Moreno Marcos, P. M., Muñoz Merino, P. J., Alario Hoyos, C., Estévez Ayres, I. & Delgado Kloos, C. (2018). Analysing the predictive power for anticipating assignment grades in a massive open online course. *Behaviour & Information Technology*, 37(10–11), pp. 1021–1036.

DOI: https://doi.org/10.1080/0144929X.2018.1458904

Analysing the predictive power for anticipating assignment grades in a Massive Open Online Course

Pedro Manuel Moreno-Marcos, Pedro J. Muñoz-Merino, Carlos Alario-Hoyos, Iria Estévez-Ayres & Carlos Delgado Kloos

The learning process in a MOOC (Massive Open Online Course) can be improved from knowing in advance learners' grades on different assignments. This would be very useful to detect problems with enough time to take corrective measures. In this work, the aim is to analyse how different course scores can be predicted, what elements or variables affect the predictions, and how much and in which way it is possible to anticipate scores. To do that, data from a MOOC about Java programming have been used. Results show the importance of indicators over the algorithms and that forum-related variables do not add power to predict grades, unlike previous scores. Furthermore, the type of task can vary the results. Regarding the anticipation, it was possible to use data from previous topics but with worse performance, although values were better than those obtained in the first seven days of the current topic.

Keywords: MOOCs; prediction; learners' grades; indicators; learning analytics; edX

1. Introduction

MOOCs have been gaining popularity since 2012 and nowadays there is an industry with millions of learners, thousands of MOOCs and hundreds of universities offering them (Christensen et al. 2013). The most important platforms worldwide are Coursera and edX (Shah 2015), although courses usually share a common structure. Grainger (2013) identified eight types of contents in MOOCs: video lectures, assessments, forums, activities, live video sessions, reading materials, additional video resources, and social media.

A relevant aspect of MOOCs is the huge amount of information that can be inferred regarding learners based on their interactions. This information can be used not only to diagnose factors explaining the observed data, but also to identify patterns and predict trends. Predictions in MOOCs have special relevance because of several reasons. Predictions can provide teachers and other stakeholders an idea of what is happening in the course to be able to anticipate problems with learners or the content. Besides, they can support interventions to improve the learning processes or make decisions that affect the design of the course. Moreover, predictions can be useful to the learners themselves because they can obtain data for their self-reflection (Greller and Drachsler 2012), which can help them to enhance their engagement and performance on the course.

In MOOCs, one of the main areas of prediction is the forecasting of dropouts, because completion rates can be very low (even lower than 10%) (Daniel 2012) and it is important to detect students at risk in advance. In studies on prediction of dropouts, timing considerations are important since most of the dropouts occur at the beginning of the MOOC (Santos et al. 2014). For this reason, several contributions tried to predict when the learner was going to drop out the course (Xing et al. 2016; Kloft et al. 2014). Nevertheless, it is noteworthy that dropouts are not necessarily a problem in all cases because many learners may enrol the course just to explore the content without intention to finish it or even without intention to take any activity (Kolowich 2013).

Another area of interest is the prediction of scores in MOOCs. Research in this field is also important to identify learners who have trouble understanding the course contents and to be able to provide frequent and effective feedbacks (Pardo et al. 2016). Furthermore, anticipating grades can be more useful than dropouts in some cases since e.g., there can be users who have interest on the course but do not manage to acquire enough skills to pass. The identification of those users and their difficulties can be done through the prediction of the final score but can be more specific if there is an analysis of every partial assignment, which will provide a better characterisation of the learner by topics. However, the most important aspect is detecting the difficulties on time because that would result in benefits both for the learner and the instructor or the institution because, for example, teachers can modify the difficulty of the course or offer complementary materials to support learners with lower initial level.

To tackle this task of forecasting learners' grades, prior to the definition of prediction algorithms, it is crucial to identify what data can be useful. For example, Ren, Rangwala, and Johri (2016) analysed different variables to predict homework grades in a MOOC and obtained the highest correlation with the number of previous quizzes attempted. Greene, Oswald, and Pomerantz (2015) also analysed variables and found a statistically significant relationship with the level of schooling and the intended hours to be spent on the MOOC. However, those contributions did not mention forum variables, which can be explained with detail to discover the relationship between grades and social behaviours.

Moreover, there might be other factors that affect the predictive power obtained when forecasting grades. Some of these factors can be the type of assignments (i.e. closed-ended vs. open-ended assignments), the time when grades are predicted or the time span when data are collected. In general, predictions can be used to detect patterns or anticipate future events. Particularly, when forecasting future scores, it is relevant to identify how much do we need to wait to obtain acceptable prediction results since the anticipation is a key aspect to change undesired behaviours on learners (Kettle and Häubl 2010). This has raised the need to analyse how predictions improve day-by-day and how older interactions can provide early diagnostics about the future. In the literature, there are not many works (Ren, Rangwala, and Johri 2016; Elbadraway et al. 2016) which predict scores quantitatively on MOOCs, and the analysis of how grades can be anticipated is an open issue.

Therefore, prediction of grades in MOOCs is at an early stage. There is a need of including new predictor variables such as forum-related ones or the type of submission. Furthermore, work on analysing how the predictive power changes along the time is required as well as the exploration and comparison of different machine learning techniques for prediction.

This work aims to address this gap and discover factors that affect learner's grades in a MOOC, such as variables (e.g., forum indicators), algorithms or the type of assignments. To cover all the purposes, the analysis has been organised in the following research questions, which are analysed in a specific MOOC:

RQ1: What is the predictive power that can be achieved to anticipate learners' grades in assignments? What machine learning methods provide better predictive power?

RQ2: Do prediction models improve with the use of forum-related variables?

RQ3: Are there any differences between predicting grades in closed-ended questions and open-ended questions?

RQ4: Do variables from previous weeks affect in the results of prediction of new ones?

RQ5: What is the moment in which the prediction power stabilises after a day-by-day analysis? What are the reasons for that?

The structure of the paper is as follows. Section 2 provides a background on what has been researched in prediction in education and, particularly in MOOCs, and how this work goes beyond the state of the art. Section 3 describes the methodology used to conduct this study. Section 4 describes the results related to the research questions, while a discussion of the main findings is presented in Section 5. Finally, the main conclusions and future research directions are provided in Section 6.

2. Related work

In this section, the existing approaches on prediction are covered and the relationship with this work. First, an introduction of prediction in education is provided. Next, a particularisation of prediction in MOOCs is presented.

2.1. Prediction in Education

Prediction is a wide area of study and there have been several contributions in the educational field, both in “closed” courses in schools or universities, and in MOOCs. In general, one of the main interests has been to predict if the student will finish (and pass) the course or not. To do that, some typical indicators are the Cumulative Grade Point Average (CGPA) (Shahiri, Husain, and Rashid 2015) or demographic variables (Aguiar et al. 2014). Lykourentzou et al. (2009) predicted students’ dropouts considering demographic information and their engagement in the learning platform as the course evolved. Results showed that the initial classification rate improved about 22-25% over the time.

Other works have focused on understanding students’ behaviours. For example, Xenos (2004) modelled behaviours using Bayesian Networks or Chen, Vorvoreanu, and Madhavan (2014) identified students’ problems, such as heavy study load, trouble sleeping, etc in students' tweets.

Apart from how students behave, there are studies about predicting learners' performance (grades). One early study (Feng, Heffernan, and Koedinger 2006) focused on predicting grades of a state exam for school students. In that work, they evaluated how the information from their intelligent tutoring system helped to predict the score and obtained a MAD (*Median Absolute Difference*) of 10.2% of the full score, which was very satisfactory.

Moreover, Ashenafi, Riccardi, and Ronchetti (2015) built a predictive model from peer-assessment activities to forecast the final grade in two computer science courses. Results showed a Root Mean Squared Error (RMSE) of 2.93 and 3.44 (in the range 18-30 as only passed students were considered, maximum grade was 30 points and 60% was required to pass) in each course, which indicates possible variances when transferring models from one course to another. Romero et al. (2013) also took forum messages and predicted students' performance based on a classification of posts considering how complete or precise the information was and what level of knowledge the learner showed with the message. It is noteworthy that forum messages include a wide variety of data since they can provide variables related to participation aspects and the text itself. However, this type of data is not always available on courses and each context should be analysed considering its particularities. Finally, the evolution of the predictive power through time should also be considered. For instance, Okubo et al. (2017) developed a model for predicting grades using neural networks, and their analysis showed how the accuracy improved from 50% in the first week to 100% in the tenth one.

