de los modelos y cuando tenga predicciones precisas, podría analizar los valores proporcionados para modificar el curso. Por ejemplo, si identifica que hay un alto porcentaje de alumnos con riesgo de suspenso, podría plantearse la creación de nuevos materiales que apoyen a los alumnos. También, en función de la combinación de diferentes predicciones (p. ej., abandono y nota final) e indicadores, se podría adaptar el aprendizaje a grupos determinados de alumnos u ofrecerles un apoyo más personalizado. Además, si se identifica que hay muchos estudiantes con un alto riesgo de abandono, se podrían adaptar las dinámicas de las clases para motivar más a los alumnos.

Asimismo, si el profesor conoce las características de los escenarios de la tesis, podría reflexionar en el diseño de los cursos para sacar un mayor potencial de las predicciones. Por ejemplo, si un profesor conoce que es posible obtener unas predicciones precisas de pruebas parciales y que las notas sumativas previas pueden ser útiles en la predicción, podría diseñar un curso con diversas pruebas para obtener predicciones no solo de lo que ocurre al final, sino lo que va ocurriendo a lo largo del curso. Estas predicciones a nivel más atómico le pueden proporcionar información adicional para adaptar el aprendizaje semana a semana.

Además, estas predicciones podrían servir para explotar mejor algunas dinámicas, como la clase invertida [303]. Para un diseño de la clase invertida, si un profesor tiene predicciones lo suficientemente precisas sobre el rendimiento de los alumnos en cada parte del curso, el profesor podría rediseñar cada sesión para sacar más rendimiento. Para una sesión en la que la predicción sea mala, el profesor debería reforzar más los conceptos básicos, antes que ponerse a cubrir conceptos más complejos. También, estas predicciones se podrían usar para detectar el rendimiento y crear grupos homogéneos en las clases presenciales a la hora de realizar actividades colaborativas [304]. Asimismo, en la línea con la relación entre variables dependientes, se podría predecir quién va a realizar las actividades de la clase invertida antes de la sesión y relacionarlo con la predicción de la nota, de tal manera que se identifique si los estudiantes tienen dificultades porque no cubran los contenidos o porque no los comprendan, y se puedan adaptar las clases en función de ellos.

En resumen, los resultados de esta tesis pueden ser utilizados por diferentes interesados. A los diseñadores de los sistemas predictivos les puede interesar conocer

los resultados para tenerlos en cuenta de los futuros sistemas, y a los profesores (o quienes usen los datos) les puede servir para comprender mejor los datos y contribuir a la mejora de los cursos con una mayor adaptación, aprendizaje personalizado, y un mejor diseño y orquestación de los cursos.

10.1.3. Limitaciones y trabajo futuro

A pesar de que en la tesis se han obtenido diversas conclusiones sobre diferentes aspectos de la predicción, hay algunas limitaciones que merece la pena destacar. En primer lugar, se destaca que los resultados obtenidos se han hecho en base al análisis de seis escenarios. Aunque se supone que estos escenarios son significativos, al cubrir cursos de diferentes etapas educativas, con diversos modos de instrucción y tipo de curso, y desarrollados en diferentes regiones (Europa, Asia y América), puede haber más contextos y puede que los resultados no sean aplicables a todos los contextos. Este hecho puede dar lugar a limitaciones específicas en relación con los diferentes aspectos tratados en la tesis.

En relación con las variables, se han indicado qué tipos de variables pueden ofrecer un mayor o menor poder predictivo. Sin embargo, estos resultados podrían cambiar en otros escenarios. Por ejemplo, las variables del foro no influyen en la predicción en los análisis realizados, pero si hubiera un curso donde la actividad del foro fuese sumativa, pudiera ser que la actividad en el foro influyese en el rendimiento de otras pruebas. Del mismo modo, la persistencia se ha analizado en actividades formativas. Si se analizase el efecto de la persistencia en actividades sumativas respecto al resultado de otras pruebas sumativas diferentes, el resultado podría variar. Por ello, aunque se utilizan diversos escenarios para obtener conclusiones en base a todos ellos, puede haber más casos posibles, y como trabajo futuro se podrían extender los análisis a nuevos casos.

Otra limitación en referencia a las variables es que, aunque en los diferentes análisis se utilicen variables similares, no todas las variables coinciden en todos los escenarios. En unos casos, esto ha sido porque el contexto no permite analizar otras variables (por ejemplo, hay plataformas que no tienen vídeos, por lo que no se puede analizar este tipo de variables) y en otros ha sido porque se ha querido focalizar sobre alguna variable en algún escenario determinado. Este hecho puede limitar el estudio, al no comprobarse todas las variables en todos los escenarios. Como trabajo futuro, se podrían analizar en concreto en más escenarios aquellas variables que se han analizado en menos contextos, como las variables de comportamiento.

En relación con este tipo de variables de comportamiento, se ha utilizado un conjunto limitado a tres variables (persistencia, eficiencia y constancia). Sobre estas variables, se podrían extender los análisis para ver si hay casos en los que podrían ser útiles en ausencia de otras variables. También, se podrían analizar más comportamientos, incluyendo comportamientos sobre la personalidad del estudiante, para ver su efecto en la predicción, si pueden tener un alto poder predictivo por sí mismos, y si pueden ser útiles en ausencia de otras variables. Además, como trabajo futuro, se podría extender el análisis del SRL para relacionar más los eventos con las estrategias de SRL y analizar el poder predictivo de cada una de ellas.

Asimismo, otra limitación relacionada con los datos es sobre el filtrado de los alumnos. En cada escenario, se ha elegido un criterio de selección de la muestra que se adaptase al estudio de cada escenario. Sin embargo, se podrían haber escogido otros criterios, que podrían haber afectado los resultados. Además, en otros escenarios

diferentes, se podrían plantear otros filtrados, por lo que un trabajo futuro podría ser el análisis de cómo afecta el filtrado en los modelos predictivos.

Aparte del filtrado de los alumnos, en muchas ocasiones, puede haber datos que lleguen erróneos de la plataforma. Por ejemplo, en algún caso, se identificó alguna discrepancia entre las notas almacenadas de los ejercicios y las notas acumuladas de la plataforma, o se identificó que algún alumno tenía un tiempo invertido en las actividades de un día superior a 24 horas, lo cual no es posible. Para los análisis, se filtraron los datos claramente erróneos, aunque se podría hacer un mayor análisis para identificar mejor estos casos, ver cómo tratarlos y cuál es su efecto en la predicción. De igual forma, es posible que en algunos cursos pueda haber estudiantes que hayan hecho trampas. Por ejemplo, en los MOOCs, Ruipérez-Valiente et al. [305] vieron que había alumnos que utilizaban dos cuentas: una para obtener las respuestas y otra para usar dichas respuestas y aprobar el curso. También, podría haber estudiantes que compartieran la misma cuenta, lo que afectaría a los modelos. Entre los trabajos futuros, en relación con estas limitaciones, sería relevante identificar valores atípicos en las trazas para ver si tienen algún efecto en la predicción. Del mismo modo, una línea futura puede centrarse en detectar qué alumnos pueden estar haciendo trampas para tratarlos de forma especial en los modelos predictivos y ver si eso mejora los modelos.

Respecto a los algoritmos, en esta tesis se han utilizado cuatro algoritmos bastante comunes en la predicción y se ha identificado que la diferencia de poder predictivo entre ellos no es muy elevada. Sin embargo, en los últimos trabajos de la literatura, se ha empezado a trabajar en nuevos modelos con redes neurales, que podrían ser más potentes. Un posible trabajo futuro sería analizar en los escenarios tratados el potencial de este tipo de algoritmos, tanto en poder predictivo como en tiempo computacional. Este último aspecto es relevante porque, aunque en la tesis no se haya puesto hincapié en él, si los modelos predictivos se usan con miles de estudiantes y se actualizan frecuentemente, es necesario que el tiempo de computación sea razonable. El análisis de estos tiempos también podría ser de interés para el futuro.

En relación con el tiempo, otra de las limitaciones relacionadas con el contexto es sobre los resultados del análisis temporal. En este caso, se ha proporcionado un rango de porcentaje del curso a partir del cual las predicciones pueden ser lo suficientemente precisas. Este rango es amplio y se ha obtenido en base a diversos escenarios heterogéneos, pero es posible que haya cursos con contextos diferentes que no ajustasen. Por ello, como trabajo futuro se podrían analizar escenarios con características muy distintas para validar este rango. También, otra de las conclusiones del análisis temporal es que el poder predictivo de las variables puede variar a lo largo del tiempo. Como trabajo futuro, podría extenderse esta análisis para entender mejor qué factores pueden hacer que unas variables tengan mayor o peor poder predictivo a lo largo del tiempo.

En cuanto a la generalización, una de las limitaciones es que se ha analizado un subconjunto posible de casos donde puede tener que producirse la generalización. Por ejemplo, en el escenario 4, los SPOCs formaban parte del mismo programa de preparación a la prueba de acceso y tenían una metodología similar. Por ello, sería relevante analizar más casos para obtener un detalle mayor sobre las condiciones necesarias para la generalización. En línea con el análisis temporal, otro aspecto que sería relevante explorar es si la generalización podría variar a lo largo del tiempo.

Por ejemplo, podría ser que un modelo fuese transferible a otro curso, pero no al inicio del mismo y el periodo necesario para que las predicciones fuesen precisas fuese mayor que cuando se usan datos del mismo curso utilizado para entrenar. Por ello, como parte del trabajo futuro se podría analizar la generalización a lo largo del tiempo.

Por último, en referencia a las variables a predecir, una de las limitaciones es que las variables, dependiendo del contexto, deben definirse de una manera u otra para que tengan el valor educativo que se requiere en cada escenario. Esto puede ser una limitación a la hora de comparar los resultados, ya que aunque el concepto de la variable sea la misma, la medida es diferente. También, como parte de los análisis, se ha analizado la relación entre algunas variables dependientes como la predicción de la nota final y la del examen final, y la predicción de la nota final y el abandono. Como trabajo futuro, sería relevante extender este análisis a nuevas variables dependientes, de modo que se pueda ver si la relación entre dichas variables puede aportar información adicional que sirva para la mejora del aprendizaje.

10.2. English version

10.2.1. Summary of results

In this section, a summary of the main results is presented in connection to the objectives of this thesis.

In relation to the first objective about the analysis of the needs of predictive systems in the educational field, it is identified that despite the fact that the level of adoption of predictive systems at institutional level is low, stakeholders (institutional leaders, experts, professors and students) recognize the importance of using predictions. Particularly, they are interested in systems that provide information about dropout and students at risk. This fact can serve to justify that the focus of the dependent variables of this thesis is on variables related to dropout and academic results.

In addition, the analyses of interviews and focus groups show that stakeholders also recognize the importance of early interventions. Given that the level of anticipation depends at some point on the accuracy of the predictive models, a need to analyze the moment when predictions are accurate enough (so they allow early interventions) is identified.

With regard to the second objective, related to the analysis of the effect of predictor variables, it is first checked that variables related to interactions with exercises can achieve a very good predictive power in the analyzed scenarios. This result corroborates findings of previous works (e.g., [23], [128]) about the predictive power of those variables. Another aspect that is observed is that variables related to the participation (e.g., how many exercises are done) can be more relevant for dropout prediction, while variables related to performance (e.g., grades of the exercises) can be more relevant to predict grades. In addition, previous summative grades can also enhance the predictive power of the models that only use formative grades (as it was observed in scenarios 1 and 2), so it can be convenient to combine formative and summative grades when developing the models.

Another kind of variables that can achieve high predictive power are variables related to interactions with videos and activity in the platform, according to the analyzed scenarios. These kinds of variables were already explored in previous

works (e.g., [126], [133]) and their good predictive power is corroborated. However, variables related to forum activity are found not to be good predictors and they do not influence predictions in the two analyzed scenarios (as it also happens in [131], [275]). Because of that, it may not worth using this kind of variables.

Another kind of variables that can achieve a high predictive power (according to the analyses in scenario 3) are the variables about SRL. In the analyses, it is observed that these variables can obtain a high predictive power by themselves, although their contribution in the models is lower when other related variables are present. Because of that, this kind of variables should be taken into consideration in the design of predictive models. In contrast, when analyzing variables related to students' behaviors, such as persistence, efficiency and constancy, it is concluded (based on the analyses of two scenarios) that these variables add little value to the models and the improvement they provide is not significant. They can only add value to the models when few variables are available and particularly, when the variables related to exercises that have higher predictive power are missing. Nevertheless, further work can be done to analyze the cases where these variables could be useful. In order to finish with the predictor variables, it is also observed (based on the results of two scenarios) that when predicting partial grades, results are generally better when they are obtained using only the interactions related to specific assessments and not all the interactions of the course.

