



---

**Universidad de Valladolid**

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

---

**Análisis y Aplicación de Estrategias de  
Arbitraje en el Mercado de Apuestas  
Deportivas**

---

Autor:

Pablo de Arriba Mendizábal

Tutores:

Agustín Mayo Íscar

Francisco Hernando Gallego

Curso 2024-2025



# Agradecimientos

A mis tutores que han dirigido este trabajo, en especial a Francisco, que cada semana ha estado pendiente de ver la evolución y dar las indicaciones necesarias para que esto haya sido posible.

A mis compañeros del doble grado, que me han acompañado durante estos cinco años y han servido de apoyo en momentos complicados, estando siempre dispuestos a ayudar.

A los profesores del Grado en Estadística, cuyos conocimientos han sido la base para realizar el presente trabajo y finalizar mis estudios.

Por último, a mi familia, en especial a mis padres y a mi hermano, que son los que están y estarán siempre, y los que aguantan mis momentos de agobio y mal humor y, a pesar de ello, no dudan en ayudarme en todo lo que necesito.



# Resumen

El presente Trabajo de Fin de Grado tiene como objetivo analizar y evaluar distintas estrategias de inversión en el mercado de las apuestas deportivas, con especial atención a las denominadas *value bets*, oportunidades en las que la probabilidad real de un resultado es mayor que la implícita en la cuota ofrecida por la casa de apuestas. A partir de un conjunto de datos reales recopilados de las casas Bet365 y Winamax mediante la herramienta Betburger, se lleva a cabo un procesamiento que incluye limpieza de datos, cálculo de cuotas implícitas y clasificación de apuestas según su resultado.

Más adelante, se aplican distintas estrategias de gestión del capital: *stake plano*, *stake progresivo*, *stake proporcional con bank fijo* y *stake proporcional con bank dinámico*. Estas se evalúan tanto sobre el conjunto de datos como mediante simulaciones bootstrap, con métricas como el beneficio final, el riesgo asumido o la relación entre ambos.

Los resultados muestran que, si bien estrategias conservadoras como el stake plano ofrecen una mayor estabilidad, enfoques más arriesgados como el stake proporcional con bank dinámico presentan un mayor potencial de beneficio, aunque con una volatilidad muy alta. Finalmente, se hace una reflexión sobre la importancia de una adecuada gestión del riesgo y la necesidad de disponer de herramientas de software adecuadas para poder obtener beneficios reales, aunque las limitaciones que imponen las casas de apuestas hacen que no sea algo seguro.



# Abstract

This Final Degree Project aims to analyze and evaluate different investment strategies in the sports betting market, with a special focus on value bets, opportunities where the actual probability of an outcome is higher than the one implied by the odds offered by a bookmaker. Using a real dataset collected from Bet365 and Winamax through the Betburger tool, a data processing phase is carried out, including data cleaning, computation of implied odds, and classification of bets according to their outcome.

Then, several capital management strategies are applied: flat stake, progressive stake, proportional stake with fixed bankroll, and proportional stake with dynamic bankroll. These are evaluated both on the dataset and through bootstrap simulations, using metrics such as final profit, risk taken, and the trade-off between both.

The results show that while conservative strategies like flat stake offer greater stability, riskier approaches such as proportional stake with dynamic bankroll present higher profit potential, but with much greater volatility. Finally, the study highlights the importance of proper risk management and the need for adequate software tools to achieve real profits, although the limitations imposed by bookmakers mean that such profits are far from guaranteed.



# Índice general

<b>1. Introducción</b>	<b>15</b>
1.1. Contextualización . . . . .	15
1.2. Motivación . . . . .	16
1.3. Objetivos . . . . .	17
<b>2. Estado del arte</b>	<b>19</b>
2.1. Situación del mercado . . . . .	19
2.2. Tecnología y tratamiento de datos . . . . .	20
2.3. Estrategias de gestión del patrimonio . . . . .	22
<b>3. Descripción del problema</b>	<b>25</b>
3.1. Identificación de Value bets . . . . .	25
3.2. Dataset . . . . .	26
3.2.1. Análisis descriptivo . . . . .	26
3.2.2. Procesamiento . . . . .	28
3.2.3. Cálculo de la cuota de Bet365 . . . . .	29
3.2.4. Eliminación de outliers . . . . .	30
3.2.5. Dataset definitivo . . . . .	33
3.3. Clasificación de las apuestas . . . . .	34

<b>4. Metodología</b>	<b>37</b>
4.1. Estrategias . . . . .	38
4.1.1. Stake plano . . . . .	39
4.1.2. Stake progresivo . . . . .	39
4.1.3. Stake proporcional . . . . .	40
4.2. Implementación . . . . .	40
4.2.1. Descripción del programa . . . . .	40
<b>5. Resultados</b>	<b>43</b>
5.1. Stake plano . . . . .	43
5.2. Stake progresivo . . . . .	46
5.3. Stake proporcional . . . . .	48
5.3.1. Bank fijo . . . . .	48
5.3.2. Bank dinámico . . . . .	50
5.4. Comparación de resultados . . . . .	52
<b>6. Conclusiones y trabajo futuro</b>	<b>57</b>
6.1. Conclusiones . . . . .	57
6.2. Limitaciones del estudio . . . . .	57
6.3. Líneas futuras . . . . .	58
<b>Bibliografía</b>	<b>61</b>

# Lista de Figuras

3.1. Presencia de cada deporte en el dataset. . . . .	28
3.2. Distribución del stake en el dataset. . . . .	31
3.3. Boxplot del stake con outliers. . . . .	32
3.4. Boxplot del stake tras eliminar outliers. . . . .	32
3.5. Convergencia del porcentaje de aciertos . . . . .	35
4.1. Diagrama de flujo del cálculo del rendimiento . . . . .	42
5.1. Beneficio y probabilidad de ganancias con stake plano . . . . .	44
5.2. Evolución del bank con stake plano . . . . .	45
5.3. Evolución del bank con stake plano (simulación) . . . . .	45
5.4. Beneficio y probabilidad de ganancias con stake progresivo . . . . .	46
5.5. Evolución del bank con stake progresivo . . . . .	47
5.6. Evolución del bank con stake progresivo (simulación) . . . . .	47
5.7. Evolución del bank con stake proporcional y bank fijo . . . . .	48
5.8. Evolución del bank con stake proporcional y bank fijo (simulación) . . . . .	49
5.9. Evolución del bank con stake proporcional y bank dinámico . . . . .	50
5.10. Evolución del bank con stake proporcional y bank dinámico (simulación) . . . . .	51
5.11. Comparativa de la evolución del bank por estrategia. . . . .	53

5.12. Comparativa de la evolución del bank para cada estrategia (simulaciones). 53

5.13. Evolución del bank para cada estrategia. . . . . 54

# Lista de Tablas

3.1. Cuota y Stake en dataset original . . . . .	27
3.2. Deporte y Tipo de Apuesta en dataset original . . . . .	27
3.3. Distribución de la nueva agrupación de apuestas . . . . .	29
3.4. Comparación cuotas Winamax y Bet365 . . . . .	30
3.5. Comparación de métodos de eliminación de outliers . . . . .	32
3.6. Cuota y Stake tras procesamiento . . . . .	33
3.7. Deporte y Tipo de Apuesta tras procesamiento . . . . .	33
3.8. Cuota Winamax y Bet365 tras procesamiento . . . . .	33
3.9. Tabla de frecuencias de fallos y aciertos . . . . .	35
5.1. Resultados de stake plano en el dataset . . . . .	43
5.2. Comparación simulación y dataset con stake plano . . . . .	45
5.3. Comparación simulación y dataset con stake progresivo . . . . .	46
5.4. Comparación simulación y dataset con stake proporcional y bank fijo .	49
5.5. Comparación simulación y dataset con stake proporcional y bank dinámico	51
5.6. Relación beneficio - riesgo de cada estrategia (datos de las simulaciones)	52
5.7. Resumen del bank final tras 10.000 apuestas por estrategia. . . . .	54
5.8. Resultados de las estrategias en el dataset . . . . .	55

5.9. Resultados de las estrategias en las simulaciones . . . . . 56

# Capítulo 1

## Introducción

### 1.1. Contextualización

Las apuestas deportivas constituyen un mercado en expansión que mueve miles de millones de dólares cada año. Según estimaciones recientes, el mercado global de apuestas deportivas alcanzó un valor aproximado de 160 mil millones de dólares (USD) en 2024, con una tasa de crecimiento anual del 9,2% [6], lo que refleja su enorme atractivo y su rápido crecimiento en los últimos años.

El estallido de este mercado se produce a principios del siglo XXI, coincidiendo con la adopción masiva de internet por parte de los usuarios y la aparición de las primeras casas de apuestas online, que se establecían principalmente en Reino Unido. Como consecuencia, la competencia entre casas de apuestas se ha multiplicado en los últimos años, en gran medida por la aparición de nuevas casas de apuestas desde Asia, obligando a las compañías tradicionales a adaptarse a un entorno en el que los márgenes de beneficio han disminuido. (Grant *et al.* 2018 [2]).

