



Universidad  
Carlos III de Madrid

Departamento de Teoría de la Señal y Comunicaciones

PROYECTO FIN DE CARRERA

# PREDICCIÓN DE RESULTADOS DE EVENTOS DEPORTIVOS: TENIS

**Autor:** Miguel Ángel Martín Mazuelo

**Tutor:** Fernando Pérez-Cruz



Título: Predicción de resultados de eventos deportivos: tenis

Autor: Miguel Ángel Martín Mazuelo

Director: Fernando Pérez-Cruz

EL TRIBUNAL

Presidente:

---

Vocal:

---

Secretario:

---

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día \_\_\_\_\_de \_\_\_\_\_de 2014 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE



## AGRADECIMIENTOS

Me gustaría agradecer el final de esta etapa principalmente a mis padres y hermana, que siempre han estado ahí cuando los he necesitado, tanto en los buenos momentos como en los malos, y me han ayudado a formarme académicamente, pero sobre todo como persona, no sólo en mis años universitarios sino a lo largo de toda mi vida. Sé que siempre me echaréis una mano, o las dos, cuando lo necesite, tal y como ha sido hasta ahora. Igualmente sabéis que siempre estaré a vuestro lado pase lo que pase.

También querría dedicárselo con especial cariño a mi pareja, por cada instante que ha estado a mi lado durante todos los años que llevamos juntos. Por soportarme en mis buenos y malos momentos, ayudarme siempre que lo he necesitado, apoyarme y animarme. Sin ti esto no habría sido posible y espero que sigamos estando unidos para siempre.

A mi tutor, por concederme la posibilidad de realizar este proyecto tan interesante, dirigirme y orientarme a través de él, por las horas que me ha dedicado y por todas las facilidades proporcionadas debido al poco tiempo del que he dispuesto para realizarlo debido a motivos laborales.

En general, gracias a todos los compañeros y profesores que me han ayudado a lo largo de estos años universitarios, que no han sido pocos.



## RESUMEN

El mundo de las apuestas ha ido evolucionando día a día hasta llegar a la situación actual, en la que está en auge. Este proceso comenzó con las clásicas apuestas en eventos en directo, posteriormente se crearon las casas de apuestas situadas en lugares físicos y, por último, se implantaron las apuestas virtuales, las cuales nos han llevado a una gran mejora de las condiciones ofrecidas a los usuarios y a una expansión que no tiene límites.

Por otro lado, el mundo del tenis, sobre todo en torneos de Grand Slam o Masters 1000, desencadena una gran expectación en todo el mundo, moviendo a una cantidad elevada de personas.

Uniendo estas dos visiones, llegamos al eje central sobre el que gira este proyecto: las apuestas en el popular deporte del tenis.

Si profundizamos un poco más y ahondamos en el contenido de este proyecto, podemos afirmar que vamos a ser capaces de crear un modelo que nos permitirá conseguir una rentabilidad económica apostando a lo largo de un torneo. Para ello, realizaremos una comparación de los porcentajes de victoria asignados por las casas de apuestas y los calculados por nosotros. En cada partido decidiremos el jugador por el que es beneficioso apostar según nuestro algoritmo y calcularemos la cantidad óptima que debemos apostar, es decir, el porcentaje de nuestra banca a invertir.

Por último, tras el análisis de los resultados que vamos alcanzando en los diferentes torneos, obtendremos una serie de conclusiones, que nos servirán para evaluar el funcionamiento del modelo desarrollado en nuestro proyecto y para introducir mejoras e ideas futuras de cara a la creación de nuevos algoritmos y proyectos.





# ÍNDICE GENERAL

<b>RESUMEN .....</b>	<b>7</b>
<b>1.INTRODUCCIÓN .....</b>	<b>12</b>
1.1. OBJETIVOS.....	13
1.2. ESTRUCTURA .....	14
<b>2.ESTADO DEL ARTE .....</b>	<b>16</b>
2.1. HISTORIA DEL TENIS .....	16
2.2. HISTORIA DE LAS APUESTAS DEPORTIVAS .....	18
2.3. TIPOS DE CASAS DE APUESTAS.....	20
2.4. TIPOS DE CUOTAS .....	22
2.5. APUESTAS EN EL TENIS .....	27
2.6. ALGORITMO DE KELLY .....	30
<b>3.MODELO FINAL.....</b>	<b>33</b>
3.1. ESTRUCTURA DEL MODELO.....	34
3.2. ELECCIÓN DE CUOTAS ENTRE CASAS DE APUESTAS.....	35
3.3. CÁLCULO DE PROBABILIDADES MEDIANTE LAS CUOTAS.....	37
3.4. BASES DE DATOS DE JUGADORES Y TORNEOS.....	41
3.5. ALGORITMO PARA EL CÁLCULO DE PROBABILIDADES DE VICTORIA .....	49
3.6. USO DEL ALGORITMO DE KELLY PARA LA REALIZACIÓN DE APUESTAS .....	51
<b>4.PRUEBAS Y RESULTADOS. ....</b>	<b>55</b>
4.1. PRUEBA 1: TORNEO MASTERS DE MADRID.....	55
4.2. PRUEBA 2: TORNEO MASTERS DE MIAMI.....	65
4.3. PRUEBA 3: TORNEO MASTERS DE MONTREAL .....	77
<b>5.CONCLUSIONES E IDEAS FUTURAS.....</b>	<b>87</b>
5.1. CONCLUSIONES .....	87
5.2. IDEAS FUTURAS .....	90
<b>6.PRESUPUESTO .....</b>	<b>92</b>
<b>7.REFERENCIAS .....</b>	<b>97</b>
7.1. PÁGINAS DE INTERNET .....	97
7.2. LIBROS Y ARTÍCULOS.....	99

# ÍNDICE DE FIGURAS

Figura 1. Estructura del modelo.....	34
Figura 2. Diagrama de estados de un juego .....	49
Figura 3. Simulación de un partido.....	50
Figura 4. Porcentaje de victoria de los jugadores que ganaron la ronda (Masters Madrid).....	58
Figura 5. Porcentaje de victoria de los jugadores que perdieron la ronda (Masters Madrid)...	58
Figura 6. Ganancia obtenida partido a partido (Masters Madrid).....	62
Figura 7. Beneficio acumulado a lo largo del torneo (Masters Madrid).....	63
Figura 8. Porcentaje de victoria de los jugadores que ganaron la ronda (Masters Miami).....	70
Figura 9. Porcentaje de victoria de los jugadores que perdieron la ronda (Masters Miami)....	70
Figura 10. Ganancia obtenida partido a partido (Masters Miami) .....	74
Figura 11. Beneficio acumulado a lo largo del torneo (Masters Miami) .....	75
Figura 12. Porcentaje de victoria de los jugadores que ganaron la ronda (Masters Montreal)	80
Figura 13. Porcentaje de victoria de los jugadores que perdieron la ronda (Masters Montreal).....	81
Figura 14. Ganancia obtenida partido a partido (Masters Montreal).....	83
Figura 15. Beneficio acumulado a lo largo del torneo (Masters Montreal).....	84

## ÍNDICE DE TABLAS

Tabla 1. Extracto de la base de datos de jugadores .....	45
Tabla 2. Extracto de la base de datos de torneos .....	47
Tabla 3. Probabilidades de las casas de apuestas y de nuestro modelo (Masters Madrid)...	57
Tabla 4. Apuestas realizadas a lo largo del torneo (Masters Madrid) .....	61
Tabla 5. Resultados obtenidos (Masters Madrid) .....	64
Tabla 6. Probabilidades de las casas de apuestas y de nuestro modelo (Masters Miami) ....	69
Tabla 7. Apuestas realizadas a lo largo del torneo (Masters Miami) .....	74
Tabla 8. Resultados obtenidos (Masters Miami) .....	76
Tabla 9. Probabilidades de las casas de apuestas y de nuestro modelo (Masters Montreal) 79	
Tabla 10. Apuestas realizadas a lo largo del torneo (Masters Montreal) .....	83
Tabla 11. Resultados obtenidos (Masters Montreal) .....	85
Tabla 12. Fases del proyecto .....	93
Tabla 13. Costes de material .....	93
Tabla 14. Gastos detallados del proyecto .....	94
Tabla 15. Presupuesto del proyecto .....	95

# 1. INTRODUCCIÓN

El tema principal de este proyecto surgió con la idea de realizar una base de datos de jugadores a partir de la cual pudiéramos realizar una predicción de resultados para eventos deportivos con la que fuera posible obtener un beneficio en el mercado de las apuestas.

Más adelante concretamos centramos en el tenis ya que, por una parte, es un deporte que considero interesante y por otra, es uno de los deportes en los que más apuestas se realizan. La unión de estas dos visiones es lo que da lugar al título del proyecto.

Posteriormente se fueron introduciendo nuevos conceptos e ideas, que se explican en detalle en el resto de capítulos, como es el caso de la creación de más de una base de datos o la elección de las cuotas de las casas de apuestas.

Por último, una vez creado el algoritmo para predecir los porcentajes de victoria, se decidió elegir el Criterio de Kelly para apostar, y siguiendo estos criterios se obtuvieron una serie de resultados, que se exponen en el capítulo cuarto.

## 1.1. OBJETIVOS

El objetivo principal de este proyecto es crear un modelo ganador, a partir de un conjunto de datos obtenidos a través de varias fuentes, que permita obtener un beneficio a largo plazo en el mundo de las apuestas de tenis.

Para lograr ese objetivo final, hemos definido los siguientes objetivos secundarios, que debemos ir cumpliendo para alcanzar el beneficio deseado:

- Conocer qué características y qué factores son más determinantes en un partido de tenis y en su resultado final, para utilizarlos en la base de datos que se va a crear.
- Ajustar al máximo el porcentaje de victoria que puede tener un jugador en un partido de tenis individual masculino en un torneo.
- Conseguir un volumen adecuado de datos para definir un modelo acertado pero sin dimensiones desproporcionadas.
- Conocer el funcionamiento de las casas de apuestas y analizar si este conocimiento puede ser útil para una posible futura inversión en las apuestas en torneos.
- Crear dos bases de datos en las que englobaremos, por un lado, información acerca de los tenistas y, por otro, los movimientos que éstos generan en el mundo de las casas de apuestas deportivas.

## 1.2. ESTRUCTURA

En este proyecto exponemos un estudio para el mundo de las apuestas en eventos deportivos, utilizando bases de datos, métodos estadísticos y un modelo de predicción. En concreto nos vamos a centrar en el tenis individual masculino.

A continuación se describe la estructura general, capítulo a capítulo, del proyecto.

- Capítulo 1: Introducción del proyecto fin de carrera, especificando la estructura y los objetivos previstos.
- Capítulo 2: Explicación del estado del arte, tanto de las casas de apuestas como la parte específica del tenis. Además, se explica el algoritmo de Kelly como método óptimo para calcular la cantidad apropiada a apostar en cada partido.
- Capítulo 3: Explicación del modelo final. Cálculo de probabilidades a través de las cuotas, elección de cuotas entre casas de apuestas, descripción del algoritmo propio para el cálculo de probabilidades de victoria de cada jugador, bases de datos y utilización del algoritmo de Kelly para apostar.
- Capítulo 4: Exposición de las pruebas realizadas a lo largo de diversos torneos del circuito ATP y los resultados obtenidos.

- Capítulo 5: Descripción de las conclusiones generales del proyecto a través de los resultados obtenidos y las posibles mejoras de cara al futuro.
  
- Capítulo 6: Descripción del presupuesto del proyecto.
  
- Capítulo 7: Descripción de las referencias utilizadas para la realización de este proyecto: páginas web, artículos y libros.

## 2. ESTADO DEL ARTE

### 2.1. HISTORIA DEL TENIS

Los primeros juegos parecidos al deporte del tenis actual se sitúan en dos civilizaciones, la griega y la romana.

Entre los siglos XIII-XIV, en Francia se comenzó a practicar el “Jeu de Paume”, que es un juego que consistía en golpear una pelota con la palma de la mano, y que fue evolucionando hasta acabar utilizando una raqueta.

Durante los siglos XVI-XVIII los juegos de pelota empezaron a ganar popularidad en Europa occidental, aunque eran más parecidos al frontón actual que al tenis.

A finales del siglo XIX, Walter Clopton Wingfield, el cual es considerado oficialmente como el inventor del tenis moderno, lo bautizó con el nombre del vocablo griego Sphairistiké (“jugando con bola”) y redactó las primeras normas del deporte.

El torneo más antiguo es Wimbledon, creado en 1877, y a éste le siguieron el resto de torneos de Gran Slam existentes: Campeonato Norteamericano en 1881 (Open de Estados Unidos), los Campeonatos franceses (Roland Garros), y el Campeonato Australiano (Open de Australia), junto a la Copa Davis y la Copa Wightman.

En sus inicios, el tenis fue un deporte elitista, sólo practicado por las clases altas en clubes privados. El británico William Renshaw, fue el primer



gran tenista de la historia y ayudó a popularizarlo y hacerlo más espectacular entre 1880 y 1890.

El tenis formó parte del programa de los primeros Juegos Olímpicos de la modernidad en 1896, que fueron disputados en Atenas. En los Juegos de París de 1900 se introdujo por primera vez la competición femenina, y en Estocolmo, en 1912, se incorporó la modalidad de dobles mixtos por países.

Podemos considerar a los países anglosajones como los iniciadores y dominadores tradicionales de las principales competiciones tenísticas, especialmente Estados Unidos y Australia, que casi siempre han contado entre sus filas con los mejores jugadores. Sin embargo, esta situación comenzó a cambiar a partir de los años setenta y ochenta con la irrupción de nuevas potencias tenísticas europeas, países que han aportado grandes campeones en los últimos tiempos y conquistado numerosos títulos por equipos.

Entre los tenistas españoles destacaron Lili Álvarez en los años veinte y Manuel Santana en los sesenta, pero el auge vino en la última década del siglo XX con la llegada de jugadores de la talla de Carlos Moyá y especialmente de nuestro icono tenístico Rafael Nadal.

## 2.2. HISTORIA DE LAS APUESTAS DEPORTIVAS

Las primeras apuestas tuvieron lugar en la civilización griega, cuyos ciudadanos se reunían en los estadios para animar a los mejores deportistas de la época.

Posteriormente, esta misma costumbre también apareció en el Imperio Romano, donde se apostaba a los combates de gladiadores y en las carreras.

En la Edad Media, las apuestas se centraban en los torneos caballerescos, en las competiciones con tiro con arco y en el lanzamiento de cuchillos.

Sin embargo, las apuestas tuvieron su verdadera revolución alrededor de los siglos XVIII y XIX, gracias a la evolución de la prensa, que comenzó a crear secciones dedicadas exclusivamente a las apuestas deportivas. En la segunda mitad del siglo XIX, las apuestas llegaron a América.

En el siglo XX las páginas de apuestas empezaron a multiplicarse. La primera abrió en Liverpool y se fue expandiendo por Europa y Norteamérica. A partir de este momento, las apuestas alcanzaron gran popularidad a nivel mundial.

Con la llegada de Internet se empezaron a crear empresas especializadas en apuestas deportivas para poder apostar en línea.

Actualmente existe una elevada competencia entre las distintas casas de apuestas, cuyo objetivo es atraer a los nuevos clientes y conseguir mantener a los ya existentes en un mercado en pleno crecimiento y expansión.

## 2.3. TIPOS DE CASAS DE APUESTAS

Las casas de apuestas más conocidas son las tradicionales, donde podemos apreciar las cotizaciones que el equipo encargado de las predicciones en la casa de apuestas ha establecido, y, en función de esas cuotas, decidir si deseamos apostar o no.

Hoy en día existe un tipo de casas de apuestas que funciona de una manera totalmente diferente al resto. Son las casas de intercambio de apuestas. En estas casas se juntan clientes que quieren apostar en determinados eventos. Por lo tanto, el apostante no cruza su apuesta con la casa de apuestas, sino que la cruza con otro apostante, siendo la casa de intercambio de apuestas un intermediario entre ambos apostantes.

La casa de intercambio de apuestas deportivas asume un papel pasivo en la elección de las cuotas y las cantidades que son apostadas, ya que son los propios apostantes quienes las deciden de mutuo acuerdo. Las casas de intercambio generan sus ingresos al cobrar una comisión, que es calculada como una cuota fija por transacción o por porcentajes de las ganancias netas de cada cliente.

Para los buenos apostantes, este tipo de casas de apuestas tiene la ventaja de que no existen límites para apostar o para ganar.

Otro de los grandes secretos del éxito de estas casas es el Trading, es decir, el arte de comprar y vender apuestas como si fuesen acciones de bolsa.

Las **ventajas** principales de las casas de apuestas frente a las casas de intercambio son:

- Cuota de salida: Las casas de apuestas ponen una cuota de salida de antemano, definida por sus expertos. Sin embargo, en las casas de intercambio son los apostantes los que definen la cuota.
- Liquidez: aunque la casa no limita la cantidad a apostar, depende de los usuarios elevar la liquidez de las mismas.

Los **inconvenientes** que se pueden encontrar en una casa de apuestas respecto a las de intercambio son:

- Sus mejores cuotas: Como en este tipo de casas cada uno apuesta contra otros, no existen los márgenes de las casas de apuestas tradicionales. Podríamos estar ganando en torno a un 15 o 20 por ciento. Si a esto le sumamos la comparación con las cuotas de varias de ellas, conseguiremos la cuota más beneficiosa de todo el mercado.
- Sin limitaciones: en las casas de apuestas tradicionales se limita la cantidad a ser apostada y las cantidades máximas de ganancias que un apostante puede tener para un periodo de tiempo determinado.

En nuestro caso particular nos centraremos en las casas de apuestas tradicionales, ya que hoy en día no están permitidas las casas de intercambio de apuestas en España.

## 2.4. TIPOS DE CUOTAS

Las cuotas ofrecidas por las casas de apuestas sirven para calcular la ganancia que se puede llegar a obtener si se gana una apuesta.

En las casas de apuestas las cuotas son fijadas por sus técnicos, los cuales las determinan según una serie de algoritmos y pronósticos. En el caso de las casas de intercambio de apuestas el valor de las cuotas es fijado por las apuestas de los propios apostantes. Como hemos dicho anteriormente, nos centraremos en las primeras, ya que las segundas no están permitidas en el territorio nacional.