The next objective of this thesis (objective 3) is related to the temporal analysis to determine when the best moment to start making predictions is. For the case of the analyzed MOOCs, it is observed that it is possible to achieve a good AUC (above 0.8) from 20-40% of the duration of the course. This result is valid for both synchronous and asynchronous courses, considering the date each learner starts the MOOC for asynchronous courses and taking that date as a reference to compute indicators. For the case of analyzed SPOCs used to support university courses and data from primary and secondary education, good predictions are obtained from 13-35% of the course duration. The only exception is the course to support the university admission test. In this case, given that the variable to predict is about an external assessment and the SPOC is conceived as support material, the predictive power is worse. Because of that, one conclusion is that, it could be possible to use 13-40% of the course duration as a reference of the period to start making predictions, provided that the contexts are similar to those analyzed in this thesis. Nevertheless, this threshold can vary for specific contexts. The positive aspect is that these values show that it is possible to achieve early predictions (as it was also reported in other works, such as [23], [24]), even before half of the course, and this is positive to enable possible corrective measures in the course.

Furthermore, another conclusion of the temporal analysis, based on scenario 6, is that the predictive power of the variables can vary over time and there can be variables that are more representatives at the beginning, and then they lose predictive power, and vice versa. For example, for dropout prediction in scenario 6, the average time to solve an exercise is more relevant at the beginning of the course and then, some predictive power is lost. In contrast, the percentage of attempted exercises becomes more relevant at the end of the course. Because of that, it is relevant to use different variables that capture students' information and to analyze which factors are more relevant in each part of the course.

Another objective of this thesis (objective 4) is to analyze the degree of generalizability of predictive models. In relation to this objective, three approaches were considered. The first approach consists on transferring one model from one course to another. This approach can be valid under certain conditions and, according to the analyses, the best case is when the course used for training and the course used to predict have the same students. For the case when the course is maintained, but students are different (i.e., transfer a model to another edition of the same course), some predictive power is lost (as happened in [30]), but the predictive power can be acceptable (as reported in [211], [212]). The worst case is when the courses and the students are different. In this case, models cannot generalize well in most of the cases (with some observed exceptions in the analyzed MOOCs). Considering the characteristics of the courses used for the analyses, some important aspects for the generalizability can be having the same students (or students with very similar behaviors) and having similar courses. The similarity of the courses involves the methodology, instruction mode, course duration, perceived importance of the course (i.e., whether the course is compulsory or just provides support materials), and the relationship between the course and the dependent variable that is measured (e.g., if the dependent variable is measured within the course or it is an external variable).

Another approach consists on analyzing the generalizability through global models, i.e., models which include data from several courses. In the conducted analyses, it is observed that it is possible to achieve a high predictive power at global level (as it was also reported in [216]), but that does not imply that the predictive power is high in all the individual courses. Particularly, it is identified that there is a significant loss of predictive power (a drop above 0.1 in AUC and 0.05 in RMSE) in about 15-18% of the analyzed courses. When variables related to the course type are used, the predictive power at global level can improve, but this does not solve the generalizability issue and there can still be courses which do not correctly fit to the global model, according to the analyses in scenario 5. Nevertheless, given that the model can work in many cases, this approach can be used, although it is important to verify that the model can fit to the specific courses where models are used.

Given that the context can play an important role in the generalizability of the models, one recommendation is using specific models for each course as much as possible, which is the third considered approach. In this case, the generalizability consists on reusing and adapting the models for each context. Depending on the characteristics of each context, it may be easier or harder to adapt to the model, but this way, generalizability would be easier to obtain. In this case, the idea is to generate a model for each individual course and transfer it to other editions, unless it is possible to generate a model from the interactions of the same course (e.g., for partial assessments, a model trained for one assessment could be used to predict the following one). This latter case would be preferable as problems related to course methodology would be alleviated.

Another objective of this thesis (objective 5) is about the analysis of variables to predict. In relation to this objective, analyses show that variables to predict can have a significant impact on the models and it is important to analyze how to specifically define each variable in each context and define which variables best fit in each context. For example, dropout may have a different meaning depending on

the context. In some contexts, it may be better to define it in relation to activity periods. In other contexts where students do not necessarily show a continuous activity over time, it may be better to select other criteria related to the completed part of the course. Moreover, some differences are observed (based on the analyses in scenario 2) between predictions of final grade and final exam grade. Particularly, better results are obtained when predicting the final exam, which indicates that interactions related to partial assessments can be more significant for that test and for the aggregated final grade, but not for the final exam grade. In addition, some differences were identified in the analyses between dropout prediction and grade prediction, and they allow identifying several students' profiles. For example, it is possible to identify students with low risk of dropout, but high risk of failure. That may indicate that these students are going to keep working on the course, but with difficulties to grasp the contents. Furthermore, it may be possible to detect students with high probability of dropout (defined as a lack of completion of activities) and moderate probability of failure, which may indicate that these students may be able to solve the activities correctly, they do not feel motivated to do them. These profiles can be useful to identify possible students' difficulties so as to provide a personalized and adaptative support.

Finally, apart from the objectives that were set for this thesis, some prediction algorithms were also analyzed. Regarding to these algorithms, the conclusion is that **RG** and **RF** generally provide better results in the conducted analyses, and **DT** usually provides a poorer predictive power. However, differences between the predictive models are not very high. Thus, it is observed that the variables used in the models can have a bigger impact than algorithms, which is coherent with [188].

10.2.2. Applicability of the results

This section aims to present the applicability of the results presented in the previous section.

First, the results can be used for the implementation of predictive systems. For the case of predictor variables, results can be used to select the most appropriate variables depending on the context when designing predictive systems. For example, if variables related to forum activity add little value to predictive values, the focus should not be put on those variables. In addition, if the context is limited, it can be possible to know that there can be some variables that can contribute to the models when other variables are not available, and they could be computed in those cases. Furthermore, insights about the predictive power of different groups of variables can help to better understand predictions. If instructors are said which variables have a higher effect on prediction and values of those predictions are given, instructors could better understand the information.

Regarding the temporal analyses, results can serve to both the systems design and the interpretability of predictions. Technicians who design the systems can use these results to decide when to start providing predictions so that predictions are not provided when they are not accurate enough. Moreover, these results can give instructors an idea of the reliability of the systems, i.e., whether or not the system is accurate enough. This way, instructors could be more cautious in their interventions when the predictions are not reliable enough.

Results related to the generalizability of the models can be also be very relevant for the design of the systems. When models are trained, data from the past are

used. However, when a predictive system is running, predictions have to be made on different courses from those used for training. Given that the context can considerably affect the predictive models, results of this thesis can guide future implementations to better determine what the best generalizability approach is for each case so that predictive models can work in different courses from those used for training.

With regard to the results about the variables to predict, their applicability depends on the stakeholders. On the one hand, designers of the predictive models can be interested in the criteria to select the dependent variables, as this fact can have an effect on the models and the interpretability other stakeholders can give on the results. On the other hand, for the stakeholders who visualize the predictions (e.g., teachers, students, managers, etc.), these results show that the combination of different dependent variables can provide further information about the models and that can serve to better understand students. For example, the combination of the predictions of dropout and the final grade can help to better understand students difficulties in the course (e.g., if a student has a low probability of dropout but a high probability of failure, it may mean that the student is going to attempt the tasks, but with difficulties in the learning process).

In addition, the combination of the results can serve to improve the design and orchestration of the courses. In order to do that, instructors may consider the precision of the models, and they can analyze the predictions when predictions are accurate enough to modify the course. For example, if instructors identify that many students are at risk of failure, they may create new materials to support their students. Moreover, the combination of different predictions (e.g., dropout and final grade) and indicators can serve to adapt learning in different group of students and offer personalized support. Furthermore, if many students are at risk of dropout, the dynamics of the class could be changed to motivate students.

Moreover, if instructors know how the scenarios used in this thesis are, they may reflect on the design of their courses to take advantage of the predictions. For example, if instructors know that it is possible to obtain accurate predictions on partial grades and summative grades can enhance predictions, they may design a course with different assessments to obtain not only predictions about the final grade, but also predictions about the different assessments of the course. These predictions can serve to provide further information of the learning process so as to adapt learning week by week.

Furthermore, predictions can serve to better exploit some dynamics, such as flipped classroom [303]. For the design of flipped classroom, if instructors have accurate predictions about students' performance in each part of the course, they may redesign sessions to make them more effective. For example, if predictions indicate negative results in one session, instructors may reinforce basic concepts before covering more complex ideas. In addition, predictions could be used to detect performance and create homogeneous groups of students in face-to-face classes to enhance collaborative learning [304]. What is more, regarding the relationship between dependent variables, it could be possible to predict the students that will carry out the activities before the class when using flipped classroom, and those predictions could be related to grades. This way, it could be possible to identify whether students have difficulties because they do not cover the contents before the class or because they do not understand them, and that could serve to adapt classes.

To sum up, results from this thesis can be used by different stakeholders. Designers of predictive systems can be interested in the results to consider them in future implementations of the systems, and instructors (or the primary use of data) can use them to better understand the information and contribute in the improvement of courses through a better adaptation, personalized learning, and better design and orchestration of the courses.

10.2.3. Limitations and future work

Despite the fact that many conclusions have been obtain related to different aspects of predictions, there are some limitations that are worth mentioning. First, results have been obtained based on the analyses of six scenarios. While these scenarios are significant and they cover courses from different levels, different instruction modes, kind of courses, and developed in different regions (Europe, Asia, and America), there can be more possible contexts and results may not be applicable to all possible contexts. This fact is relevant for other specific limitations related to the aspects covered in this thesis.

With regard to the variables, results indicate which variables can offer higher or lower predictive power. However, these results may change in other scenarios. For example, variables related to forum activity do not influence predictive models in the analyzed scenarios. However, if this issue had been addressed in another course where forum activity was part of summative evaluation, it may happen that forum activity influences the performance of other assessments. Similarly, persistence has been analyzed using formative activities. If the effect of persistence in summative activities was analyzed to predict other summative activities, results may change. Because of that, despite the fact several scenarios are used to gather global conclusions based on them, more scenarios could be used, and future work could focus on extending the analyses on new cases.

Another limitation related to the variables is that despite the fact that variables are similar in different scenarios, there are some differences among them. In some cases, these differences are due to the differences of the contexts (e.g., some platforms do not contain videos, so these variables cannot be retrieved). In other cases, some variables have been selected to be analyzed just in specific contexts. This fact can limit the study, as not all variables are analyzed in all scenarios. As future work, the variables that have been less analyzed could be further explored, such as variables related to students' behaviors.

In relation to variables about students' behaviors, a subset of three variables (persistence, efficiency and constancy) has been used. In this case, more analyses could be done to analyze whether or not there are situations when these variables could be useful when other variables are not present. Moreover, more students' behaviors could be analyzed, including those related to students' personality. These analyses should focus on the effect of the variables on prediction, whether or not behaviors can be good predictors by themselves, and whether or not they can be useful when other variables are not available. In addition, future work could be done to extend the analyses of SRL to relate events with SRL strategies and analyze the predictive power of each of them.

Another limitation is related to the way students are filtered. In each scenario, the sample selection criteria have been chosen so as the criteria are adapted to the scenario. However, other criteria could have been chosen and that could affect

the results. In addition, in other scenarios, other filtering criteria could be chosen. Because of that, further work could be done to analyze how filtering can affect predictive models.

Apart from the students' filtering, in many occasions, there can be errors in the traces collected from the platform. For example, in some cases, some discrepancies were identified between individual grades stored in the platform and average grades stored in the platform. Moreover, it was identified that some students had spent more than 24 hours in a day interacting with exercises, which is not possible. For the analyses, data with clear errors were removed, although further analyses could be done to better identify errors, analyze how they should be treated, and their effect on prediction. Similarly, it is possible that some students have cheated in their courses. For example, in the area of MOOCs, Ruipérez-Valiente et al. [305] observed that some students used two accounts to solve the courses. They used one account to obtain the answers and another account to pass the course. Moreover, it could be possible that several students shared the same account and that could affect the models. Among future works, in relation to these limitations, it would be relevant to identify atypical values in the traces to analyze their effect on prediction. Similarly, future work could focus on cheating detection to treat students who cheat differently in the predictive models, and to analyze how cheating can affect predictions.

Regarding the algorithms, this thesis has used four common prediction algorithms and it has been identified that the differences between them in terms of predictive power is not very high. However, recent works in the literature have started working in models using neural networks that could offer better results. Future work could be done to analyze how powerful this kind of algorithms can be in the scenarios of this thesis, both in predictive power and computational time. This last aspect can also be relevant because, although this thesis has not focused on this issue, if models are used with hundreds of students and predictions are updated frequently, it would be necessary to achieve a reasonable computational time. The analysis of computational time could be also of interest for future work.

In relation to time, another limitation related to the context is about the temporal analysis. In this case, a range of percentage of the course duration is provided as the best moment to start making accurate predictions. This range is wide and is obtained from several heterogeneous scenarios, but it is possible to have courses with different contexts that do not adjust to this range. Because of that, future work could focus on analyzing more scenarios with different characteristics to validate this range. Moreover, another conclusion of the temporal analysis is that the predictive power of variables can vary over time. As future work, more analyses could be done to better understand which factors can affect the evolution of the predictive power of variables over time.