2.2. Prediction in MOOCs

In the context of prediction, MOOCs have special characteristics which make them different from other types of courses. Firstly, their volume of data is greater than in traditional courses, since there can be thousands of learners enrolled in the same course. Secondly, the profile of users may be very heterogeneous (in terms of culture, education, personality, etc.), which produces a wider variety of data. Thirdly, the intensive use of videos and social interactions is a characteristic of MOOCs and not only produces more data, but also new indicators, unlike traditional courses where these interactions are not normally present.

Regarding social interactions in MOOCs, there have been approaches that identified the top contributors in the forums (Alario-Hoyos et al. 2016), the personality of learners (Chen et al. 2016), user's confusions (Yang, Kraut, and Rose 2016), the participation of learners in peer reviews (Er et al., 2017) or if there will be intervention from an instructor in a forum message or not (Chaturvedi, Goldwasser, and Daumé III 2014). All of them also have in common the use of forum data, which have also yielded some contributions to classify forum messages. For example, Brinton et al. (2014) analysed 73 MOOCs from Coursera to classify posts according to their relevance, and Ramesh et al. (2015) suggested a classification of messages according to different aspects (course, video, sentiments, etc.).

Although forum can be analysed specifically, it can provide indicators to be used for other prediction purposes and one of the most typical ones is the study of dropouts. Halawa, Greene, and Mitchell (2014) analysed learner activity features and found that absence times over three weeks were a strong indicator of dropout. Ramesh et al. (2014) also developed a model to predict course completion and found that watching video lectures activities were the most consistent in the predictions over different phases of the course.

A drawback of analysing course completion (or dropouts) is that a learner might complete the course but not pass it. Because of that, there is also an interest on predicting scores. In this area, MOOCs are also different as the way of learning may differ as well as the assessment methods. In MOOCs, predictor variables are usually related to the platform use, forum activity, video-watching activities (Ye and Biswas 2014), and the results of the assignments, although a combination of features is also possible. For example, Yang et al. (2017) used a time series neural network to predict grades based only on previous grades, and combining these with clickstream data.

Results showed how the combination of these features served to improve the RMSE. However, it is also possible to explore and define new variables. For example, Kennedy et al. (2015) focused on how many times learners switched from an exercise to another and added special assignments to evaluate their effect on the final score.

In this line, Elbadrawy et al. (2016) predicted assignment scores using different regression models and improved the RMSE from 0.225 in the first reported assignment to 0.145 in their sixth and last one. Pérez-Lemonche, Martínez-Muñoz, and Pulido-Cañabate (2017) also predicted grades in a MOOC about Android programming and achieved a MAD of around 10% with Random Forest and Neural Networks. Moreover, Ren, Rangwala, and Johri (2016) predicted intermediate assignments with RMSEs between 0.14 and 0.20 when using interactions from all the assignments excepting the used for prediction. However, many contributions predict only categorical variables (e.g., pass/fail) instead of a continuous grade (e.g., grade between 0-10). For example, Sinha and Cassell (2015) discretised grades into four groups on equal frequency: Low, Medium, High and Very High. Similarly, Li, Wang, and Wang (2017) used four levels (A to D) using N-gram features (i.e. sequences of actions) in clickstream data.

Other approaches regarding scores prediction have focused on predicting if the learner was going to answer Correctly at their First Attempt (CFA) (Brinton et al. 2016; Brinton and Chiang 2015) or performed a binary classification to forecast the correct and incorrect answer in a questionnaire using the Bayesian Knowledge Tracing model (Wang et al. 2016). Finally, there have been contributions who have only focused on what variables show a stronger correlation with the final score. For example, Vu, Pattison, and Robins (2015) found a strong correlation with the total video view time.

Despite all the efforts to develop the previous contributions, there is a lot of room for research in prediction in MOOCs. Authors (2018) reviewed the state-of-the-art in prediction in MOOCs and found some future research directions. These shall include the need to analyse more heterogeneous courses in terms of platforms, thematic areas and durations. Furthermore, it will also be useful to focus on new variables to predict (e.g., learners' expectances, efficiency, constancy, etc.), new features as predictor variables (e.g., self-regulated learning variables, forum-related ones, etc.), the analysis of the best moments for making the prediction, or the improvement of the predictive power of current models through new indicators or the enhancement of algorithms.

This paper will show a study on how scores can be predicted on a MOOC raising the best results possible, which will be contrasted with current contributions in Section 4. This study will be different to other works because it will analyse the effect of new variables that have not been used when predicting assignment scores, like forum ones. Besides, the area of predicting scores is in its early stages in MOOCs, and the analysis of intermediate assignments instead of only a final grade, together with the specific context of this course, are particular of this contribution. Furthermore, the study will also conduct a temporal analysis to discover how grades can be anticipated, which is a novel contribution in MOOCs, aimed at discovering patterns that will be useful for the improvement of the design and development of future courses. Finally, it is important to note how this contribution serves to transform current research. Researchers will be able to identify the relationships between the features/algorithms and the predictive power; and they will be able to apply that knowledge in their prediction. Moreover, the findings related with timing considerations will serve researchers to guide their design processes when preparing appropriate interventions (e.g., warning systems) to make effect on learners.

3. Methodology

3.1. Case study

This study was conducted on a MOOC about programming called *Introduction to Programming with Java – Part 1: Starting to Program in Java* (see characteristics in Table 1). This was the first one of a trilogy of courses developed by Universidad Carlos III de Madrid (UC3M) for learning Java from scratch. The MOOC was organised in five weeks where learners had to visualise videos and take activities prior to weekly graded assignments. In each week (i), there was a close-ended graded test (T_i), which represented the 15% of the final grade (FG), and there were also two (open-ended) graded programming assignments (P_i) in weeks 3 and 5, which counted 10% and 15% respectively. This type of assignments was assessed based on peer-review following a common scoring rubric.

Table 1. Characteristics of the MOOC

Platform	edX
Duration	5 weeks
Dates	28/04/2015 – 30/06/2015
Methodology	Instructor-based
Time between assessment is opened and deadline	2 weeks (except week 1 where there are 3 weeks)
Passing grade	60%
Enrolled learners	95,555
Leaners who played at least a video	24,055
Passed learners	1,507
Syllabus	Week 1: Programming basics and making decision when implementing algorithms Week 2: Representation of data and programs, and conditional repetitions Week 3: Organisation of code in methods and recursion Week 4: Object-oriented programming Week 5: Packaging

Regarding the topics, although they are related, there is certain independency among them; some of them require more computational thinking on how to develop algorithms, while others require more abstraction abilities. All these characteristics are of special importance because results of this analysis may be extended to courses with similar contexts.

3.2. Data collection

This work used the available information provided from edX. Particularly, four files of data have been used. Except for the grades, the rest of them were retrieved from the Database Data (edX 2017). The information of these files is as follows:

- *{org}-{course}-{run}-course_structure-{site}-analytics.json* contains the structure of the course with the reference of all its items. It has been used to map course components to the week they belong to be able to filter data per week.
- *{org}-{course}-{run}-courseware_studentmodule-{site}-analytics.sql* contains the state of each learner in each of the course components that has been accessed. It has been used to retrieve variables about exercises and video interactions.
- *{org}-{course}-{run}-{site}.mongo* contains the forum messages and the characteristics of course discussion interactions. It has been used to retrieve variables about how learners engage with the forum.

The last file used is the grade report available from the instructor dashboard of the course. This file is named with format *{course_id}-grade_report_{datetime}.csv* and contains the grades (dependent variables in this work) of all graded assignments.

Data from these files provide information of almost 100,000 learners. However, most of them have not taken assignments and have not been engaged in the course. If all of them were taken for the analysis, results would be significantly biased. Because of that, it was necessary to design a sample selection criteria which reduced considerably the number of learners (cases in the experiments). These criteria can be explained in two parts. Firstly, as one part of the analysis will focus on forum-related variables and the number of participants in the forum reduces significantly the total number of users, only learners who have contributed at least once in the forum (by posting a message) have been considered. This will also be useful to focus on learners who have been active on the course as Anderson et al. (2014) found that 90% of active users in the MOOC are also active in the forum.