Las cuotas determinadas para cada jugador son inversamente proporcionales a su probabilidad de victoria, es decir, si apostamos a favor de un jugador que tiene una cuota baja, tendrá una probabilidad elevada de victoria, mientras que si apostamos a favor de uno que tiene una cuota alta, su probabilidad de victoria será mucho menor.

En las casas de apuestas podemos encontrarnos cuotas de varios tipos.

La más frecuente es el **formato decimal**, que son un tipo de cuotas utilizadas principalmente en Europa. La cuota se presenta en forma de número entero o decimal. Una manera rápida y fácil de calcular una cuota es dividiendo 1 por la probabilidad de victoria que se estime a ese jugador.

Por ejemplo, si las casas de apuestas asignan a un jugador una probabilidad de victoria del 50%, la cuota que aparecerá será 2, siempre que consideremos el caso ideal en que la casa de apuestas no se guarda un porcentaje de ganancia, que se obtiene a partir de la división de  $1 / 0.5$ .

Más adelante hablaremos del porcentaje de ganancia de las casas de apuestas.

Por lo tanto, como ya hemos explicado, si queremos calcular nuestra cuota propia, no tenemos más que dividir 1 entre la probabilidad de victoria del jugador.

Cuando el porcentaje ofrecido por la casa de apuestas sea menor que el calculado o estimado por nosotros, será el momento oportuno para realizar la apuesta.

Imaginemos que apostamos 10€ a favor de un jugador cuya cuota es de 3. Dicha cuota nos permite conocer cuánto podemos ganar si el tenista por el que apostamos se impone a su rival.

Para calcular la ganancia, en caso de acierto, aplicamos la siguiente fórmula:

$$\text{Ganancia} = \text{Cantidad apostada} * \text{Cuota}$$

Si lo que deseamos saber es el beneficio:

$$\text{Beneficio} = \text{Cantidad apostada} * (\text{Cuota} - 1)$$

Por tanto, si nuestro jugador vence, nuestra ganancia será de 30 euros, mientras que el beneficio neto será de 20. Como bien sabemos, se cumple que el beneficio es igual a la ganancia menos la cantidad apostada, en nuestro caso 20 ( $30 - 10 = 20$ ).

Por otra parte, en las casas de apuestas británicas, tenemos el **formato fraccionario**, en el cual las cuotas tienen forma de fracción. En ellas, el numerador indica las unidades de ganancia y el denominador las unidades apostadas.

Para calcular el beneficio utilizando cuotas fraccionarias, hemos de aplicar la siguiente fórmula:

$$\text{Beneficio} = \text{Cantidad apostada} * \text{Cuota fraccionaria}$$

Veamos un ejemplo. Se disputa un partido entre Ferrer y Federer. Roger Federer tiene una cuota fraccionaria de 8/10. Si apostamos 10 euros a favor del suizo y gana el partido, nuestros beneficios serán de 8€ (10€\*8/10).

Existe una fórmula sencilla para convertir una cuota fraccionaria en decimal:

$$\text{Cuota decimal} = \text{Cuota fraccionaria} + 1$$

En el ejemplo anterior, la cuota fraccionaria 8/10 de Federer se convertiría de la siguiente forma a cuota decimal:  $(8/10) + 1 = 1,80$ . Y si calculamos el beneficio neto con la fórmula para cuotas decimales observamos que obtenemos la misma cantidad de 8€ (10€\*0.80).

Existe un tercer formato de cuotas utilizado en las casas de apuestas estadounidenses, denominado **formato americano**. Son habituales en la



mayoría de los deportes típicos norteamericanos. Se pueden expresar en forma de número positivo o negativo. Las cuotas negativas indican la cantidad que hay que apostar para conseguir un beneficio de 100 unidades y las positivas muestran el beneficio que se obtiene con una apuesta de 100 unidades.

Si realizamos una comparación de este formato con el decimal, las cuotas positivas, en el formato americano, son las mayores que 2, en el formato decimal, y las negativas son menores que 2.

La fórmula para calcular el beneficio en este tipo de cuotas es la siguiente:

$$\text{Beneficio} = (\text{cantidad apostada} * \text{cuota}) / 100$$

Las cuotas negativas nunca deben llevar el signo menos al hacer el cálculo.

Veamos la utilización de este tipo de cuotas con un ejemplo. En el partido de Ferrer y Federer, la cuota del español sería +110 y la del suizo -80. Si apostamos 10€ por Ferrer y gana, nuestro beneficio sería de 11€ ( $10 * 110 / 100$ ). En cambio, si apostamos 10€ por Federer y vence, nuestro beneficio sería de 8€ ( $10 * 80 / 100$ ).

Existe la siguiente fórmula para pasar una cuota de formato americano a decimal:

$$\text{Cuota decimal} = 1 + (\text{cuota americana} / 100)$$

En esta fórmula, al igual que en la del cálculo del beneficio, los valores de las cuotas americanas negativas no llevan el signo menos.

La cuota de Ferrer en formato decimal sería 2,1 ( $1 + (110/100)$ ). Y la de Federer 1,8 ( $1 + (80/100)$ ).

## 2.5. APUESTAS EN EL TENIS

El tenis es un deporte popular en la mayor parte del mundo y los apostantes se intentan especializar en él porque son muchos los factores de los que depende la victoria de un jugador.

Las principales razones de que el tenis sea uno de los deportes principales en el mundo de las apuestas son las siguientes:

- No se juega contra un cronómetro sino contra un objetivo.
- Su estilo de puntuación hace que se produzcan muchos cambios de ventajas en el partido, ideal para variaciones en directo, es decir, es ideal para el trading por su excesiva fluctuación.
- No hay factor de empate.
- Los mini-descansos son ideales para operar.
- Es un deporte individual en el que los desequilibrios psicológicos pueden hacer cambiar drásticamente el desarrollo de un partido.

Para comenzar a apostar en el tenis hay que tener en cuenta ciertos factores para analizar un partido:

- Es un **deporte individual**, en el que a diferencia de los deportes de equipo, influye mucho: si tienes un mal día, una mala mentalidad, un mal estado físico, te relajas, etc.
- La **superficie** en la que se juega el partido hace que varíen mucho los resultados de los encuentros. Hay jugadores que son especialistas en una superficie determinada.

- El **tipo de torneo** influye en factores como la superficie, el tipo de bola, la altitud de la pista, la humedad y temperatura, que afectan directamente al juego de los tenistas.
  
- **Estado de ánimo y motivación.** Cuanta mayor motivación, más difícil va a ser que se pierda un jugador considerado como favorito.
  
- **Estilos de juego:** Podríamos diferenciar dos grandes grupos de jugadores, los que juegan al ataque y los de perfil más pasivo o defensivo. También es importante conocer si los jugadores son grandes sacadores, e igualmente si tienen capacidad para restar bien, sobre todo en determinados tipos de pista.
  
- **Número de Sets y duración de los partidos:** En el caso de los Grand Slam y de la Copa Davis los partidos son a 5 sets y en el resto de torneos son a 3 sets. Esto es importante porque es mucho más difícil sorprender a un cabeza de serie en un partido largo a 5 sets. Además, cuanto más largo sea un partido, más influye el estado físico del jugador.
  
- **Cansancio acumulado:** El rendimiento de los tenistas baja cuando no están a bien físicamente, sobre todo en los que tienen menor preparación física o un determinado estilo de juego.
  
- Las **lesiones.** En el caso de que un jugador se lesione, pierde el partido automáticamente. Los tenistas que juegan con molestias tienen una dificultad añadida para ganar.

Un sistema de apuestas muy utilizado es el Trading, que se basa en aprovecharse de las tendencias del mercado para especular con las cuotas.

Consiste en apostar por una cuota (apostar a favor de un jugador) y hacer lo contrario cuando la cuota haya bajado (apostar en contra o apostar por el otro participante) y así se obtiene un beneficio sin tener que esperar al final del partido. De esta forma nos da igual quién sea el ganador porque obtendremos un beneficio igualmente.

El trading no sólo se utiliza para asegurar beneficios, sino que también puede ser usado para minimizar pérdidas. Una persona con experiencia es capaz de saber cuándo va perdiendo y cuando no necesita arriesgar más.

Esta técnica se puede aplicar en las casas de apuestas especializadas en el tenis, en primer lugar por su similitud con el mercado de valores y en segundo lugar por ser un deporte individual en el que las cuotas cambian mucho y muy rápidamente.

## 2.6. ALGORITMO DE KELLY

“El Algoritmo de Kelly” fue desarrollado por John Kelly en 1956, revolucionando el mundo de las apuestas. Originalmente fue pensado para apostar a los caballos, ya que era una de sus grandes pasiones, pero es válido para cualquier tipo de apuesta.

Se trata de un criterio que fue diseñado para maximizar nuestra banca disponible a largo plazo, calculando las cantidades óptimas a jugar en cada apuesta y que a día de hoy sigue siendo una herramienta muy útil a la hora de gestionar nuestro presupuesto.

Con el método de Kelly se puede calcular la apuesta óptima en función de la cuota de la apuesta y su pronóstico. Apostar más es un riesgo innecesario y apostar menos conlleva un menor rendimiento.

La fórmula para eventos con únicamente dos resultados posibles es la siguiente:

$$A = \frac{\alpha p - 1}{\alpha - 1} * 100$$

En la cual:

- A es el porcentaje de dinero de nuestra banca destinado a apostar.
- $\alpha$  es la cuota que ofrece la casa de apuesta por la victoria del jugador.
- p es nuestra estimación de la probabilidad de victoria que tiene el jugador al que se va a apostar, expresada en tanto por ciento (%).

La utilización de este algoritmo tiene las siguientes características:

- El cálculo de la probabilidad es algo totalmente subjetivo del apostante, por lo que la eficiencia de este algoritmo va a residir en la capacidad que tengamos de poder estimar una probabilidad adecuada al evento.
- La clave de este algoritmo es calcular una probabilidad que se ajuste más a la realidad que las estimadas por las casas de apuestas, las cuales determinan sus propias cuotas.
- Dependiendo de la casa de apuesta, estas cuotas se impondrán mediante un algoritmo propio o a través de las apuestas de los usuarios. En España, por ley, las cuotas las impone la propia casa de apuestas.
- En el caso de la cuota asignada por la casa de apuestas, no vamos a centrarnos en una sola, sino que elegiremos la casa de apuestas que más nos convenga en cada caso para así obtener un beneficio máximo.
- El porcentaje de banca destinado para cada apuesta, en teoría será el óptimo, pero puede que bajo determinados factores pueda llegar a ser un riesgo, por ejemplo en el caso de que las primeras apuestas no vayan bien y disminuya nuestra cantidad de dinero total destinada a apostar. En estos casos a veces es necesaria la utilización de multiplicadores que suavicen esa cantidad a apostar calculada previamente.

A continuación se presenta un ejemplo del uso de este algoritmo para el tema que nos compete:

- Tenemos un partido de tenis entre Rafa Nadal y Roger Federer.
- Disponemos de una banca para las apuestas de 1.000 €.
- Las mejores cuotas ofrecidas por las casas de apuestas son las siguientes: Nadal 1,7 y Federer 2,1.

A través de nuestro modelo estadístico calculamos que Federer tiene un 50% de posibilidades de ganar el encuentro. Por lo tanto la cuota de 2,1 ofrecida por las casas de apuestas está totalmente subestimada y nos interesa apostar por su victoria en este encuentro.

El método de Kelly permite calcular la apuesta óptima:

$$\% \text{ Apostar} = (2.1 * 0.5 - 1) / (2.1 - 1) = 0.0454$$

Por lo tanto, hay que apostar el 4,54%, que son 45,4 €, es decir el 4,54% de 1.000 € = 45,4€.

Una variación de este método es utilizar la misma fórmula pero reduciendo las apuestas máximas para no perder grandes cantidades. Hacer una apuesta máxima (por ejemplo entre un 1% y un 5% de su cartera de apuestas) o multiplicar los porcentajes del método de Kelly por un coeficiente entre 0 y 0,5).

Aunque se pueda apostar más, es necesario limitarse a apostar las cantidades que indica este sistema, ya que se dispone de una cantidad de dinero para apostar y, si nos quedamos sin ella, nuestro método ya no serviría.

El método de Kelly aporta beneficios a largo plazo y sólo tiene sentido si somos capaces de pronosticar el resultado de un evento mejor que las casas de apuestas.



### 3. MODELO FINAL

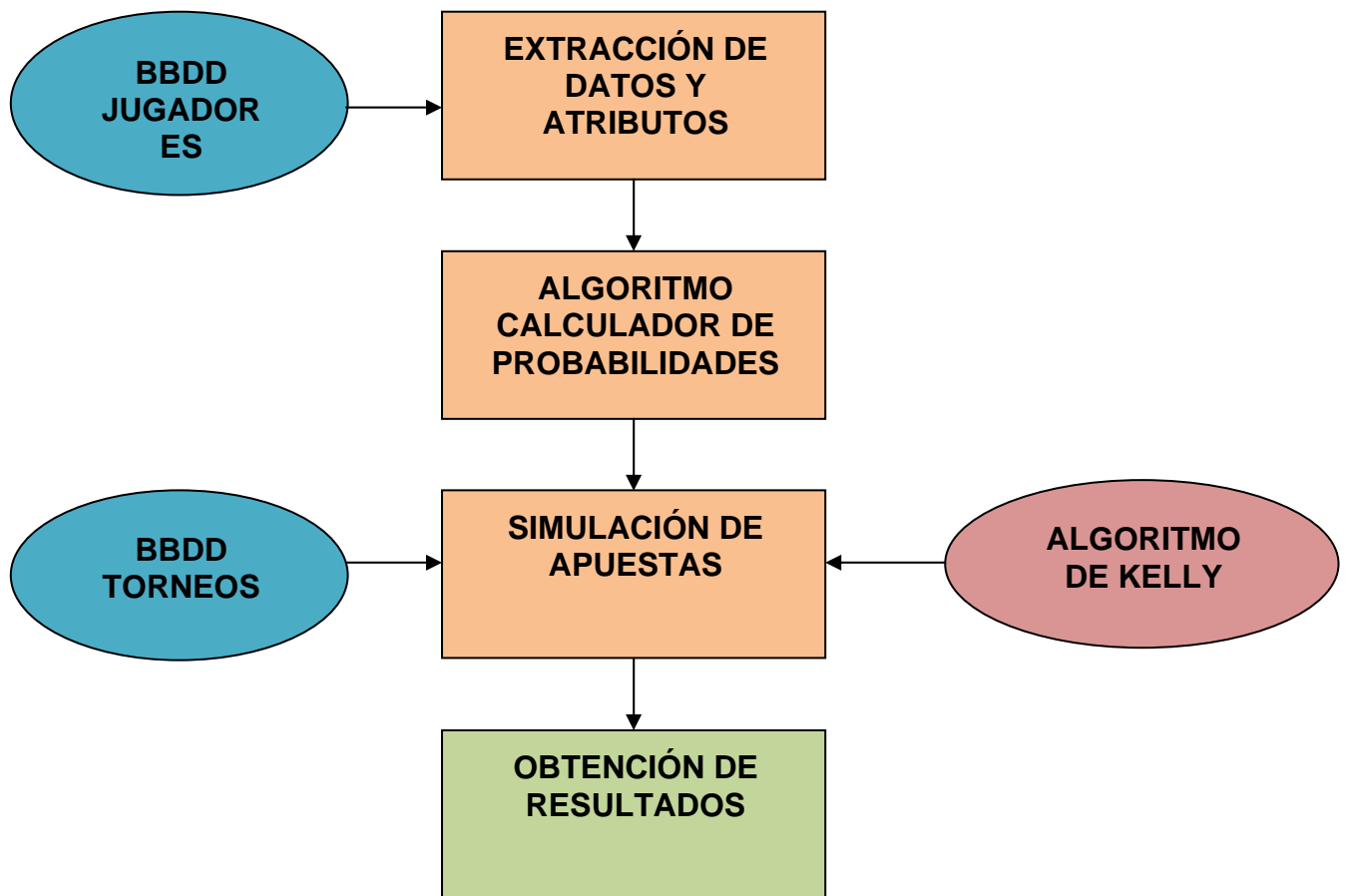
En este capítulo nos centraremos en describir cada uno de los pasos fundamentales que se han realizado durante este proyecto, desde el inicio, en el que tenemos un partido de tenis en el que nos interesa saber quién va a ser el ganador, hasta el momento en el que decidimos apostar por el jugador elegido gracias al cálculo de sus probabilidades de victoria.

Más en concreto, dichos pasos son: la elección de cuotas, el cálculo de probabilidades mediante las cuotas de las casas de apuestas, la creación de las bases de datos de jugadores y torneos, el modelo mediante el cual calculamos la probabilidad de victoria entre cada jugador y el algoritmo de Kelly, que nos indica la cantidad a apostar por el jugador seleccionado.

En el siguiente capítulo describiremos los resultados obtenidos en varios torneos del circuito ATP en pista rápida mediante esta política de apuestas.

### 3.1. ESTRUCTURA DEL MODELO

A continuación se muestra un diagrama de estados que especifica la estructura y funcionalidad de nuestro modelo.



*Figura 1. Estructura del modelo*

## 3.2. ELECCIÓN DE CUOTAS ENTRE CASAS DE APUESTAS

Hoy en día existen múltiples casas de apuestas en la red para que el usuario pueda consultar datos de eventos deportivos y apostar si así lo desea.

Cada una de ellas ofrece unas cuotas muy variadas y totalmente diferentes por cada resultado que pueda ocurrir. Esto es debido a que estas casas tienen unos analistas especializados y utilizan diferentes algoritmos para asignar dichas cuotas.

Si sumamos los porcentajes ofrecidos, la cantidad total que se ofrece siempre es menor al 100%. Esto es así porque siempre se quedan con un porcentaje, que es lo que asegura que siempre obtengan un beneficio.

Si comparamos entre sí esta diversidad de cuotas, nos damos cuenta de que existe una gran dispersión de estas.