As for the generalizability, one of the limitations is that this issue has been analyzed in a subset of possible cases where generalizability could be needed. For example, all the SPOCs from scenario 4 were part of the same program to prepare the admission test and they all shared a similar methodology. Because of that, it would be relevant to analyze more cases to obtain further details about the conditions needed to achieve the generalizability of predictive models. In relation to the temporal analysis, another relevant aspect for future work would be exploring whether or not generalizability could vary over time. For example, it may happen that one model can be transferred to another course, but not at the beginning of the

course. Moreover, the period needed to achieve accurate predictions can be longer than the period obtained when using data from the course used for training. Because of that, further work could be done to analyze the generalizability of predictive models over time.

Finally, in relation to the variables to predict, one limitation is that these variables are defined differently depending on the context so as they have the appropriate educational meaning in each scenario. However, this can limit the comparison of results because although the dependent variables are conceptually the same, their measurement is different. Furthermore, as part of the analyses, the relationship between some variables, such as the predictions of dropout and final grade, and the predictions of final exam grade and final grade, have been analyzed. As future work, these analyses could be extended to new dependent variables to discover if the relationship between other variables can also add additional information that could serve to improve learning.

Referencias

- [1] R. S. Baker and P. S. Inventado, “Educational Data Mining and Learning Analytics,” in *Learning Analytics*, J. A. Larusson and B. White, Eds. New York, NY, USA: Springer, 2014, pp. 61–75.
- [2] G. Siemens, D. Gasevic, C. Haythornthwaite, S. Dawson, S. B. Shum, R. Ferguson, E. Duval, K. Verbert, and R. S. J. d. Baker, “Open Learning Analytics: an integrated & modularized platform,” *Society for Learning Analytics Research*, 2011. [Online]. Available: <https://www.solaresearch.org/wp-content/uploads/2011/12/OpenLearningAnalytics.pdf>.
- [3] A. M. Kaplan and M. Haenlein, “Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster,” *Business Horizons*, vol. 59, no. 4, pp. 441–450, Jul./Aug. 2016.
- [4] A. Fox, “From MOOCs to SPOCs,” *Communications of the ACM*, vol. 56, no. 12, pp. 38–40, Dec. 2013.
- [5] J. Bergmann and A. Sams, *Flip your classroom: Reach every student in every class every day*, 1st ed. Eugene, OR, USA: International Society for Technology in Education, 2012.
- [6] R. Nkambou, R. Mizoguchi, and J. Bourdeau, *Advances in Intelligent Tutoring Systems*, 1st ed. Berlin ,Germany: Springer-Verlag, 2010, vol. 308.
- [7] N. A. Adzharuddin and L. H. Ling, “Learning management system (LMS) among university students: Does it work,” *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 3, no. 3, pp. 248–252, Jun. 2013.
- [8] H. Lasi, P. Fettke, H. Kemper, T. Feld, and M. Hoffmann, “Industry 4.0,” *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, Jun. 2014.
- [9] C. Delgado Kloos, C. Alario-Hoyos, P. J. Muñoz-Merino, M.-B. Ibáñez, I. Estévez-Ayres, and R. M. Crespo-García, “What Can You Do with Educational Technology that is Getting More Human?” in *Proc. 2019 IEEE Global Engineering Education Conference*, Dubai, United Arab Emirates, Apr. 2019, pp. 1480–1487.
- [10] F. Martin-Sanchez and K. Verspoor, “Big Data in Medicine Is Driving Big Changes,” *Yearbook of Medical Informatics*, vol. 23, no. 1, pp. 14–20, Aug. 2014.
- [11] D. Enke and S. Thawornwong, “The use of data mining and neural networks for forecasting stock market returns,” *Expert Systems with Applications*, vol. 29, no. 4, pp. 927–940, Nov. 2005.
- [12] P. Mikulecký, “Smart Environments for Smart Learning,” in *Proc. 9th International Scientific Conference on Distance Learning in Applied Informatics*, Štúrovo, Slovakia, May 2012, pp. 213–222.

- [13] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos, “Learning analytics dashboard applications,” *American Behavioral Scientist*, vol. 57, no. 10, pp. 1500–1509, Feb. 2013.
- [14] L. P. Prieto, M. Holenko Dlab, I. Gutiérrez, M. Abdulwahed, and W. Balid, “Orchestrating technology enhanced learning: a literature review and a conceptual framework,” *International Journal of Technology Enhanced Learning*, vol. 3, no. 6, pp. 583–598, Jan. 2011.
- [15] R. F. Kizilcec and G. L. Cohen, “Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 17, pp. 4348–4353, Apr. 2017.
- [16] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, and C. Delgado Kloos, “Prediction in MOOCs: A review and future research directions,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 3, pp. 384–401, Jul./Sep. 2019.
- [17] M. L. Bote-Lorenzo and E. Gómez-Sánchez, “Predicting the decrease of engagement indicators in a MOOC,” in *Proc. 7th International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, Mar. 2017, pp. 143–147.
- [18] G. Chen, D. Davis, C. Hauff, and G.-J. Houben, “On the Impact of Personality in Massive Open Online Learning,” in *Proc. 2016 Conference on User Modeling Adaptation and Personalization*, Halifax, NS, Canada, Jul. 2016, pp. 121–130.
- [19] D. Ifenthaler, “Are higher education institutions prepared for learning analytics?” *TechTrends*, vol. 61, no. 4, pp. 366–371, Jul. 2017.
- [20] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai, “Learning analytics methods, benefits, and challenges in higher education: A systematic literature review.” *Online Learning*, vol. 20, no. 2, pp. 13–29, Jun. 2016.
- [21] A. Pardo, “Designing learning analytics experiences,” in *Learning analytics*, J. A. Larusson and B. White, Eds. New York, NY, USA: Springer, 2014, pp. 15–38.
- [22] M. A. Chatti, V. Lukarov, H. Thüs, A. Muslim, A. M. F. Yousef, U. Wahid, C. Greven, A. Chakrabarti, and U. Schroeder, “Learning Analytics: Challenges and Future Research Directions,” *eleed*, vol. 10, no. 1, Dec. 2014.
- [23] J. A. Ruipérez-Valiente, R. Cobos, P. J. Muñoz-Merino, Á. Andujar, and C. Delgado Kloos, “Early Prediction and Variable Importance of Certificate Accomplishment in a MOOC,” in *Proc. 5th European MOOCs Stakeholders Summit*, Madrid, Spain, May 2017, pp. 263–272.
- [24] W. Xing and D. Du, “Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention,” *Journal of Educational Computing Research*, vol. 57, no. 3, pp. 547–570, Jun. 2019.

- [25] J. Figueroa-Cañas and T. Sancho-Vinuesa, “Predicting early dropout students is a matter of checking completed quizzes: the case of an online statistics module,” in *Proc. Learning Analytics Summer Institute Spain 2019*, Vigo, Spain, Jun. 2019, pp. 100–111.
- [26] S. Halawa, D. Greene, and J. Mitchell, “Dropout Prediction in MOOCs using Learner Activity Features,” in *Proc. 2nd European MOOCs Stakeholders Summit*, Lausanne, Switzerland, Feb. 2014, pp. 58–65.
- [27] A. Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis, and H. Rangwala, “Predicting Student Performance Using Personalized Analytics,” *Computer*, vol. 49, no. 4, pp. 61–69, Apr. 2016.
- [28] Z. Ren, H. Rangwala, and A. Johri, “Predicting Performance on MOOC Assessments Using Multi-Regression Models,” in *Proc. 9th International Conference on Educational Data Mining*, Raleigh, NC, USA, Jul. 2016, pp. 484–489.
- [29] P. M. Moreno-Marcos, T. De Laet, P. J. Muñoz-Merino, C. Van Soom, T. Broos, K. Verbert, and C. Delgado Kloos, “Generalizing Predictive Models of Admission Test Success Based on Online Interactions,” *Sustainability*, vol. 11, no. 18, 4940, Sep. 2019.
- [30] S. Boyer and K. Veeramachaneni, “Transfer Learning for Predictive Models in Massive Open Online Courses,” in *Proc. 17th International Conference on Artificial Intelligence in Education*, Madrid, Spain, Jun. 2015, pp. 54–63.
- [31] G. Siemens and R. S. d. Baker, “Learning analytics and educational data mining: towards communication and collaboration,” in *Proc. 2nd International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, Apr. 2012, pp. 252–254.
- [32] Z. Papamitsiou and A. A. Economides, “Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence,” *Journal of Educational Technology & Society*, vol. 17, no. 4, pp. 49–64, Oct. 2014.
- [33] C. Romero and S. Ventura, “Data mining in education,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, Jan. 2013.
- [34] T. Elias, “Learning Analytics: Definitions, processes and potential,” 2011. [Online]. Available: https://landing.athabasca.ca/mod/file/download.php?file_guid=43709.
- [35] C. Steiner, M. Kickmeier-Rust, and M. Türker, “Review Article About LA and EDM Approaches,” TU Graz, Tech. Rep. Deliverable D3.1, 2014. [Online]. Available: <http://leas-box.cognitive-science.at/downloads/D3.1.pdf>.
- [36] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, “A reference model for learning analytics,” *International Journal of Technology Enhanced Learning*, vol. 4, no. 5-6, pp. 318–331, Feb. 2013.

- [37] W. Greller and H. Drachsler, “Translating Learning into Numbers: A Generic Framework for Learning Analytics,” *Journal of Educational Technology & Society*, vol. 15, no. 3, pp. 42–57, Jul. 2012.
- [38] H. Drachsler, T. Bogers, R. Vuorikari, K. Verbert, E. Duval, N. Manouselis, G. Beham, S. Lindstaedt, H. Stern, M. Friedrich, and M. Wolpers, “Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning,” *Procedia Computer Science*, vol. 1, no. 2, pp. 2849–2858, Sep. 2010.
- [39] M. Khalil and M. Ebner, “What is learning analytics about? A survey of different methods used in 2013-2015,” in *Proc. 2016 Conference on Smart Learning*, Dubai, United Arab Emirates, Mar. 2016, pp. 294–304.
- [40] C. Conati, A. Gertner, and K. Vanlehn, “Using Bayesian Networks to Manage Uncertainty in Student Modeling,” *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 371–417, Nov. 2002.
- [41] R. Burke, “Hybrid recommender systems: Survey and experiments,” *User modeling and user-adapted interaction*, vol. 12, no. 4, pp. 331–370, Nov. 2002.
- [42] R. Bodily and K. Verbert, “Review of research on student-facing learning analytics dashboards and educational recommender systems,” *IEEE Transactions on Learning Technologies*, vol. 10, no. 4, pp. 405–418, Oct./Dec. 2017.
- [43] C.-M. Chen, “Intelligent web-based learning system with personalized learning path guidance,” *Computers & Education*, vol. 51, no. 2, pp. 787–814, Sep. 2008.
- [44] C. Shi, S. Fu, Q. Chen, and H. Qu, “VisMOOC: Visualizing video clickstream data from massive open online courses,” in *Proc. 2015 IEEE Pacific Visualization Symposium*, Hangzhou, China, Apr. 2015, pp. 159–166.
- [45] J. A. Ruiperez-Valiente, P. J. Munoz-Merino, J. A. Gascon-Pinedo, and C. Delgado Kloos, “Scaling to Massiveness With ANALYSE: A Learning Analytics Tool for Open edX,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 909–914, Dec. 2017.
- [46] B. A. Schwendimann *et al.*, “Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research,” *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 30–41, Jan./Mar. 2017.
- [47] P. J. Muñoz-Merino, J. A. Ruipérez Valiente, and C. Delgado Kloos, “Inferring higher level learning information from low level data for the Khan Academy platform,” in *Proc. 3rd International Conference on Learning Analytics and Knowledge*, Leuven, Belgium, Apr. 2013, pp. 112–116.
- [48] P. Blikstein, “Multimodal learning analytics,” in *Proc. 3rd International Conference on Learning Analytics and Knowledge*, Leuven, Belgium, Apr. 2013, pp. 102–106.
- [49] R. Martinez-Maldonado, T. Power, C. Hayes, A. Abdiprano, T. Vo, C. Axisa, and S. Buckingham Shum, “Analytics meet patient manikins: challenges