Secondly, as the aim is predicting grades, users who have unenrolled the course, and thus there are not scores about them, have been excluded, although they may have participated in the forum. For the rest of the learners, the grade report always provides a grade that may be zero if the learner has not doing anything and the intention is to predict also if the learner will not complete the assignment. After this filter, there is a total of 4,358 learners out of the 95,555 users mentioned before, so a big proportion of learners who have not been engaged with the course has been removed.

3.3. Variables and techniques

3.3.1. Selected variables

Once data from edX have been retrieved, high-level variables have been obtained by processing that information to be used for prediction. These variables are collected from the beginning of the course or from the beginning of a specific week (depending on the experiment), but they are always collected before the deadline of the assignment considered for prediction. Table 2 shows the list of these variables (all of them are continuous):

Table 2. Features used for the assignment prediction

FORUM VARIABLES	
Number of participations	No. messages a learner posted
Number of threads	No. threads a user stared
Number of comments	No. answers in threads a user posted
Number of received votes	No. votes a learner received from other users
Number of emitted votes	No. positive votes a learner emitted to messages from others
Number of endorsed messages	No. messages flagged by the instructor or the message originator because of their value/relevance
Average length	Average number of characters of the messages posted by the learner considered
Contribution percentile	Percentile obtained after sorting learners according to the number of messages posted
Sentiment profile	Value between 0 (negative) and 1 (positive) indicating the average positivity of messages. This indicator has been obtained using Random Forest and three indicators: message length, difference of days between the message date and the beginning of the course, and an *orientation* variable. This variable increased or decreased when positive/negative words were found in dictionaries compiled by Hu and Liu (2004), and considered negations and emoticons' meanings.
EXERCISES VARIABLES (from non-grades activities)	
Average grade	Average grade of non-graded activities in the week(s) considered. Non-attempted activities count as 0.
Exercises attempted	No. exercises attempted over the total in the week(s) considered
Exercises opened	No. exercises access (although not submitted) over the total in the week(s) considered
Average number of attempts	Average number of attempts in the exercises attempted (not only accessed) of the week(s) considered
Number of correct exercises	No. exercises with 100% correct over the total in the week(s) considered
CFA (Correct at First Attempt)	No. 100% correct exercises in the first attempt over the total in the week(s) considered
VIDEO VARIABLES	
Percentage of videos opened	Percentage of videos the learner started watching over the total in the week(s) considered
GRADES VARIABLES	
Previous grades	Grades from previous graded assignments (tests or peer-review assignments) whenever available (e.g., week 1 has no previous grades)

The dependent variables will be the results of the five graded tests, the two peer-review programming assignments and the final grade. All these variables are continuous in the range 0-1. Furthermore, the final grade will be analysed as a categorical variable with two binary values: pass (grade above 0.6) or fail.

3.3.2. Techniques and metrics

Predictor variables can be useful to forecast grades, but a subjacent model is also needed to make the relationship between indicators and scores. In this work, four algorithms have been considered and their results have been contrasted to know which one provides the best result in each case. These algorithms are the following: (1) Regression (RG), (2) Support Vector Machines (SVM), (3) Decision Trees (DT), and (4) Random Forest (RF).

Once results are obtained using 10-fold cross validation with these algorithms, it is important to identify how to assess their performance. In Section 2, most common metrics for predicting continuous grades were MAD and RMSE. However, Pelánek (2015) criticised the first one as it prefers models which are biased towards the majority result and suggested the use of RMSE. Because of that, results will be reported with RMSE, although MAD is also mentioned to compare the results with other works.

Finally, when the binary case of predicting if learners pass or fail is considered, the same algorithms will be used, but in their classification version. For the case of metrics, F-Score and AUC will be used as they are the most common in the literature and they have also been accepted in Pelánek's work.

4. Results

This section is divided in five parts to address each of the five research questions that were stated in Section 1. Each subsection explains the analysis carried out to solve the question and a discussion about the applicability of the findings.

4.1. RQ1: Predictive power in course assignments and machine learning techniques

The first part of the analysis consists of a model to predict grades for each of the seven graded assignments of the 4,358 learners using exercise-related variables, video-related variables and previous grades (see Table 2), when available. In this case, each assignment has been predicted using data from interactions of that task, so activities have been filtered to only consider those related to the specific assignment and that happen before the deadline. This period normally covers 15 days, which is the time span between the release date and the deadline (activities are usually released on Tuesdays and closed after two weeks including the first and last Tuesday, which means there are 15 days and not 14). Regarding previous grades, as they are not non-graded interactions, their effect will be assessed separately.

Considering the four algorithms mentioned in Section 3.3.2, results of prediction, evaluated through the RMSE, are presented in Table 3. In this table, values have been separated depending on if only exercise and video variables are used without previous grades (Model A) or if all these three types of variables are used (Model B). An initial observation is that earlier assignments are more difficult to predict. There is only an RMSE of 0.26 in the first test of the course, while the metric improves to 0.13 in the last test (using RF with previous grades in both cases). As there is only data about the specific week intended to predict, this result implies that people who interact at the beginning of the course are not necessarily committed to doing the summative assignments, considering in each case the best result for the four machine learning methods used. However, people who remain in the latter weeks may have more predictable behaviours. Moreover, results show that previous grades always improve the results (when available). This entails that learners might tend to have a similar level of achievement among assignments.

Table 3. Predictive power in course assignment using different machine learning techniques (indicated with RMSE)

Method		T1	T2	T3	T4	T5	P3	P5	FG
Model A	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**
Model B	RG	0.27	0.21	**0.18**	**0.15**	**0.13**	**0.24**	**0.19**	**-**
	SVM	0.27	**0.20**	**0.18**	**0.15**	**0.13**	**0.24**	0.20	-
	DT	0.34	0.26	0.23	0.20	0.17	0.32	0.26	-
	RF	**0.26**	**0.20**	**0.18**	**0.15**	**0.13**	**0.24**	**0.19**	**-**

Regarding the algorithms, RG, SVM and RF have similar performance and better than DT. Among these three, after performing paired tests, there is not statistically significant difference between RG and RF (p-value = 0.14), but both are significantly better than SVM, although the difference is relatively small. Therefore, it is better focusing on data rather than on the algorithms, as there is a low difference of results between them, while the variances are higher when modifying the indicators.

As for the values, it is not possible to compare directly with other contributions from the literature because the contexts are different and, in most cases, they do not refer to MOOCs; still it is interesting to check what others have achieved to take them as a reference. In the most similar works (Ren, Rangwala, and Johri 2016; Elbadraway et al. 2016), authors managed to get an RMSE between 0.14-0.15 in their best assignments. Here, it is only possible to reach those values in tests 4 and 5, but the RMSE is better in the last one. In terms of MAD, although less recommendable (Pelánek 2015), it is possible to get 0.09 with the final grade and 0.05 in test 5, which are also better than the 0.10 presented by Feng, Heffernan, and Koedinger (2006). Therefore, although differences with other works could be affected by differences in contexts, results in this study are more than acceptable.

Finally, this analysis has also considered the final grade as a dependent variable. In this case, previous grades have always been excluded because this grade is the average of them, but it is possible to see how learners' interactions can predict the final grade. The best result was 0.14 (RG and RF), which is better than the one for any single assignment when previous grades are excluded.

Furthermore, the grade was also discretised in *pass* and *fail*, and a binary classification was performed using the four algorithms. The best model was obtained with SVM (AUC = 0.97, F-score = 0.82, accuracy = 0.93). Ruipérez-Valiente et al. (2017) analysed how these metrics evolved over seven weeks when predicting certificate earners by adding data successively. While they got an AUC of nearly 1.0 and an F-score of 0.95 after seven weeks, they only obtained an AUC of 0.95 and F-score of 0.78 after the fifth week, which is the last one in the case study presented here. Therefore, the duration of the course can be also a factor which affects prediction results. In fact, in this case, AUC was only 0.83 after week 1 and 0.95 after week 3; while F-Score was 0.43 and 0.70, respectively, as it is shown in Figure 1. Nevertheless, final results show that, in most cases, it is possible to forecast who will pass the course, as it is the intention.