Este trabajo busca centrarse en un aspecto concreto de las apuestas deportivas: el arbitraje. El arbitraje es un concepto que proviene de los mercados financieros, y se basa en sacar provecho de una discrepancia de precios de un mismo activo entre distintos mercados de forma que se obtiene un beneficio económico sin riesgo.

En las apuestas deportivas, el concepto va en la misma línea. El arbitraje, también conocido como *sure bets*, consiste en identificar discrepancias en las cuotas de distintas casas de apuestas con el objetivo de obtener un beneficio asegurado apostando a todos los resultados posibles de un evento (Vlastakis *et al.* 2009 [20]). De esta forma, el beneficio es independiente del resultado del evento. Esta posibilidad surge debido a la diferencia de opiniones, las diferencias en los algoritmos de cálculo de probabilidades

o distintas estrategias comerciales adoptadas por cada casa de apuestas.

Si bien el arbitraje en las apuestas deportivas es una estrategia que puede generar beneficios seguros, no está exenta de riesgos. En primer lugar, las casas de apuestas suelen limitar o incluso cerrar las cuentas de los jugadores que realizan arbitraje de forma sistemática, lo que puede dificultar la aplicación de esta estrategia a largo plazo. Además, para la detección de oportunidades de arbitraje son necesarias herramientas que permitan comparar cuotas en múltiples casas de apuestas en tiempo real, lo que puede suponer un coste adicional en términos tecnológicos y económicos.

Más allá del arbitraje, existen otras estrategias que se alejan de la idea de “beneficio seguro”. Estas se basan en la identificación de valor en las cuotas a través del análisis de probabilidades. Concretamente, este trabajo busca centrarse en las apuestas cuya probabilidad real es superior a la probabilidad implícita en la cuota establecida por la casa de apuestas. Estas apuestas se conocen como *value bets* o apuestas de valor. A diferencia del arbitraje, que garantiza un beneficio inmediato, las *value bets* buscan un retorno esperado positivo a largo plazo.

El estudio y aplicación de estrategias como el arbitraje y las *value bets* han captado el interés de apostadores profesionales, analistas de datos y economistas, ya que combinan aspectos matemáticos, estadísticos y tecnológicos con el fin de realizar predicciones precisas en un contexto de gran incertidumbre.

Este trabajo se estructura de la siguiente forma: en el apartado 2 se van a repasar los avances que se han realizado en los últimos años en este campo y todo lo relacionado con el estado del arte. En el punto 3, entramos en materia detallando el dataset del que se dispone, qué modificaciones necesita y cómo se va a emplear para el estudio. El punto 4 explica la metodología del trabajo: qué estrategias se aplican, en qué consisten y cómo se llevan a cabo. También se explica en profundidad el código desarrollado para ello. Finalmente, en el apartado 5 se muestran los resultados que arrojan cada una de las estrategias, evaluando y comparando su desempeño, y mostrando sus puntos fuertes y débiles. Al final se detallan las conclusiones que se extraen del estudio y se exponen las limitaciones del mismo. Además, se proponen futuras líneas de investigación y mejora.

## 1.2. Motivación

Realizar un Trabajo de Fin de Grado sobre arbitraje en apuestas deportivas en el ámbito de la estadística tiene muchas motivaciones. En primer lugar, la presencia de conceptos probabilísticos aplicados en un contexto real es constante, ya que es la forma de identificar *value bets*. Además, el componente económico lo convierte en un tema con gran atractivo para la mayoría de las personas.

Otro aspecto a destacar es la necesidad de utilizar datos reales y técnicas de aná-

lisis de datos con el fin de comprobar la viabilidad de las distintas estrategias. El análisis estadístico es fundamental para extraer conclusiones de los resultados de cada estrategia y permite tomar decisiones respecto a la utilización de cada una de ellas.

Por último, se repasan los últimos avances tecnológicos y los métodos más avanzados que se están utilizando para conseguir mejores predicciones, que van desde la aplicación de modelos de regresión lineal o logística, hasta técnicas de *deep learning*.

### 1.3. Objetivos

El objetivo principal de este trabajo es estudiar el rendimiento que se puede extraer de las *value bets* analizando diferentes enfoques, desde los más conservadores que minimizan el riesgo sacrificando parte del beneficio, hasta otros más arriesgados que buscan un mayor rendimiento. Se aplicarán diversas estrategias y se evaluarán y compararán desde un punto de vista estadístico, pudiendo sacar conclusiones acerca de los resultados obtenidos.

Este objetivo principal engloba otros objetivos más específicos:

- Estudiar el mercado de las apuestas deportivas, su funcionamiento, evolución y perspectivas futuras.
- Comprender qué son las *value bets* y por qué pueden reportar beneficios a largo plazo.
- Analizar datos reales de *value bets* y obtener resultados a partir de ellos.
- Interpretar y analizar los resultados obtenidos y estudiar la viabilidad de las diferentes estrategias extrayendo conclusiones significativas.
- Proponer mejoras y posibles líneas de investigación futuras que completen este estudio.



## Capítulo 2

# Estado del arte

Las estrategias aplicadas en el arbitraje de apuestas deportivas han evolucionado con el tiempo, especialmente en los últimos años, y así se muestra en la literatura. Durante años se han propuesto diferentes enfoques para sacar provecho a este mercado. La mayoría de ellos van en la línea de mejorar los modelos predictivos y hacer un uso adecuado de los datos, que cada vez son más y requieren de mayor cuidado, aunque cabe destacar la importancia de una buena gestión del patrimonio como premisa principal para sacar rendimiento a este mercado.

### 2.1. Situación del mercado

En primer lugar, Constantinou y Fenton (2013) [5] analizaron datos de 14 ligas europeas de fútbol durante 7 temporadas. En su estudio muestran cómo las cuotas que establecen las casas de apuestas no siempre se ajustan a las probabilidades reales del evento, concluyendo que el mercado de las apuestas deportivas es un mercado que presenta ineficiencias que pueden ser explotadas mediante las técnicas adecuadas. Este hecho ha motivado el desarrollo de herramientas de software que facilitan la identificación de oportunidades de arbitraje, por lo que, si continúa esta tendencia, las casas de apuestas deben adaptarse para evitar riesgos sustanciales.

En la misma línea, diversos estudios han profundizado en las consecuencias del crecimiento exponencial del mercado de las apuestas deportivas. Gómez-González y del Corral (2018) [3] destacan cómo la disminución de los márgenes de beneficio de las casas ha abierto la puerta a nuevas oportunidades para los usuarios, particularmente con estrategias como el arbitraje y las apuestas de valor (*value bets*). Además, se prevé que esta tendencia al alza del número de casas de apuestas continúe en los próximos años.

## 2.2. Tecnología y tratamiento de datos

Franck *et al.* [8] [7] dan otro enfoque a su estudio, siendo los primeros en introducir el concepto de arbitraje entre mercados (*Inter-market Arbitrage*). Estos analizan una muestra de 12.782 partidos de fútbol, encontrando oportunidades de arbitraje en un 0,8% de los partidos cuando se estudian únicamente casas de apuestas tradicionales, conocidas como *bookmakers*. Sin embargo, ampliando el foco a los *betting exchanges*, el porcentaje de partidos con oportunidades de arbitraje aumenta hasta el 19,2%. Los *betting exchanges* son aquellas casas de apuestas donde los usuarios apuestan contra otros usuarios, en lugar de contra la propia casa.

En el dominio de las *value bets*, el enfoque en la literatura varía respecto al arbitraje. En el arbitraje, las oportunidades suelen ser escasas y más difíciles de encontrar, pues se requiere de al menos 2 casas de apuestas para aprovecharse de las ineficiencias. Sin embargo, con las *value bets* es posible sacar rendimiento con una única casa de apuestas, lo cual facilita el proceso, aunque no reporta beneficios asegurados. Es por esto que en los últimos años la literatura se ha centrado en gran medida en buscar métodos que mejoren las predicciones de las casas para encontrar valor en las apuestas.

En una época donde los datos están a la orden del día, la tendencia es utilizar grandes cantidades de datos históricos para ajustar modelos predictivos. Recientemente, se han buscado nuevas métricas que puedan servir de utilidad para predecir resultados de eventos deportivos. Ramirez *et al.* (2023) [22] propusieron la “sabiduría de masas” (*wisdom of crowds*) como factor para mejorar la capacidad predictiva de algunos modelos de apuestas. Este concepto se basa en la idea de que la inteligencia colectiva de un grupo de personas puede ser superior a la de un solo experto en la toma de decisiones [11].

En su estudio, muestran cómo factores como las visitas a la página de Wikipedia de un jugador de tenis (*WikiBuzz*) o la diferencia de ranking entre los dos jugadores que se enfrentan (*RankDist*), pueden ser determinantes a la hora de encontrar cuotas sobreestimadas en las casas de apuestas. Se aplicó este modelo a más de 5000 apuestas de tenis entre 2019 y 2020 exclusivamente en Bet365, llegando a obtener retornos de la inversión (ROIs) de hasta el 29,38%, superando con creces los resultados obtenidos en estudios previos, donde se han obtenido ROIs del 3,56% (Angelini *et al.*, 2023 [9]) hasta el 4,35% (Sipko, 2015 [18]).