- in an authentic small-group healthcare simulation classroom,” in *Proc. 7th International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, Mar. 2017, pp. 90–94.
- [50] S. Crossley, L. Paquette, M. Dascalu, D. S. McNamara, and R. S. Baker, “Combining click-stream data with NLP tools to better understand MOOC completion,” in *Proc. 6th International Conference on Learning Analytics and Knowledge*, Edinburgh, UK, Apr. 2016, pp. 6–14.
- [51] R. Ferguson, “Learning analytics: drivers, developments and challenges,” *International Journal of Technology Enhanced Learning*, vol. 4, no. 5/6, pp. 304–317, Jan. 2012.
- [52] A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth, “Unsupervised modeling for understanding MOOC discussion forums: a learning analytics approach,” in *Proc. 5th International Conference on Learning Analytics and Knowledge*, Poughkeepsie, NY, USA, Mar. 2015, pp. 146–150.
- [53] K. M. Kapp, *The gamification of learning and instruction: game-based methods and strategies for training and education*, 1st ed. Hoboken, NJ, USA: John Wiley & Sons, 2012.
- [54] C. Holman, S. Aguilar, and B. Fishman, “Gradecraft: What can we learn from a game-inspired learning management system?” in *Proc. 3rd International Conference on Learning Analytics and Knowledge*, Leuven, Belgium, Apr. 2013, pp. 260–264.
- [55] Á. Serrano-Laguna, J. Torrente, P. Moreno-Ger, and B. Fernández-Manjón, “Application of Learning Analytics in educational videogames,” *Entertainment Computing*, vol. 5, no. 4, pp. 313–322, Dec. 2014.
- [56] C. Alonso-Fernández, I. Martínez-Ortiz, R. Caballero, M. Freire, and B. Fernández-Manjón, “Predicting students’ knowledge after playing a serious game based on learning analytics data: A case study,” *Journal of Computer Assisted Learning*, vol. 36, no. 3, pp. 350–358, Jun. 2020.
- [57] S. Joksimović, V. Kovanović, and S. Dawson, “The Journey of Learning Analytics,” *HERDSA Review of Higher Education*, vol. 6, pp. 27–63, Jul. 2019.
- [58] C. Romero and S. Ventura, “Guest editorial: Special issue on early prediction and supporting of learning performance,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 145–147, Apr./Jun. 2019.
- [59] A. Ahmed and I. S. Elaraby, “Data Mining: A prediction for Student’s Performance Using Classification Method,” *World Journal of Computer Application and Technology*, vol. 2, no. 2, pp. 43–47, Feb. 2014.
- [60] H. S. Nwana, “Intelligent tutoring systems: an overview,” *Artificial Intelligence Review*, vol. 4, no. 4, pp. 251–277, Dec. 1990.
- [61] T. Murray, “Authoring intelligent tutoring systems: An analysis of the state of the art,” *Journal of Artificial Intelligence in Education*, vol. 10, pp. 98–129, Jan. 1999.

- [62] M. Virvou and C. Troussas, “Knowledge-based authoring tool for tutoring multiple languages,” in *Proc. 4th International Conference on Intelligent Interactive Multimedia Systems and Services*, Piraeus-Athens, Greece, Jul. 2011, pp. 163–175.
- [63] P. Albacete, P. Jordan, D. Lusetich, I. A. Chounta, S. Katz, and B. M. McLaren, “Providing Proactive Scaffolding During Tutorial Dialogue Using Guidance from Student Model Predictions,” in *Proc. 19th International Conference on Artificial Intelligence in Education*, London, United Kingdom, Jun. 2018, pp. 20–25.
- [64] T. Käser, S. Klingler, A. G. Schwing, and M. Gross, “Dynamic Bayesian Networks for Student Modeling,” *IEEE Transactions on Learning Technologies*, vol. 10, no. 4, pp. 450–462, Oct./Dec. 2017.
- [65] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, “Individualized Bayesian Knowledge Tracing Models,” in *Proc. 16th International Conference on Artificial Intelligence in Education*, Memphis, TN, USA, Jul. 2013, pp. 171–180.
- [66] R. S. d Baker, Z. A. Pardos, S. M. Gowda, B. B. Nooraei, and N. T. Heffernan, “Ensembling Predictions of Student Knowledge within Intelligent Tutoring Systems,” in *Proc. 19th International Conference on User Modeling, Adaptation, and Personalization*, Girona, Spain, Jul. 2011, pp. 13–24.
- [67] W. Canfield, “ALEKS: A Web-based intelligent tutoring system,” *Mathematics and Computer Education*, vol. 35, no. 2, pp. 152–158, Apr. 2001.
- [68] M. J. Timms, “Using Item Response Theory (IRT) to select hints in an ITS,” in *Proc. 13th International Conference on Artificial Intelligence in Education*, vol. 158, Los Angeles, CA, USA, Jun. 2007, pp. 213–221.
- [69] I. Roll, V. Alevan, B. M. McLaren, and K. R. Koedinger, “Improving students’ help-seeking skills using metacognitive feedback in an intelligent tutoring system,” *Learning and instruction*, vol. 21, no. 2, pp. 267–280, Apr. 2011.
- [70] M. C. Desmarais and R. S. d Baker, “A review of recent advances in learner and skill modeling in intelligent learning environments,” *User Modeling and User-Adapted Interaction*, vol. 22, no. 1-2, pp. 9–38, Apr. 2012.
- [71] S. Crossley, S. Karumbaiah, J. Ocumpaugh, M. J. Labrum, and R. S. Baker, “Predicting Math Success in an Online Tutoring System Using Language Data and Click-Stream Variables: A Longitudinal Analysis,” in *Proc. 2nd Conference on Language, Data and Knowledge*, Dagstuhl, Germany, May 2019, pp. 25:1–25:13.
- [72] C. Romero, R. Cerezo, A. Bogarín, and M. Sánchez-Santillán, “Educational Process Mining: A Tutorial and Case Study using Moodle Data Sets,” in *Data Mining and Learning Analytics: Applications in Educational Research*, S. ElAtia, D. Ipperciel, and O. R. Zaïane, Eds. Hoboken, NJ, USA: Wiley Online Library, 2016, pp. 1–28.

- [73] R. J. Quinn and G. Gray, “Prediction of student academic performance using Moodle data from a Further Education setting,” *Irish Journal of Technology Enhanced Learning*, vol. 5, no. 1, 2020.
- [74] M. H. Baturay, “An overview of the world of MOOCs,” *Procedia-Social and Behavioral Sciences*, vol. 174, no. 12, pp. 427–433, Feb. 2015.
- [75] G. Christensen, A. Steinmetz, B. Alcorn, A. Bennett, D. Woods, and E. J. Emanuel, “The MOOC phenomenon: Who takes massive open online courses and why?” *SSRN*, 2013. [Online]. Available: <https://ssrn.com/abstract=2350964>.
- [76] G. Siemens, “MOOCs are really a platform,” *Elearnspace: Learning, Networks, Knowledge, Technology, Community*, 2012. [Online]. Available: <http://www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform/>.
- [77] M. Kesim and H. Altınpulluk, “A theoretical analysis of MOOCs types from a perspective of learning theories,” in *Proc. of 5th World Conference on Learning, Teaching and Educational Leadership*, Prague, Czech Republic, Oct. 2014, pp. 15–19.
- [78] S. Kocdar, A. Karadeniz, A. Bozkurt, and K. Buyuk, “Measuring Self-Regulation in Self-Paced Open and Distance Learning Environments,” *The International Review of Research in Open and Distributed Learning*, vol. 19, no. 1, pp. 25–43, Feb. 2018.
- [79] B. Grainger, “Massive Open Online Course (MOOC) Report 2013,” *University of London*, 2013. [Online]. Available: <http://www.practiqa.pl/downloads/mooc/BR1-MOOC-report-2013.pdf>.
- [80] P. M. Moreno-Marcos, P. J. Muñoz-Merino, C. Alario-Hoyos, and C. Delgado Kloos, “Analyzing Students’ Persistence using an Event-Based Model,” in *Proc. Learning Analytics Summer Institute Spain 2019*, Vigo, Spain, Jun. 2019, pp. 56–70.
- [81] M. Munoz Organero, C. Delgado Kloos, and P. Munoz Merino, “Personalized service-oriented e-learning environments,” *IEEE Internet Computing*, vol. 14, no. 2, pp. 62–67, Mar./Apr. 2010.
- [82] D. Dagger, A. O’Connor, S. Lawless, E. Walsh, and V. P. Wade, “Service-oriented e-learning platforms: From monolithic systems to flexible services,” *IEEE internet computing*, vol. 11, no. 3, pp. 28–35, May/Jun. 2007.
- [83] J. M. Spector, “Conceptualizing the emerging field of smart learning environments,” *Smart learning environments*, vol. 1, no. 2, Oct. 2014.
- [84] G.-J. Hwang, “Definition, framework and research issues of smart learning environments—a context-aware ubiquitous learning perspective,” *Smart Learning Environments*, vol. 1, no. 4, Nov. 2014.
- [85] A. Carruana Martín, C. Alario-Hoyos, and C. Delgado Kloos, “Smart education: A review and future research directions,” in *Proc. 13th International*

Conference on Ubiquitous Computing and Ambient Intelligence, vol. 31, no. 1, 57, Toledo, Spain, Dec. 2019.

- [86] S. J. Yang, T. Okamoto, and S.-S. Tseng, “Context-aware and ubiquitous learning (Guest editorial),” *Journal of Educational Technology & Society*, vol. 11, no. 2, pp. 1–2, Apr. 2008.
- [87] A. Downes *et al.*, “xAPI Technical Specification,” *Advanced Distributed Learning*, 2014. [Online]. Available: <https://github.com/adlnet/xAPI-Spec>.
- [88] R. Ball *et al.*, “Advanced Distributed Learning Sharable Content Object Reference Model Version 1.2 The SCORM Content Aggregation Model,” *Advanced Distributed Learning*, 2001. [Online]. Available: http://www.vsscorm.net/docs/SCORM_1.2.CAM.pdf.
- [89] A. Whyte, V. Haag, L. Feng, M. Gylling, M. Ashbourne, W. LaMarche, and E. Pelaprat, “Caliper Analytics® Specification, version 1.1,” *IMS Global Learning Consortium, Inc.*, 2018. [Online]. Available: <https://www.imsglobal.org/sites/default/files/caliper/v1p1/caliper-spec-v1p1/caliper-spec-v1p1.html>.
- [90] P. Ferrer Ocaña, S. Arjona, F. Busquets, S. Valldeoriola, M. Grau, I. Cardona Subiela, and A. Ribas, “Marsupial,” *Ministry of Education of the Government of Catalonia*, 2014. [Online]. Available: <https://github.com/projectestac/marsupial>.
- [91] P. M. Moreno-Marcos, P. J. Muñoz-Merino, J. Maldonado-Mahauad, M. Pérez-Sanagustín, C. Alario-Hoyos, and C. Delgado Kloos, “Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs,” *Computers & Education*, vol. 145, 103728, Feb. 2020.
- [92] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, “Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses,” *Computers & education*, vol. 104, pp. 18–33, Jan. 2017.
- [93] M. Boekaerts, “Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students,” *Learning and instruction*, vol. 7, no. 2, pp. 161–186, Jun. 1997.
- [94] N. Michinov, S. Brunot, O. Le Bohec, J. Juhel, and M. Delaval, “Procrastination, participation, and performance in online learning environments,” *Computers & Education*, vol. 56, no. 1, pp. 243–252, Jan. 2011.
- [95] S. P. Lajoie and R. Azevedo, “Teaching and Learning in Technology-Rich Environments,” in *Handbook of educational psychology*, P. Alexander and P. Winne, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers, 2006, pp. 803–821.
- [96] J. Wong, M. Baars, D. Davis, T. Van Der Zee, G.-J. Houben, and F. Paas, “Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review,” *International Journal of Human-Computer Interaction*, vol. 35, no. 4-5, pp. 356–373, Mar. 2019.

- [97] R. F. Kizilcec and E. Schneider, “Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale,” *ACM Transactions on Computer-Human Interaction*, vol. 22, no. 2, pp. 1–24, Mar. 2015.
- [98] S. Zheng, M. B. Rosson, P. C. Shih, and J. M. Carroll, “Understanding student motivation, behaviors and perceptions in MOOCs,” in *Proc. 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, Vancouver, BC, Canada, Feb. 2015, pp. 1882–1895.
- [99] J. Broadbent, “Comparing online and blended learner’s self-regulated learning strategies and academic performance,” *The Internet and Higher Education*, vol. 33, pp. 24–32, Apr. 2017.
- [100] M. Richardson, C. Abraham, and R. Bond, “Psychological correlates of university students’ academic performance: A systematic review and meta-analysis,” *Psychological bulletin*, vol. 138, no. 2, pp. 353–387, Mar. 2012.
- [101] N. Hood, A. Littlejohn, and C. Milligan, “Context counts: How learners’ contexts influence learning in a MOOC,” *Computers & Education*, vol. 91, pp. 83–91, Dec. 2015.
- [102] Z. Papamitsiou, A. A. Economides, I. O. Pappas, and M. N. Giannakos, “Explaining learning performance using response-time, self-regulation and satisfaction from content: an fsQCA approach,” in *Proc. 8th International Conference on Learning Analytics and Knowledge*, Sydney, NSW, Australia, Mar. 2018, pp. 181–190.
- [103] D. Tempelaar, B. Rienties, and Q. Nguyen, “Investigating learning strategies in a dispositional learning analytics context: the case of worked examples,” in *Proc. 8th International Conference on Learning Analytics and Knowledge*, Sydney, NSW, Australia, Mar. 2018, pp. 201–205.
- [104] Y. Lee, J. Choi, and T. Kim, “Discriminating factors between completers of and dropouts from online learning courses,” *British Journal of Educational Technology*, vol. 44, no. 2, pp. 328–337, Mar. 2013.
- [105] Z. Sun, K. Xie, and L. H. Anderman, “The role of self-regulated learning in students’ success in flipped undergraduate math courses,” *The Internet and Higher Education*, vol. 36, pp. 41–53, Jan. 2018.
- [106] A. Bandura, “Comments on the crusade against the causal efficacy of human thought,” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 26, no. 3, pp. 179–190, Sep. 1995.
- [107] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, “Recommending self-regulated learning strategies does not improve performance in a MOOC,” in *Proc. 3rd ACM Conference on Learning@ Scale*, Edinburgh, UK, Apr. 2016, pp. 101–104.
- [108] P. M. Moreno-Marcos, P. J. Muñoz-Merino, C. Alario-Hoyos, and C. Delgado Kloos, “Re-Defining, Analyzing and Predicting Persistence Using Student Events in Online Learning,” *Applied Sciences*, vol. 10, no. 5, 1722, Mar. 2020.