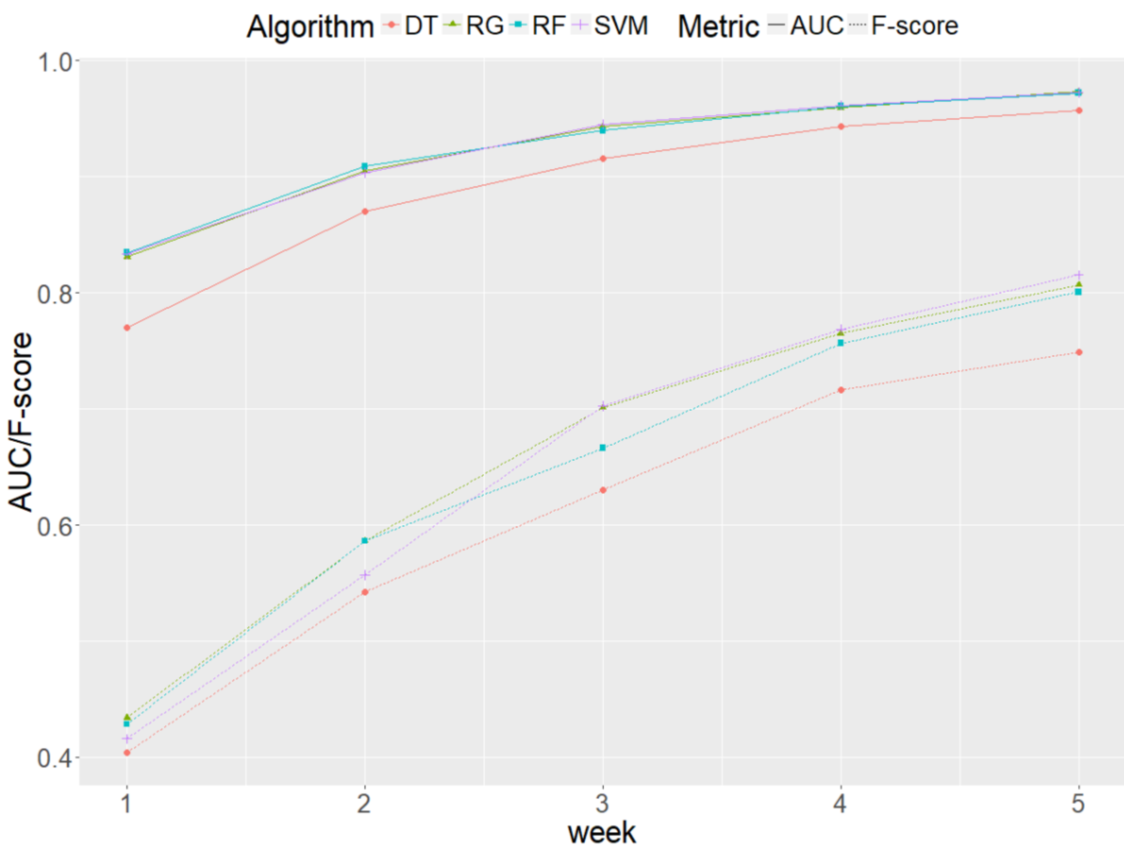

Figure 1. Evaluation of results of final grades prediction over time.

4.2. RQ2: Effect of forum-related variables in prediction

The models presented in Section 4.1 only included interactions related to videos and exercises. However, MOOCs provide much more information, and forum interactions can be also considered to check their effect in the final grades. Previous contributions in the line of predicting test scores did not focus on forum indicators and only variables related to the number of contributions were occasionally mentioned. In this case, three models (for each of the four algorithms) will be developed to see the predictive power when using forum-related variables.

The first model (model C) only includes forum-related variables (see Table 2). This model is enhanced with exercises and video variables in the second model (model D), and with previous grades in the third model (model E). Results of these models are presented in Table 4. It should be noted that results of final grade have been removed in model E because that model includes previous grades, which cannot be used because their weighted average is directly the final grade.

Table 4. Predictive power in course assignment after adding forum-related variables (indicated with RMSE)

Method		T1	T2	T3	T4	T5	P3	P5	FG
Model C	RG	0.41	0.36	0.33	0.31	0.27	0.34	0.24	0.26
	SVM	0.46	0.40	0.35	0.33	0.28	0.34	0.26	0.28
	DT	0.46	0.40	0.35	0.33	0.30	0.36	0.28	0.27
	RF	0.42	0.37	0.33	0.31	0.27	0.34	0.25	0.25
Model D	RG	0.26	0.22	0.20	0.18	0.16	0.25	0.20	**0.14**
	SVM	0.26	0.22	0.21	0.19	0.17	0.26	0.21	0.15
	DT	0.34	0.28	0.26	0.23	0.19	0.32	0.28	0.17
	RF	0.25	0.21	0.20	0.19	0.16	0.25	0.20	**0.14**
Model E	RG	0.26	0.21	**0.18**	**0.15**	**0.13**	**0.24**	**0.19**	-
	SVM	0.26	**0.20**	**0.18**	**0.15**	0.14	0.25	0.20	-
	DT	0.34	0.26	0.23	0.20	0.17	0.32	0.26	-
	RF	**0.25**	**0.20**	**0.18**	**0.15**	**0.13**	**0.24**	**0.19**	-

Results show that forum-related variables have a weak predictive power when forecasting grades. The best result obtained is in peer-review 5, where the RMSE is 0.24, which is far from what it was obtained in Section 4.1. To corroborate the poor predictive power mentioned, two baseline values were computed. The first one consists on assigning random values to each grade, which produced an RMSE between 0.49 and 0.53, which is worse than the performance of forum variables. However, the second baseline consisted on assigning the average grade to all learners. Results gave RMSEs in the range 0.34-0.40, which implies that forum variables do not improve so much (if any) those baselines.

Regarding models D and E, their results can be compared with those presented in Table 3 (models A and B) because the only difference is the inclusion of forum variables. The comparison reveals that most of the values are equal (and even the best algorithms are the same) and a paired t-test shows that there is not statistically significant difference between including forum-related variables or not (p-value = 0.98). Therefore, the conclusion is that forum activity is not related to learners' grades.

This finding can be justified with contributions which state that there are many good students that tend to be individualist (Silverman 1990; Valadez Sierra et al. 2015). If this occurs, learners may not contribute in the forum despite they are understanding the contents and achieving passing grades. In the MOOC analysed in this paper, from those who contributed at least once in the forum, only 754 passed the course (out of 1,507 learners who passed in total). This means that 49.97% of users who passed the course did not write a single message and 3,604 users wrote messages but failed. Because of that, it is reasonable to think that although there can be many forum-related variables, they do not provide a significant effect when predicting grades.

4.3. RQ3: Differences between predicting grades on close-ended and open-ended questions

Third and fifth weeks of the course, unlike the others, have two assignments on the same week. One of them is a closed-ended task, which consists on a test with theoretical questions and small practical exercises about the content explained on the videos. The format of these questions can be very different and some common types of problems are multiple-choice questions, checkbox questions, numerical input questions, etc. In addition, there is an open-ended activity, which consists on a practical programming task where learners implement a program in Java and submit the code to be reviewed by peers using a scoring rubric; the grade is obtained from the scores given by peers.

Therefore, because of these differences, it is worth analysing the difference in prediction for both types of assignments.

As both assignments belong to the same week, prediction models use the same interactions for forecasting both grades. To analyse the differences between the models, the best results of prediction for each assignment are summarised in Table 5, where the first column is taken from model C, and the second and third columns use models A and B, respectively, as it was shown that forum variables did not improve significantly those models (as analysed in models D and E).

Table 5. Comparison between RMSEs obtained in closed-ended and open-ended assignments in weeks 3 and 5

Assignment	Forum (Model C)	Problems and video (Model A)	Problems, video and grades (Model B)
Test 3	0.33	0.20	**0.18**
Peer-review 3	0.34	0.25	0.24
Test 5	0.27	**0.16**	0.13
Peer-review 5	**0.25**	0.20	0.19

Firstly, the predictive power using forum-related variables is considered. Although RMSE values are poor when using those indicators (as explained in Section 4.2), it is noteworthy that there is no sign that indicates differences among types of assignments. However, when considering the rest of variables excluding forum ones (categories 2, 3 and 4), results show a clear improvement of the predictive power when forecasting closed-ended activities. A possible explanation is that learners can be more likely to ask questions when they have trouble doing a practical open-ended activity because it is more difficult to use a trial and error strategy than in a typical closed-ended exercise (i.e. multiple-choice question) where learners can try different options until they get the correct one.