Más adelante, Clegg y Cartlidge (2024) [15] replicaron este mismo estudio, obteniendo importantes conclusiones en cuanto a la importancia de la calidad de los datos. Por un lado, confirmaron la existencia de ineficiencias en las casas de apuestas que ya habían mostrado Ramirez *et al.* [22] anteriormente; sin embargo, descubrieron una inconsistencia en el dataset que había sido objeto de estudio, donde una cuota errónea había inflado los resultados artificialmente. Una vez corregido este error, el ROI disminuyó drásticamente, quedando en un 12,44% siempre y cuando se consideraran únicamente los partidos más competitivos cuyas probabilidades estimadas estuvieran

entre el 40% y el 60%. Para partidos con probabilidades entre el 20% y el 80% se obtuvieron ROIs negativos tras la corrección del error. Esto muestra cómo las casas de apuestas cometen más errores en sus cuotas en eventos igualados frente a otros eventos más desequilibrados, lo que puede servir como base para diseñar futuras estrategias en el marco de las apuestas de valor.

En el estudio se evaluaron dos estrategias distintas:

- Best odds: Utiliza las mejores cuotas disponibles
- Bet365: Utiliza únicamente las cuotas de Bet365

Para comprobar que los resultados no eran fruto de la aleatoriedad, se realizaron 100.000 simulaciones aleatorias, de las cuales únicamente el 0,2% mejoró los retornos obtenidos por el modelo de apuestas basado en Bet365. Como conclusión, se puede asegurar que los resultados son significativamente mejores que los obtenidos de forma aleatoria.

Finalmente, se analizó la evolución del modelo con datos posteriores a 2020 y se encontró que, a partir de esta fecha, los coeficientes WikiBuzz y RankDist ya no eran significativos para el modelo utilizado. Este descubrimiento refuerza la idea de que las casas de apuestas han mejorado sus estrategias para fijar cuotas en los últimos años, dificultando (y motivando) la labor de los investigadores en este campo. Además, queda en evidencia que los modelos de apuestas son muy sensibles a errores en los datos, donde un solo error puede provocar una gran variación en los resultados, por lo que es crucial asegurar la calidad de la información utilizada.

En materia tecnológica, uno de los grandes avances es el uso de inteligencia artificial (IA) y modelos de *machine learning* en las apuestas deportivas. Esta tecnología ha permitido desarrollar nuevas estrategias de detección de ineficiencias y predicción de resultados, más rápidas y precisas, que en muchos casos mejoran los resultados que se habían obtenido con modelos estadísticos tradicionales. Sin embargo, la evaluación de los modelos y la calidad de los datos que se utilizan siguen siendo algunos de los grandes problemas que surgen, y requieren de nuevas investigaciones para garantizar la robustez y fiabilidad de estos modelos, de ahí que numerosos estudios hayan explorado el uso de *machine learning* en la predicción de resultados.

Kollar (2020) [13] destaca la enorme cantidad de datos que se genera cada día en el deporte, incluyendo información sobre atletas, equipos, eventos o temporadas, entre otros aspectos. También muestra distintas técnicas de IA aplicables en la predicción de eventos deportivos. En primer lugar, las Redes Neuronales Artificiales (ANN), concretamente el **perceptrón multicapa (MLP)**, que requiere de 4 fases para su desarrollo:

1. Selección de parámetros de predicción

2. Diseño de un modelo de red neuronal apropiado
3. Recopilación de datos
4. Predicción de resultados

En este caso se emplean 10 parámetros de predicción, que servirán como entrada para un MLP con 2 capas ocultas y 20 neuronas por capa, que lidiarán con la complejidad de este problema [24].

Por otro lado, Kollar muestra interés en la utilización de *Support Vector Machines* (SVM) para hacer predicciones. Este algoritmo se basa en buscar el mejor hiperplano que divide dos grupos, actuando como una línea de decisión. Esto se combina con árboles de decisión, ya que los SVM no generan reglas de decisión explícitas.

Un modelo basado en SVM puede predecir el resultado de un partido, pero no proporciona una interpretación clara sobre cómo se ha llegado a esa predicción. Esto limita su utilidad cuando se busca comprender los factores que afectan, por ejemplo, al rendimiento de un equipo. Por el contrario, los árboles de decisión generan reglas explícitas, lo que permite identificar patrones. Por ejemplo, un árbol de decisión podría establecer que si un equipo mantiene más del 60 % de posesión y alcanza un porcentaje de acierto en tiros superior al 50 %, entonces la probabilidad de victoria es del 80 %.

Para aprovechar las fortalezas de ambos enfoques, se propone la combinación de SVM con árboles de decisión mediante el modelo **HSVMDT** (*Hybrid Support Vector Machine and Decision Tree*) [23].

Por último, Kollar explica brevemente cómo pueden emplearse **Cadenas de Markov** en este contexto. La idea, de forma simplificada, se basa en calcular la probabilidad de un equipo considerando únicamente el resultado de su partido anterior, lo que se conoce como Cadenas de Markov de primer orden.

## 2.3. Estrategias de gestión del patrimonio

Uno de los factores más importantes para sacar rendimiento con las apuestas deportivas, más allá de los avances tecnológicos y la mejora de las predicciones, es la gestión del patrimonio (*bank*).

Durante años, se han realizado estudios, teóricos y experimentales, que buscan optimizar estrategias de gestión del *bank* con el fin de maximizar los beneficios y reducir los riesgos. En este sentido, Uhrín *et al.* (2021) [17] afirma que un peor modelo predictivo con una buena gestión del *bank*, puede mejorar los resultados de un mejor modelo con una mala gestión. Además, destaca cómo muchos apostadores utilizan

estrategias informales (triviales) de gestión del bank, como el *flat betting*, donde se apuesta siempre la misma cantidad sin tener en cuenta otros factores.

Por otro lado, existen estrategias formales basadas principalmente en 2 corrientes: la teoría de Markowitz (*Modern Portfolio Theory* (MPT)) [16] y el criterio de Kelly [12]. La primera se basa en la idea de que un portfolio es mejor que otro si tiene, al menos, el mismo beneficio esperado y, como máximo, el mismo riesgo asociado. Por otro lado, el criterio de Kelly establece que un portfolio es elegido de forma que el valor del patrimonio en el largo plazo se maximiza. Este criterio también tiene en cuenta el riesgo de forma implícita, ya que el patrimonio de un portfolio será menor en el largo plazo si tiene más riesgo.

Algunas de las estrategias informales más usadas, basadas en heurísticas simples, son:

- Apostar un porcentaje fijo en cuotas favorables.
- Apostar un porcentaje fijo en la cuota con el valor esperado máximo.
- Apostar un porcentaje igual a la discrepancia absoluta entre la estimación del jugador y la de la casa de apuestas.
- Apostar un porcentaje igual a la discrepancia relativa entre la estimación del jugador y la de la casa de apuestas.
- Apostar un porcentaje igual a la probabilidad estimada de ganar.

Sin embargo, estas estrategias han sido probadas como inferiores a las estrategias formales de MPT y el criterio de Kelly [10].

Otro tipo de estrategias se basan en alcanzar un objetivo. Por ejemplo, partiendo de 100€, alcanzar un bank de 1000€. En esto se centra la Conjetura de L. Breiman [14], que afirma la existencia de un umbral que marca un cambio de estrategia dependiendo de si se está por encima o por debajo de dicho umbral. Según Breiman, esta estrategia optimiza los resultados para estrategias basadas en umbral; sin embargo, se ha demostrado la inverosimilitud de esta conjetura en estudios posteriores [25].

En resumen, el análisis de la literatura demuestra que existen oportunidades de inversión en el mercado de las apuestas deportivas. Sin embargo, parece evidente la necesidad de tener unos datos de calidad para ajustar modelos de predicción adecuados, y aquí residirá la clave de futuras líneas de investigación que tienden con claridad hacia el uso de modelos de *machine learning* y técnicas vanguardistas de inteligencia artificial que emplean grandes volúmenes de datos. De la misma forma, a raíz de lo visto en la literatura, podemos afirmar que el diseño de una estrategia de gestión del patrimonio es, como mínimo, tan importante como tener un modelo predictivo de calidad. Y en esto se enfocará este trabajo.



## Capítulo 3

# Descripción del problema

La idea principal de este trabajo es analizar las estrategias basadas en value bets aplicadas a datos reales de dos casas de apuestas, Bet365 y Winamax. Como material sobre el que trabajar, vamos a utilizar un dataset proporcionado por la empresa Betburger [1], una plataforma que ofrece servicios para la identificación de oportunidades de arbitraje y *value bets*.

### 3.1. Identificación de Value bets

Betburger, en su documentación, proporciona la siguiente fórmula para identificar una *value bet*:

$$\text{Stake} = \left( \frac{1}{BC} \times P \right) - \left( \frac{1}{BD} \times 100 \right) \quad (3.1)$$

Donde:

- **BC:** Cuota de la casa usada de referencia (bet365).
- **BD:** Cuota de la casa usada para realizar la *value bet* (Winamax).
- **P:** Porcentaje del dinero que se entrega como premios (payout).
- **Stake:** Valor de la apuesta y cantidad a apostar.