- [109] D. Lufi and A. Cohen, “A Scale for Measuring Persistence in Children,” *Journal of personality assessment*, vol. 51, no. 2, pp. 178–185, Feb. 1987.
- [110] A. Silvervarg, M. Haake, and A. Gulz, “Perseverance Is crucial for learning. “OK! but Can I take a break?”,” in *Proc. 19th International Conference on Artificial Intelligence in Education*, London, UK, Jun. 2018, pp. 532–544.
- [111] D. S. Eley, J. Leung, B. A. Hong, K. M. Cloninger, and C. R. Cloninger, “Identifying the Dominant Personality Profiles in Medical Students: Implications for Their Well-Being and Resilience,” *PLoS One*, vol. 11, no. 8, Aug. 2016.
- [112] R. Scherer and J.-E. Gustafsson, “The relations among openness, perseverance, and performance in creative problem solving: A substantive-methodological approach,” *Thinking Skills and Creativity*, vol. 18, pp. 4–17, Dec. 2015.
- [113] M. Ventura, V. Shute, and W. Zhao, “The relationship between video game use and a performance-based measure of persistence,” *Computers & Education*, vol. 60, no. 1, pp. 52–58, Jan. 2013.
- [114] P. M. Moreno-Marcos, D. Martínez de la Torre, G. González Castro, P. J. Muñoz-Merino, and C. Delgado Kloos, “Should we consider Efficiency and Constancy for Adaptation in Intelligent Tutoring Systems?” in *Proc. 16th International Conference on Intelligent Tutoring Systems*, Athens, Greece, Jun. 2020.
- [115] S. R. Hiltz and J. Arbaugh, “Improving quantitative research methods in studies of asynchronous learning networks (ALN),” in *Elements of Quality Online Education, Practice and Direction*, J. Bourne and J. C. Moore, Eds. Needham, MA, USA: The Sloan Consortium, 2003, vol. 4, pp. 59–72.
- [116] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, D. Leony, and C. Delgado Kloos, “ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform,” *Computers in Human Behavior*, vol. 47, pp. 139–148, Jun. 2015.
- [117] Z. A. Pardos, S. Tang, D. Davis, and C. V. Le, “Enabling real-time adaptivity in MOOCs with a personalized next-step recommendation framework,” in *Proc. 4th ACM Conference on Learning@ Scale*, Cambridge, MA, USA, Apr. 2017, pp. 23–32.
- [118] D. Jaramillo-Morillo, M. Solarte Sarasty, G. Ramírez González, and M. Pérez-Sanagustín, “Follow-Up of Learning Activities in Open edX: A Case Study at the University of Cauca,” in *Proc. 5th European MOOCs Stakeholders Summit*, Madrid, Spain, May 2017, pp. 217–222.
- [119] I. Lykourantzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, “Dropout prediction in e-learning courses through the combination of machine learning techniques,” *Computers & Education*, vol. 53, no. 3, pp. 950–965, Nov. 2009.

- [120] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, “Machine learning approaches to predict learning outcomes in Massive open online courses,” in *Proc. 2017 International Joint Conference on Neural Networks*, Anchorage, AK, USA, Jul. 2017, pp. 713–720.
- [121] J. A. Greene, C. A. Oswald, and J. Pomerantz, “Predictors of Retention and Achievement in a Massive Open Online Course,” *American Educational Research Journal*, vol. 52, no. 5, pp. 925–955, Oct. 2015.
- [122] C. Brooks, C. Thompson, and S. Teasley, “Who You Are or What You Do: Comparing the Predictive Power of Demographics vs. Activity Patterns in Massive Open Online Courses (MOOCs),” in *Proc. 2nd ACM Conference on Learning@ Scale*, Vancouver, BC, Canada, Mar. 2015, pp. 245–248.
- [123] S. Ritter, A. Joshi, S. Fancsali, and T. Nixon, “Predicting Standardized Test Scores from Cognitive Tutor Interactions,” in *Proc. 6th International Conference on Educational Data Mining*, Memphis, TN, USA, Jul. 2013, pp. 169–176.
- [124] C. G. Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, “Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance,” *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3677–3692, Jul. 2016.
- [125] T.-Y. Yang, C. G. Brinton, C. Joe-Wong, and M. Chiang, “Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 716–728, Aug. 2017.
- [126] J. Ran, G. Zhang, T. Zheng, and W. Wang, “Logistic Regression Analysis on Learning Behavior and Learning Effect Based on SPOC Data,” in *Proc. 13th International Conference on Computer Science & Education*, Colombo, Sri Lanka, Aug. 2018, pp. 330–334.
- [127] M. M. Ashenafi, G. Riccardi, and M. Ronchetti, “Predicting students’ final exam scores from their course activities,” in *Proc. 2015 IEEE Frontiers in Education Conference*, El Paso, TX, USA, Dec. 2015, pp. 1–9.
- [128] M. Feng, J. Beck, N. Heffernan, and K. Koedinger, “Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standardized Test?” in *Proc. 1st International Conference on Educational Data Mining*, Montréal, QC, Canada, Jun. 2008, pp. 107–116.
- [129] M. Klüsener and A. Fortenbacher, “Predicting students’ success based on forum activities in MOOCs,” in *Proc. 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, vol. 2, Warsaw, Poland, Sep. 2015, pp. 925–928.
- [130] J. Arguello and K. Shaffer, “Predicting speech acts in MOOC forum posts,” in *Proc. 9th International AAAI Conference on Web and Social Media*, Oxford, UK, May 2015, pp. 2–11.

- [131] R. Cobos and V. Macías Palla, “edX-MAS: Model Analyzer System,” in *Proc. 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*, Cádiz, Spain, Oct. 2017, pp. 1–7.
- [132] J. Liang, C. Li, and L. Zheng, “Machine learning application in MOOCs: Dropout prediction,” in *Proc. 11th International Conference on Computer Science & Education*, Nagoya, Japan, Aug. 2016, pp. 52–57.
- [133] A. Alamri, M. Alshehri, A. Cristea, F. D. Pereira, E. Oliveira, L. Shi, and C. Stewart, “Predicting MOOCs dropout using only two easily obtainable features from the first week’s activities,” in *Proc. 15th International Conference on Intelligent Tutoring Systems*, Kingston, Jamaica, Jun. 2019, pp. 163–173.
- [134] H. Wan, K. Liu, Q. Yu, and X. Gao, “Pedagogical Intervention Practices: Improving Learning Engagement Based on Early Prediction,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 278–289, Apr./Jun. 2019.
- [135] J. Xie, A. Essa, S. Mojarad, R. S. Baker, K. Shubeck, and X. Hu, “Student Learning Strategies and Behaviors to Predict Success in an Online Adaptive Mathematics Tutoring System,” in *Proc. 10th International Conference on Educational Data Mining*, Wuhan, China, Jun. 2017, pp. 460–465.
- [136] M. O. Z. San Pedro, J. Ocumpaugh, R. S. Baker, and N. T. Heffernan, “Predicting STEM and Non-STEM College Major Enrollment from Middle School Interaction with Mathematics Educational Software,” in *Proc. 7th International Conference on Educational Data Mining*, London, UK, Jul. 2014, pp. 276–279.
- [137] V. Matzavela, K. Chrysafiadi, and E. Alepis, “Questionnaires and artificial neural networks: A literature review on modern techniques in education,” in *Proc. 2017 IEEE Global Engineering Education Conference*, Athens, Greece, Apr. 2017, pp. 1700–1704.
- [138] S.-H. Zhong, Y. Li, Y. Liu, and Z. Wang, “A computational investigation of learning behaviors in MOOCs,” *Computer Applications in Engineering Education*, vol. 25, no. 5, pp. 693–705, May 2017.
- [139] N. M. Hicks *et al.*, “Integrating analytics and surveys to understand fully engaged learners in a highly-technical STEM MOOC,” in *Proc. 2016 IEEE Frontiers in Education Conference*, Erie, PA, USA, Oct. 2016, pp. 1–9.
- [140] A. Pardo, F. Han, and R. A. Ellis, “Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance,” *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 82–92, Jan./Mar. 2016.
- [141] B. Hong, Z. Wei, and Y. Yang, “Discovering learning behavior patterns to predict dropout in MOOC,” in *Proc. 12th International Conference on Computer Science and Education*, Houston, TX, USA, Aug. 2017, pp. 700–704.

- [142] X. Chen, M. Vorvoreanu, and K. Madhavan, “Mining Social Media Data for Understanding Students’ Learning Experiences,” *IEEE Transactions on Learning Technologies*, vol. 7, no. 3, pp. 246–259, Jul./Sep. 2014.
- [143] B. Xu and D. Yang, “Motivation Classification and Grade Prediction for MOOCs Learners,” *Computational Intelligence and Neuroscience*, vol. 2016, Jan. 2016.
- [144] A. Pigeau, O. Aubert, and Y. Prié, “Success Prediction in MOOCs: A Case Study,” in *Proc. 12th International Conference on Educational Data Mining*, Montréal, QC, Canada, Jul. 2019, pp. 390–395.
- [145] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, “Models for early prediction of at-risk students in a course using standards-based grading,” *Computers & Education*, vol. 103, pp. 1–15, Dec. 2016.
- [146] K. Kelly, I. Arroyo, and N. Heffernan, “Using ITS Generated Data to Predict Standardized Test Scores,” in *Proc. 6th International Conference on Educational Data Mining*, Memphis, TN, USA, Jul. 2013, pp. 322–323.
- [147] J. Daniel, “Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility,” *Journal of Interactive Media in Education*, vol. 2012, no. 3, Dec. 2012.
- [148] F. Jiang and W. Li, “Who Will Be the Next to Drop Out? Anticipating Dropouts in MOOCs with Multi-View Features,” *International Journal of Performability Engineering*, vol. 13, no. 2, pp. 201–210, Mar. 2017.
- [149] N. Wu, L. Zhang, Y. Gao, M. Zhang, X. Sun, and J. Feng, “CLMS-Net: dropout prediction in MOOCs with deep learning,” in *Proc. 2019 ACM Turing Celebration Conference*, Chengdu, China, May 2019, pp. 1–6.
- [150] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, “Predicting MOOC Dropout over Weeks Using Machine Learning Methods,” in *Proc. EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, Doha, Qatar, Oct. 2014, pp. 60–65.
- [151] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, “Data mining for modeling students’ performance: A tutoring action plan to prevent academic dropout,” *Computers & Electrical Engineering*, vol. 66, pp. 541–556, Feb. 2018.
- [152] F. D. Pereira, E. Oliveira, A. Cristea, D. Fernandes, L. Silva, G. Aguiar, A. Alamri, and M. Alshehri, “Early Dropout Prediction for Programming Courses Supported by Online Judges,” in *Proc. 20th International Conference on Artificial Intelligence in Education*, vol. 2, Chicago, IL, USA, Jun. 2019, pp. 67–72.
- [153] M. Ding, K. Yang, D.-Y. Yeung, and T.-C. Pong, “Effective Feature Learning with Unsupervised Learning for Improving the Predictive Models in Massive Open Online Courses,” in *Proc. 9th International Conference on Learning Analytics and Knowledge*, Tempe, AZ, USA, Mar. 2019, pp. 135–144.