Despite the small differences with forum variables, results show in both weeks (using models A and B) a statistically significant improvement in the predictive power when forecasting a graded test (p-value < 0.05). Regarding exercises variables, it is reasonable that they have a stronger relationship with graded tests because they are taken from closed-ended formative activities, with similar format to the summative closed-ended tests. For the case of practical assignments, learners can practise on coding using an external tool called Codeboard, which is seamlessly but loosely integrated in edX, and they do not have the option to submit and assess the code automatically. The use of such external tool precludes from matching users across the tool and the edX platform, which supposes a lack of exercises variables related to these formative activities. For these formative activities, learners used the forum to ask about the solution, which makes that forum variables have similar predictive power in both closed-ended and open-ended tasks.

Another fact which facilitates that closed-ended grades can be predicted better is the way both types of tasks are assessed. For the case of graded tests, there is an objective evaluation, which means equal answers receive the same grade. However, programming assignments are graded based on a scoring rubric, which tries to be as objective as possible, but there is always some implicit subjectivity in the process. Furthermore, graders are learners instead of teachers. Sajjadi, Alamgir, and von Luxburg (2016) showed that in peer-review activities, there can be a high variance among grades and the wide sources of grading errors may decrease the reliability of grades. Because of that, the grade in open-ended questions will be harder to predict than in closed-ended ones.

Finally, it is noteworthy that results of this research question can be more dependent on the context than those from previous research questions as these results highly depend on the characteristics of the activities. For example, if open-ended assignments could be somehow assessed automatically, subjectivity errors would be reduced. In the opposite way, if the task was even more subjective like assessing a poetry or a narrative story, the variance between grades when assessing equal works could be higher. Therefore, it is important to consider the MOOC context to identify possible changes in behaviour to what it has been obtained here.

4.4. RQ4: Effect of variables from previous weeks to predict grades

In the previous analysis, data used to create the predictive models included all the interactions before the deadline of the assignment corresponding only to the week (actually 15 days because contents are released one week in advance) where learners had to do the task. However, waiting until the very last moment is not desirable because there will not be time to react in case of problems. Because of that, it can be interesting to analyse how data from other topics (other weeks in the course) can be useful to predict assignments.

Apart from that, there is a question regarding what interval of time should be considered to gather data for making predictions. If the aim is to forecast the score of the first week test, there is no other option than taking data from that week. However, when predicting the score of the last test, there are two possibilities: using only data since the previous assignment until the current one (as done in previous sections) or using all the information available since the beginning of the course.

Regarding this question, an analysis considering all the assignments has been conducted to discover if it is better to use all available data (cumulative mode) or just data from the previous topic (non-cumulative). It is important to mention that previous data will not add new variables but will modify the values of the indicators instead. In this analysis, exercises and video variables have been used as forum variables do not improve significantly the models, as discussed in Section 4.2. Previous grades have been removed to be fair in the number of variables (in some weeks there can be more previous grades than in others). Thereby, indicators will have different values but the number of them will be constant. Considering the variables used, the non-cumulative mode is equivalent to model A and the cumulative mode with the new model F. The four algorithms presented have been used to predict grades for each assignment (and the final grade) in both modes and the results are presented in Table 6.

Table 6. Predictive power (in RMSE) using only data of current week (model A) or data since the beginning of the course (model F)

Method		**T1**	**T2**	**T3**	**T4**	**T5**	**P3**	**P5**	**FG**
Model A Non-cumulative	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
Model F Cumulative	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**

Results show that the RMSE is lower in the non-cumulative mode in all assignments (excluding test 1 where data are the same for both modes). This means that obtaining the indicators with more data is not always better and, in this case, it is preferable to use only interactions in the current topic. However, when predicting the final grade, the result varies among algorithms and values are very similar in both modes. A possible explanation is that as the final grade is the weighted average of all the assignments, there are better predictions using all data. Nevertheless, the latest interactions are more relevant as there are learners who start doing activities but then they drop out, which makes non-cumulative mode has a similar predictive power respect to the cumulative mode.

Another observation is that the difference between modes is higher as weeks evolve. The reason might be that in the first weeks, previous data are more recent and may have a stronger relationship than in the last weeks, where the first topics took place some time ago. In terms of content, results mean that there is certain independence among units and a learner can master some contents much better than others. In this course, this sounds reasonable because there is a gap between the first three topics and the last two. However, the overall content is incremental (as it is the difficulty) and results might be different in another course with more (in)dependency among topics.

Regarding the algorithms, results show that in the cumulative mode, the best algorithm is RF in all situations. The difference with the rest is statistically significant (p-value < 0.01) after comparing with each algorithm, although the RMSEs of RG are similar. As RF is also one of the best options in non-cumulative mode, the rest of this analysis will focus only on this algorithm and results will not be significantly affected.

If the initial aim is to anticipate learner grades, it is possible to take data from only previous topics to forecast later grades. The performance will not be as higher as it should be using the current topic, but results will be available earlier. When doing that, one can also consider adding data from more than one week, if available. Previous analysis showed that the non-cumulative mode was better, but it might be different when all data belong to previous weeks. To analyse this, the RMSE has been computed using each week to predict grades of later weeks. Furthermore, there is a cumulative mode in which data are used since the beginning of the course until the current week considered to forecast a later assignment. Figure 2 shows the results of this analysis and how RMSE evolve over time.

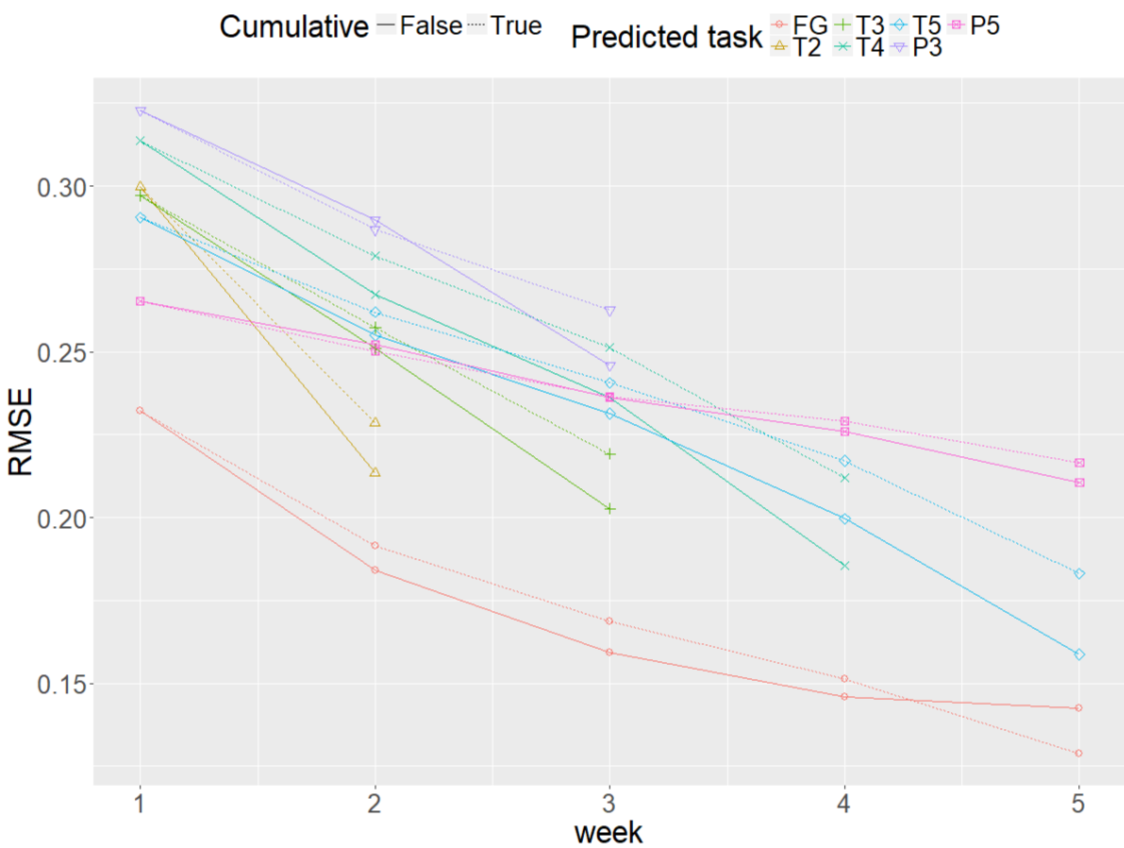

Figure 2. Grades prediction using data from previous topics

In the figure, when predicting closed-ended assignments, the best option is using previous topics without considering all data since the beginning. This might suppose that immediate previous contents have a stronger relationship and they are more useful to predict grades. In contrast, in open-ended assignments, it is better using as much previous data at the beginning, but there is a point where results improve by using only data from the last week. Regarding the final grade, it is surprising that the non-cumulative mode has better results (except in week 5). A possible reason is that there can be many learners at the beginning who do not continue the course and these interactions may affect negatively in predictions. The conclusion after this analysis is that data from previous topics can be used (although their performance is obviously worse) to anticipate what learners will achieve. Nevertheless, as topics might have certain independency and the engagement of users over the time can vary considerably, results would be better if data are collected only in the time span where the assignment to be predicted is opened for submissions. However, this may be different in another course with a stronger relationship among topics.