Las probabilidades que asignan las casas de apuestas a los resultados de un evento vienen dadas por el inverso de la cuota. La suma de estas probabilidades será mayor que 1, siendo el excedente el margen de beneficio de la casa (*overround*). Esta fórmula compara la probabilidad asignada por Bet365 con la probabilidad asignada por Winamax. Estas probabilidades implícitas en las cuotas no reflejan la probabilidad real del evento debido al *overround*, lo que supone que las probabilidades implícitas estén infladas. Es por esto que en la fórmula 3.1 se aplica una corrección a la probabilidad de Bet365 multiplicándola por el payout, de forma que se obtiene una mejor aproximación a la probabilidad real.

Basándose en esto, existirá una *value bet* si y solo si el stake es mayor que 0. Además, un mayor stake indica una mayor confianza en la apuesta. Con esta premisa, es posible diseñar estrategias basadas en el stake, las cuales veremos más adelante y comprobaremos el desempeño de cada una.

## 3.2. Dataset

El dataset proporcionado por Betburger contiene información sobre apuestas en distintos eventos deportivos. Concretamente, recopila mensajes de Telegram que contienen toda la información necesaria sobre *value bets* existentes en dichos eventos.

El dataset empleado en este estudio ha sido puesto a disposición pública para garantizar la transparencia y la reproducibilidad de los resultados. Puede consultarse en el siguiente enlace: [GitHub - TFG Value Bets Dataset](#)

### 3.2.1. Análisis descriptivo

El dataset consta de un total de 2333 instancias correspondientes a apuestas de valor identificadas de manera automática por el software de Betburger.

Contiene las siguientes columnas:

- **Timestamp Mensaje:** Fecha y hora en la que se envió el mensaje.
- **Partido:** Nombre de los equipos/jugadores que se enfrentan y competición a la que pertenece.
- **Deporte:** Deporte al que pertenece el evento.
- **Fecha del Evento:** Fecha en la que se celebra el evento.
- **Tipo de Apuesta:** Tipo de apuesta que se debe realizar (resultado final, número de goles, diferencia de puntos...).

- **Cuota:** Cuota ofrecida por Winamax.
- **Stake (€):** Valor de la apuesta de Winamax frente a la de Bet365. Indica la confianza que se tiene en dicha apuesta.
- **Enlace:** Enlace a la página de Winamax donde se puede realizar la apuesta.

Las columnas 'Cuota' y 'Stake (€)' representan las variables numéricas que se emplearán en este estudio. El resto de las variables, todas ellas categóricas, servirán para identificar el evento y poder clasificar cada una de las apuestas.

Para tener una visión global del dataset, se han realizado resúmenes descriptivos de las variables más relevantes: Deporte, Tipo de Apuesta, Cuota y Stake. Las variables Timestamp Mensaje, Partido, Fecha del Evento y Enlace no se incluyen en este análisis, ya que o bien no aportan información de interés para el estudio (Timestamp Mensaje y Enlace) o bien su valor para cada instancia es único (Partido y Fecha del Evento), por lo que no es de utilidad resumirlas. En la tabla 3.1 se muestra un resumen de las variables numéricas. Las variables categóricas se encuentran en la tabla 3.2.

	Cuota	Stake (€)
count	2333	2333
mean	2.52	5.172
std	0.53	7.714
min	1.70	0.000
median	2.30	3.076
max	3.95	66.923

Tabla 3.1: Cuota y Stake en dataset original

	Deporte	Tipo de Apuesta
count	2333	2333
unique	17	595
top	Basketball	WIN 1 0.0 (regular time)
freq	1223	168

Tabla 3.2: Deporte y Tipo de Apuesta en dataset original

Se puede ver también en el gráfico 3.1 cuáles son los deportes donde se encuentra el mayor número de value bets en este estudio.

A partir de él, se puede concluir que el deporte donde más oportunidades de value bets se encuentran es el baloncesto, con una frecuencia de casi el doble que su perseguidor, el fútbol.

En este análisis se puede observar la presencia de posibles errores y aspectos llamativos en los datos, por lo que se va a proceder a explicar dichos errores y a realizar una limpieza y procesamiento del dataset.

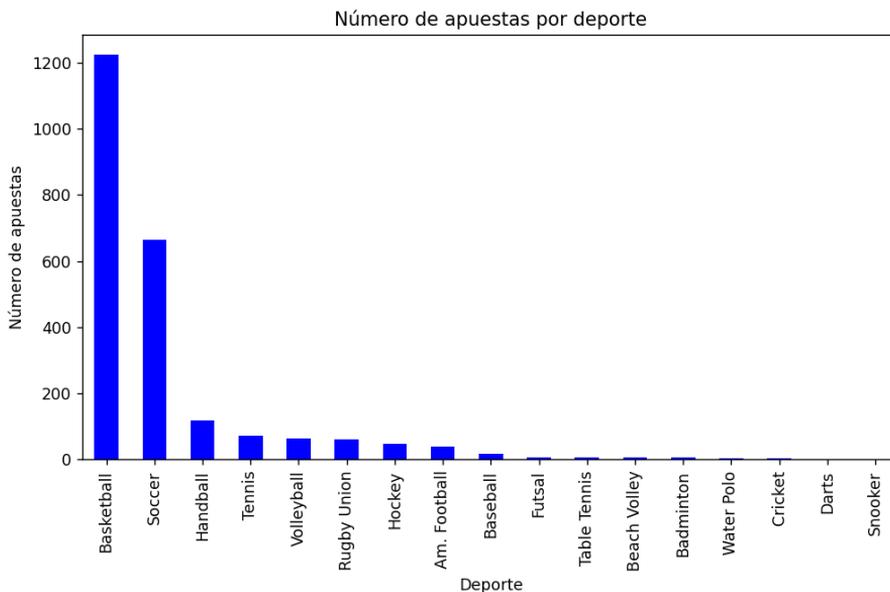


Figura 3.1: Presencia de cada deporte en el dataset.

### 3.2.2. Procesamiento

Con el análisis descriptivo realizado se han observado algunas anomalías en los datos. En primer lugar, llama la atención el stake mínimo con valor 0 en la tabla 3.1. Un stake de 0 implica que no hay valor en esa apuesta y, por tanto, no se debe apostar. Por este motivo, se deben eliminar las instancias de apuestas con stake 0.

Por otro lado, la tabla 3.2 muestra un total de 595 valores únicos para la variable Tipo de Apuesta. Para poder interpretar esta variable, es necesario agrupar las apuestas en 3 tipos:

- **Asian Handicap.** Es un tipo de apuesta que otorga una ventaja (o desventaja) a uno de los participantes. Por ejemplo, Asian Handicap1() +1.5 da una ventana de 1.5 puntos al participante 1, de forma que la apuesta será ganadora si y solo si el participante 1 no pierde de más de 2 puntos.
- **Over/Under.** Se apuesta a que el número total de goles, puntos u otro parámetro será mayor (Over) o menor (Under) que un valor determinado por la casa de apuestas. Por ejemplo, la apuesta Total Over() 2.5 resulta ganadora si hay más de 2.5 goles en el partido.
- **1X2.** Es la apuesta tradicional al resultado del partido, donde se elige entre la victoria del equipo local (1), empate (X) o victoria del equipo visitante (2).

Con este método de agrupación se cubre el total de las instancias, de forma que apuestas como Asian Handicap1() +1.5 y Asian Handicap1() -5.5, que antes se tomaban por separado, ahora se incluyen en el mismo grupo. En la tabla (3.3) se muestra la nueva distribución de la variable Tipo de Apuesta.

Grupo	Frecuencia
Asian Handicap	898
Over/Under	898
1X2	537
<b>Total</b>	<b>2333</b>

Tabla 3.3: Distribución de la nueva agrupación de apuestas

### 3.2.3. Cálculo de la cuota de Bet365

La información de la que se dispone es suficiente para aplicar estrategias de *value betting* y comprobar su viabilidad; sin embargo, nos interesa conocer también la cuota ofrecida por Bet365 para profundizar más en el análisis. Para ello, podemos utilizar la información disponible en el dataset para calcular la cuota de Bet365 a partir de la cuota de Winamax y el stake.

A partir de la fórmula (3.1), se despeja la cuota de Bet365:

$$BC = \frac{P}{\left(\text{Stake} + \frac{100}{BD}\right)} \quad (3.2)$$

Con la fórmula (3.2) podemos obtener la cuota de bet365 a partir de la cuota de Winamax y el stake. Para ello, necesitamos conocer el payout de la apuesta. El payout es el porcentaje de dinero que la casa de apuestas devuelve a los jugadores en forma de premios. Para una apuesta individual, en un evento con 2 resultados posibles, el payout se calcula de la siguiente forma:

$$P = \frac{1}{\frac{1}{WIN1} + \frac{1}{WIN2}} \quad (3.3)$$

Es fácil comprobar la relación entre payout y overround, siendo el payout el inverso del overround.