Si para cualquier apuesta de un evento deportivo elegimos las mejores cuotas para cada uno de los resultados posibles y hacemos que el porcentaje total a ganar supere el 100%, diremos que hay una operación de arbitraje, es decir, se ganará independientemente del resultado. En el caso de nuestro proyecto, en ningún caso superamos este porcentaje.

A continuación procederemos a describir el proceso de elección de cuotas dentro del ámbito de nuestro proyecto.

En primer lugar se realizó una búsqueda exhaustiva de información entre las casas de apuestas: las cuotas, los tipos de apuestas, las reglas de cada una de ellas...

A continuación elegimos las cuotas más convenientes en cada caso, es decir, las más elevadas para cada jugador, ya que esto nos reporta el mayor beneficio para cada una de nuestras apuestas. Esto es esencial durante este punto del proyecto, ya que mediante este procedimiento calculamos varios de los atributos de la base de datos de torneos para posteriormente realizar las simulaciones.

Por último, se realizaron una serie de pruebas, en las que no se apostó dinero de forma real, pero se simuló todo el proceso de apuestas. En general, los resultados fueron muy satisfactorios. Su descripción y análisis se detallan en el siguiente capítulo del proyecto.

### 3.3. CÁLCULO DE PROBABILIDADES MEDIANTE LAS CUOTAS

Una vez que hemos elegido las cuotas más apropiadas y rentables para los partidos, es necesario calcular las probabilidades que representan para cada uno de los jugadores, para posteriormente compararlas con las calculadas con nuestro modelo.

Una de las formas de calcular las probabilidades de las casas de apuestas es a partir de las cuotas ofrecidas. Estas cuotas reflejan la cantidad que la casa de apuestas nos dará por cada euro apostado.

Veamos mediante un ejemplo la explicación de lo que significan las cuotas.

- Supongamos que tenemos un partido entre Roger Federer y Andy Murray, y las cuotas ofrecidas por las casas de apuestas son de 1.7 y 2.1 respectivamente.
- En el caso de que gane el partido Roger Federer, si apostamos un euro por él, obtendríamos 1.7 euros y si apostamos un euro por Andy Murray y gana el partido, obtendríamos 2.1 euros.
- Si apostamos 2 euros por Federer y gana, entonces obtendríamos 3.4 euros, mientras que si hacemos lo mismo con Murray y gana, cobraríamos 4.2 euros.

A continuación se explica cómo calcular las probabilidades a partir de las cuotas ofrecidas por las casas de apuestas. Para ello existen varias ecuaciones.

En el caso de apostar al ganador en un partido de tenis, únicamente existe una cuota para cada jugador, por lo tanto las ecuaciones de sus probabilidades son las siguientes:

$$p_1 = \frac{w_2}{w_1 + w_2}$$

$$p_2 = \frac{w_1}{w_1 + w_2}$$

Donde  $p_1$  y  $p_2$  son las probabilidades de victoria de cada uno de los jugadores y  $w_1$  y  $w_2$  son las cuotas ofrecidas por cada jugador respectivamente.

Además de estas probabilidades ofrecidas, la casa de apuestas dispone de un factor, denominado Factor Q, que nos indica el porcentaje que se va a repartir entre los acertantes. A partir de dicho factor podemos extraer el beneficio que se lleva la casa de apuestas en cada resultado.

Disponemos de la siguiente ecuación:

$$w = \frac{Q}{p}$$

Donde  $w$  es la cuota de la casa de apuestas,  $Q$  es el factor Q y  $p$  es la probabilidad de que ocurra el suceso relacionado con la cuota.

Si despejamos:

$$Q = w p$$

En el mundo del tenis, existen las siguientes ecuaciones para el cálculo de Q:

$$Q = w_1 p_1$$

$$Q = w_2 p_2$$

$$Q = w_2(1 - p_1)$$

$$Q = w_1(1 - p_2)$$

Es decir, el Factor Q es igual a la cuota ofrecida multiplicado por el porcentaje que la casa de apuestas reparte.

Podemos explicar la metodología seguida mediante un ejemplo.

Se juega un partido entre Ferrer y Nalbandian, siendo sus cuotas 1.7 y 2.28 respectivamente. Si calculamos las probabilidades para cada jugador y el factor Q, según las fórmulas anteriores tenemos:

$$w_1 = 1.7$$

$$w_2 = 2.28$$

$$p_1 = \frac{2.28}{1.7 + 2.28} = 0.57$$

$$p_2 = \frac{1.7}{1.7 + 2.28} = 0.43$$

$$Q = 1.7 \times 0.57 = 0.974$$

$$Q = 2.28 \times 0.43 = 0.974$$

Por lo tanto, el porcentaje de victoria, según las casas de apuestas, para Ferrer es del 57% y para Nalbandian del 43%. Además de esto, el factor Q es del 97.4% y el porcentaje que se quedaría la casa de apuestas es  $1 - Q$ , es decir 0.026%.

Este procedimiento explicado es el que utilizo en mi modelo de política de apuestas para, en primer lugar, descubrir cuáles son los porcentajes de victoria de cada jugador en las casas de apuestas, y, posteriormente, tras calcularlos con mis algoritmos, realizar una comparación y decidir por qué jugador debo apostar. Estos detalles los explicaré en los siguientes puntos.



### 3.4. BASES DE DATOS DE JUGADORES Y TORNEOS

Otra de las tareas que se realizó durante este proyecto fue la construcción de dos bases de datos, una con los datos de las estadísticas de los jugadores y otra con los resultados y las cuotas en varios torneos para, posteriormente, realizar las pruebas.

Tras realizar una búsqueda intensiva en Internet a través de numerosas páginas, se obtuvieron una gran cantidad de datos útiles. En este caso, se ha expandido la visión a múltiples fuentes para poder dar una mayor fiabilidad a las bases de datos.

Se ha considerado que lo idóneo para crear nuestras bases de datos es extraer la información utilizando diversas fuentes, en vez de solo una, ya que esta opción proporcionaría una mayor riqueza y, pese a que aumentaría de forma considerable el esfuerzo realizado, haría que nuestras bases de datos fueran más fiables y seguras.

Se han utilizado dos bases de datos, una con los primeros jugadores de la ATP, y otra con torneos disputados a lo largo del año en pista rápida, para poder realizar simulaciones y pruebas.

Por recomendación de mi tutor, una de las casas de apuestas más adecuadas para iniciarme en el proyecto sería Betfair, ya que ofrecía unas cuotas bastante aceptables. Sin embargo, basándome en esta, posteriormente decidimos expandirnos a otras, para conseguir las mejores cuotas. La utilización de esta técnica hizo que fuera lo más beneficioso posible a la hora de apostar y realizar las pruebas y simulaciones.

Esta búsqueda de las mejores cuotas fue lo más laborioso, pero a la vez más gratificante, tras observar en las pruebas finales que se obtenía un buen beneficio en cada torneo.

La información necesaria para la construcción de las bases de datos, se extrajo principalmente de la base de datos de Betfair, que además de las cuotas, disponía de información en ficheros muy extensos divididos en pequeños periodos de tiempo. Dichos ficheros se pueden descargar en <http://data.betfair.com/> disponiendo de un usuario y de una serie de puntos por haber realizado alguna apuesta en la casa. Al ser tan extensos, la extracción de datos es muy costosa pero a la vez muy completa y efectiva.

Posteriormente se amplió la búsqueda a todo el rango de casas de apuestas existentes para las cuotas y la página de la ATP, además de otras páginas de gran interés en la materia para las estadísticas de los jugadores. Dicha información se extrajo de las respectivas bases de datos, de archivos estadísticos y manualmente.

A continuación se detallan los principales atributos de la base de datos de Betfair:

- **EVENT\_ID**: Identificador del evento por el que se apuesta (en nuestro caso, un partido de tenis).
- **FULL\_DESCRIPTION**: Descripción completa del evento en el que se apuesta (en nuestro caso, un partido de un torneo de tenis determinado del circuito ATP).
- **SCHEDULED\_OFF**: Fecha y hora en las que dará inicio el evento.
- **EVENT**: Descripción del evento por el que se realiza la apuesta.
- **ACTUAL\_OFF**: Fecha y hora en la que el evento se inició realmente.
- **SELECTION**: Tipo de apuesta que se realiza (en nuestro caso es una apuesta por el jugador ganador del partido).
- **SETTLED\_DATE**: Fecha y hora en la que el evento se terminó.
- **ODDS**: Cuota a la que estaba el evento en el momento en el que se realizó la apuesta.
- **LATEST\_TAKEN**: Fecha y hora en que se hizo la última apuesta.
- **FIRST\_TAKEN**: Fecha y hora en la que se hizo la primera apuesta.
- **IN\_PLAY**: Tipo de apuesta realizada a tiempo real.

- **NUMBER\_BETS**: Apuestas realizadas en el mismo evento.
- **VOLUME\_MATCHED**: Volumen de apuestas durante el partido.
- **SPORTS\_ID**: Deporte al que pertenece el evento (en nuestro caso filtraremos por el tenis).
- **SELECTION\_ID**: Identificador numérico por el que se guarda la apuesta.
- **WIN\_FLAG**: Identificador binario que indica si la apuesta fue ganada o perdida.

Como vemos, la mayoría de los datos son útiles para el proyecto, y como se trata de la recolección de una cantidad muy elevada de datos, el siguiente paso sería la ordenación de los mismos.

Se unieron todos los datos en dos conjuntos distintos, uno para las estadísticas de los jugadores y otro para todos los datos de las cuotas de los partidos de cada uno de los torneos. El trabajo de ordenación fue muy costoso debido a la magnitud de los datos obtenidos.

Una vez ordenados se procedió a organizarlos y corregir los posibles errores que pudieran contener.

Tras estos pasos se construyeron las bases de datos finales, mediante el cálculo de estadísticas y la creación de una serie de columnas o variables que fueran útiles para, posteriormente, realizar predicciones y simular un conjunto de pruebas de apuestas en cada torneo usando nuestro modelo.

A continuación describiremos el contenido de cada una de las bases de datos con sus principales variables:

## BASE DE DATOS DE JUGADORES

En esta tabla se recoge la información principal de todos los jugadores y sus estadísticas. Es la base de datos fundamental para el cálculo de las probabilidades de victoria de los jugadores, ya que el algoritmo creado por nosotros para ello extrae todas las estadísticas necesarias a partir de esta.

Las principales columnas de esta base de datos de jugadores son:

- **Ranking:** Atributo numérico en el que se almacena el ranking ATP de cada uno de los jugadores disponibles en la base de datos.
- **Nombre:** Nombre y apellidos del jugador. Es una cadena de caracteres.
- **Primer Servicio:** Atributo numérico que expresa el porcentaje medio de primeros servicios que logra el jugador.
- **Segundo Servicio:** Atributo numérico que expresa el porcentaje medio de segundos servicios que logra el jugador.
- **Puntos Ganados PS:** Atributo numérico que expresa el porcentaje medio de puntos ganados al primer servicio por el jugador.
- **Puntos Ganados SS:** Atributo numérico que expresa el porcentaje medio de puntos ganados al segundo servicio por el jugador.
- **Puntos Ganados Resto PS:** Atributo numérico que expresa el porcentaje medio de puntos ganados al resto con el primer servicio del jugador contrario.

- **Puntos Ganados Resto SS:** Atributo numérico que expresa el porcentaje medio de puntos ganados al resto con el segundo servicio del jugador contrario.

A continuación se muestra un extracto de la base de datos de jugadores, para poder apreciar las columnas de las que consta.

Ranking	Jugador	1er Servicio %	2do Servicio %	Puntos Ganados PS %	Puntos Ganados SS %	Puntos Ganados Resto PS %	Puntos Ganados Resto SS %
1	Rafael Nadal	67%	99%	75%	60%	31%	55%
2	Roger Federer	62%	99%	78%	56%	34%	51%
3	Novak Djokovic	64%	99%	71%	52%	34%	54%
4	Andy Murray	54%	99%	77%	53%	32%	55%
5	Robin Soderling	60%	99%	78%	52%	30%	52%
6	Tomas Berdych	56%	99%	79%	55%	31%	52%
7	David Ferrer	65%	99%	70%	55%	33%	54%
8	Andy Roddick	69%	99%	79%	57%	25%	49%
9	Fernando Verdasco	67%	99%	71%	49%	32%	53%
10	Mikhail Youzhny	63%	99%	71%	52%	29%	52%
11	Jurgen Melzer	61%	99%	72%	55%	30%	49%
12	Gael Monfils	63%	99%	73%	51%	31%	49%
13	Jo-Wilfried Tsonga	62%	99%	76%	53%	29%	47%
14	Marin Cilic	57%	99%	74%	53%	30%	50%
15	Nicolas Almagro	60%	99%	74%	54%	29%	49%
16	Mardy Fish	55%	99%	80%	53%	29%	51%
17	Ivan Ljubicic	59%	99%	76%	49%	29%	48%
18	Sam Querrey	61%	99%	79%	50%	27%	48%
19	John Isner	69%	99%	76%	55%	22%	42%
20	Marcos Baghdatis	52%	99%	76%	52%	31%	50%

*Tabla 2. Extracto de la base de datos de jugadores*

## BASE DE DATOS DE TORNEOS

En esta tabla se recoge la información principal de los torneos de pista rápida en forma de enfrentamientos, resultados y cuotas. Es la base de datos fundamental para la realización de apuestas, ya que a partir de ella extraemos todos los datos útiles para hacer las apuestas necesarias a lo largo del torneo y, posteriormente, elaborar una serie de pruebas con los resultados obtenidos para saber si se ha conseguido un beneficio y observar si nuestro sistema es rentable o no.

Las principales columnas de esta base de datos de jugadores son:

- **Torneo:** Nombre del torneo que se está disputando.
- **Fecha:** Fecha en la que se disputa el torneo. Es una cadena de caracteres.
- **Superficie:** Superficie sobre la que se disputa el torneo. Es una cadena de caracteres.
- **Ronda:** Ronda del torneo a la que pertenece el partido
- **Sets:** Atributo numérico en el que se especifica el número de sets máximo que se pueden disputar en los partidos correspondientes a este torneo.
- **Jugador A:** Jugador que ha disputado el partido obteniendo como resultado la victoria.
- **Jugador B:** Jugador que ha disputado el partido obteniendo como resultado la derrota.
- **Rank A:** Posición que ocupa en el ranking de la ATP el jugador A.

- **Rank B:** Posición que ocupa en el ranking de la ATP el jugador B.
- **Cuota A:** Atributo numérico que expresa la mayor cuota encontrada en las casas de apuestas para el jugador A.
- **Cuota B:** Atributo numérico que expresa la mayor cuota encontrada en las casas de apuestas para el jugador B.

A continuación se muestra un extracto de la base de datos de torneos, para poder apreciar las columnas de las que consta.

Torneo	Fecha	S	R	Sets	Jugador A	Jugador B	Rank A	Rank B	Cuota A	Cuota B
Montreal	08/08	Dura	1	3	Dolgopolov O.	Chvojka E.	21	290	1,08	11,66
Montreal	08/08	Dura	1	3	Wawrinka S.	Nalbandian D.	17	51	1,88	2,1
Montreal	08/08	Dura	1	3	Anderson K.	Andujar P.	35	46	1,2	5,61
Montreal	08/08	Dura	1	3	Karlovic I.	Monaco J.	125	39	1,83	2,15
Montreal	08/08	Dura	1	3	Gulbis E.	Ferrero J.C.	55	104	1,53	2,75
Montreal	08/08	Dura	1	3	Bogomolov A.	Mannarino A.	57	54	1,6	2,6
Montreal	08/08	Dura	1	3	Bellucci T.	Golubev A.	36	99	1,36	3,7
Montreal	08/08	Dura	1	3	Dodig I.	Chardy J.	41	61	1,98	2
Montreal	08/08	Dura	1	3	Davydenko N.	Cipolla F.	30	113	1,33	4,58
Montreal	08/08	Dura	1	3	Cilic M.	Seppi A.	29	52	1,34	6
Montreal	08/08	Dura	1	3	Del Potro J.M.	Nieminen J.	19	50	1,16	6,87
Montreal	09/08	Dura	1	3	Tomic B.	Lu Y.H.	68	66	1,6	2,82
Montreal	09/08	Dura	1	3	Tsonga J.W.	Fognini F.	16	34	1,14	9,24
Montreal	09/08	Dura	1	3	Gasquet R.	Mayer F.	13	23	1,47	3,05
Montreal	09/08	Dura	1	3	Petzschner P.	Simon G.	80	11	3,8	1,33
Montreal	09/08	Dura	1	3	Stakhovsky S.	Kohlschreiber P.	43	42	2,62	1,61
Montreal	09/08	Dura	1	3	Russell M.	Montanes A.	100	47	1,72	2,44
Montreal	09/08	Dura	1	3	Tipsarevic J.	Falla A.	24	109	1,31	4
Montreal	09/08	Dura	1	3	Pospisil V.	Chela J.I.	127	22	3,01	1,58
Montreal	09/08	Dura	1	3	Llodra M.	Youzhny M.	31	14	3	1,47
Montreal	09/08	Dura	1	3	Troicki V.	Yani M.	15	221	1,12	9,41
Montreal	10/08	Dura	1	3	Verdasco F.	Kamke T.	20	95	1,33	4
Montreal	10/08	Dura	1	3	Isner J.	Baghdatis M.	28	40	1,5	2,9
Montreal	10/08	Dura	1	3	Lopez F.	Stepanek R.	25	27	2,2	1,87

Tabla 2. Extracto de la base de datos de torneos

Una vez descritas todas las bases de datos del proyecto y sus principales columnas, debemos comentar que algunas de ellas son producto de la ordenación y organización de datos y otras han sido calculadas a partir de otra serie de datos o columnas, lo cual eleva la complejidad en su realización.

Por último, una vez finalizadas ambas bases de datos, se realizó una inspección exhaustiva de estas para conseguir la mayor calidad de los datos y que no provocaran problemas a la hora de ser utilizados por nuestros algoritmos. Analizamos si estaban completos, si eran correctos o si tenían cualquier tipo de error y, en caso de tenerlos, corregirlos.