4.5. RQ5. Stabilisation of predictive power in a day-by-day analysis

In the previous question, it was analysed how it was possible to anticipate results using data from previous topics. However, RMSEs are lower when only interactions from the current topic are used. The problem is that it is not very useful to wait until the week finishes because in that moment, actual grades will be available. Because of that, the intention is to analyse how the predictive power improves day-by-day in the week where the topic is covered.

This analysis has been conducted by adding data successively and computing the RMSE. Firstly, data are gathered from the first day of the week and then data are added day-by-day until the deadline. As graded activities are usually opened with one week in advance and for a time span of two weeks (except in the first topic where there are three weeks of margin), there are normally 15 days to analyse the evolution. It does not matter if a learner finishes the graded assignment before the deadline because it is assumed that this is the last activity the learner will be doing in that week, which sounds logical, as the instructor presents a learning sequence with ends up with a summative activity.

In this analysis, it is also interesting to see when there is enough data in the week considered to create a model with better predictive power than the model obtained with data from previous weeks. Figure 3 shows the evolution of the RMSE day-by-day and the moment where data from the current topic improve the model obtained using data from the previous topic. As it was justified in Section 4.4, the algorithm used is RF and data include exercises and video-related variables.

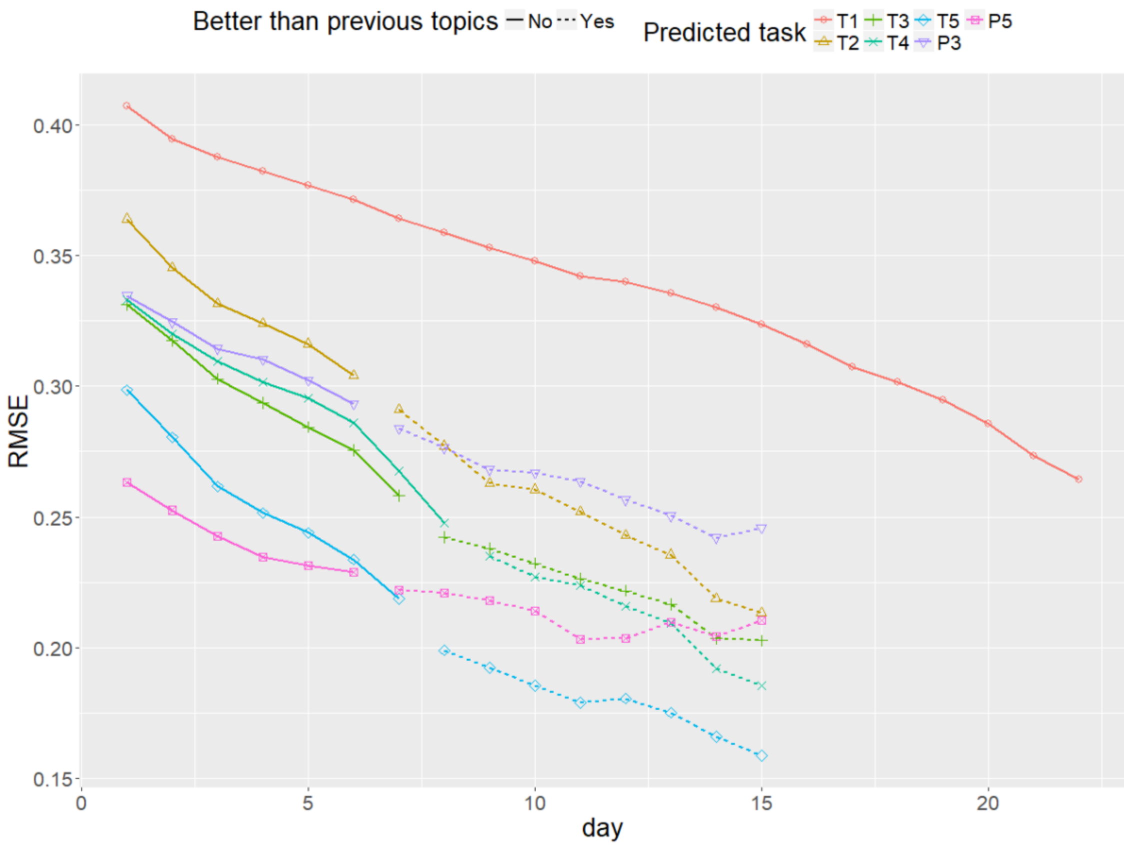

Figure 3. Evolution of the predictive power day-by-day

Results show a continuous improvement as days go by. In all cases, the moment where RMSE improves the results obtained with previous data is between day 7 and day 9. This means that the first week when materials are opened is not very useful for predicting grades. This is reasonable because this first week overlaps with the second (and last) week for the previous topic. Therefore, learners can be more focused on finishing the tasks with the nearest deadline than on working in the materials for the new released topic.

Another observation is that there is not a clear stabilisation of the predictive power, and values progressively improve until the end of the analysed period. This improvement is nearly linear in some cases like in the first topic, although there are single days where differences are low. Consequently, within the period considered, it is possible to anticipate grades, but the period is short enough to cause that more data still make an improvement; this means that there is a trade-off between having to wait more to obtain better results and the improvement of the predictive power. Furthermore, the fact that learners can take the assignments in the last moment might contribute to this fact. However, in open-ended assignments, the best model does not include the whole period. This implies that interactions in the last minute to finish programming tasks might not be enough to make them work.

5. Discussion

5.1. Lessons learned

The analysis of the five research questions has served to obtain some findings related to how to predict assignment grades in a MOOC. The summary of these findings is as follows:

(1) **Early assignments are harder to predict**. Predictive models improve considerably when assignments of the course are towards the end. This means that interactions at the beginning do not necessarily reflect what learners will achieve in the graded assignments.

(2) **Machine learning algorithms are important, but data make the difference**. Results show that differences in performances are not high, particularly with RF and RG, which are the best algorithms in the study. In contrast, modifications in the data used can vary significantly the results. Therefore, more efforts should be put on designing models with new indicators rather than on creating algorithms.

(3) **Previous grades always improve the predictive power of the models**. The RMSE always decrease when previous grades are available. This supposes that learners tend to have a similar level of achievement among assignments and that these variables should be used whenever possible.

(4) **Forum-related variables do not improve significantly the predictive models.** Although several variables related to forum interactions were collected, they do not enhance the predictive power. Furthermore, half of learners who passed did not post any message, so it is not worth spending a lot of time gathering this type of variables.

(5) **It is easier to predict closed-ended assignments than open-ended assignments.** As non-graded exercises are typically closed-ended assignments, models to forecast tests achieve a higher predictive power than those used for open-ended assignments with the same data.

(6) **Adding interactions from previous topics to create the indicators makes predictive models worse**. If data are not available from the current topic, previous data can be useful to anticipate results. However, models whose indicators only include data from the current topic are better than those which mix interactions from the current week and the previous ones.

(7) **Data from nearest previous weeks have stronger relationship with current grades**. When there are only data from previous topics, the best option is to choose only interaction from the last available week instead of using all the available interactions. The possible reasons are the variances in engagement over the time and the independence among topics.

(8) **Interactions from the current weeks become relevant after approximately seven days.** Although data from the current topic can achieve the best results,

there are not enough data until seven days of the release of materials to be able to improve the performance obtained using data from previous topics. For this result, it must be considered that contents are opened two weeks before the deadline, so seven days is half of the time to complete the assignment.

The abovementioned findings can be useful not only to advance current research, but also to transform related educational practices. Particularly, our findings can be applied to detect learners who have difficulties in the course and knowing the predictive power at the different moments of a course. If the predictions are carried out at the right time, it is possible to warn students at risk, or design other interventions (e.g., personalised support or proper methodologies) that help students achieve the expected goals. Therefore, instructors should use the predictive models in real courses (while the course is being developed) and have an effect on their learners.