Para nuestro caso, al no tener información para calcular el payout de cada apuesta, vamos a utilizar el payout medio de Bet365. Para mercados grandes, el payout se sitúa en torno al 94 %, mientras que puede descender hasta el 90 % para mercados con menor liquidez [4].

El dataset que se va a utilizar contiene eventos de todo tipo de deportes y ligas, abarcando tanto mercados grandes como pequeños, aunque son mayoría las ligas y deportes minoritarios. Esto parece lógico si tenemos en cuenta que la falta de información suele llevar a mayor discrepancia entre las casas de apuestas y, como consecuencia, a la aparición de *value bets*. Por esta razón, las cuotas de Bet365 se van a calcular con un payout del 92%, un punto intermedio en el rango que maneja esta casa.

Tras esto, se ha utilizado un programa en python para el cálculo de las cuotas y se ha añadido esta nueva variable al dataset. La diferencia entre las cuotas de Winamax y Bet365 se observa en la tabla 3.4.

	Cuota Winamax	Cuota Bet365
mean	2.52	2.07
std	0.53	0.44
min	1.70	0.96
median	2.30	1.93
max	3.95	3.50

Tabla 3.4: Comparación cuotas Winamax y Bet365

#### 3.2.4. Eliminación de outliers

En el proceso de revisión del dataset se han observado algunos valores de stake que pueden resultar atípicamente altos y, tras el cálculo de las cuotas de Bet365, es evidente la presencia de errores en los datos. Debido al stake tan alto asignado a algunas apuestas, la cuota de Bet365 obtenida es inferior a 1, lo cual carece de toda lógica, ya que, aunque la apuesta fuese ganadora, el retorno sería negativo.

El dataset contiene apuestas con stake de más de 60. Stakes demasiado elevados suponen estrategias muy arriesgadas al estar poniendo en juego gran parte de nuestro capital en una sola apuesta. Por otro lado, un stake excesivamente alto nos tiene que hacer sospechar que puede haber algún error en los datos, ya que es consecuencia de una notable diferencia entre las cuotas, por lo que un stake elevado indicaría una discrepancia de opiniones entre las casas de apuestas demasiado grande, lo cual es extremadamente improbable.

Esta premisa hace necesario realizar una limpieza de los datos, eliminando aquellos valores de stake que presenten grandes diferencias respecto al resto (*outliers*).

Existen diversos métodos para la detección de *outliers*, cada uno con características y enfoques distintos. La elección del método adecuado dependerá en gran medida de la naturaleza de los datos con los que se esté trabajando. Los datos pueden variar significativamente en términos de distribución o escala, lo que determinará el desempeño de cada uno de los métodos. Por lo tanto, es esencial tener en cuenta la naturaleza del conjunto de datos al seleccionar el método más adecuado.

Algunos de los métodos más utilizados son el criterio de 3 sigma [19], que usa la media como pivote, o el método de Tukey [21], que emplea el rango intercuartílico (IQR) para establecer los umbrales de eliminación.

En el gráfico 3.2 se puede observar cómo se distribuye el stake. Es una distribución con gran asimetría a la derecha, donde se pueden observar valores atípicos cercanos a 60.

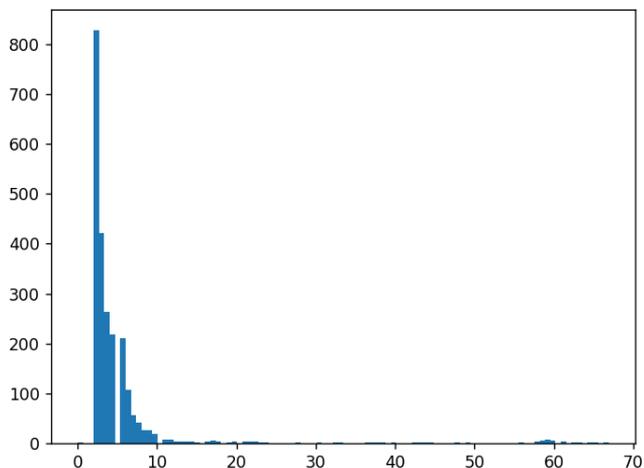


Figura 3.2: Distribución del stake en el dataset.

En una distribución sesgada a la derecha, la media se ve afectada por los valores extremos en la cola, por lo que el criterio de 3 sigma no es adecuado. Al utilizar la mediana como medida central, estamos evitando que estos valores influyan en la identificación de outliers, ya que el uso de la mediana supone un enfoque más robusto frente a valores atípicos extremos.

Por otro lado, el método de Tukey funciona mejor cuando los datos siguen una distribución aproximadamente simétrica o cuando la asimetría no es muy pronunciada. En distribuciones sesgadas a la derecha, este método puede ser menos efectivo, ya que los valores extremos pueden influir significativamente en los cuartiles.

### Método basado en la mediana

Un método adecuado a la naturaleza de nuestros datos puede ser utilizar un múltiplo de la mediana como umbral. La mediana es una medida muy robusta en comparación con la media, por lo que no se ve afectada por valores extremadamente grandes

o pequeños. También es preferible a la utilización del IQR al funcionar mejor con distribuciones poco simétricas.

El uso de un múltiplo de la mediana nos permite fijar esta magnitud de manera arbitraria, de forma que puede adaptarse a los datos que tenemos. Por ejemplo, para distribuciones sesgadas, como en este caso, es preferible fijar múltiplos más altos que para distribuciones más simétricas. Además, es un enfoque intuitivo que facilita la interpretación del análisis.

Es común emplear múltiplos como 5 o 10 veces la mediana. En este caso, para una mediana de 3.0769, parece adecuado emplear como umbral 5 veces la mediana. Si escogiéramos 10, aceptaríamos stakes del orden de 30, lo que me parece arriesgado y poco probable de ocurrir. Por tanto, la elección del umbral es de 5 veces la mediana. En la tabla 3.5 se puede ver el desempeño de cada método sobre los datos.

Método	Umbral	Instancias eliminadas
Tukey	14.6154	87
3 Sigma	28.3092	51
5 Mediana	15.3846	83

Tabla 3.5: Comparación de métodos de eliminación de outliers

Empleando el criterio ya mencionado, y mediante un programa en Python, se han eliminado las instancias que tenían un stake superior a 5 veces la mediana.

En las figuras 3.3 y 3.4 se puede ver la diferencia antes y después de la eliminación de los outliers. Siguen existiendo algunos, pero no se deben eliminar de acuerdo con el criterio elegido.

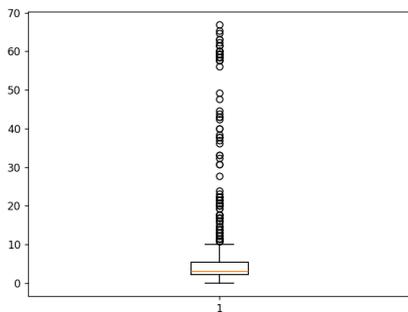


Figura 3.3: Boxplot del stake con outliers.

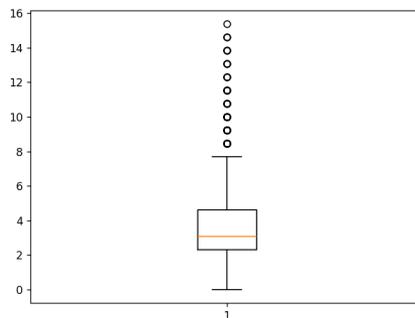


Figura 3.4: Boxplot del stake tras eliminar outliers.

### 3.2.5. Dataset definitivo

Una vez se ha realizado el procesamiento correspondiente, se ha calculado la cuota de Bet365 y se han eliminado los valores atípicos, tenemos una mejor perspectiva para interpretar el dataset y, más adelante, analizar las posibles estrategias a aplicar sobre estos datos.

El nuevo dataset contiene una variable adicional respecto al original, correspondiente a la cuota de Bet365. Además, posee 83 instancias menos, las cuales se trataban de outliers en la variable Stake.

En las tablas 3.6 y 3.7 se encuentran los resúmenes estadísticos correspondientes a las variables de interés del nuevo dataset.

	Cuota	Stake
mean	2.51	3.896
std	0.52	1.992
min	1.70	2.307
median	2.30	3.076
max	3.95	15.384

Tabla 3.6: Cuota y Stake tras procesamiento

	Deporte	Tipo de Apuesta
count	2250	2250
unique	16	3
top	Basketball	Asian Handicap
freq	1207	876

Tabla 3.7: Deporte y Tipo de Apuesta tras procesamiento

En el dataset original había 17 deportes, mientras que en el nuevo hay únicamente 16, ya que durante el procesamiento se ha eliminado el snooker, que tan solo contaba con 1 instancia.

En la tabla 3.8 vemos la nueva comparativa entre las cuotas de Winamax y Bet365.