### 3.5. ALGORITMO PARA EL CÁLCULO DE PROBABILIDADES DE VICTORIA

El algoritmo para el cálculo de probabilidades de victoria para cada uno de los jugadores que se enfrentan en un partido se fundamenta en la simulación del partido mediante la comparación de una serie de variables aleatorias con otras variables que tenemos almacenadas en la base de datos de jugadores.

La simulación del partido tiene como parte más compleja, y que más se repite a su vez, la sucesión de juegos. El siguiente diagrama de estados explica el desarrollo de cada juego.

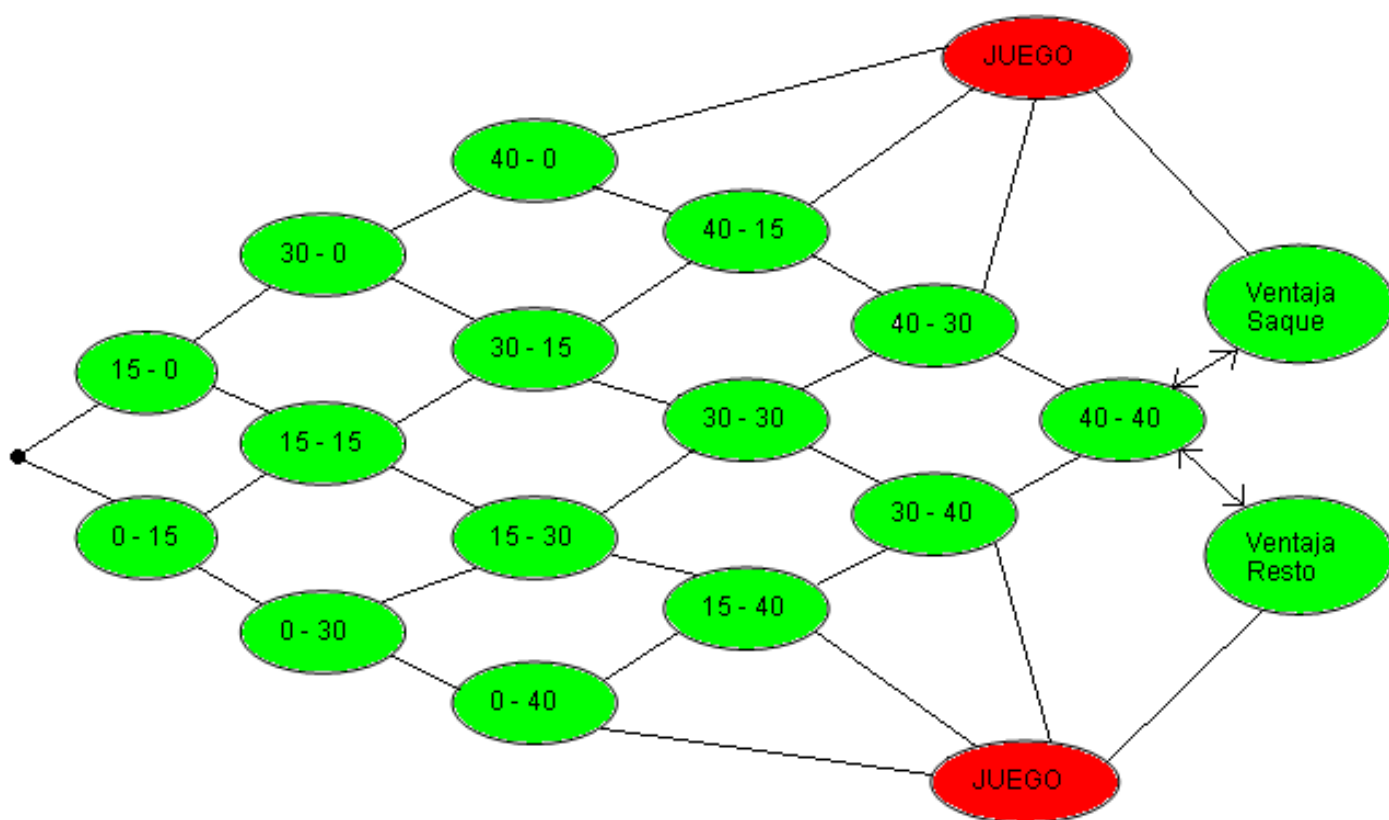


Figura 2. Diagrama de estados de un juego

En cada uno de los estados, tanto en el primer como en segundo servicio, se realiza la comparación de varias variables aleatorias con los datos correspondientes, seleccionados a partir de la base de datos de jugadores, y a partir de estas comparaciones se va decidiendo el ganador de cada punto. Así se va progresando a lo largo del diagrama de estados hasta llegar a la decisión de quién es el ganador del juego.

Una vez progresamos con esta misma política, vamos sumando juegos para cada uno de los jugadores hasta, según las leyes del tenis y del torneo específico (número de sets, disputa o no de tie break...), elegir el ganador del set y, de forma recursiva, decidir el ganador del partido.

Cuando sabemos si el ganador es el sacador o el restador, vamos acumulando el resultado hasta llegar a un ganador del set. Esto se repite set a set, hasta conocer quién es el ganador del partido.



*Figura 3. Simulación de un partido*

Posteriormente realizamos esta misma simulación del partido pero con una cantidad lo suficientemente elevada de veces, 10.000 simulaciones por cada encuentro, del tal forma que podamos tener una estadística fiable de cuál es la probabilidad de victoria para cada uno de los dos jugadores.

Tendremos un porcentaje de victoria para los dos jugadores, que será el que compararemos con los datos ofrecidos por las casas de apuestas, y veremos por quién nos interesará apostar.

### 3.6. USO DEL ALGORITMO DE KELLY PARA LA REALIZACIÓN DE APUESTAS

Una vez hemos realizado estas predicciones y cálculos, analizamos todos los datos conseguidos para saber, comparándolo con las predicciones de las casas de apuestas, por qué jugador debemos apostar.

El fundamento de nuestro modelo se basa en calcular el porcentaje de victoria y apostar por el que tenga un porcentaje mayor al de la casa de apuestas, sea el favorito para ganar o no. De hecho, es un modelo rentable gracias a las primeras rondas, en la que jugadores muy poco favoritos dan la sorpresa y eso hace que nuestra ganancia aumente cuantitativamente.

Por ejemplo, si la casa de apuestas da unas probabilidades de 80% para el jugador 1 y del 20% para el jugador 2, y nuestro modelo nos aporta un 70% a 30%, entonces nosotros apostaremos por el jugador número 2, ya que en nuestro modelo se ha obtenido un porcentaje mayor, pese a que se haya predicho que dicho jugador es más posible que pierda.

Una vez hemos decidido por qué jugador apostar es importante el cálculo de la cantidad a apostar en cada partido, que como ya hemos comentado anteriormente, lo haremos mediante el algoritmo de Kelly, el cual también ha sido implementado e incluido en nuestro modelo.

El Algoritmo de Kelly tiene como función principal comparar las probabilidades extraídas de las casas de apuestas y las calculadas mediante nuestro algoritmo. Mediante esta comparación, el algoritmo de Kelly se encarga de calcular la cantidad óptima para cada apuesta y de determinar si es rentable apostar por un determinado jugador o no.

Para poder aplicarlo necesitamos tener las probabilidades de victoria para cada jugador, estimadas por nuestro algoritmo, y las cuotas ofrecidas por las casas de apuestas.

El algoritmo consta de los siguientes pasos:

1. Multiplicación de las probabilidades estimadas por sus cuotas respectivas en las casas de apuestas. En el caso de que el resultado sea mayor que 1, será rentable apostar y seguiremos con el proceso. En el caso de que sea menor entonces no nos va a resultar conveniente apostar ninguna cantidad y pasaríamos al siguiente caso para apostar.
2. Ordenación de mayor a menor de los resultados obtenidos en el paso anterior, en el caso de que haya varios. En nuestro caso, como solo queremos apostar a un suceso, que es el ganador del partido por un jugador determinado, no es necesario ordenarlo.
3. Aplicación de la siguiente ecuación. Con ella obtendremos el valor de  $b$ , el cual necesitamos para el siguiente paso.

$$F_i = \frac{1 - p_i}{1 - \frac{1}{\alpha_i}}$$

Dicho valor de  $b$  se correspondería con el mínimo de todos los valores calculados mediante esta fórmula. En nuestro caso disponemos de un único caso, por lo tanto, coincide con el de  $F$ .

4. Obtención de la cantidad a apostar mediante la siguiente ecuación, que se corresponde con la explicada en el capítulo anterior.

$$a_i = p_i - \frac{b}{\alpha_i}$$

En nuestro caso, como solo tenemos una apuesta, la podemos describir de la siguiente manera:

$$a = p - \frac{b}{\alpha}$$

Y si desarrollamos, y sustituimos la b por su valor...

$$b = \frac{1 - p}{1 - \frac{1}{\alpha}}$$

vemos que coincide con la fórmula de Kelly, explicada en el capítulo anterior.

$$a = \frac{\alpha p - 1}{\alpha - 1}$$

Si multiplicamos a por 100 obtenemos el tanto por ciento de nuestra banca que debemos apostar.

Para explicarlo mejor, veamos un ejemplo.

Imaginemos que queremos apostar por un jugador cuya cuota es de 3,55 y su porcentaje de victoria es de 0,25.

En primer lugar, si multiplicamos ambas obtenemos un resultado de 1,0295, por lo tanto sí que nos va a ser rentable apostar por la victoria de dicho jugador. En el caso de que dicho resultado fuera menor que 1, no seguiríamos con el proceso.

Posteriormente procedemos a calcular la F, que en nuestro caso coincide con b. Si sustituimos valores obtenemos:

$$b = \frac{1 - p}{1 - \frac{1}{\alpha}} = 0,9884$$

Por último, procedemos a calcular la cantidad a apostar:

$$a = p - \frac{b}{\alpha} = \frac{\alpha p - 1}{\alpha - 1} = 0,0116$$

Por lo tanto debemos apostar un 1,16% de la banca que tengamos disponible por la victoria del jugador en cuestión.

Una vez que la ecuación nos indica la cantidad a apostar, se puede respetar o, si consideramos que es demasiado arriesgado para nuestra banca, podemos apostar respetando las proporciones pero dividiéndolo entre 10 o 100, disminuyendo así el riesgo de quedarnos sin dinero disponible para seguir apostando.

Como ya indicamos anteriormente, únicamente vamos a utilizar el Criterio de Kelly en aquellos jugadores en los que nuestras probabilidades estimadas son mayores que las de las casas de apuestas.

El modelo se basa principalmente en las victorias inesperadas que se producen y en la mejor aproximación del porcentaje de victoria de los jugadores, y es por esto que utilizamos esta política de apuestas.

Como podremos apreciar en las pruebas, casi siempre vamos a lograr un mayor beneficio en las primeras rondas que posteriormente se irá reduciendo poco a poco hasta la ronda final, lo cual ratifica que esta forma de apostar obtiene un mayor beneficio con los resultados inesperados.

## 4. PRUEBAS Y RESULTADOS.

En este apartado se presentan los resultados obtenidos al realizar apuestas en tres de los torneos de pista rápida más importantes en el circuito ATP: El Masters de Madrid, de Miami y de Montreal, siguiendo la política de nuestro modelo.

### 4.1. PRUEBA 1: TORNEO MASTERS DE MADRID

En primer lugar calculamos las probabilidades de victoria para cada jugador a partir de las cuotas de las casas de apuestas, siempre eligiendo la más favorable.

Posteriormente calculamos las mismas probabilidades utilizando nuestro modelo. Para ello, realizamos una simulación de 10000 resultados por cada partido del torneo.

En la tabla que aparece a continuación se muestra la comparación de los resultados obtenidos en ambos casos.

Probabilidades de la casa de apuestas.		Probabilidades de nuestro modelo con 10000 resultados por partido.	
Jugador A (ganador)	Jugador B (perdedor)	Jugador A (ganador)	Jugador B (perdedor)
0,3385	0,6615	0,4785	0,5215
0,3712	0,6288	0,4076	0,5924
0,3046	0,6954	0,5374	0,4626
0,6674	0,3326	0,6487	0,3513
0,2305	0,7695	0,5661	0,4339

0,4401	0,5599	0,3238	0,6762
0,6585	0,3415	0,2128	0,7872
0,3377	0,6623	0,3405	0,6595
0,1928	0,8072	0,4335	0,5665
0,3138	0,6862	0,725	0,275
0,4401	0,5599	0,4065	0,5935
0,6818	0,3182	0,6645	0,3355
0,3578	0,6422	0,1276	0,8724
0,6554	0,3446	0,731	0,269
0,1852	0,8148	0,5285	0,4715
0,2049	0,7951	0,449	0,551
0,5296	0,4704	0,5007	0,4993
0,6579	0,3421	0,5469	0,4531
0,5316	0,4684	0,5189	0,4811
0,8211	0,1789	0,4513	0,5487
0,4878	0,5122	0,4259	0,5741
0,2705	0,7295	0,5147	0,4853
0,0884	0,9116	0,1986	0,8014
0,5510	0,4490	0,5197	0,4803
0,7069	0,2931	0,7454	0,2546
0,3810	0,6190	0,3798	0,6202
0,5732	0,4268	0,6879	0,3121
0,9528	0,0472	0,7349	0,2651
0,3304	0,6696	0,593	0,407
0,6629	0,3371	0,3748	0,6252
0,5721	0,4279	0,3326	0,6674
0,9501	0,0499	0,7506	0,2494
0,6897	0,3103	0,5199	0,4801
0,7378	0,2622	0,6378	0,3622
0,6279	0,3721	0,8068	0,1932
0,6709	0,3291	0,3793	0,6207
0,9724	0,0276	0,6127	0,3873
0,7114	0,2886	0,5599	0,4401
0,8535	0,1465	0,4931	0,5069



0,2404	0,7596	0,3373	0,6627
0,9059	0,0941	0,4672	0,5328
0,9091	0,0909	0,6714	0,3286
0,7054	0,2946	0,4451	0,5549
0,6785	0,3215	0,6101	0,3899
0,8380	0,1620	0,6326	0,3674
0,1797	0,8203	0,3327	0,6673
0,4874	0,5126	0,3491	0,6509
0,9098	0,0902	0,7948	0,2052
0,3348	0,6652	0,3738	0,6262
0,9840	0,0160	0,7434	0,2566
0,6606	0,3394	0,3796	0,6204
0,7225	0,2775	0,5361	0,4639
0,8681	0,1319	0,6943	0,3057
0,8885	0,1115	0,7773	0,2227
0,2615	0,7385	0,4885	0,5115

*Tabla 3. Probabilidades de las casas de apuestas y de nuestro modelo (Masters Madrid)*

En el gráfico que se muestra a continuación se reflejan los porcentajes de victoria para los jugadores que ganaron la ronda. La línea azul muestra los resultados obtenidos por la casa de apuestas y la línea roja los obtenidos con nuestro modelo. Como podemos observar, en muchos casos esos porcentajes se asemejan mucho, como en el partido 4, y en otros casos esos porcentajes son muy dispares, como en el caso de los partidos 20 y 22.

Esta disparidad de criterios en ciertos partidos es la que vamos a aprovechar a la hora de apostar y posiblemente nos reporte un gran beneficio.

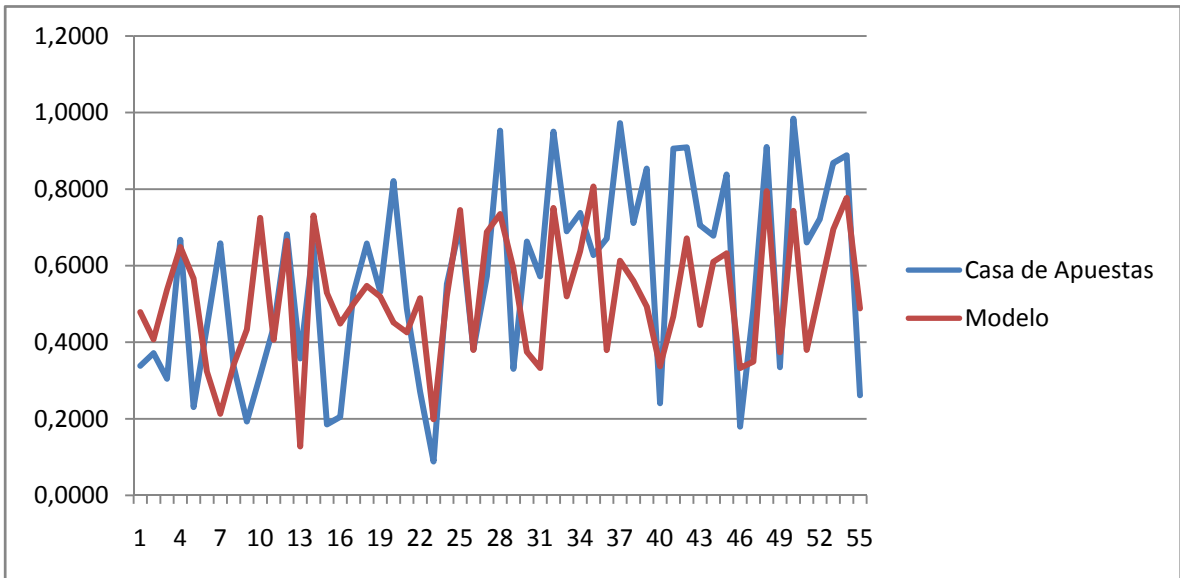


Figura 4. Porcentaje de victoria de los jugadores que ganaron la ronda (Masters Madrid)

En el gráfico que se muestra a continuación se reflejan los porcentajes de victoria de los jugadores que perdieron la ronda. La línea azul muestra los resultados obtenidos por la casa de apuestas y la línea roja los obtenidos con nuestro modelo. Obviamente, al igual que en el caso anterior, en muchos casos esos porcentajes se asemejan mucho, como en el partido 4, y en otros casos esos porcentajes son muy dispares, como en el caso de los partidos 20 y 22.