- [154] C. Yu, “SPOC-MFLP: A Multi-feature Learning Prediction Model for SPOC Students Using Machine Learning,” *Journal of Applied Science and Engineering*, vol. 21, no. 2, pp. 279–290, Jun. 2018.
- [155] M. Feng, N. T. Heffernan, and K. R. Koedinger, “Predicting State Test Scores Better with Intelligent Tutoring Systems: Developing Metrics to Measure Assistance Required,” in *Proc. 8th International Conference on Intelligent Tutoring Systems*, Jhongli, Taiwan, Jun. 2006, pp. 31–40.
- [156] S. E. Fancsali, G. Zheng, Y. Tan, S. Ritter, S. R. Berman, and A. Galyardt, “Using embedded formative assessment to predict state summative test scores,” in *Proc. 8th International Conference on Learning Analytics and Knowledge*, Sydney, NSW, Australia, Mar. 2018, pp. 161–170.
- [157] R. Jena, “Predicting students’ learning style using learning analytics: a case study of business management students from India,” *Behaviour & Information Technology*, vol. 37, no. 10-11, pp. 978–992, Oct. 2018.
- [158] A. Tato, R. Nkambou, and R. Ghali, “Towards Predicting Attention and Workload During Math Problem Solving,” in *Proc. 15th International Conference on Intelligent Tutoring Systems*, Kingston, Jamaica, Jun. 2019, pp. 224–229.
- [159] C. Peters, I. Arroyo, W. Burleson, B. Woolf, and K. Muldner, “Predictors and Outcomes of Gaming in an Intelligent Tutoring System,” in *Proc. 14th International Conference on Intelligent Tutoring Systems*, Montreal, QC, Canada, Jun. 2018, pp. 366–372.
- [160] M. R. Barrick and M. K. Mount, “The Big Five personality dimensions and job performance: A meta-analysis,” *Personnel Psychology*, vol. 44, no. 1, pp. 1–26, Mar. 1991.
- [161] A. Bakharia, “Towards Cross-domain MOOC Forum Post Classification,” in *Proc. 3rd ACM Conference on Learning@ Scale*, Edinburgh, UK, Apr. 2016, pp. 253–256.
- [162] N. Souza and G. Perry, “Identification of Affective States in MOOCs: A Systematic Literature Review,” *International Journal for Innovation Education and Research*, vol. 6, no. 12, pp. 39–55, Dec. 2018.
- [163] N. Jaques, C. Conati, J. M. Harley, and R. Azevedo, “Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System,” in *Proc. 12th International Conference on Intelligent Tutoring Systems*, Honolulu, HI, USA, Jun. 2014, pp. 29–38.
- [164] D. Vu, P. Pattison, and G. Robins, “Relational event models for social learning in MOOCs,” *Social Networks*, vol. 43, pp. 121–135, Oct. 2015.
- [165] D. Yang, D. Adamson, and C. P. Rosé, “Question recommendation with constraints for massive open online courses,” in *Proc. 8th ACM Conference on Recommender systems*, Foster City, CA, USA, Oct. 2014, pp. 49–56.

- [166] C. Romero, M.-I. López, J.-M. Luna, and S. Ventura, “Predicting students’ final performance from participation in on-line discussion forums,” *Computers & Education*, vol. 68, pp. 458–472, Oct. 2013.
- [167] M. Xenos, “Prediction and assessment of student behaviour in open and distance education in computers using bayesian networks,” *Computers & Education*, vol. 43, no. 4, pp. 345–359, Dec. 2004.
- [168] B. Makhabel, *Learning Data Mining with R*. Birmingham, UK: Packt Publishing Ltd, 2015.
- [169] A. F. Wise, Y. Cui, W. Jin, and J. Vytasek, “Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling,” *The Internet and Higher Education*, vol. 32, pp. 11–28, Jan. 2017.
- [170] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. Delgado Kloos, “Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators,” *Expert Systems*, vol. 35, no. 6, e12298, 2018.
- [171] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY, USA: Springer Science+Business Media, 2000.
- [172] M. Fei and D.-Y. Yeung, “Temporal Models for Predicting Student Dropout in Massive Open Online Courses,” in *Proc. 2015 IEEE International Conference on Data Mining Workshop*, Atlantic City, NJ, USA, Feb. 2015, pp. 256–263.
- [173] J. Macina, I. Srba, J. J. Williams, and M. Bielikova, “Educational Question Routing in Online Student Communities,” in *Proc. 11th ACM Conference on Recommender Systems*, Chicago, IL, USA, Aug. 2017, pp. 47–55.
- [174] S. Huang and N. Fang, “Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models,” *Computers & Education*, vol. 61, pp. 133–145, Feb. 2013.
- [175] J. R. Quinlan, “Simplifying decision trees,” *International Journal of Man-Machine Studies*, vol. 27, no. 3, pp. 221–234, Sep. 1987.
- [176] M. Kumar, A. Singh, and D. Handa, “Literature survey on student’s performance prediction in education using data mining techniques,” *International Journal of Education and Management Engineering*, vol. 7, no. 6, pp. 42–49, Nov. 2017.
- [177] W. Xing, X. Chen, J. Stein, and M. Marcinkowski, “Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization,” *Computers in Human Behavior*, vol. 58, pp. 119–129, May 2016.
- [178] A. Pardo, N. Mirriahi, R. Martinez-Maldonado, J. Jovanovic, S. Dawson, and D. Gašević, “Generating actionable predictive models of academic performance,” in *Proc. 6th International Conference on Learning Analytics and Knowledge*, Edinburgh, UK, Apr. 2016, pp. 474–478.

- [179] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, “Random forest: a classification and regression tool for compound classification and QSAR modeling,” *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1947–1958, Nov. 2003.
- [180] R. N. Laveti, S. Kuppili, J. Ch, S. N. Pal, and N. S. C. Babu, “Implementation of learning analytics framework for MOOCs using state-of-the-art in-memory computing,” in *Proc. 5th National Conference on E-Learning & E-Learning Technologies*, Hyderabad, India, Aug. 2017, pp. 1–6.
- [181] C. Ye, J. S. Kinnebrew, G. Biswas, B. J. Evans, D. H. Fisher, G. Narasimham, and K. A. Brady, “Behavior Prediction in MOOCs using Higher Granularity Temporal Information,” in *Proc. 2nd ACM conference on Learning@ Scale*, Vancouver, BC, Canada, Mar. 2015, pp. 335–338.
- [182] P. Cortez and A. M. Gonçalves Silva, “Using Data Mining to Predict Secondary School Student Performance,” in *Proc. 5th Annual Future Business Technology Conference*, Porto, Portugal, Apr. 2008, pp. 5–12.
- [183] C. Imrie, S. Durucan, and A. Korre, “River flow prediction using artificial neural networks: generalisation beyond the calibration range,” *Journal of Hydrology*, vol. 233, no. 1-4, pp. 138–153, Jun. 2000.
- [184] Á. Pérez-Lemonche, G. Martínez-Muñoz, and E. Pulido-Cañabate, “Analysing Event Transitions to Discover Student Roles and Predict Grades in MOOCs,” in *Proc. 26th International Conference on Artificial Neural Networks*, Alghero, Italy, Sep. 2017, pp. 224–232.
- [185] A. Joshi *et al.*, “Affect-driven Learning Outcomes Prediction in Intelligent Tutoring Systems,” in *Proc. 14th IEEE International Conference on Automatic Face & Gesture Recognition*, Lille, France, May 2019, pp. 1–5.
- [186] W. Lu, T. Wang, M. Jiao, X. Zhang, S. Wang, X. Du, and H. Chen, “Predicting Student Examinee Rate in Massive Open Online Courses,” in *Proc. 22th International Conference on Database Systems for Advanced Applications*, Suzhou, China, Mar. 2017, pp. 340–351.
- [187] D. Agarwal, N. Babel, and R. Baker, “Contextual Derivation of Stable BKT Parameters for Analysing Content Efficacy,” in *Proc. 11th International Conference on Educational Data Mining*, Buffalo, NY, USA, Jul. 2018, pp. 596–601.
- [188] J. Gardner and C. Brooks, “Student success prediction in MOOCs,” *User Modeling and User-Adapted Interaction*, vol. 28, no. 2, pp. 127–203, Jun. 2018.
- [189] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [190] J. Qiu, J. Tang, T. X. Liu, J. Gong, C. Zhang, Q. Zhang, and Y. Xue, “Modeling and Predicting Learning Behavior in MOOCs,” in *Proc. 9th ACM International Conference on Web Search and Data Mining*, San Francisco, CA, USA, Feb. 2016, pp. 93–102.

- [191] A. F. Botelho, A. Varatharaj, T. Patikorn, D. Doherty, S. A. Adjei, and J. E. Beck, “Developing Early Detectors of Student Attrition and Wheel Spinning Using Deep Learning,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 158–170, Apr./Jun. 2019.
- [192] A.-D. Mezaour, “Filtering Web Documents for a Thematic Warehouse Case Study: eDot a Food Risk Data Warehouse (extended),” in *Proc. International IIS: Intelligent Information Processing and Web Mining 2005 Conference*, Gdansk, Poland, Jun. 2005, pp. 269–278.
- [193] C. A. Coleman, D. T. Seaton, and I. Chuang, “Probabilistic Use Cases: Discovering Behavioral Patterns for Predicting Certification,” in *Proc. 2nd ACM Conference on Learning@ Scale*, Vancouver, BC, Canada, Mar. 2015, pp. 141–148.
- [194] G. Kostopoulos, S. Karlos, and S. Kotsiantis, “Multiview Learning for Early Prognosis of Academic Performance: A Case Study,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 212–224, Apr./Jun. 2019.
- [195] A. Ramesh, S. H. Kumar, J. Foulds, and L. Getoor, “Weakly Supervised Models of Aspect-Sentiment for Online Course Discussion Forums,” in *Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, vol. 1, Beijing, China, Jul. 2015, pp. 74–83.
- [196] D. Baneres, M. E. Rodríguez-Gonzalez, and M. Serra, “An Early Feedback Prediction System for Learners At-Risk Within a First-Year Higher Education Course,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 249–263, Apr./Jun. 2019.
- [197] J. Brownlee, “Machine Learning Evaluation Metrics in R,” *Machine Learning Mastery*, 2016. [Online]. Available: <https://machinelearningmastery.com/machine-learning-evaluation-metrics-in-r/>.
- [198] P. J. Muñoz-Merino, R. González Novillo, and C. Delgado Kloos, “Assessment of skills and adaptive learning for parametric exercises combining knowledge spaces and item response theory,” *Applied Soft Computing*, vol. 68, pp. 110–124, Jul. 2018.
- [199] M.-R. Vázquez, F. P. Romero, J. A. Olivas, E. Orbe, and J. Serrano-Guerrero, “An approach to academic performance prediction in tutoring systems based on fuzzy deformable prototypes,” *Progress in Artificial Intelligence*, vol. 5, no. 1, pp. 55–64, Feb. 2016.
- [200] D. S. Chaplot, E. Rhim, and J. Kim, “Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks,” in *Proc. Workshops 17th International Conference on Artificial Intelligence in Education*, vol. 3, Madrid, Spain, Jun. 2015, pp. 7–12.
- [201] G. Kennedy, C. Coffrin, P. De Barba, and L. Corrin, “Predicting success: how learners’ prior knowledge, skills and activities predict MOOC performance,”

in *Proc. 5th International Conference on Learning Analytics and Knowledge*, Poughkeepsie, NY, USA, Mar. 2015, pp. 136–140.

- [202] R. Pelánek, “Metrics for Evaluation of Student Models,” *Journal of Educational Data Mining*, vol. 7, no. 2, pp. 1–19, Jun. 2015.
- [203] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American statistical Association*, vol. 102, no. 477, pp. 359–378, Mar. 2007.
- [204] L. A. Jeni, J. F. Cohn, and F. De La Torre, “Facing Imbalanced Data—Recommendations for the Use of Performance Metrics,” in *Proc. 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva, Switzerland, Sep. 2013, pp. 245–251.
- [205] R. Ferguson, D. Clow, L. Macfadyen, A. Essa, S. Dawson, and S. Alexander, “Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption,” in *Proc. 4th International Conference on Learning Analytics and Knowledge*, Indianapolis, IN, USA, Mar. 2014, pp. 251–253.
- [206] J. Ocumpaugh, R. Baker, S. Gowda, N. Heffernan, and C. Heffernan, “Population validity for Educational Data Mining models: A case study in affect detection,” *British Journal of Educational Technology*, vol. 45, no. 3, pp. 487–501, May 2014.
- [207] D. M. Olivé, D. Q. Huynh, M. Reynolds, M. Dougiamas, and D. Wiese, “A Quest for a One-Size-Fits-All Neural Network: Early Prediction of Students at Risk in Online Courses,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 171–183, Apr./Jun. 2019.
- [208] A. Merceron, “Educational Data Mining/Learning Analytics: Methods, Tasks and Current Trends,” in *Proc. DeLFI Workshops 2015 co-located with 13th e-Learning Conference of the German Computer Society*, München, Germany, Sep. 2015, pp. 101–109.
- [209] K. D. Strang, “Beyond engagement analytics: which online mixed-data factors predict student learning outcomes?” *Education and Information Technologies*, vol. 22, no. 3, pp. 917–937, May 2017.
- [210] B. Schneider and P. Blikstein, “Unraveling Students’ Interaction Around a Tangible Interface Using Multimodal Learning Analytics,” *Journal of Educational Data Mining*, vol. 7, no. 3, pp. 89–116, Oct. 2015.
- [211] J. He, J. Bailey, B. I. Rubinstein, and R. Zhang, “Identifying At-Risk Students in Massive Open Online Courses,” in *Proc. 29th AAAI Conference on Artificial Intelligence*, Austin, TX, USA, Jan. 2015, pp. 1749–1755.
- [212] N. Gitinabard, Y. Xu, S. Heckman, T. Barnes, and C. F. Lynch, “How Widely Can Prediction Models Be Generalized? Performance Prediction in Blended Courses,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 184–197, Apr./Jun. 2019.