	Cuota Winamax	Cuota Bet365
mean	2.51	2.10
std	0.52	0.42
min	1.70	1.42
median	2.30	1.93
max	3.95	3.50

Tabla 3.8: Cuota Winamax y Bet365 tras procesamiento

Por último, para comprobar si existe una diferencia significativa entre las cuotas de ambas casas, se ha realizado un contraste de medias con muestras pareadas:

$$H_0 : \mu_{\text{diferencias}} \leq 0$$

$$H_1 : \mu_{\text{diferencias}} > 0$$

cuyo estadístico se define como:

$$t = \frac{\bar{d}}{S_d/\sqrt{n}} \sim t_{n-1}$$

donde:

- $\bar{d}$  es la media de las diferencias.
- $S_d$  es la desviación estándar de las diferencias.
- $n$  es el número de pares de observaciones.

El test arroja un p-valor de aproximadamente 0, por lo que se puede concluir que las cuotas de Winamax son significativamente superiores a las de Bet365, de ahí que estas apuestas tengan valor.

### 3.3. Clasificación de las apuestas

Con el fin de evaluar el rendimiento de las distintas estrategias, será necesario etiquetar cada apuesta en función de su resultado. Para ello, se utilizará una variable de tipo factor con 3 niveles, donde cada nivel corresponde a uno de los resultados posibles de una apuesta. Estos niveles son:

- **Acierto.** El jugador gana dinero con la apuesta.
- **Fallo.** El jugador pierde el dinero de la apuesta.
- **Nulo.** El dinero de la apuesta es devuelto al jugador.

Un resultado nulo se da en casos en los que la apuesta contempla esa posibilidad. Por ejemplo, en apuestas como Asian Handicap1() 2.0, resultará nula si el equipo 1 pierde exactamente de 2 puntos. Además, en el caso de Winamax, en caso de retirada de un jugador por lesión (en deportes individuales, por ejemplo, el tenis), cualquier apuesta se considera nula, a excepción de las que ya hubieran resultado ganadoras o

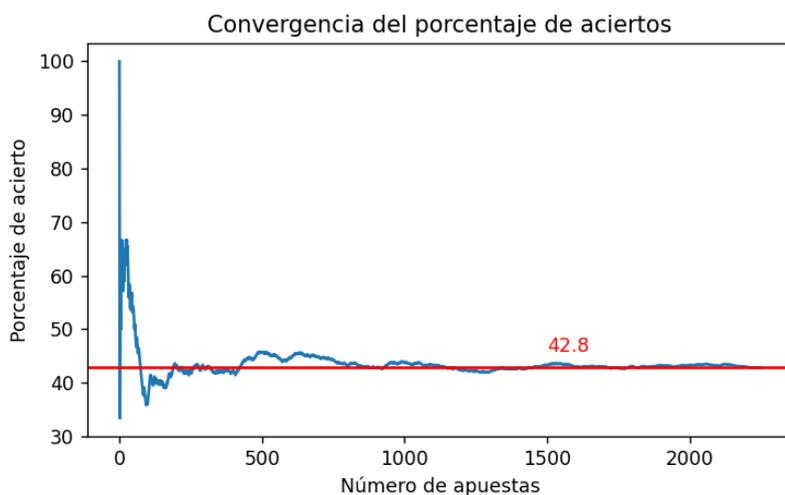


Figura 3.5: Convergencia del porcentaje de aciertos

perdedoras antes de la retirada. Lo mismo sucede con eventos que hayan sido aplazados o cancelados.

La clasificación de cada apuesta se lleva a cabo comprobando su resultado en el enlace proporcionado en el dataset. Cada apuesta tiene su enlace, por lo que será necesario visitarlos uno por uno. Para agilizar la tarea, he desarrollado un script en Python que accede automáticamente a cada uno de los enlaces uno por uno y permite etiquetar las apuestas de forma rápida, introduciendo '1' si es un acierto, '2' si es un fallo o '3' si la apuesta es nula.

De esta forma se han clasificado las 2250 apuestas, pudiendo calcular las tasas de acierto y fallo, dando una primera vista de los resultados que ofrece esta herramienta. Las frecuencias se muestran en la tabla 3.9.

	<b>Acierto</b>	<b>Fallo</b>	<b>Nula</b>
Total	963	1239	48
Tasa	42,8 %	55,066 %	2,133 %

Tabla 3.9: Tabla de frecuencias de fallos y aciertos

En el gráfico 3.5 se muestra la convergencia de la tasa de aciertos, lo que indica que el tamaño de la muestra es suficiente para tener un buen estimador de la probabilidad de acierto.



## Capítulo 4

# Metodología

Una vez se tienen los datos necesarios, el objetivo es comprobar si es posible obtener rentabilidad a partir de las *value bets*. Como ya hemos visto, los dos pilares fundamentales para lograr esto son unas predicciones de calidad y una buena gestión del bank. De lo primero se encarga el software de Betburger, por lo que es necesario abordar el segundo problema.

Para una buena gestión del bank, durante años se han diseñado diversas estrategias, las cuales ofrecerán mejores o peores resultados en función de diferentes aspectos, como el tipo de apostador o el bank del que se dispone. Una estrategia puede adaptarse mejor a un apostador recurrente, el cual dedique más tiempo y realice varias apuestas al día, mientras que otra estrategia puede dar mejores resultados para alguien que lo haga con menor frecuencia. Es por ello que existe gran cantidad de estrategias, y será objetivo de cada jugador encontrar aquella que mejor se adapte a su estilo y su bank.

En el marco de este trabajo, las estrategias que se van a estudiar son las siguientes:

1. **Stake plano.** La cantidad apostada es constante.
2. **Stake progresivo.** La cantidad apostada es proporcional al stake.
3. **Stake proporcional con bank fijo.** La cantidad apostada es un porcentaje del bank inicial.
4. **Stake proporcional con bank dinámico.** La cantidad apostada es un porcentaje del bank actual.

La primera estrategia es la más simple y, por tanto, por la que se decanta la gran mayoría de apostadores inexpertos. La segunda, sin embargo, tiene más dificultad al

tener en cuenta el stake (valor de la apuesta), pero previsiblemente arrojará mejores resultados al dar más peso a las apuestas con más valor. Las dos últimas representan la misma estrategia, aunque desde dos enfoques distintos. En la primera de ellas, el porcentaje se aplica siempre sobre el bank inicial, mientras que en la segunda se aplica sobre el bank actual.

El bank dinámico es un enfoque que puede multiplicar exponencialmente los beneficios. De hecho, el beneficio sigue una distribución exponencial, lo cual hace que aumente enormemente la volatilidad y exista una gran diferencia entre los valores más bajos y los más altos. Por este motivo, a la hora de realizar simulaciones con bank dinámico no se debe utilizar la media, sino que usaremos la mediana para tener más robustez.

Para hacer una comparación justa de las estrategias, se van a aplicar bajo las mismas condiciones:

- Mismas apuestas para cada estrategia
- Mismo orden de apuestas
- Bank inicial de 1000€

Además, cada una de las estrategias será evaluada y comparada utilizando las mismas métricas:

- Bank final
- Beneficio
- Bank máximo
- Bank mínimo
- Cantidad total apostada
- Gráficas de la evolución del bank

## 4.1. Estrategias

Fijadas las condiciones y los criterios de evaluación, hay que conocer a fondo las estrategias que se han aplicado.

### 4.1.1. Stake plano

La estrategia del stake plano es la más simple de todas. Consiste en apostar un mismo porcentaje del bank en todas las apuestas. Este porcentaje suele estar entre el 1% y el 10%, dependiendo del riesgo que busque asumir el jugador. El objetivo será encontrar el stake óptimo que maximice los beneficios asumiendo el mínimo riesgo. Para ello, se realizarán simulaciones que permitan graficar la evolución del riesgo y del beneficio a medida que aumenta el porcentaje, y con ello se podrá ver el punto donde se alcance la relación riesgo-beneficio óptima.

Este enfoque tiene pros y contras. En primer lugar, a diferencia de las demás estrategias que veremos más adelante, aquí no se tiene en cuenta el nivel de confianza o el valor que hay en la apuesta para determinar la cantidad apostada. Esto puede derivar en una disminución de los beneficios y un aumento del riesgo, lo cual es recomendable evitar. Sin embargo, su propia naturaleza la convierte en la estrategia más sencilla de aplicar, ya que evitas hacer cálculos respecto a la cantidad a apostar, siendo muchas veces la estrategia preferida de los apostadores inexpertos.

De cada estrategia se mostrarán los resultados obtenidos con los datos que tenemos y los resultados de realizar 10000 simulaciones *bootstrap*, para así comparar el caso concreto con la generalización. Más adelante, veremos los resultados de esta estrategia y los compararemos con los de las demás.

### 4.1.2. Stake progresivo

Esta estrategia tiene en cuenta la confianza en la apuesta, que está representada por el stake. La idea se basa en fijar una cantidad de dinero por unidad de stake, de forma que la cantidad total apostada viene dada por el producto del stake multiplicado por dicha cantidad fijada. En este caso, también se debe encontrar la cantidad fijada óptima y, a partir de ahí, obtener el rendimiento de esta estrategia. Para ello, se sigue el mismo procedimiento que en la estrategia de stake plano.