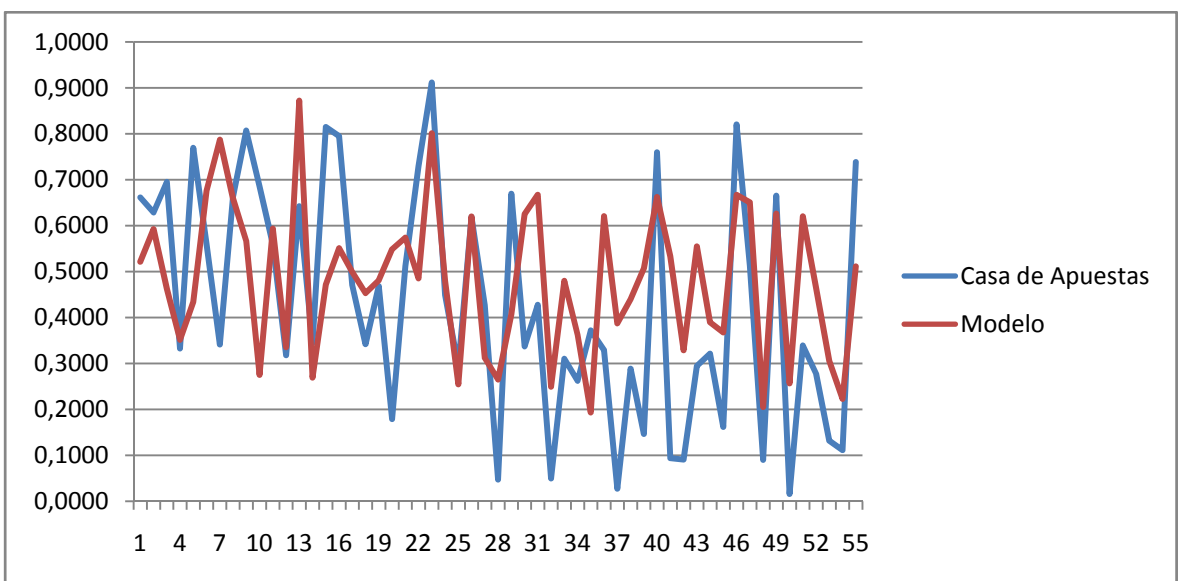


Figura 5. Porcentaje de victoria de los jugadores que perdieron la ronda (Masters Madrid)

Una vez que ya tenemos tanto los porcentajes de las casas de apuestas como los calculados por nuestro modelo, podemos dar el siguiente paso y comenzar a apostar.

A la hora de realizar nuestras apuestas, siempre vamos a hacerlo sobre el jugador en el que la probabilidad de victoria calculada por nosotros es mayor que la de la casa de apuestas. Por ejemplo, si la casa de apuestas le da un porcentaje de victoria a un jugador de 0,65 y el calculado por nosotros es de 0,80, apostaremos por ese jugador. Del mismo modo, si la casa de apuestas le da a un jugador un 0,20 y nosotros un 0,35, apostaremos por ese jugador aunque no sea favorito para ganar el partido, ya que nuestro porcentaje calculado es mucho mayor que el de la casa de apuestas.

Hemos fijado una banca de 100 euros en total para las apuestas, y, siguiendo el criterio de Kelly, tal y como explicamos anteriormente, hemos calculado la cantidad a apostar en cada uno de los partidos.

En la tabla que se muestra a continuación podemos observar la cantidad que apostamos en cada uno de los partidos (columna 1), el jugador por el que tenemos que apostar (columna 2), la ganancia que obtenemos (columna 3), el beneficio obtenido en ese partido, que es la diferencia entre la ganancia y la cantidad apostada (columna 4), y, por último, el beneficio total, que es el beneficio acumulado (columna 5), que en el caso de este torneo es 5,38115 euros.

Cantidad a apostar	Jugador por el que apostamos	Ganancia	Beneficio	Beneficio Total
0,21644	A	0,64715	0,43072	0,43072
0,06117	A	0,16576	0,10460	0,53531

0,33714	A	1,11593	0,77879	1,31411
0,02856	B	0,00000	-0,02856	1,28554
0,43619	A	1,89306	1,45687	2,74242
0,27145	B	0,00000	-0,27145	2,47097
0,67807	B	0,00000	-0,67807	1,79289
0,01075	A	0,03225	0,02150	1,81439
0,29699	A	1,52952	1,23253	3,04692
0,60439	A	1,98239	1,37800	4,42492
0,08538	B	0,00000	-0,08538	4,33954
0,02643	B	0,00000	-0,02643	4,31311
0,64454	B	0,00000	-0,64454	3,66857
0,22345	A	0,34188	0,11843	3,78700
0,42372	A	2,33047	1,90675	5,69375
0,30588	A	1,48353	1,17765	6,87140
0,06391	B	0,00000	-0,06391	6,80749
0,17965	B	0,00000	-0,17965	6,62784
0,00937	B	0,00000	-0,00937	6,61847
0,45059	B	0,00000	-0,45059	6,16788
0,14820	B	0,00000	-0,14820	6,01968
0,33087	A	1,20438	0,87351	6,89319
0,12117	A	1,37528	1,25411	8,14730
0,07109	B	0,00000	-0,07109	8,07621
0,12442	A	0,17544	0,05101	8,12722
0,00000	B	0,00000	0,00000	8,12722
0,27177	A	0,47559	0,20383	8,33105
0,22836	B	0,00000	-0,22836	8,10269
0,38950	A	1,16850	0,77900	8,88169
0,43495	B	0,00000	-0,43495	8,44675
0,44567	B	0,00000	-0,44567	8,00108
0,20989	B	0,00000	-0,20989	7,79118
0,24378	B	0,00000	-0,24378	7,54740
0,13195	B	0,00000	-0,13195	7,41546
0,48480	A	0,77568	0,29088	7,70634
0,44671	B	0,00000	-0,44671	7,25963

0,36979	B	0,00000	-0,36979	6,88983
0,21614	B	0,00000	-0,21614	6,67369
0,42039	B	0,00000	-0,42039	6,25330
0,12625	A	0,52267	0,39642	6,64972
0,48459	B	0,00000	-0,48459	6,16514
0,26146	B	0,00000	-0,26146	5,90368
0,36944	B	0,00000	-0,36944	5,53424
0,09373	B	0,00000	-0,09373	5,44050
0,24088	B	0,00000	-0,24088	5,19962
0,18668	A	1,03982	0,85314	6,05276
0,27552	B	0,00000	-0,27552	5,77724
0,12651	B	0,00000	-0,12651	5,65073
0,06070	A	0,18210	0,12140	5,77213
0,24456	B	0,00000	-0,24456	5,52757
0,42269	B	0,00000	-0,42269	5,10488
0,26895	B	0,00000	-0,26895	4,83592
0,20508	B	0,00000	-0,20508	4,63085
0,12554	B	0,00000	-0,12554	4,50531
0,30839	A	1,18423	0,87584	5,38115
		19,62566	5,38115	

*Tabla 4. Apuestas realizadas a lo largo del torneo (Masters Madrid)*

En la gráfica que se muestra a continuación podemos observar la ganancia en euros obtenida partido a partido.

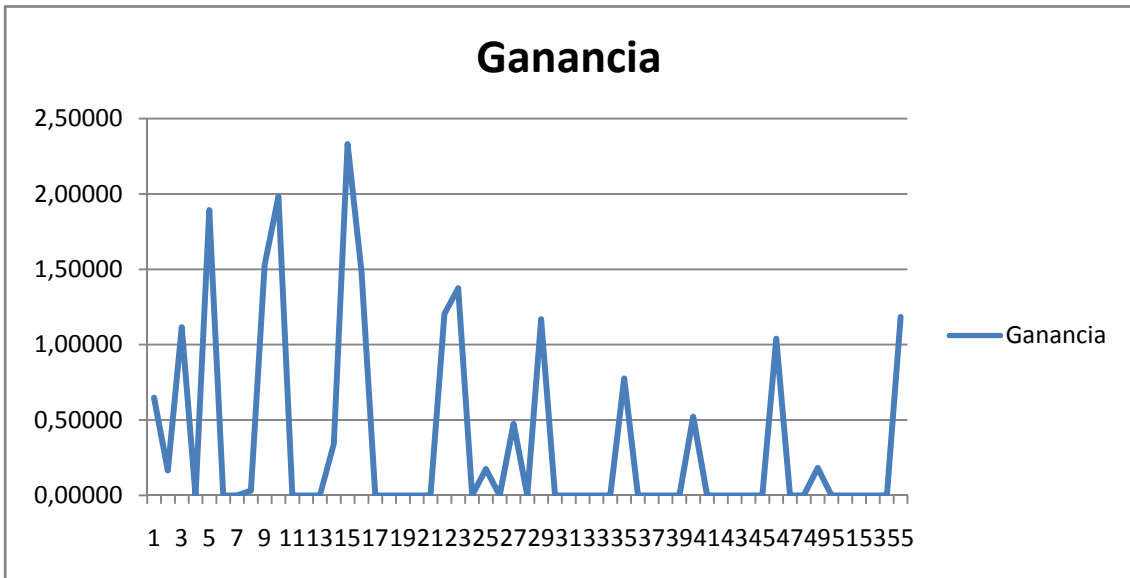
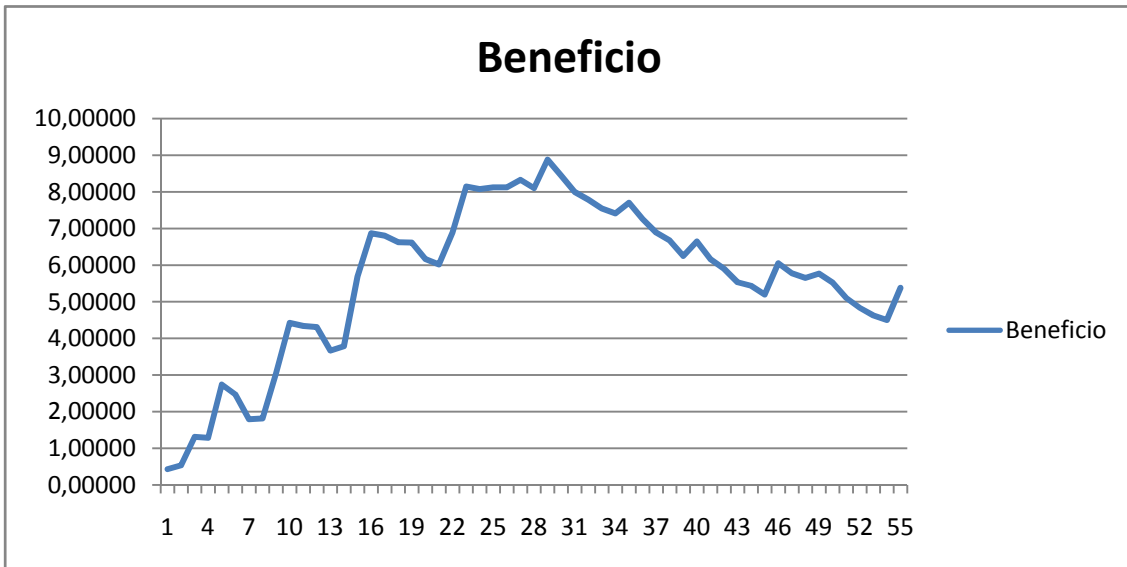


Figura 6. Ganancia obtenida partido a partido (Masters Madrid)

Los 40 primeros partidos corresponden a la primera y segunda ronda del torneo, siendo estas en las que se consigue una mayor ganancia, especialmente en la primera, debido a que no siempre los jugadores favoritos vencen, y son esos partidos los que aprovechamos para aumentar nuestras ganancias.

Como vemos en la gráfica, los picos producidos son la base del éxito de nuestro modelo de apuestas, ya que obtenemos ganancias en pocos partidos, pero estas son muy grandes, y esto es lo que nos reporta al final un beneficio.

La gráfica que se muestra a continuación muestra el beneficio acumulado, partido a partido, a lo largo del torneo.



*Figura 7. Beneficio acumulado a lo largo del torneo (Masters Madrid)*

Como podemos observar, el beneficio es creciente durante las dos primeras rondas, obteniendo su máximo al final de la segunda ronda aproximadamente. Esto no hace más que ratificar que el periodo idóneo para esta política de apuestas son las dos primeras rondas. En el caso de que solamente hubiéramos apostado durante estas, habríamos obtenido un beneficio de unos 9 euros. Sin embargo, el beneficio final es de 5,38115 euros ya que realizamos apuestas durante todo el torneo.

En esta última tabla vamos a dividir la información según los porcentajes, trabajando en rangos del 10% (columna 1). En la columna 2 se muestran los partidos ganados por los jugadores que tenían un porcentaje calculado por nuestro modelo dentro de su rango correspondiente. En la columna 3 se aplica el mismo criterio pero para los partidos perdidos. En la columna 4 calculamos la proporción entre partidos ganados y partidos totales. En las columnas 5, 6, y 7 se establecen los mismos criterios que en las columnas 2, 3 y 4 respectivamente, pero teniendo en cuenta los datos de las casas de apuestas.

Porcentaje	Ganados Modelo	Perdidos Modelo	Ganados/Totales Modelo	Ganados Casa Apuestas	Perdidos Casa Apuestas	Ganados/Totales Casa Apuestas
0 – 10 %	0	0	0/0	1	7	1/8 = 12%
10 – 20 %	2	1	2/3 = 67%	3	5	3/8 = 37%
20 – 30 %	1	8	1/9 = 11%	5	5	5/10 = 50%
30 – 40 %	11	9	11/20 = 55%	9	11	9/20 = 45%
40 – 50 %	11	12	11/23 = 48%	4	5	4/9 = 44%
50 – 60 %	12	11	12/23 = 52%	5	4	5/9 = 55%
60 – 70 %	9	11	9/20 = 45%	11	9	11/20 = 55%
70 – 80 %	8	1	8/9 = 89%	5	5	5/10 = 50%
80 – 90 %	1	2	1/3 = 33%	5	3	5/8 = 62%
90 – 100 %	0	0	0/0	7	1	7/9 = 78%

*Tabla 5. Resultados obtenidos (Masters Madrid)*

Si nos fijamos detenidamente en los rangos medios de porcentajes, es decir, del 40% al 60%, el valor numérico de la columna de **Ganados/Totales** está incluido en el rango de la de **Porcentaje**. Existen varias franjas que no tienen los suficientes datos para ser representativos, por ejemplo entre el 0% y el 20%, o entre el 80% y el 100%.



## 4.2. PRUEBA 2: TORNEO MASTERS DE MIAMI

Al igual que en el caso anterior, calculamos las probabilidades de victoria para cada jugador a partir de las cuotas de las casas de apuestas, siempre eligiendo la más favorable.

Posteriormente calculamos las mismas probabilidades utilizando nuestro modelo. Para ello, realizamos una simulación de 10.000 resultados por cada partido del torneo.

En la tabla que aparece a continuación se muestra la comparación de los resultados obtenidos en ambos casos.

Probabilidades de la casa de apuestas.		Probabilidades de nuestro modelo con 10000 resultados por partido.	
Jugador A (ganador)	Jugador B (perdedor)	Jugador A (ganador)	Jugador B (perdedor)
0,4175	0,5825	0,4988	0,5012
0,3297	0,6703	0,537	0,463
0,2188	0,7813	0,4648	0,5352
0,2681	0,7319	0,6939	0,3061
0,7813	0,2188	0,2612	0,7388
0,3794	0,6206	0,2377	0,7623
0,1337	0,8663	0,7096	0,2904
0,4775	0,5225	0,511	0,489
0,5250	0,4750	0,7095	0,2905
0,7505	0,2495	0,7962	0,2038

0,9447	0,0553	0,652	0,348
0,5387	0,4613	0,4789	0,5211
0,4447	0,5553	0,6916	0,3084
0,5000	0,5000	0,844	0,156
0,5674	0,4326	0,6706	0,3294
0,4403	0,5597	0,1641	0,8359
0,8062	0,1938	0,7072	0,2928
0,6250	0,3750	0,5688	0,4312
0,6308	0,3692	0,7132	0,2868
0,4581	0,5419	0,3382	0,6618
0,4885	0,5115	0,7067	0,2933
0,6623	0,3377	0,5365	0,4635
0,7222	0,2778	0,5139	0,4861
0,5831	0,4169	0,3816	0,6184
0,1609	0,8391	0,4166	0,5834
0,5161	0,4839	0,56	0,44
0,5303	0,4697	0,672	0,328
0,5841	0,4159	0,4799	0,5201
0,6000	0,4000	0,4481	0,5519
0,4181	0,5819	0,2763	0,7237
0,3813	0,6187	0,4669	0,5331
0,3776	0,6224	0,6717	0,3283
0,3776	0,6224	0,5433	0,4567
0,3462	0,6538	0,6815	0,3185

0,8587	0,1413	0,5678	0,4322
0,6811	0,3189	0,7298	0,2702
0,6185	0,3815	0,6227	0,3773
0,8345	0,1655	0,5421	0,4579
0,4512	0,5488	0,0886	0,9114
0,0876	0,9124	0,395	0,605
0,1195	0,8805	0,2997	0,7003
0,1468	0,8532	0,1417	0,8583
0,8977	0,1023	0,5745	0,4255
0,6696	0,3304	0,5475	0,4525
0,8635	0,1365	0,624	0,376
0,5847	0,4153	0,5553	0,4447
0,9284	0,0716	0,7047	0,2953
0,7547	0,2453	0,477	0,523
0,7285	0,2715	0,4498	0,5502
0,8412	0,1588	0,8581	0,1419
0,4248	0,5752	0,521	0,479
0,5316	0,4684	0,534	0,466
0,7021	0,2979	0,6266	0,3734
0,9391	0,0609	0,4416	0,5584
0,3759	0,6241	0,3953	0,6047
0,6545	0,3455	0,4864	0,5136
0,4296	0,5704	0,2822	0,7178
0,7856	0,2144	0,421	0,579

0,1815	0,8185	0,2748	0,7252
0,1099	0,8901	0,1017	0,8983
0,7607	0,2393	0,5922	0,4078
0,3651	0,6349	0,4254	0,5746
0,9314	0,0686	0,6148	0,3852
0,9379	0,0621	0,6932	0,3068
0,4129	0,5871	0,4457	0,5543
0,7967	0,2033	0,6412	0,3588
0,8134	0,1866	0,4022	0,5978
0,4347	0,5653	0,2275	0,7725
0,7356	0,2644	0,5344	0,4656
0,5493	0,4507	0,4844	0,5156
0,5841	0,4159	0,4626	0,5374
0,9602	0,0398	0,9513	0,0487
0,9248	0,0752	0,79	0,21
0,5861	0,4139	0,6603	0,3397
0,2000	0,8000	0,4518	0,5482
0,4541	0,5459	0,479	0,521
0,9489	0,0511	0,5776	0,4224
0,9396	0,0604	0,6582	0,3418
0,7621	0,2379	0,6602	0,3398
0,5725	0,4275	0,4701	0,5299
0,2228	0,7772	0,5221	0,4779
0,8553	0,1447	0,8751	0,1249

0,4265	0,5735	0,5807	0,4193
0,6842	0,3158	0,5187	0,4813
0,9396	0,0604	0,6514	0,3486
0,8762	0,1238	0,5987	0,4013
0,5791	0,4209	0,5126	0,4874
0,9675	0,0325	0,5295	0,4705
0,3363	0,6637	0,4254	0,5746
0,9432	0,0568	0,6529	0,3471
0,8526	0,1474	0,5594	0,4406
0,8713	0,1287	0,6758	0,3242
0,8871	0,1129	0,6139	0,3861
0,5665	0,4335	0,7014	0,2986
0,4761	0,5239	0,4898	0,5102

*Tabla 6. Probabilidades de las casas de apuestas y de nuestro modelo (Masters Miami)*

En el gráfico que se muestra a continuación se reflejan los porcentajes de victoria para los jugadores que ganaron la ronda. La línea azul muestra los resultados obtenidos por la casa de apuestas y la línea roja los obtenidos con nuestro modelo. Como podemos observar, en muchos casos esos porcentajes se asemejan mucho, como en los partidos 8 y 12, y en otros casos esos porcentajes son muy dispares, como en el caso de los partidos 4 y 5.