5.2. Limitations

Despite being able to obtain the findings mentioned in Section 5.1, this work is not exempt of some limitations that are worth mentioning. This subsection outlines the main limitations and the way they have been addressed, or the way they should be tackled in the future.

One limitation related to data was that tracking logs, which includes all the information of learners' events in the platform, were not completely available for this course, so there was not a complete information about users' interactions. For example, information about video replays was not available. Therefore, the prediction is done based on indicators from all the available sources that can be calculated. The inclusion of new indicators might increase the prediction power and it would be worth analysing how the results might change. However, in quite a few occasions not all the information about the interactions might be available and prediction using only these data is very relevant.

Apart from data, another restriction was the sample selection criteria. This work considered data from learners who participated at least once in the forum. Therefore, this work presents prediction results for students that interacted in the forum but not for the rest of users. This was also a way to filter users who did not interact enough in the course and that could bias the results. However, it is likely that there are learners who post messages but do not participate in assignments (or vice-versa) as it has been showed that the predictive power of forum variables is not high. Because of that, after performing the analysis, it seems the sample selection criteria might be modified by using a different filter of learners. This gives opportunities for further research on analysing who the real active learners are, and their relationship with grades.

It is also noteworthy that these findings have been obtained in a specific MOOC about Java programming, with an instructor-based methodology. This means that the conclusions of this work can be extended to online courses with similar characteristics, but there might be differences in other contexts. The problem of generalisation is common in learning analytics, and because of that, further research can be done to consider new educational scenarios (e.g., changing the topics of the course, the duration or the methodology). Nonetheless, this work includes novel approaches and considerations regarding how to predict grades that are useful to know the best ways to anticipate what learners will achieve in the course.

6. Conclusions and future work

In this work, an analysis on prediction of grades in a MOOC has been conducted. Predictor variables were calculated using data files provided by edX. These indicators included variables related to the forum, exercises, videos and previous grades. Models that predict scores in different assignments of the course, and the final grade have also been proposed with a best RMSE of 0.126 in the case of assignments (test 5 with RF and model E) and 0.129 in the case of the final grade (with RF and model F), which are more than acceptable results considering what has been previously achieved in other educational contexts.

Nevertheless, from this work, new opportunities for research also arise. As it has been mentioned there might be a variance of results when using MOOCs from other contexts, it will be interesting to assess the applicability of these findings in other courses with different characteristics. For example, self-paced courses where there are not limited periods to complete tasks, courses where open-ended assignments are assessed automatically, where the evaluation is based only on a final exam, or where topics are very different among them.

Furthermore, as future work, it will be interesting to develop models with learners who do not participate in the forum but are engaged in other types of activities to make a comparison with current results. This will allow establishing better understanding on who an active student is indeed, and the relationship among students' active participation through the MOOC and their final performance.

Moreover, it will be useful to enhance models to differentiate learners who fail because they have trouble understanding course contents and learners who are enrolled without commitment to do the activities. In addition, it will be relevant to put the prediction models into practice in a real course while it is under development and analyse how anticipating grades can be useful to improve learning outcomes and to design suitable interventions that can improve current educational practices.

Apart from that, as there were limitations with data from edX, one possible work would be using courses where there is control about all traces of the platform, which will allow designing new indicators based on user events, like how many times the user accesses the course, how many time learners spend doing activities, and so on. This will be useful to enhance models and add new dimensions to the current discussions, which have served to discover how grades can be predicted on a MOOC, and how anticipation can be done more efficiently to advance in the state of the art and achieve a better understanding of the MOOC and its development.

7. References

Aguiar, Everaldo, G. Alex Ambrose, Nitesh V. Chawla, Victoria Goodrich, and Jay Brockman. 2014. "Engagement vs Performance: Using Electronic Portfolios to Predict First Semester Engineering Student Persistence". *Journal of Learning Analytics* 1 (3): 7-33. doi:10.18608/jla.2014.13.3.

Alario-Hoyos, Carlos, Pedro J. Muñoz-Merino, Mar Pérez-Sanagustín, Carlos Delgado Kloos, and Hugo A. Parada G. 2016. "Who Are the Top Contributors in a MOOC? Relating Participants' Performance and Contributions". *Journal of Computer Assisted Learning* 32 (3): 232-243. doi:10.1111/jcal.12127.

Anderson, Ashton, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. "Engaging with Massive Online Courses". In *Proceedings of the 23rd International Conference on World Wide Web*, 687-698. New York: ACM.

Ashenafi, Michael Mogessie, Giuseppe Riccardi, and Marco Ronchetti. 2015. "Predicting Students' Final Exam Scores from their Course Activities". In *Proceedings of the 45th Annual Frontiers in Education Conference*, 1-9. Piscataway: IEEE.

Autors (Blinded for peer-review)

Brinton, Christopher G., Swapna Buccapatnam, Mung Chiang, and H. Vincent Poor. 2016. "Mining MOOC Clickstreams: Video-Watching Behavior Vs. In-Video Quiz Performance". *IEEE Transactions on Signal Processing* 64 (14): 3677-3692. doi:10.1109/tsp.2016.2546228.

Brinton, Christopher G., and Mung Chiang. 2015. "MOOC Performance Prediction via Clickstream Data and Social Learning Networks. In *Proceedings of the 34th Annual IEEE Conference on Computer Communications*, 2299-2307. Piscataway: IEEE.

Brinton, Christopher G., Mung Chiang, Shaili Jain, Henry Lam, Zhenming Liu, and Felix Ming Fai Wong. 2014. "Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model". *IEEE Transactions on Learning Technologies* 7 (4): 346-359. doi:10.1109/tlt.2014.2337900.

Chaturvedi, Snigdha, Dan Goldwasser, and Hal Daumé III. 2014. "Predicting Instructor's Intervention in MOOC forums". In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 1501-1511. Stroudsburg: Association for Computational Linguistics (ACL).

Chen, Guanliang, Dan Davis, Claudia Hauff, and Geert-Jan Houben. 2016. "On the Impact of Personality in Massive Open Online Learning". In *Proceedings of the 24th Conference on User Modeling Adaptation and Personalization*, 121-130. New York: ACM.

Chen, Xin, Mihaela Vorvoreanu, and Krishna P.C. Madhavan. 2014. "Mining Social Media Data for Understanding Students' Learning Experiences". *IEEE Transactions on Learning Technologies* 7 (3): 246-259. doi:10.1109/tlt.2013.2296520.

Christensen, Gayle, Andrew Steinmetz, Brandon Alcorn, Amy Bennett, Deirdre Woods, and Ezekiel J. Emanuel. 2013. "The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why?". *SSRN Electronic Journal*. doi:10.2139/ssrn.2350964.

Daniel, John. 2012. "Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility". *Journal of Interactive Media in Education* 2012 (3): 18. doi:10.5334/2012-18.

edX. 2017. "EdX Research Guide Release". *edX*, August 25. https://media.readthedocs.org/pdf/devdata/latest/devdata.pdf

Elbadrawy, Asmaa, Agoritsa Polyzou, Zhiyun Ren, Mackenzie Sweeney, George Karypis, and Huzefa Rangwala. 2016. "Predicting Student Performance Using Personalized Analytics". *Computer* 49 (4): 61-69. doi:10.1109/mc.2016.119.

Er, Erkan, Miguel L. Bote-Lorenzo, Eduardo Gómez-Sánchez, Yannis A. Dimitriadis, Juan Ignacio Asensio-Pérez. 2017. "Predicting Student Participation in Peer Reviews in MOOCs". In *Proceedings of the 5th European MOOCs Stakeholders Summit*, 65-70. Cham: Springer.

Feng, Mingyu, Neil Heffernan, and Kenneth Koedinger. 2006. "Predicting State Test Scores Better with Intelligent Tutoring Systems: Developing Metrics to Measure

Assistance Required". In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 31-40. Berlin: Springer.

Grainger, Barney. 2013. "Massive Open Online Course (MOOC) Report 2013". *University of London*.

Greller, Wolfgang, and Hendrik Drachsler. 2012. "Translating Learning into Numbers: A Generic Framework for Learning Analytics". *Journal of Educational Technology & Society* 15 (3): 42.