Dependiendo de la cantidad fijada, estaremos adoptando un enfoque más agresivo o más conservador. Esta cantidad puede variar entre un 1% y un 3% del bank. El porcentaje del bank será menor que en la estrategia de stake plano, ya que, como es obvio, la cantidad final apostada aumentará al estar multiplicada por el stake.

Al final, se calcularán las métricas de evaluación para el porcentaje óptimo obtenido y se comparará su desempeño frente al resto de estrategias.

### 4.1.3. Stake proporcional

La última estrategia sigue la línea del stake progresivo. El stake proporcional es una estrategia que se rige estrictamente por el valor de la apuesta, sin dejar ningún parámetro a la libre arbitrariedad del apostador (excepto el bank, lógicamente). Consiste en tomar el stake como el porcentaje del bank que se debe apostar, y representa la estrategia más popular entre los usuarios.

Esta estrategia, aunque es simple, requiere hacer cálculos para todas las apuestas que se realicen, ya que el stake para cada una será diferente. Por otro lado se encuentra la variante de bank dinámico, en la que, tras cada apuesta, el bank de referencia será distinto. Esto no requiere cálculos extra, sino que únicamente es necesario actualizar el bank sobre el que se aplica el stake. Además, no es necesario calcular una cantidad óptima, como sí ocurría en las estrategias anteriores, ya que no existen parámetros arbitrarios, lo que de cierto modo supone una ventaja frente a las demás estrategias.

Por estas razones, el stake proporcional parece, a priori, una de las estrategias que mejores resultados puede dar, y una de las más atractivas para muchos usuarios debido a su sencillez.

## 4.2. Implementación

Para el cálculo de las métricas asociadas a las estrategias de apuestas, se ha desarrollado un programa en Python que realiza esta tarea. Se ha empleado el entorno interactivo *Jupyter Notebook*, que permite ejecutar código de forma modular y es especialmente útil en el campo del análisis de datos. Además, se han utilizado las librerías *pandas*, *numpy* y *matplotlib*, que facilitan la manipulación de datos, los cálculos numéricos y la visualización de resultados, respectivamente.

### 4.2.1. Descripción del programa

El programa toma como entrada el conjunto de datos en formato CSV que contiene las 2250 apuestas, cada una etiquetada como ganada (1), perdida (2) o nula (3), y se recorre cada una de ellas. Si una apuesta es nula, se pasa a la siguiente iteración sin hacer ningún cálculo. Si la apuesta no es nula, se calcula el stake en función de la estrategia que estamos aplicando y se actualiza el bank sumando o restando la cantidad correspondiente dependiendo del resultado de la apuesta. Tras cada iteración, se almacena el bank en una lista que servirá para observar su evolución. También se calcula la suma acumulada del stake, que indicará la cantidad total de dinero apostado.

Al finalizar las 2250 iteraciones, se muestra el bank final, el beneficio obtenido, el

bank mínimo y máximo almacenado en la lista de banks y el dinero total que ha sido apostado.

El flujo de trabajo sucede de la siguiente forma:

1. **Cálculo del stake óptimo (estrategias 1 y 2).** En primer lugar, para las estrategias de stake plano y stake progresivo, debemos calcular el stake óptimo. Se realizan 1000 simulaciones con cada valor de stake, calculando el beneficio medio y la probabilidad de obtener beneficios, y representándolos en un gráfico de líneas. El punto donde corten las líneas será el valor de stake óptimo.
2. **Cálculo de métricas para el dataset.** Con el valor óptimo calculado, se calcula el rendimiento de la estrategia para nuestros datos. Además del rendimiento, se calculan otras métricas como el bank mínimo y máximo alcanzados, o el dinero total apostado. La figura 4.1 muestra el diagrama de flujo del cálculo del rendimiento para las apuestas del dataset. Adicionalmente, para la estrategia que emplea bank dinámico, sería necesario actualizar el bank tras cada apuesta.
3. **Cálculo de métricas con simulaciones *bootstrap*.** Finalmente, para estimar la variabilidad de las métricas obtenidas, se realizan 10000 simulaciones de muestras aleatorias del tamaño del dataset (2250), calculando las métricas para cada una de las simulaciones y mostrando la media de todas ellas (la mediana en el caso del bank dinámico). Es importante mencionar que las medias son calculadas teniendo en cuenta únicamente las simulaciones que no acaban en bancarrota.

Adicionalmente, se muestran gráficos de la evolución del bank para cada una de las estrategias, tanto para el dataset como para las simulaciones.

Flow diagram

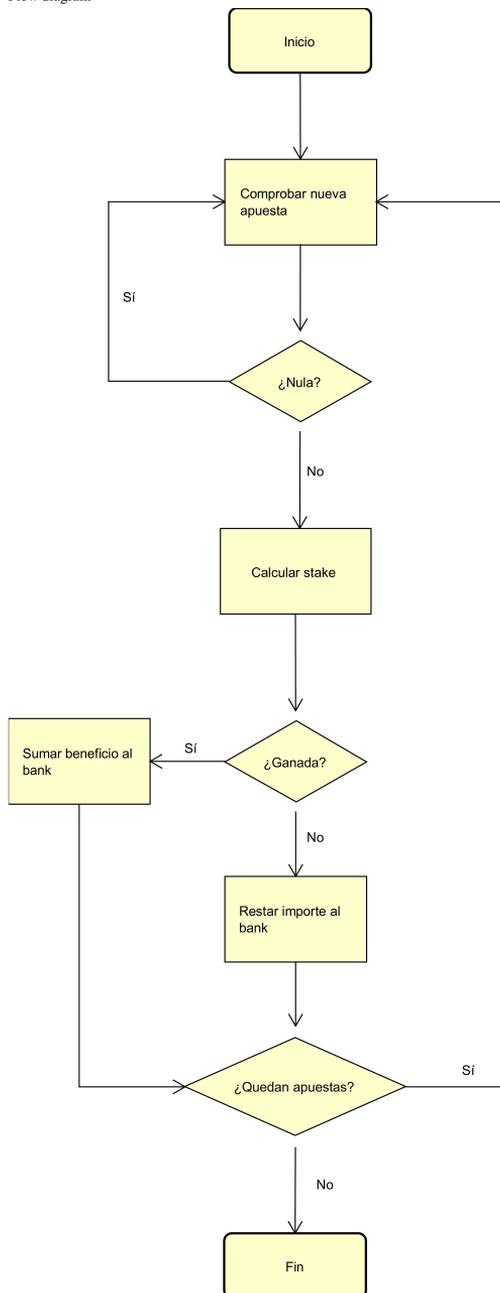


Figura 4.1: Diagrama de flujo del cálculo del rendimiento

## Capítulo 5

# Resultados

### 5.1. Stake plano

Con el fin de conocer el porcentaje óptimo, se ha calculado el rendimiento para cada unidad de porcentaje desde 1 hasta 10. Los resultados se muestran en la tabla 5.1.

Stake	Bank final	Beneficio (€)	Beneficio (%)
1 %	2426.5	1426.5	142.65
2 %	3853.0	2853.0	285.30
3 %	5279.5	4279.5	427.95
4 %	6706.0	5706.0	570.60
5 %	8132.5	7132.5	713.25
6 %	9559.0	8559.0	855.90
7 %	10985.5	9985.5	998.55
8 %	12412.0	11412.0	1141.20
9 %	0.0	-1000.0	-100.00
10 %	0.0	-1000.0	-100.00

Tabla 5.1: Resultados de stake plano en el dataset

Se puede observar cómo el beneficio aumenta notablemente con el porcentaje apostado, hasta alcanzar el 8 %. A partir de ese nivel, la estrategia no es factible y se llega a la bancarrota. Este resultado indica que el stake óptimo es 8 % para nuestros datos; sin embargo, es probable que no sea el caso general.

Para poder generalizar estos resultados, se han realizado 1.000 simulaciones de muestras con reemplazamiento de tamaño 1.000 de nuestro dataset para cada unidad

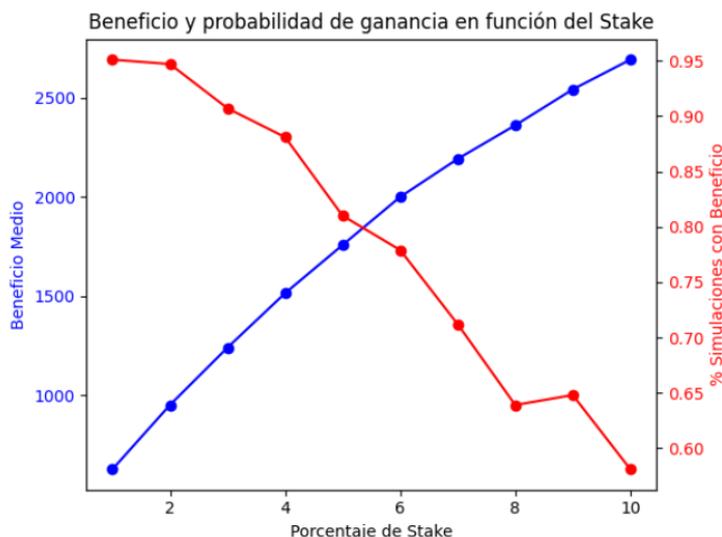


Figura 5.1: Beneficio y probabilidad de ganancias con stake plano

de porcentaje, calculando el beneficio promedio. Con esto se logra reducir el componente de aleatoriedad, ya que en muchas ocasiones el orden de las apuestas determina la ocurrencia o no de bancarrota.