- [213] J.-L. Hung, B. E. Shelton, J. Yang, and X. Du, “Improving predictive modeling for at-risk student identification: a multistage approach,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 148–157, Apr./Jun. 2019.
- [214] J. López-Zambrano, J. A. Lara, and C. Romero, “Towards Portability of Models for Predicting Students’ Final Performance in University Courses Starting from Moodle Logs,” *Applied Sciences*, vol. 10, no. 1, 354, Jan. 2020.
- [215] L. Kidzinsk, K. Sharma, M. S. Boroujeni, and P. Dillenbourg, “On Generalizability of MOOC Models,” in *Proc. 9th International Conference on Educational Data Mining*, Raleigh, NC, USA, Jul. 2016, pp. 406–411.
- [216] R. F. Kizilcec and S. Halawa, “Attrition and Achievement Gaps in Online Learning,” in *Proc. 2nd ACM Conference on Learning@ Scale*, Vancouver, BC, Canada, Mar. 2015, pp. 57–66.
- [217] M. L. Bote-Lorenzo and E. Gómez-Sánchez, “An Approach to Build in situ Models for the Prediction of the Decrease of Academic Engagement Indicators in Massive Open Online Courses,” *Journal of Universal Computer Science*, vol. 24, no. 8, pp. 1052–1071, Aug. 2018.
- [218] J. Whitehill, K. Mohan, D. Seaton, Y. Rosen, and D. Tingley, “MOOC dropout prediction: How to measure accuracy?” in *Proc. 4th ACM Conference on Learning@ Scale*, Cambridge, MA, USA, Apr. 2017, pp. 161–164.
- [219] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, “Analyzing and Predicting Students’ Performance by Means of Machine Learning: A Review,” *Applied Sciences*, vol. 10, no. 3, 1042, Feb. 2020.
- [220] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, May/Jun. 2020.
- [221] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, Dec. 2007.
- [222] B. Kuechler and V. Vaishnavi, “On theory development in design science research: anatomy of a research project,” *European Journal of Information Systems*, vol. 17, no. 5, pp. 489–504, Sep. 2008.
- [223] edX, “EdX Research Guide Release,” 2018. [Online]. Available: <https://media.readthedocs.org/pdf/devdata/latest/devdata.pdf>.
- [224] J. Maldonado-Mahauad, M. Pérez-Sanagustín, R. F. Kizilcec, N. Morales, and J. Munoz-Gama, “Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses,” *Computers in Human Behavior*, vol. 80, pp. 179–196, Mar. 2018.
- [225] C. Taylor, K. Veeramachaneni, and U.-M. O’Reilly, “Likely to stop? predicting stopout in massive open online courses,” 2014. [Online]. Available: <https://arxiv.org/pdf/1408.3382.pdf>.

- [226] K. Veeramachaneni, U.-M. O'Reilly, and C. Taylor, "Towards feature engineering at scale for data from massive open online courses," *arXiv preprint arXiv:1407.5238*, 2014. [Online]. Available: <https://arxiv.org/pdf/1407.5238.pdf>.
- [227] J. A. Cunningham, "Predicting Student Success in a Self-Paced Mathematics MOOC," Ph.D. dissertation, Ariz. St. Unive., Tempe, AZ, USA, 2017.
- [228] V. Kovanović, D. Gašević, S. Dawson, S. Joksimović, R. S. Baker, and M. Hatala, "Penetrating the black box of time-on-task estimation," in *Proc. 5th International Conference on Learning Analytics and Knowledge*, Poughkeepsie, NY, USA, Mar. 2015, pp. 184–193.
- [229] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, H. A. Parada G., M. Muñoz-Organero, and A. Rodríguez-de-las Heras, "Analysing the Impact of Built-In and External Social Tools in a MOOC on Educational Technologies," in *Proc. 8th European Conference on Technology Enhanced Learning*, Paphos, Cyprus, Sep. 2013, pp. 5–18.
- [230] S. Mak, R. Williams, and J. Mackness, "Blogs and forums as communication and learning tools in a MOOC," in *Proc. 7th International Conference on Networked Learning*, Aalborg, Denmark, May 2010, pp. 275–285.
- [231] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, I. Estévez-Ayres, and C. Delgado Kloos, "A Learning Analytics Methodology for Understanding Social Interactions in MOOCs," *IEEE Transactions on Learning Technologies*, vol. 12, no. 4, pp. 442–455, Oct./Dec. 2019.
- [232] B. Grawemeyer, M. Mavrikis, W. Holmes, A. Hansen, K. Loibl, and S. Gutiérrez-Santos, "The Impact of Feedback on Students' Affective States," in *Proc. Workshops at the 17th International Conference on Artificial Intelligence in Education*, vol. 7, Madrid, Spain, Jun. 2015, pp. 4–13.
- [233] X. Wang, D. Yang, M. Wen, K. Koedinger, and C. P. Rosé, "Investigating How Student's Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains," in *Proc. 8th International Conference on Educational Data Mining*, Madrid, Spain, Jun. 2015, pp. 226–233.
- [234] D. F. O. Onah, J. E. Sinclair, and R. Boyatt, "Exploring the use of MOOC discussion forums," in *Proc. London International Conference on Education 2014*, London, UK, Nov. 2014, pp. 1–4.
- [235] C. Alario-Hoyos, P. J. Muñoz-Merino, M. Pérez-Sanagustín, C. Delgado Kloos, and H. A. Parada G, "Who are the top contributors in a MOOC? Relating participants' performance and contributions," *Journal of Computer Assisted Learning*, vol. 32, no. 3, pp. 232–243, Jun. 2016.
- [236] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, USA, Aug. 2004, pp. 168–177.

- [237] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, I. Estévez-Ayres, and C. Delgado Kloos, “Sentiment analysis in MOOCs: A case study,” in *Proc. 2018 IEEE Global Engineering Education Conference*, Santa Cruz de Tenerife, Spain, Apr. 2018, pp. 1489–1496.
- [238] H. Wang and J. A. Castanon, “Sentiment expression via emoticons on social media,” in *Proc. 2015 IEEE International Conference on Big Data*, Santa Clara, CA, USA, Oct./Nov. 2015, pp. 2404–2408.
- [239] W.-H. Wu, Y.-C. J. Wu, C.-Y. Chen, H.-Y. Kao, C.-H. Lin, and S.-H. Huang, “Review of trends from mobile learning studies: A meta-analysis,” *Computers & Education*, vol. 59, no. 2, pp. 817–827, Sep. 2012.
- [240] B. Macarini, L. Antonio, C. Cechinel, M. F. Batista Machado, V. Faria Culmant Ramos, and R. Munoz, “Predicting Students Success in Blended Learning—Evaluating Different Interactions Inside Learning Management Systems,” *Applied Sciences*, vol. 9, no. 24, 5523, Dec. 2019.
- [241] Y. Lee and J. Choi, “A review of online course dropout research: Implications for practice and future research,” *Educational Technology Research and Development*, vol. 59, no. 5, pp. 593–618, Oct. 2011.
- [242] R. Ferguson and D. Clow, “Where is the evidence? A call to action for learning analytics,” in *Proc. 7th International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, Mar. 2017, pp. 56–65.
- [243] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, “The current landscape of learning analytics in higher education,” *Computers in Human Behavior*, vol. 89, pp. 98–110, Dec. 2018.
- [244] H. Waheed, S.-U. Hassan, N. R. Aljohani, and M. Wasif, “A bibliometric perspective of learning analytics research landscape,” *Behaviour & Information Technology*, vol. 37, no. 10-11, pp. 941–957, Oct. 2018.
- [245] C. Colvin, T. Rogers, A. Wade, S. Dawson, D. Gasevic, S. Buckingham Shum, K. Nelson, S. Alexander, L. Lockyer, G. Kennedy, L. Corrin, and J. Fisher, “Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement,” *Australian Government Office for Learning and Teaching*, 2016. [Online]. Available: https://opus.lib.uts.edu.au/bitstream/10453/117173/1/AUS_OLT_LearningAnalytics_2016.pdf.
- [246] P. Arroway, G. Morgan, M. O’Keefe, and R. Yanosky, “Learning analytics in higher education,” *EDUCAUSE Center for Analysis and Research*, 2015. [Online]. Available: <https://library.educause.edu/~media/files/library/2016/2/ers1504la>.
- [247] G. Siemens, S. Dawson, and G. Lynch, “Improving the Quality and Productivity of the Higher Education Sector,” *Society for Learning Analytics Research*, 2013. [Online]. Available: https://solaresearch.org/wp-content/uploads/2017/06/SoLAR_Report_2014.pdf.

- [248] Y.-S. Tsai, D. Gašević, A. Whitelock-Wainwright, P. J. Muñoz-Merino, P. M. Moreno-Marcos, A. Rubio Fernández, C. Delgado Kloos, M. Scheffel, I. Jivet, H. Drachsler, K. Tammets, A. Ruiz Calleja, and K. Kollom, “SHEILA: Supporting Higher Education to Integrate Learning Analytics,” 2018. [Online]. Available: <http://sheilaproject.eu/wp-content/uploads/2018/11/SHEILA-research-report.pdf>.
- [249] Y.-S. Tsai, D. Rates, P. M. Moreno-Marcos, P. J. Muñoz-Merino, I. Jivet, M. Scheffel, H. Drachsler, C. Delgado Kloos, and D. Gasevic, “Learning Analytics in European Higher Education-Trends and Barriers,” *Computers & Education*, vol. 155, 103933, Oct. 2020.
- [250] K. Kollom, M. Scheffel, K. Tammets, I. Jivet, Y.-S. Tsai, P. J. Muñoz-Merino, P. M. Moreno-Marcos, A. Whitelock-Wainwright, A. Ruiz Calleja, D. Gasevic, C. Delgado Kloos, H. Drachsler, and T. Ley, “A four-country cross-case analysis of academic staff expectations about learning analytics in higher education,” unpublished.
- [251] Y.-S. Tsai, P. M. Moreno-Marcos, K. Tammets, K. Kollom, and D. Gašević, “SHEILA policy framework: informing institutional strategies and policy processes of learning analytics,” in *Proc. 8th International Conference on Learning Analytics and Knowledge*, Sydney, NSW, Australia, Mar. 2018, pp. 320–329.
- [252] Y.-S. Tsai, P. M. Moreno-Marcos, I. Jivet, M. Scheffel, K. Tammets, K. Kollom, and D. Gašević, “The SHEILA framework: informing institutional strategies and policy processes of learning analytics,” *Journal of Learning Analytics*, vol. 5, no. 3, pp. 5–20, Nov. 2018.
- [253] L. P. Prieto, M. J. Rodríguez-Triana, R. Martínez-Maldonado, Y. Dimitriadis, and D. Gašević, “Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level,” *Australasian Journal of Educational Technology*, vol. 35, no. 4, pp. 14–33, Aug. 2019.
- [254] M. Pérez Sanagustín, I. Hilliger, J. Maldonado, R. Pérez, L. Ramírez, P. J. Muñoz-Merino, Y.-S. Tsai, M. Ortiz, T. Broos, P. Pesantez, E. Sheihing, and A. Whitelock-Wainwright, “LALA Framework,” LALA Project Team, Tech. Rep. WPD2.O.3, 2018. [Online]. Available: <https://www.lalaproject.org/wp-content/uploads/2019/01/LALS-FW-2.0.pdf>.
- [255] P. Leitner, M. Ebner, and M. Ebner, “Learning Analytics Challenges to Overcome in Higher Education Institutions,” in *Utilizing Learning Analytics to Support Study Success*, D. Ifenthaler, D.-K. Mah, and J. Y.-K. Yau, Eds. Cham, Switzerland: Springer, 2019, pp. 91–104.
- [256] I. Jivet, M. Scheffel, M. Specht, and H. Drachsler, “License to evaluate: preparing learning analytics dashboards for educational practice,” in *Proc. 8th International Conference on Learning Analytics and Knowledge*, Sydney, NSW, Australia, Mar. 2018, pp. 31–40.