Greene, Jeffrey A., Christopher A. Oswald, and Jeffrey Pomerantz. 2015. "Predictors of Retention and Achievement in a Massive Open Online Course". *American Educational Research Journal* 52 (5): 925-955. doi:10.3102/0002831215584621.

Halawa, Sherif, Daniel Greene, and John Mitchell. 2014. "Dropout Prediction in MOOCs using Learner Activity Features". *Experiences and best practices in and around MOOCs* 7: 3-12.

Hu, Minqing, and Bing Liu. 2004. "Mining and summarizing customer reviews". In *Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining*, 168-177. New York: ACM.

Kennedy, Gregorl, Carleton Coffrin, Paula de Barba, and Linda Corrin. 2015. "Predicting Success: How Learners' Prior Knowledge, Skills and Activities Predict MOOC Performance". In *Proceedings of the 5th International Conference on Learning Analytics and Knowledge*, 136-140. New York: ACM

Kettle, Keri L., and Gerald Häubl. 2010. "Motivation by Anticipation: Expecting Rapid Feedback Enhances Performance". *Psychological Science* 21 (4): 545-547. doi:10.1177/0956797610363541.

Kloft, Marius, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. 2014. "Predicting MOOC Dropout over Weeks using Machine Learning Methods". In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 60-65. Taberg: Association for Computational Linguistics (ACL).

Kolowich, Steve. 2013. "Coursera Takes a Nuanced View of MOOC Dropout Rates". *The chronicle of higher education*.

Li, Xiao, Ting Wang, and Huaimin Wang. 2017. "Exploring N-gram Features in Clickstream Data for MOOC Learning Achievement Prediction". In *Proceedings of the 22nd International Conference on Database Systems for Advanced Applications*, 328-339. Cham: Springer.

Lykourentzou, Ioanna, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. "Dropout Prediction in e-Learning Courses through the Combination of Machine Learning Techniques". *Computers & Education* 53 (3): 950-965. doi:10.1016/j.compedu.2009.05.010.

Okubo, Fumiya, Takayoshi Yamashita, Atsushi Shimada, and Hiroaki Ogata. 2017. "A neural network approach for students' performance prediction". In *Proceedings of the 7th International Conference on Learning Analytics and Knowledge,* 598-599. New York: ACM.

Pardo, Abelardo, Negin Mirriahi, Roberto Martinez-Maldonado, Jelena Jovanovic, Shane Dawson, and Dragan Gašević. 2016. "Generating actionable predictive models of academic performance". In *Proceedings of the 6th International Conference on Learning Analytics and Knowledge*, 474-478. New York: ACM.

Pelánek, Radek. 2015. "Metrics for evaluation of student models". *Journal of Educational Data Mining* 7 (2): 1-19.

Pérez-Lemonche, Ángel, Gonzalo Martínez-Muñoz, Estrella Pulido-Cañabate. 2017. "Analysing Event Transitions to Discover Student Roles and Predict Grades in MOOCs". In *Proceedings of the 26th International Conference on Artificial Neural Networks and Machine Learning*, 224-232. Cham: Springer.

Ramesh, Arti, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Gettor. 2014. "Learning Latent Engagement Patterns of Students in Online Courses". In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 1272-1278. Palo Alto: AAAI Press.

Ramesh, Arti, Shachi H. Kumar, James R. Foulds, and Lise Getoor. 2015. "Weakly Supervised Models of Aspect-Sentiment for Online Course Discussion Forums". In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, 74-83. New York: Association for Computational Linguistics (ACL).

Ren, Zhiyun, Huzefa Rangwala, and Aditya Johri. 2016. "Predicting Performance on MOOC Assessments using Multi-Regression Models". arXiv preprint arXiv:1605.02269.

Romero, Cristóbal, Manuel-Ignacio López, Jose-María Luna, and Sebastián Ventura. 2013. "Predicting Students' Final Performance from Participation in On-Line Discussion Forums". *Computers & Education* 68: 458-472. doi:10.1016/j.compedu.2013.06.009.

Ruipérez-Valiente, José A., Ruth Cobos, Pedro J. Muñoz-Merino, Álvaro Andujar, and Carlos Delgado Kloos. 2017. "Early Prediction and Variable Importance of Certificate Accomplishment in a MOOC". In *Proceedings of the 5th European Conference on Massive Open Online Courses*, 263-272. Cham: Springer.

Sajjadi, Mehdi S.M., Morteza Alamgir, and Ulrike von Luxburg. 2016. "Peer Grading in a Course on Algorithms and Data Structures: Machine Learning Algorithms do not Improve over Simple Baselines". In *Proceedings of the 3rd ACM Conference on Learning @ Scale*, 369-278. New York: ACM.

Santos, José Luis, Joris Klerkx, Erik Duval, David Gago, and Luis Rodríguez. 2014. "Success, Activity and Drop-outs in MOOCs an Exploratory Study on the UNED COMA Courses". In *Proceedings of the 4th International Conference on Learning Analytics and Knowledge*, 98-102. New York: ACM.

Shah, Dhawal. 2015. "By the Numbers: MOOCs in 2015". *Class Central*.

Shahiri, Amirah Mohamed, Wahidah Husain, and Nur'aini Abdul Rashid. 2015. "A Review on Predicting Student's Performance Using Data Mining Techniques". *Procedia Computer Science* 72: 414-422. doi:10.1016/j.procs.2015.12.157.

Silverman, Linda Kreger. 1990. "Social and Emotional Education of the Gifted: The Discoveries of Leta Hollingworth". *Roeper Review*, 12 (3): 171-178. doi: 10.1080/02783199009553265.

Sinha, Tanmay, and Justine Cassell. 2015. "Connecting the Dots: Predicting Student Grade Sequences from Bursty MOOC Interactions over Time". In *Proceedings of the 2nd ACM Conference on Learning@ Scale*, 249-252. New York: ACM.

Universidad Carlos III de Madrid. 2015. "Introduction to Programming with Java. Part I: Starting to program in Java. Syllabus". *edX*, May 21. https://d37djvu3ytnwxt.cloudfront.net/assets/courseware/v1/dd292391c487859d7be2b11b746d7afa/c4x/UC3Mx/IT.1.1x/asset/IT.1x-Syllabus.pdf

Valadez Sierra, María Dolores, Angel Alberto Valdés Cuervo, Teodoro Rafael Wendlandt Amezaga, Ana Carolina Reyes Sánchez, Rogelio Zambrano Guzmán, and Juan Pedro Navarro Agraz. 2015. "Differences in Achievement Motivation and Academic and Social Self-Concept in Gifted Students of Higher Education". *Journal of Curriculum And Teaching* 4 (1): 83. doi:10.5430/jct.v4n1p83.

Vu, Duy, Philippa Pattison, and Garry Robins. 2015. "Relational Event Models for Social Learning in MOOCs". *Social Networks* 43: 121-135. doi:10.1016/j.socnet.2015.05.001.

Wang, Zhuo, Jile Zhu, Xiang Li, Zhiting Hu, and Ming Zhang. 2016. "Structured Knowledge Tracing Models for Student Assessment on Coursera". In *Proceedings of the 3rd ACM Conference on Learning@ Scale*, 209-212. New York: ACM.

Xenos, Michalis. 2004. "Prediction and Assessment of Student Behaviour in Open and Distance Education in Computers Using Bayesian Networks". *Computers & Education* 43 (4): 345-359. doi:10.1016/j.compedu.2003.09.005.

Xing, Wanli, Xin Chen, Jared Stein, and Michael Marcinkowski. 2016. "Temporal Predication of Dropouts in MOOCs: Reaching the Low Hanging Fruit Through Stacking Generalization". *Computers in Human Behavior* 58: 119-129. doi:10.1016/j.chb.2015.12.007.

Yang, Diyi, Robert Kraut, and Carolyn Rose. 2016. "Exploring the Effect of Student Confusion in Massive Open Online Courses". *Journal of Educational Data Mining* 8 (1): 52-83.

Yang, Tsung-Yen, Christopher G. Brinton, Carlee Joe-Wong, and Mung Chiang. 2017. "Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks". *IEEE Journal of Selected Topics in Signal Processing* 11 (5): 716-728.

Ye, Cheng, and Gautam Biswas. 2014. "Early Prediction of Student Dropout and Performance in MOOCs Using Higher Granularity Temporal Information". *Journal of Learning Analytics* 1 (3): 169-172. doi:10.18608/jla.2014.13.14.