Esta disparidad de criterios en ciertos partidos es la que vamos a aprovechar a la hora de apostar y posiblemente nos reporte un gran beneficio.

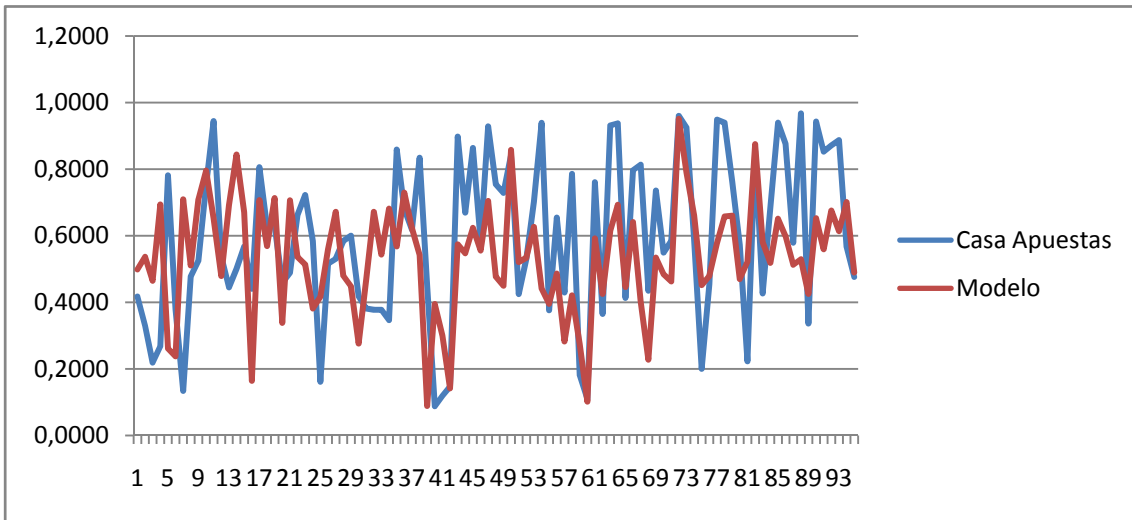


Figura 8. Porcentaje de victoria de los jugadores que ganaron la ronda (Masters Miami)

En el gráfico que se muestra a continuación se reflejan los porcentajes de victoria de los jugadores que perdieron la ronda. La línea azul muestra los resultados obtenidos por la casa de apuestas y la línea roja los obtenidos con nuestro modelo. Obviamente, al igual que en el caso anterior, en muchos casos esos porcentajes se asemejan mucho, como en los partidos 8 y 12, y en otros casos esos porcentajes son muy dispares, como en el caso de los partidos 4 y 5.

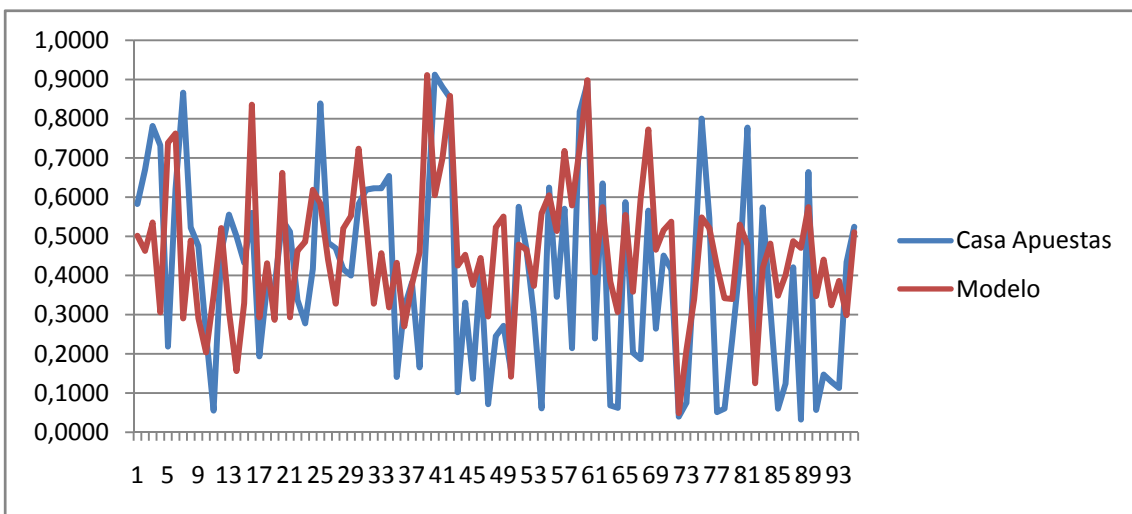


Figura 9. Porcentaje de victoria de los jugadores que perdieron la ronda (Masters Miami)

Una vez que ya tenemos tanto los porcentajes de las casas de apuestas como los calculados por nuestro modelo, podemos dar el siguiente paso y comenzar a apostar.

En la tabla que se muestra a continuación podemos observar la cantidad que apostamos en cada uno de los partidos (columna 1), el jugador por el que tenemos que apostar (columna 2), la ganancia que obtenemos (columna 3), el beneficio obtenido en ese partido, que es la diferencia entre la ganancia y la cantidad apostada (columna 4), y, por último, el beneficio total, que es el beneficio acumulado (columna 5), que en el caso de este torneo es 8,69614 euros.

Cantidad a apostar	Jugador por el que apostamos	Ganancia	Beneficio	Beneficio Total
0,14080	A	0,33792	0,19712	0,19712
0,31115	A	0,94900	0,63785	0,83497
0,31189	A	1,40349	1,09160	1,92657
0,58218	A	2,17737	1,59519	3,52176
0,66417	B	0,00000	-0,66417	2,85758
0,37891	B	0,00000	-0,37891	2,47867
0,66458	A	4,95110	4,28652	6,76519
0,06238	A	0,13037	0,06799	6,83318
0,38672	A	0,73477	0,34805	7,18123
0,17862	A	0,23757	0,05895	7,24018
0,30987	B	0,00000	-0,30987	6,93031
0,10826	B	0,00000	-0,10826	6,82205
0,45618	A	1,05378	0,59760	7,41965
0,68800	A	1,37600	0,68800	8,10765
0,27373	A	0,50093	0,22720	8,33485
0,62278	B	0,00000	-0,62278	7,71206
0,12442	B	0,00000	-0,12442	7,58764

0,08647	B	0,00000	-0,08647	7,50117
0,21872	A	0,34557	0,12686	7,62803
0,26854	B	0,00000	-0,26854	7,35948
0,41630	A	0,83677	0,42047	7,77995
0,19525	B	0,00000	-0,19525	7,58470
0,29144	B	0,00000	-0,29144	7,29326
0,34967	B	0,00000	-0,34967	6,94359
0,30221	A	1,84347	1,54126	8,48485
0,09684	A	0,18884	0,09200	8,57685
0,29060	A	0,54052	0,24992	8,82677
0,18451	B	0,00000	-0,18451	8,64226
0,27358	B	0,00000	-0,27358	8,36869
0,33455	B	0,00000	-0,33455	8,03414
0,15515	A	0,42044	0,26530	8,29944
0,48083	A	1,30785	0,82702	9,12647
0,26983	A	0,72044	0,45061	9,57708
0,51118	A	1,46708	0,95591	10,53298
0,33741	B	0,00000	-0,33741	10,19557
0,15491	A	0,22771	0,07281	10,26838
0,00418	A	0,00672	0,00255	10,27093
0,34948	B	0,00000	-0,34948	9,92145
0,80716	B	0,00000	-0,80716	9,11428
0,33598	A	3,77973	3,44375	12,55803
0,20416	A	1,70066	1,49650	14,05453
0,02477	B	0,00000	-0,02477	14,02976
0,36167	B	0,00000	-0,36167	13,66810
0,18412	B	0,00000	-0,18412	13,48398
0,27552	B	0,00000	-0,27552	13,20846
0,06173	B	0,00000	-0,06173	13,14673
0,24109	B	0,00000	-0,24109	12,90563
0,36400	B	0,00000	-0,36400	12,54163
0,38046	B	0,00000	-0,38046	12,16117
0,06977	A	0,08232	0,01256	12,17373
0,17136	A	0,40613	0,23477	12,40850



0,00000	A	0,00000	0,00000	12,40850
0,10097	B	0,00000	-0,10097	12,30753
0,52991	B	0,00000	-0,52991	11,77762
0,03749	A	0,10085	0,06336	11,84098
0,25488	B	0,00000	-0,25488	11,58610
0,35131	B	0,00000	-0,35131	11,23480
0,46491	B	0,00000	-0,46491	10,76989
0,11364	A	0,62504	0,51140	11,28129
0,11599	B	0,00000	-0,11599	11,16530
0,21677	B	0,00000	-0,21677	10,94853
0,09326	A	0,25460	0,16134	11,10987
0,33932	B	0,00000	-0,33932	10,77055
0,26059	B	0,00000	-0,26059	10,50996
0,06604	A	0,16246	0,09642	10,60639
0,19439	B	0,00000	-0,19439	10,41200
0,50742	B	0,00000	-0,50742	9,90458
0,49840	B	0,00000	-0,49840	9,40618
0,28445	B	0,00000	-0,28445	9,12173
0,12178	B	0,00000	-0,12178	8,99995
0,21390	B	0,00000	-0,21390	8,78605
0,00921	B	0,00000	-0,00921	8,77684
0,14498	B	0,00000	-0,14498	8,63186
0,19496	A	0,33728	0,14232	8,77417
0,31475	A	1,57375	1,25900	10,03317
0,04483	A	0,09863	0,05380	10,08697
0,39118	B	0,00000	-0,39118	9,69580
0,29934	B	0,00000	-0,29934	9,39646
0,13728	B	0,00000	-0,13728	9,25918
0,18676	B	0,00000	-0,18676	9,07241
0,38556	A	1,73501	1,34945	10,42186
0,04243	A	0,04880	0,00636	10,42823
0,27686	A	0,65893	0,38207	10,81030
0,25077	B	0,00000	-0,25077	10,55953
0,30657	B	0,00000	-0,30657	10,25295

0,31577	B	0,00000	-0,31577	9,93718
0,11595	B	0,00000	-0,11595	9,82123
0,45285	B	0,00000	-0,45285	9,36838
0,13224	A	0,39142	0,25918	9,62757
0,30734	B	0,00000	-0,30734	9,32023
0,34737	B	0,00000	-0,34737	8,97286
0,22258	B	0,00000	-0,22258	8,75029
0,30739	B	0,00000	-0,30739	8,44289
0,30851	A	0,54297	0,23446	8,67736
0,01739	A	0,03618	0,01878	8,69614
		34,29248	8,69614	

Tabla 7. Apuestas realizadas a lo largo del torneo (Masters Miami)

En la gráfica que se muestra a continuación podemos observar la ganancia en euros obtenida partido a partido.

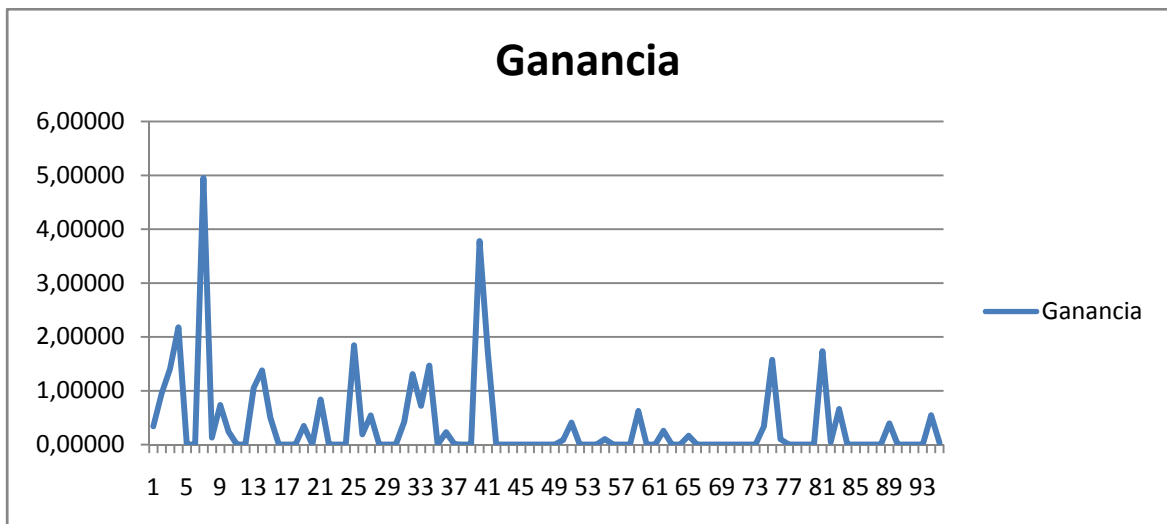


Figura 10. Ganancia obtenida partido a partido (Masters Miami)

Los 48 primeros partidos corresponden a la primera ronda del torneo, siendo esta en las que se consigue una mayor ganancia, debido a que no

siempre los jugadores favoritos vencen, y son esos partidos los que aprovechamos para aumentar nuestras ganancias.

La gráfica que se muestra a continuación muestra el beneficio acumulado durante todos los partidos a lo largo del torneo.

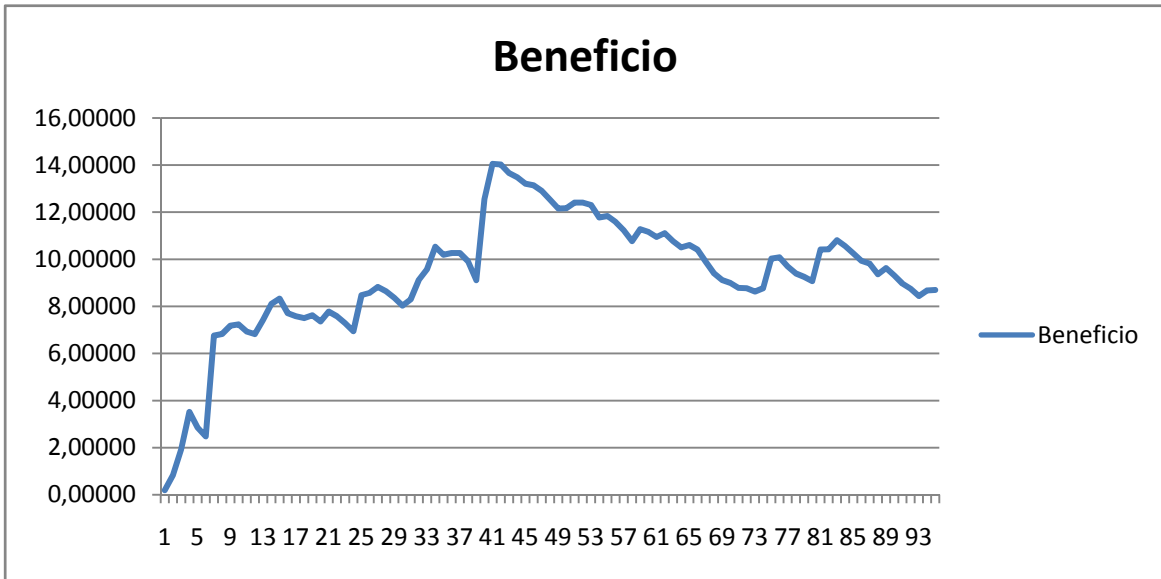


Figura 11. Beneficio acumulado a lo largo del torneo (Masters Miami)

Como podemos observar, el beneficio es creciente durante la primera ronda, obteniendo su máximo al final de la primera ronda aproximadamente. En el caso de que solamente hubiéramos apostado durante dicha ronda, habríamos obtenido un beneficio de unos 14 euros. Sin embargo, el beneficio final es de 8,69614 euros ya que realizamos apuestas durante todo el torneo.

En esta última tabla vamos a dividir la información según los porcentajes, trabajando en rangos del 10% (columna 1). En la columna 2 se muestran los partidos ganados por los jugadores que tenían un porcentaje calculado por nuestro modelo dentro de su rango correspondiente. En la

columna 3 se aplica el mismo criterio pero para los partidos perdidos. En la columna 4 calculamos la proporción entre partidos ganados y partidos totales. En las columnas 5, 6, y 7 se establecen los mismos criterios que en las columnas 2, 3 y 4 respectivamente, pero teniendo en cuenta los datos de las casas de apuestas.