- [257] L. P. Macfadyen, S. Dawson, A. Pardo, and D. Gašević, “Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge,” *Research & Practice in Assessment*, vol. 9, pp. 17–28, Oct. 2014.
- [258] H. Drachsler and W. Greller, “Privacy and analytics: it’s a DELICATE issue a checklist for trusted learning analytics,” in *Proc. 6th International Conference on Learning Analytics & Knowledge*, Edinburgh, UK, Apr. 2016, pp. 89–98.
- [259] A. Pardo, J. Jovanovic, S. Dawson, D. Gašević, and N. Mirriahi, “Using learning analytics to scale the provision of personalised feedback,” *British Journal of Educational Technology*, vol. 50, no. 1, pp. 128–138, Jan. 2019.
- [260] P. J. Muñoz-Merino, P. M. Moreno-Marcos, A. Rubio-Fernández, Y.-S. Tsai, D. Gašević, and C. Delgado Kloos, “A Systematic Analysis of Learning Analytics using Multi-source Data in the Context of Spain,” unpublished.
- [261] P. M. Moreno-Marcos, P. J. Muñoz-Merino, C. Alario-Hoyos, I. Estévez-Ayres, and C. Delgado Kloos, “Analysing the predictive power for anticipating assignment grades in a massive open online course,” *Behaviour & Information Technology*, vol. 37, no. 10-11, pp. 1021–1036, Oct. 2018.
- [262] P. M. Moreno Marcos, “Analysing the predictive power for anticipating assignment grades in a Massive Open Online Course,” M.S. thesis, Dept. Telematics Eng., Univ. Carlos III Madrid, Madrid, Spain, 2017.
- [263] P. M. Moreno-Marcos, T.-C. Pong, P. J. Muñoz-Merino, and C. Delgado Kloos, “Analysis of the Factors Influencing Learners’ Performance Prediction With Learning Analytics,” *IEEE Access*, vol. 8, pp. 5264–5282, Jan. 2020.
- [264] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, “Engaging with massive online courses,” in *Proc. 23rd International Conference on World Wide Web*, Seoul, Korea, Apr. 2014, pp. 687–698.
- [265] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu, “PeakVizor: Visual Analytics of Peaks in Video Clickstreams from Massive Open Online Courses,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 10, pp. 2315–2330, Oct. 2016.
- [266] J. Maldonado-Mahauad, M. Pérez-Sanagustín, P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, and C. Delgado-Kloos, “Predicting Learners’ Success in a Self-paced MOOC Through Sequence Patterns of Self-regulated Learning,” in *Proc. 13th European Conference on Technology Enhanced Learning*, Leeds, UK, Sep. 2018, pp. 355–369.
- [267] F. Okubo, T. Yamashita, A. Shimada, and H. Ogata, “A neural network approach for students’ performance prediction,” in *Proc. 7th International Learning Analytics and Knowledge Conference*, Vancouver, BC, Canada, Mar. 2017, pp. 598–599.

- [268] M. Vitiello, S. Walk, D. Helic, V. Chang, and C. Guetl, “User Behavioral Patterns and Early Dropouts Detection: Improved Users Profiling through Analysis of Successive Offering of MOOC,” *Journal of Universal Computer Science*, vol. 24, no. 8, pp. 1131–1150, Aug. 2018.
- [269] P. M. Moreno-Marcos, T. De Laet, P. J. Muñoz-Merino, C. Van Soom, T. Broos, K. Verbert, and C. Delgado Kloos, “Predicting admission test success using SPOC interactions,” in *Companion Proc. 9th International Conference of Learning Analytics and Knowledge*, Tempe, AZ, USA, Mar. 2019, pp. 924–934.
- [270] L. Roggemans and B. Spruyt, “Toelatingsproef (tand)arts: een sociografische schets van de deelnemers en geslaagden,” *Vrije Universiteit Brussel*, 2014. [Online]. Available: http://socipc1.vub.ac.be/torwebdat/publications/t2014_29.pdf.
- [271] M. Pérez-Sanagustín, K. Sharma, R. Pérez-Álvarez, J. Maldonado-Mahauad, and J. Broisin, “Analyzing learners’ behavior beyond the MOOC: An exploratory study,” in *Proc. 14th European Conference on Technology Enhanced Learning*, Delft, The Netherlands, Sep. 2019, pp. 40–54.
- [272] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, “Visualizing patterns of student engagement and performance in MOOCs,” in *Proc. 4th International Conference on Learning Analytics and Knowledge*, Indianapolis, IN, USA, Mar. 2014, pp. 83–92.
- [273] L. K. Silverman, “Social and emotional education of the gifted: The discoveries of Leta Stetter Hollingworth,” *Roeper Review*, vol. 12, no. 3, pp. 171–178, Mar. 1990.
- [274] M. D. Valadez Sierra, A. A. Valdés Cuervo, T. R. Wendlandt Amezaga, A. C. Reyes Sánchez, R. Zambrano Guzmán, and J. P. Navarro Agraz, “Differences in Achievement Motivation and Academic and Social Self-Concept in Gifted Students of Higher Education,” *Journal of Curriculum and Teaching*, vol. 4, no. 1, pp. 83–90, May 2015.
- [275] R. Conijn, A. Van den Beemt, and P. Cuijpers, “Predicting student performance in a blended MOOC,” *Journal of Computer Assisted Learning*, vol. 34, no. 5, pp. 615–628, Oct. 2018.
- [276] M. Bannert, “Promoting Self-Regulated Learning Through Prompts,” *Zeitschrift für Pädagogische Psychologie*, vol. 23, no. 2, pp. 139–145, Jan. 2009.
- [277] I. Molenaar and C. A. Knoop-van Campen, “How Teachers Make Dashboard Information Actionable,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 3, pp. 347–355, Jul./Sep. 2019.
- [278] C. Paul, W. Mason, D. McCaffrey, and S. A. Fox, “What Should We Do About Missing Data? (A Case Study Using Logistic Regression with Missing Data on a Single Covariate),” *California Center for Population Research*, 2003. [Online]. Available: <http://papers.ccpr.ucla.edu/index.php/pwp/article/view/PWP-CCPR-2003-028/521>.

- [279] I. Bjelland, S. Krokstad, A. Mykletun, A. A. Dahl, G. S. Tell, and K. Tambs, “Does a higher educational level protect against anxiety and depression? The HUNT study,” *Social Science & Medicine*, vol. 66, no. 6, pp. 1334–1345, Mar. 2008.
- [280] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, “Understanding variable importances in forests of randomized trees,” in *Proc. 26th Advances in Neural Information Processing Systems Conference*, Lake Tahoe, NV, USA, Dec. 2013, pp. 431–439.
- [281] C. A. Farrington, M. Roderick, E. Allensworth, J. Nagaoka, T. S. Keyes, D. W. Johnson, and N. O. Beechum, *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance—A Critical Literature Review*. Chicago, IL, USA: University of Chicago Consortium on Chicago School Research, 2012.
- [282] H. Wei, H. Li, M. Xia, Y. Wang, and H. Qu, “Predicting Student Performance in Interactive Online Question Pools Using Mouse Interaction Features,” in *Proc. 10th International Conference on Learning Analytics and Knowledge*, Frankfurt, Germany, Mar. 2020, pp. 645–654.
- [283] F. Dalipi, A. S. Imran, and Z. Kastrati, “MOOC dropout prediction using machine learning techniques: Review and research challenges,” in *Proc. 2018 IEEE Global Engineering Education Conference*, Santa Cruz de Tenerife, Spain, Apr. 2018, pp. 1007–1014.
- [284] S. Jiang, A. Williams, K. Schenke, M. Warschauer, and D. O’Dowd, “Predicting MOOC performance with week 1 behavior,” in *Proc. 7th International Conference on Educational Data Mining*, London, UK, Jul. 2014, pp. 273–275.
- [285] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, “Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success,” *The Internet and Higher Education*, vol. 28, pp. 68–84, Jan. 2016.
- [286] P. J. Guo, J. Kim, and R. Rubin, “How video production affects student engagement: an empirical study of MOOC videos,” in *Proc. 1st ACM Conference on Learning@ Scale*, Atlanta, GA, USA, Mar. 2014, pp. 41–50.
- [287] M. Ortiz, A. Jimenez, R. Maya, P. J. Muñoz-Merino, P. M. Moreno-Marcos, J. Imaz Marín, C. Delgado Kloos, M. A. Zúñiga Prieto, M. Ulloa, R. Pérez, M. Pérez-Sanagustín, V. Henríquez, J. Guerra, R. Ferreira Leite de Mello, T. Broos, and M. Millecamp, “Design of learning analytics tools,” LALA Project Team, Tech. Rep. WPD3.O.4, 2018. [Online]. Available: https://www.lalaproject.org/wp-content/uploads/2019/04/Deliverable-WP3_English_April12.pdf.
- [288] Y. Luo and Z. A. Pardos, “Diagnosing University Student Subject Proficiency and Predicting Degree Completion in Vector Space,” in *Proc. 32nd AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, Feb. 2018, pp. 7920–7927.

- [289] A. Tekin, “Early Prediction of Students’ Grade Point Averages at Graduation: A Data Mining Approach,” *Eurasian Journal of Educational Research*, vol. 54, pp. 207–226, Feb. 2014.
- [290] Y. Chen, A. Johri, and H. Rangwala, “Running out of STEM: a comparative study across STEM majors of college students at-risk of dropping out early,” in *Proc. 8th International Conference on Learning Analytics and Knowledge*, Sydney, NSW, Australia, Mar. 2018, pp. 270–279.
- [291] C. Olivares-Rodríguez, P. M. Moreno-Marcos, E. Scheihing, P. J. Muñoz-Merino, and C. Delgado Kloos, “A configuration model to predict dropout based on academic trajectory features,” unpublished.
- [292] S. Donoso D., G. Donoso T., and C. Frites C., “La experiencia chilena de retención de estudiantes en la universidad,” *Revista Ciencia y Cultura*, vol. 17, no. 30, pp. 141–171, Jun. 2013.
- [293] E. Er, E. Gómez-Sánchez, M. L. Bote-Lorenzo, Y. Dimitriadis, and J. I. Asensio-Pérez, “Generating actionable predictions regarding MOOC learners’ engagement in peer reviews,” *Behaviour & Information Technology*, pp. 1–18, Sep. 2019.
- [294] V. A. Kumar, D. D’Souza, R. Lindén, and M.-J. Laakso, “Prediction of Student Final Exam Performance in an Introductory Programming Course: Development and Validation of the Use of a Support Vector Machine-Regression Model,” *Asian Journal of Education and e-Learning (ISSN: 2321–2454)*, vol. 7, no. 1, pp. 1–14, Feb. 2019.
- [295] P. A. Jensen and J. N. Barron, “Midterm and first-exam grades predict final grades in biology courses,” *Journal of College Science Teaching*, vol. 44, no. 2, pp. 82–89, Nov./Dic. 2014.
- [296] P. Ramanathan and E. Fernandez, “Can Early-Assignment Grades Predict Final Grades in IT Courses?” in *Proc. 2017 ASEE Annual Conference & Exposition*, Columbus, Ohio, USA, Jun. 2017.
- [297] S. Rovira, E. Puertas, and L. Igual, “Data-driven system to predict academic grades and dropout,” *PLoS one*, vol. 12, no. 2, Feb. 2017.
- [298] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, *Handbook of educational data mining*, 1st ed. Boca Ratón, FL, USA: CRC press, 2010.
- [299] F. Castro, A. Vellido, À. Nebot, and J. Minguillón, “Finding Relevant Features to Characterize Student Behavior on an e-Learning System,” in *Proc. 2005 International Conference on Frontiers in Education: Computer Science and Computer Engineering*, Las Vegas, NV, USA, Jun. 2005, pp. 210–216.
- [300] P. Mylonas, P. Tzouveli, and S. Kollias, “Intelligent content adaptation in the framework of an integrated e-learning system,” in *Proc. 16th ACM Conference on Hypertext & Hypermedia-Workshop on Combining Intelligent & Adaptive Hypermedia Methods/Techniques in Web-Based Education Systems*, Salzburg, Austria, Sep. 2005, pp. 59–66.

- [301] D. Koller, A. Ng, C. Do, and Z. Chen, “Retention and intention in massive open online courses: In depth,” *Educause review*, vol. 48, no. 3, pp. 62–63, Jun. 2013.
- [302] D. R. Bacon and K. A. Stewart, “How fast do students forget what they learn in consumer behavior? a longitudinal study,” *Journal of Marketing Education*, vol. 28, no. 3, pp. 181–192, Dec. 2006.
- [303] A. Rubio-Fernández, P. M. Moreno-Marcos, P. J. Muñoz-Merino, and C. Delgado Kloos, “An Initial Analysis of Prediction Techniques as a Support for the Flipped Classroom,” in *Proc. Learning Analytics Summer Institute Spain 2020*, Vigo, Spain, Jun. 2020.
- [304] A. Rubio-Fernández, P. J. Muñoz-Merino, and C. Delgado Kloos, “Analyzing the group formation process in intelligent tutoring systems,” in *Proc. 15th International Conference on Intelligent Tutoring Systems*, Kingston, Jamaica, Jun. 2019, pp. 34–39.
- [305] J. A. Ruiperez-Valiente, P. J. Munoz-Merino, G. Alexandron, and D. E. Pritchard, “Using Machine Learning to Detect ‘Multiple-Account’ Cheating and Analyze the Influence of Student and Problem Features,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 1, pp. 112–122, Jan./Mar. 2019.