Porcentaje	Ganados Modelo	Perdidos Modelo	Ganados/Totales Modelo	Ganados Casa Apuestas	Perdidos Casa Apuestas	Ganados/Totales Casa Apuestas
0 – 10 %	1	1	1/2=50%	1	12	1/13=8%
10 – 20 %	3	3	3/6=50%	6	12	6/18=33%
20 – 30 %	7	10	7/17=41%	3	11	3/14=21%
30 – 40 %	4	20	4/24=17%	9	8	9/17=53%
40 – 50 %	22	24	22/46=48%	15	15	15/30=50%
50 – 60 %	24	22	24/46=52%	15	15	15/30=50%
60 – 70 %	20	4	20/24=83%	8	9	8/17=47%
70 – 80 %	10	7	10/17=59%	11	3	11/14=79%
80 – 90 %	3	3	3/6=50%	12	6	12/18=67%
90 – 100 %	1	1	1/2=50%	12	1	12/13=92%

*Tabla 8. Resultados obtenidos (Masters Miami)*

Si nos fijamos detenidamente en los rangos medios de porcentajes, es decir, del 40% al 60%, el valor numérico de la columna de **Ganados/Totales** está incluido en el rango de la de **Porcentaje**. Existen varias franjas que no tienen los suficientes datos para ser representativos, por ejemplo entre el 0% y el 20%, o entre el 80% y el 100%.

### 4.3. PRUEBA 3: TORNEO MASTERS DE MONTREAL

En la tabla que aparece a continuación se muestra la comparación de las probabilidades de victoria para cada jugador a partir de las cuotas de las casas de apuestas y las calculadas por nuestro modelo.

Probabilidades de la casa de apuestas.		Probabilidades de nuestro modelo con 10000 resultados por partido.	
Jugador A (ganador)	Jugador B (perdedor)	Jugador A (ganador)	Jugador B (perdedor)
0,9152	0,0848	0,9328	0,0672
0,5276	0,4724	0,4816	0,5184
0,8238	0,1762	0,5089	0,4911
0,5402	0,4598	0,5264	0,4736
0,6425	0,3575	0,2062	0,7938
0,6190	0,3810	0,5319	0,4681
0,7312	0,2688	0,5756	0,4244
0,5025	0,4975	0,4413	0,5587
0,7750	0,2250	0,7762	0,2238
0,8174	0,1826	0,6104	0,3896
0,8555	0,1445	0,6725	0,3275
0,6380	0,3620	0,7685	0,2315
0,8902	0,1098	0,5642	0,4358
0,6748	0,3252	0,5137	0,4863

0,2593	0,7407	0,4289	0,5711
0,3806	0,6194	0,7328	0,2672
0,5865	0,4135	0,3297	0,6703
0,7533	0,2467	0,7441	0,2559
0,3442	0,6558	0,487	0,513
0,3289	0,6711	0,4107	0,5893
0,8936	0,1064	0,9352	0,0648
0,7505	0,2495	0,7659	0,2341
0,6591	0,3409	0,4677	0,5323
0,4595	0,5405	0,5467	0,4533
0,1151	0,8849	0,3598	0,6402
0,6391	0,3609	0,4116	0,5884
0,9709	0,0291	0,4957	0,5043
0,4370	0,5630	0,5522	0,4478
0,7143	0,2857	0,5747	0,4253
0,7784	0,2216	0,7518	0,2482
0,5520	0,4480	0,4135	0,5865
0,8012	0,1988	0,5318	0,4682
0,9299	0,0701	0,6183	0,3817
0,7692	0,2308	0,7003	0,2997
0,2664	0,7336	0,3428	0,6572
0,4478	0,5522	0,576	0,424
0,0587	0,9413	0,3453	0,6547
0,4442	0,5558	0,5755	0,4245

0,6457	0,3543	0,254	0,746
0,6424	0,3576	0,7602	0,2398
0,7160	0,2840	0,4471	0,5529
0,6100	0,3900	0,6048	0,3952
0,8677	0,1323	0,6755	0,3245
0,6546	0,3454	0,559	0,441
0,4989	0,5011	0,4505	0,5495
0,6047	0,3953	0,9323	0,0677
0,2816	0,7184	0,639	0,361
0,3401	0,6599	0,3799	0,6201
0,5938	0,4063	0,7161	0,2839
0,4097	0,5903	0,5953	0,4047
0,7414	0,2586	0,6393	0,3607
0,8893	0,1107	0,6782	0,3218
0,5755	0,4245	0,6669	0,3331
0,7670	0,2330	0,566	0,434
0,8910	0,1090	0,4452	0,5548

*Tabla 9. Probabilidades de las casas de apuestas y de nuestro modelo (Masters Montreal)*

En el gráfico que se muestra a continuación se reflejan los porcentajes de victoria de los jugadores que ganaron la ronda. Como podemos observar, en muchos casos esos porcentajes se asemejan mucho, como en los partidos 1 y 9, y en otros casos esos porcentajes son muy dispares, como en el caso de los partidos 23 y 24.

Esta disparidad de criterios en ciertos partidos es la que vamos a aprovechar a la hora de apostar y posiblemente nos reporte un gran beneficio.

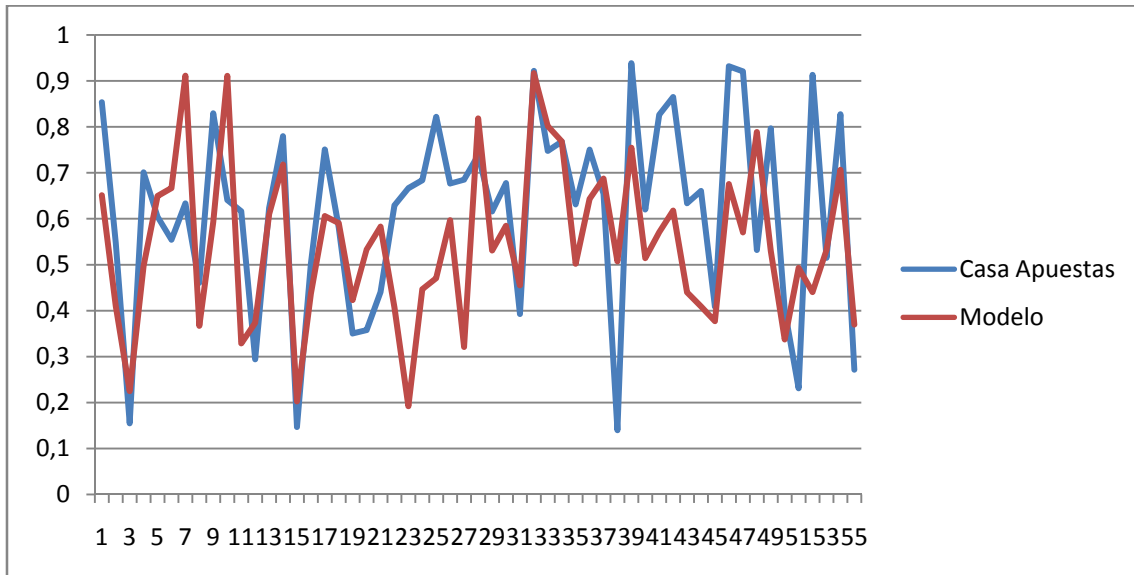


Figura 12. Porcentaje de victoria de los jugadores que ganaron la ronda (Masters Montreal)

En el gráfico que se muestra a continuación se reflejan los porcentajes de victoria de los jugadores que perdieron la ronda. Obviamente, al igual que en el caso anterior, en muchos casos esos porcentajes se asemejan mucho, como en los partidos 1 y 9, y en otros casos esos porcentajes son muy dispares, como en el caso de los partidos 23 y 24.



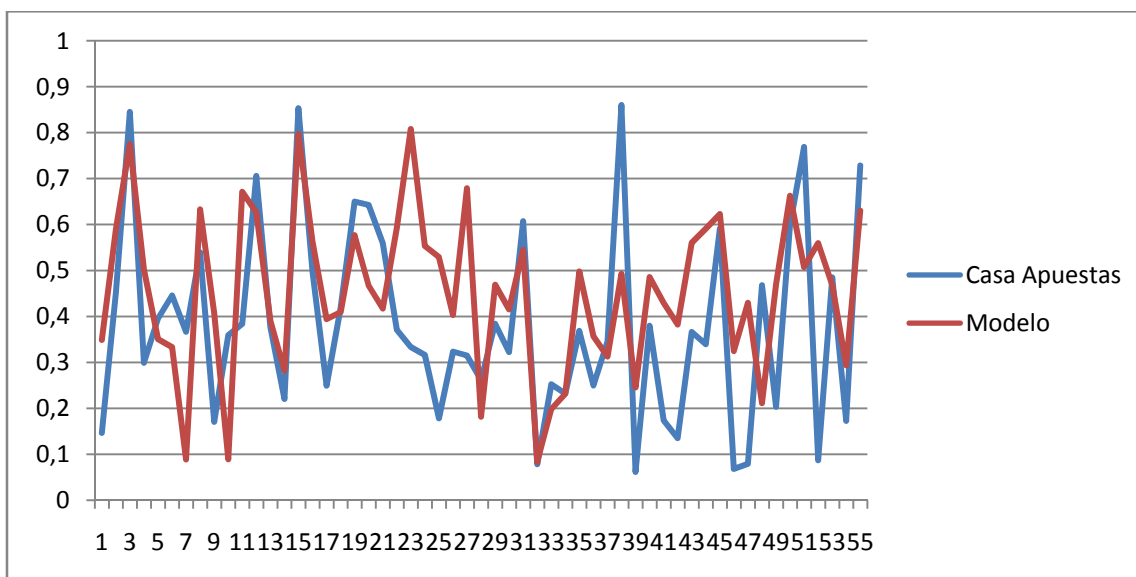


Figura 13. Porcentaje de victoria de los jugadores que perdieron la ronda (Masters Montreal)

Una vez calculados tanto los porcentajes de las casas de apuestas como los calculados por nuestro modelo, podemos comenzar a apostar.

En la tabla que se muestra a continuación podemos observar la cantidad que apostamos en cada uno de los partidos (columna 1), el jugador por el que tenemos que apostar (columna 2), la ganancia que obtenemos (columna 3), el beneficio obtenido en ese partido (columna 4), y, por último, el beneficio total (columna 5), que es el beneficio acumulado, que en el caso de este torneo es 5,6894 euros.

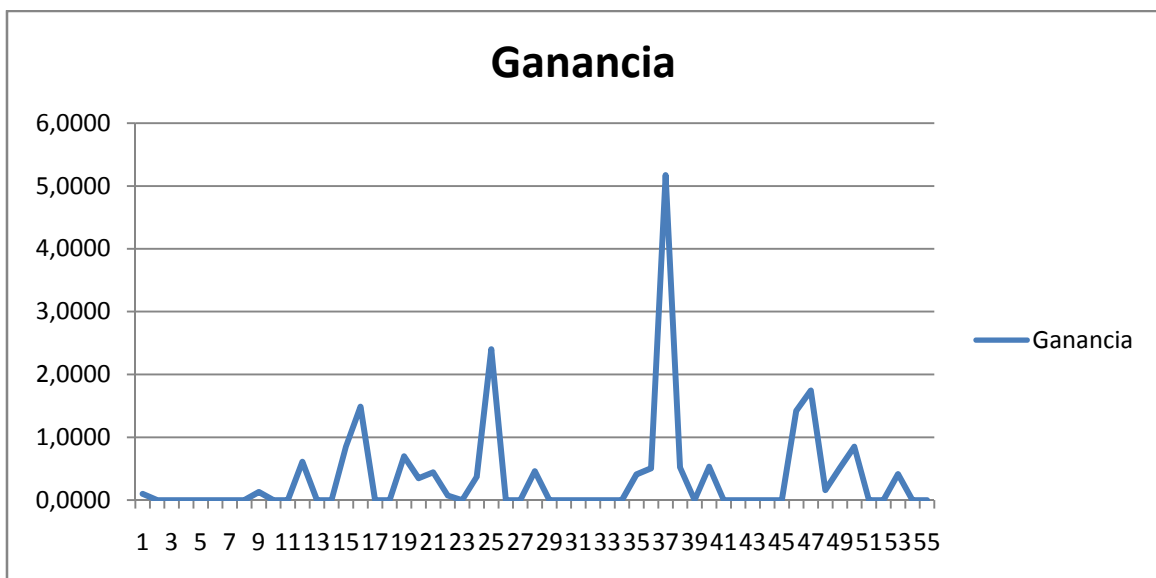
Cantidad a apostar	Jugador por el que apostamos	Ganancia	Beneficio	Beneficio Total
0,0928	A	0,1002	0,0074	0,0074
0,0806	B	0,0000	-0,0806	-0,0732
0,3807	B	0,0000	-0,3807	-0,4539
0,0159	B	0,0000	-0,0159	-0,4697
0,6760	B	0,0000	-0,6760	-1,1457
0,1357	B	0,0000	-0,1357	-1,2814
0,2112	B	0,0000	-0,2112	-1,4926

0,1174	B	0,0000	-0,1174	-1,6100
0,0980	A	0,1304	0,0323	-1,5776
0,2675	B	0,0000	-0,2675	-1,8452
0,2129	B	0,0000	-0,2129	-2,0581
0,3827	A	0,6123	0,2296	-1,8285
0,3673	B	0,0000	-0,3673	-2,1958
0,2357	B	0,0000	-0,2357	-2,4315
0,2249	A	0,8548	0,6298	-1,8017
0,5679	A	1,4878	0,9199	-0,8818
0,4413	B	0,0000	-0,4413	-1,3231
0,0079	B	0,0000	-0,0079	-1,3310
0,2318	A	0,6976	0,4659	-0,8651
0,1161	A	0,3482	0,2321	-0,6330
0,3952	A	0,4426	0,0474	-0,5856
0,0565	A	0,0752	0,0186	-0,5669
0,2861	B	0,0000	-0,2861	-0,8531
0,1690	A	0,3717	0,2027	-0,6503
0,2765	A	2,4032	2,1267	1,4763
0,3572	B	0,0000	-0,3572	1,1192
0,4893	B	0,0000	-0,4893	0,6299
0,2024	A	0,4614	0,2590	0,8889
0,1954	B	0,0000	-0,1954	0,6935
0,0309	B	0,0000	-0,0309	0,6626
0,2503	B	0,0000	-0,2503	0,4122
0,3416	B	0,0000	-0,3416	0,0707
0,3375	B	0,0000	-0,3375	-0,2669
0,0875	B	0,0000	-0,0875	-0,3544
0,1081	A	0,4107	0,3026	-0,0517
0,2285	A	0,5072	0,2787	0,2270
0,3044	A	5,1745	4,8701	5,0971
0,2332	A	0,5223	0,2891	5,3862
0,6109	B	0,0000	-0,6109	4,7753
0,3395	A	0,5330	0,1935	4,9688
0,3796	B	0,0000	-0,3796	4,5892

0,0050	B	0,0000	-0,0050	4,5842
0,2212	B	0,0000	-0,2212	4,3630
0,1468	B	0,0000	-0,1468	4,2162
0,1948	B	0,0000	-0,1948	4,0214
0,8356	A	1,4205	0,5849	4,6064
0,4957	A	1,7450	1,2493	5,8556
0,0552	A	0,1607	0,1055	5,9611
0,3047	A	0,5149	0,2102	6,1713
0,3342	A	0,8522	0,5180	6,6894
0,1403	B	0,0000	-0,1403	6,5491
0,2370	B	0,0000	-0,2370	6,3121
0,2343	A	0,4147	0,1804	6,4925
0,3024	B	0,0000	-0,3024	6,1901
0,5008	B	0,0000	-0,5008	5,6894
		20,2410	5,6894	

*Tabla 10. Apuestas realizadas a lo largo del torneo (Masters Montreal)*

En la gráfica que se muestra a continuación podemos observar la ganancia en euros obtenida partido a partido.



*Figura 14. Ganancia obtenida partido a partido (Masters Montreal)*

Los partidos centrales corresponden a la segunda y tercera ronda del torneo, siendo estas en las que se consigue una mayor ganancia, especialmente en la segunda, debido a que no siempre los jugadores favoritos vencen, y son esos partidos los que aprovechamos para aumentar nuestras ganancias.

Como vemos en la gráfica, los picos producidos son la base del éxito de nuestro modelo de apuestas, en este torneo hay un pico especialmente alto que nos produce un beneficio de más de 5 euros. Obtenemos ganancias en pocos partidos, pero éstas son muy grandes, reportándonos al final un beneficio.

La gráfica que se muestra a continuación muestra el beneficio acumulado durante todos los partidos a lo largo del torneo.

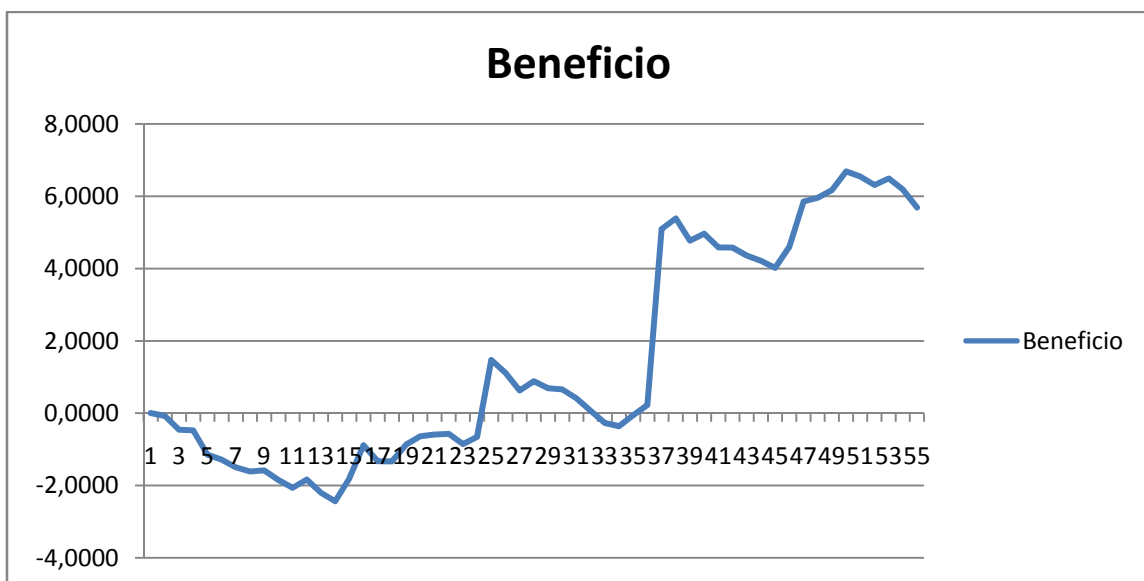


Figura 15. Beneficio acumulado a lo largo del torneo (Masters Montreal)

Como podemos observar, el beneficio es creciente especialmente durante la segunda y tercera ronda, obteniendo su máximo prácticamente hasta llegar a cuartos de final. En este torneo se produce la situación especial de que tenemos pérdidas en la primera ronda, pero se produce un

cambio radical en la segunda y en la tercera, en las cuales se obtienen beneficios muy elevados. En el caso de que solamente hubiéramos apostado durante las tres primeras rondas, habríamos obtenido un beneficio de unos 7 euros. El beneficio apostando durante todo el torneo es de 5,6894 euros.

En esta última tabla vamos a dividir la información según los porcentajes, trabajando en rangos del 10% (columna 1). En la columna 2 se muestran los partidos ganados por los jugadores que tenían un porcentaje calculado por nuestro modelo dentro de su rango correspondiente. En la columna 3 se aplica el mismo criterio pero para los partidos perdidos. En la columna 4 calculamos la proporción entre partidos ganados y partidos totales. En las columnas 5, 6, y 7 se establecen los mismos criterios que en las columnas 2, 3 y 4 respectivamente, pero teniendo en cuenta los datos de las casas de apuestas.

Porcentaje	Ganados Modelo	Perdidos Modelo	Ganados/Totales Modelo	Ganados Casa Apuestas	Perdidos Casa Apuestas	Ganados/Totales Casa Apuestas
0 – 10 %	0	3	0/4 = 0%	1	3	1/4 = 25%
10 – 20 %	0	0	0/0	1	9	1/10 = 10%
20 – 30 %	2	9	2/11 = 18%	3	10	3/13 = 23%
30 – 40 %	5	9	5/14 = 36%	4	11	4/15 = 27%
40 – 50 %	12	15	12/27 = 44%	6	7	6/13 = 46%
50 – 60 %	15	12	15/27 = 56%	7	6	7/13 = 54%
60 – 70 %	9	5	9/14 = 64%	11	4	11/15 = 73%
70 – 80 %	9	2	9/11 = 82%	10	3	10/13 = 77%
80 – 90 %	0	0	0/0	9	1	9/10 = 90%
90 – 100 %	3	0	3/3 = 100%	3	1	3/4 = 75%
	55	55		55	55	

*Tabla 11. Resultados obtenidos (Masters Montreal)*

Si nos fijamos detenidamente en los rangos medios de porcentajes, es decir, del 30% al 60%, el valor numérico de la columna de **Ganados/Totales** está incluido en el rango de la de **Porcentaje**. Existen varias franjas que no tienen los suficientes datos para ser representativos, por ejemplo entre el 0% y el 20%, o entre el 80% y el 100%.

## 5. CONCLUSIONES E IDEAS FUTURAS

### 5.1. CONCLUSIONES

Con la realización de este proyecto de fin de carrera lo que se ha pretendido es aplicar técnicas de minería de datos y simulación para la predicción de resultados en eventos de tenis, de tal manera que estos puedan ser útiles para obtener un beneficio a largo plazo apostando en casas de apuestas y con un manejo de la banca apropiado.

Poder saber cuál va a ser el ganador de un evento deportivo es una tarea compleja, ya que intervienen múltiples factores que hemos analizado con anterioridad y que no pueden ser medidos mediante un análisis lógico, estadístico o determinista.

El proyecto se centró en la predicción de los resultados de una modalidad de un evento deportivo. Más concretamente lo que se intentó fue predecir el ganador de cada uno de los partidos de tenis de la categoría individual masculina.

Este tipo de partidos eran disputados únicamente por dos jugadores, de los cuales disponíamos de una gran cantidad de información acerca de sus estadísticas de juego a lo largo del año, en diversos años, en diferentes periodos de tiempo y también datos relacionados con la pista, el torneo, la ronda, el tipo de superficie, enfrentamientos individuales, cuotas de mercado...

Pese a disponer de una cantidad elevada de información, existían múltiples factores que no podíamos controlar, debido a que se trata de un deporte y no es algo que podamos calcular de una manera determinista ni lógica. Un ejemplo, por citar alguno, podría ser el estado de ánimo de cada jugador, su situación personal, su estado físico, su mentalidad, su determinación...

Comentar que es imposible tener una precisión meridiana a la hora de determinar el porcentaje de victoria de un jugador debido a que hoy en día todo está muy igualado y además intervienen factores externos que hacen que un resultado pueda ser prácticamente impredecible. Lo que se pretendía desde un principio con este proyecto conseguir un beneficio a largo plazo mediante apuestas de tenis. Para ello, era necesario lograr un buen manejo de la banca a través de nuestras predicciones de los resultados de los partidos mediante el uso del algoritmo de Kelly

Desde el inicio del proyecto se sabía de la dificultad que conllevaría conseguir un modelo que obtuviese porcentajes de aciertos muy elevados. Sin embargo, aunque no se lograran unos porcentajes excesivamente altos, sí permitió realizar inversiones en el mercado de las apuestas con rentabilidad positiva en casi todos los torneos.

La tarea más difícil y extenuante a lo largo del proyecto fue la preparación de la información para la creación de las dos bases de datos. Por un lado, los datos fueron obtenidos de varias fuentes distintas con lo que tenían distintos formatos y distintas formas de representar los datos. Por tanto, ésta fue la tarea más complicada del proyecto ya que hubo que dedicar mucho tiempo para poder construir unas bases de datos válidas y útiles.

Observando los resultados de nuestro modelo se consiguió demostrar que es posible obtener una rentabilidad introduciéndonos en el mundo de las apuestas. La principal clave de este buen resultado es conseguir un porcentaje más o menos acertado de victoria, para después compararlo no solo con las cuotas de una casa de apuestas sino con las de las más representativas, y apostar en la que más beneficio nos pueda aportar. Esto nos proporciona un margen extenso para conseguir ganancias sobre todo a largo plazo.



La clave de nuestro modelo se centra en determinados partidos, cuyo resultado es inesperado y nos proporciona una ganancia enorme. Estos picos harán que se compensen las pérdidas de otros partidos en los que no se acierte el resultado para concluir el torneo con un balance final positivo.

Finalmente, resaltar que se consiguieron los objetivos marcados y explicados al inicio del proyecto. Se creó una base de datos minuciosa con la elección de una serie de variables claves para darnos la mayor cantidad de información acerca del partido sin tener que trabajar con un volumen de datos demasiado elevado, otra base de datos con los resultados y las cuotas de cada una de las casas de apuestas para cada uno de los partidos de los torneos analizados. Después se creó un modelo de clasificación y predicción para determinar quién tiene más probabilidades de ser el ganador de un partido. Además, utilizando las diversas cuotas de las casas de apuestas que hemos comentado y utilizando el control y gestión de nuestra banca mediante el algoritmo de Kelly, se consiguió mejorar el beneficio final a lo largo de todos los torneos.

## 5.2. IDEAS FUTURAS

A partir del propósito fundamental utilizado en este proyecto podríamos plantear otras opciones que ampliaran la idea original o bien expandieran nuestros objetivos hacia otros horizontes.

A continuación se exponen algunas ideas para la diversificación y mejora de este proyecto fin de carrera:

- La primera idea, que surge a colación de este proyecto y que podría ser una muy posible continuación, sería ampliar el mismo estudio que realizamos aquí a otras modalidades del tenis, como el tenis femenino o la modalidad de dobles, es decir, abarcar por completo este deporte y no centrarnos solo en una categoría.
- Realizar una generalización de pistas y superficies, de tal forma que nuestro modelo y bases de datos no se fijen únicamente en pistas rápidas sino en todo el circuito desde principio a fin de año.
- Continuar con el estudio realizado en el ámbito de otros deportes que puedan ser más beneficiosos o que tengan un margen mayor para obtener ganancias en nuestras apuestas.
- Centrándonos de nuevo en el tenis como deporte, podríamos realizar unas bases de datos mucho más extensas y dinámicas, incluyendo mayor número de datos y cuotas de más casas de apuestas.
- Expandir el modelo realizado de tal forma que podamos analizar mayor número de variables y factores que puedan alterar el resultado final del partido.
- Realizar un estudio para predecir de forma interactiva, es decir, poder detectar durante el transcurso del partido en qué momento se producen

los instantes críticos y apropiados de las cuotas para poder apostar. De esta forma podríamos apostar, en tiempo real, en el momento que fuera más rentable por un jugador u otro.

- Otros campos que podríamos tener en cuenta dentro de este mismo mundo del tenis es ampliar el estudio realizado de forma que no se intentase únicamente predecir el ganador del partido sino otros tipos de apuestas como podrían ser el resultado completo por sets, el ganador de un set, número de sets, juego a juego, ganador del tie break, número de juegos, tiempo de partido... con lo cual expandiríamos el espectro de apuestas disponibles en cada partido.

- Por último, en este proyecto realizamos una elección de cuotas entre las casas de apuestas, de tal forma que elegimos los valores más favorables escogiendo entre las diversas cuotas. Sin embargo podríamos avanzar un paso más en este campo, realizando una elección interactiva, eligiendo durante el transcurso del partido los picos más favorables en un conjunto de casas de apuestas dado. Así lograríamos una rentabilidad máxima, tanto en el desarrollo del partido como en la elección de la casa de apuestas.

## 6. PRESUPUESTO

En este capítulo vamos a explicar detalladamente el presupuesto del proyecto fin de carrera. Se van a especificar los costes tanto a nivel de personal como de material, siendo los primeros mucho más cuantiosos que los segundos.

En primer lugar vamos a detallar las fases en las que se ha trabajado para la realización del proyecto:

- Fase 1: Documentación.
- Fase 2: Creación y diseño de las bases de datos.
- Fase 3: Implementación del modelo.
- Fase 4: Simulaciones y resultados.
- Fase 5: Redacción de la memoria del proyecto.

Para la finalización de este proyecto han sido necesarios 18 meses, en los cuales no se ha podido trabajar de manera uniforme, sin embargo vamos a suponer un esfuerzo continuo para un total de 800 horas.

A continuación se realiza una especificación del presupuesto de personal en el que se detallan las horas por cada una de las fases de elaboración del proyecto, los cargos, costes por hora y total:

Fase	Cargo	Nº Horas	Coste Hora	Total
Fase 1	Documentación	150 horas	15 €	2.250 €
Fase 2	Diseño	150 horas	25 €	3.750 €
Fase 3	Implementación	200 horas	25 €	5.000 €
Fase 4	Pruebas	200 horas	20 €	4.000 €
Fase 5	Redacción	100 horas	15 €	1.500 €
<b>Total</b>		<b>800 horas</b>		<b>16.500 €</b>

*Tabla 12. Fases del proyecto*

Como vemos en la tabla, el coste total de personal se eleva a 16500€.

Ahora especificamos los costes de material, entre los que se incluyen el uso del PC, la licencia para Matlab y otros (impresión, desplazamientos, llamadas...) con sus respectivos costes.

Descripción	Coste Total
PC Samsung	350 €
Licencia Matlab	300 €
Otros	200 €
<b>Total</b>	<b>850 €</b>

*Tabla 13. Costes de material*

Como apreciamos en la tabla, el coste total de material es de 850€.

Por último, en esta tabla realizamos una recopilación de todos los gastos para dar con el resultado final del presupuesto incluyendo los costes indirectos del 20% y el IVA del 21%.

Concepto	Importe Total
Costes de Personal	16.500 €
Costes de Material	850 €
Costes Indirectos (20%)	3.470 €
Total sin IVA (21%)	20.820 €
IVA (21%)	4.372 €
<b>Total con IVA (21%)</b>	<b>25.192 €</b>

*Tabla 14. Gastos detallados del proyecto*



**UNIVERSIDAD CARLOS III DE MADRID**  
**Escuela Politécnica Superior**

**PRESUPUESTO DE PROYECTO**

**1.- Autor:**

Miguel Ángel Martín Mazuelo

**2.- Departamento:**

Teoría de la Señal y Comunicaciones

**3.- Descripción del Proyecto:**

- Título: Predicción de resultados de eventos deportivos: tenis  
- Duración (meses): 24  
Tasa de costes indirectos: 20%

**4.- Presupuesto total del Proyecto (valores en Euros):**

25.192,00 Euros

**5.- Desglose presupuestario (costes directos)**

**PERSONAL**

Apellidos y nombre	N.I.F. (no rellenar - solo a título informativo)	Categoría	Dedicación (meses) <sup>a)</sup>	(hombres)	Coste hombre mes	Coste (Euro)	Firma de conformidad
Martín Mazuelo, Miguel Ángel		Documentación	1,14		1.973,68	2.250,00	
Martín Mazuelo, Miguel Ángel		Diseñador	1,14		3.289,47	3.750,00	
Martín Mazuelo, Miguel Ángel		Programador	1,52		3.289,47	5.000,00	
Martín Mazuelo, Miguel Ángel		Técnico de Prueba	1,52		2.631,58	4.000,00	
Martín Mazuelo, Miguel Ángel		Redactor	0,76		1.973,68	1.500,00	
<b>Hombres mes 6,08</b>					<b>Total</b>	<b>16.500,00</b>	

<sup>a)</sup> 1 Hombre mes = 131,25 horas. Máximo anual de dedicación de 12 hombres mes (1575 horas)  
Máximo anual para PDI de la Universidad Carlos III de Madrid de 8,8 hombres mes (1.155 horas)

**EQUIPOS**

Descripción	Coste (Euro)	% Uso dedicado proyecto	Dedicación (meses)	Periodo de depreciación	Coste imputable <sup>d)</sup>
PC Samsung	875,00	100	24	60	350,00
<b>Total</b>					<b>350,00</b>

<sup>d)</sup> Fórmula de cálculo de la Amortización:

$$\frac{A}{B} \times C \times D$$

A = nº de meses desde la fecha de facturación en que el equipo es utilizado  
B = periodo de depreciación (60 meses)  
C = coste del equipo (sin IVA)  
D = % del uso que se dedica al proyecto (habitualmente 100%)

**SUBCONTRATACIÓN DE TAREAS**

Descripción	Empresa	Coste imputable
<b>Total</b>		<b>0,00</b>

**OTROS COSTES DIRECTOS DEL PROYECTO<sup>e)</sup>**

Descripción	Empresa	Costes imputable
Software		300,00
Material fungible		200,00
<b>Total</b>		<b>500,00</b>

<sup>e)</sup> Este capítulo de gastos incluye todos los gastos no contemplados en los conceptos anteriores, por ejemplo: fungible, viajes y dietas.

**6.- Resumen de costes**

Presupuesto Costes Totales	Presupuesto Costes Totales
Personal	16.500
Amortización	350
Subcontratación de tareas	0
Costes de funcionamiento	500
Costes Indirectos	3.470
<b>Total</b>	<b>20.820</b>

Tabla 15. Presupuesto del proyecto

El presupuesto total del proyecto asciende a 25.192 €.

El ingeniero proyectista,

Fdo. Miguel Ángel Martín Mazuelo



## 7. REFERENCIAS

### 7.1. PÁGINAS DE INTERNET

- [1] Atp. Página Web. <http://es.atpworldtour.com>
  
- [2] Betfair. Base de Datos. <http://data.betfair.com>
  
- [3] Betfair. Página Web. <http://sports.betfair.es>
  
- [4] Mi Apuesta. Página Web. <http://www.miapuesta.es>
  
- [5] William Hill. Página Web. <http://www.williamhill.es>
  
- [6] Bwin. Página Web. <https://www.bwin.es>
  
- [7] LBApuestas. Página Web. <http://www.lbapuestas.es>
  
- [8] PAF. Página Web. <https://www.paf.es>
  
- [9] Betclíc. Página Web. <https://es.betclíc.com>
  
- [10] Interwetten. Página Web. <https://www.interwetten.es>
  
- [11] Bet365. Página Web. <http://www.bet365.es>.
  
- [12] Base de datos principal. Página Web. <http://tennis.matchstat.com>.
  
- [13] Estadísticas partidos y torneos. Página Web. <http://www.tennis-data.co.uk>.
  
- [14] El Comparador. Página Web. <http://www.elcomparador.com>

[15] Sport Trader. Página Web. <http://www.sportytrader.es>

[16] OnCourt. <http://www.oncourt.info>

[17] Mira Mi Apuesta. Página Web. <http://miramiapuesta.com>

## 7.2. LIBROS Y ARTÍCULOS

- [18] T. Barnett and S. R. Clarke. *Combining player statistics to predict outcomes of tennis matches*. IMA Journal of Management Mathematics, 16:113–120, 2005.
- [19] T. Barnett, A. Brown and S. Clarke. *Developing a model that reflects outcomes of tennis matches*. Australasian Conference on Mathematics and Computers in Sport, 178-188, 2006.
- [20] S.R. Clarke, D. Dyte. *Using official ratings to simulate major tennis tournaments*, International Transactions in Operational Research 7 585-594. 2000.
- [21] J.L. Kelly, *A New Interpretation of Information Rate*. Bell System Technical Journal. 35 : 917-926
- [22] L. MacLean, E. O. Thorp, Y. Zhao, and W. T. Ziemba. *Medium term simulations of Kelly and fractional Kelly strategies*. 2009.
- [23] P. K. Newton and J. B. Keller. *Probability of winning at tennis i. theory and data*. Studies in applied Mathematics, 114:241–269, 2005.
- [24] E. Thorp. *The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market*. Paper presented at: The 10th International Conference on Gambling and Risk Taking. 1997.
- [25] I. Witten. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 2009.