Los resultados obtenidos para cada porcentaje del bank apostado se encuentran en el gráfico 5.1. Este gráfico muestra la relación entre beneficio obtenido y probabilidad de obtener beneficio tras 1000 apuestas. Se observa una clara proporcionalidad inversa entre las dos variables, como era esperable. El punto donde se cruzan las líneas marca el porcentaje de stake con mejor relación riesgo - beneficio, situándose entre el 5% y el 6%. Decidimos escoger un stake del 6%, ya que del 5% al 6% hay un aumento notable del beneficio, y la disminución de la probabilidad de obtener ganancias no es muy grande.

Con este valor de stake, se obtiene un bank final de **9559€**, lo que supone un incremento del **855.9%** del bank inicial para el dataset. En el caso de las simulaciones, el bank final medio es de **10259€**. Por otro lado, el bank máximo es de 11134€ en el dataset, y de 10922.9€ en las simulaciones. Estos resultados se pueden ver en la tabla 5.2. Además, las simulaciones indican que hay una probabilidad estimada del **22.9%** de perder la totalidad de tu bank tras 2250 apuestas.

Se puede observar que los resultados obtenidos con el dataset son ligeramente inferiores a los de las simulaciones. La mayor diferencia se observa en el bank mínimo, con un valor de 260€ para el dataset y de 632€ en las simulaciones. Esto indica que puede haber ocurrido una racha larga de apuestas perdidas, lo cual en las simulaciones se palía al converger hacia la media teórica de la distribución.

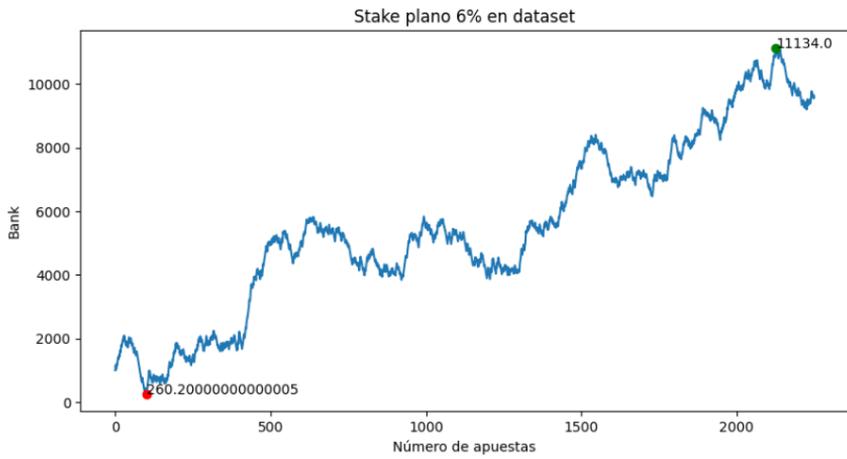


Figura 5.2: Evolución del bank con stake plano

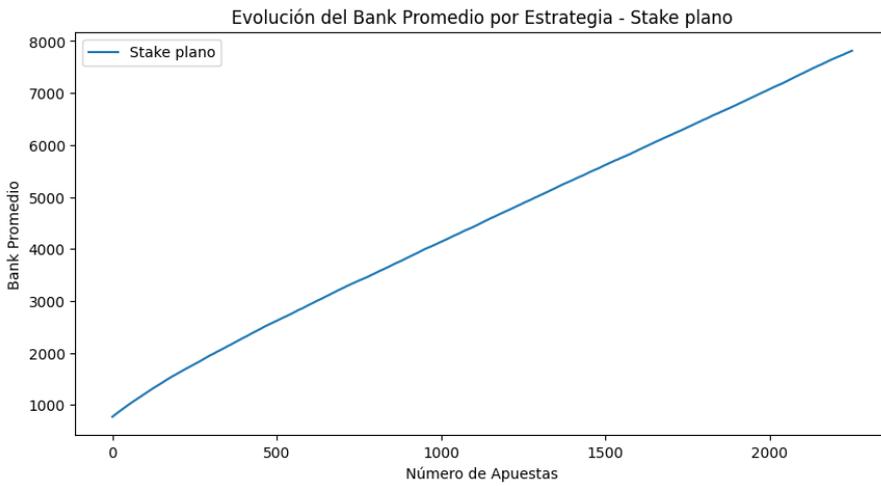


Figura 5.3: Evolución del bank con stake plano (simulación)

	Datset	Simulación
Bank final	9559	10259.55
% Beneficio	855.9	925.95
Bank máximo	11134	10922.9
Bank mínimo	260.20	632.76
Total apostado	135000	135000

Tabla 5.2: Comparación simulación y dataset con stake plano

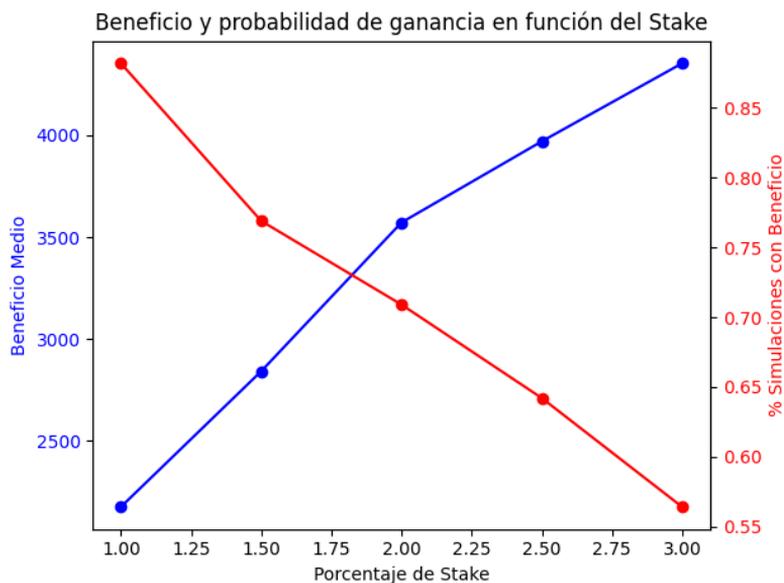


Figura 5.4: Beneficio y probabilidad de ganancias con stake progresivo

	Dataset	Simulación
Bank final	13196.92	14590.87
% Beneficio	1219.7	1359
Bank máximo	14870	15499.73
Bank mínimo	17.85	606.77
Total apostado	175353.84	175289.81

Tabla 5.3: Comparación simulación y dataset con stake progresivo

## 5.2. Stake progresivo

Para esta estrategia debemos calcular también el porcentaje óptimo del bank a utilizar. En este caso se han hecho simulaciones con 5 valores entre 1% y 3% con paso 0.5, cuyos resultados se encuentran en el gráfico 5.4. El punto de corte de ambas gráficas se encuentra entre el 1.5% y el 2% y, como en la estrategia anterior, la decisión es la de escoger el valor mayor (2%).

Con un stake del 2%, se obtiene un bank final de **13196.92€**, lo que supone un incremento del **1219.7%** del bank inicial para nuestros datos. Las simulaciones dan mejores resultados también en este caso: el bank final medio es de **14590.87€**, teniendo un beneficio del **1359%**. Estos resultados se pueden ver en la tabla 5.5. En este caso, la probabilidad de bancarrota es del **36.7%**. Es superior a la del stake plano

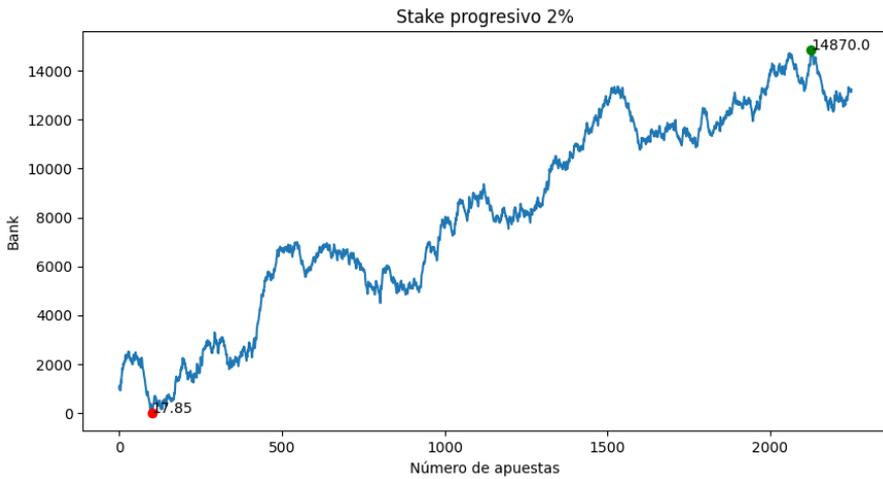


Figura 5.5: Evolución del bank con stake progresivo

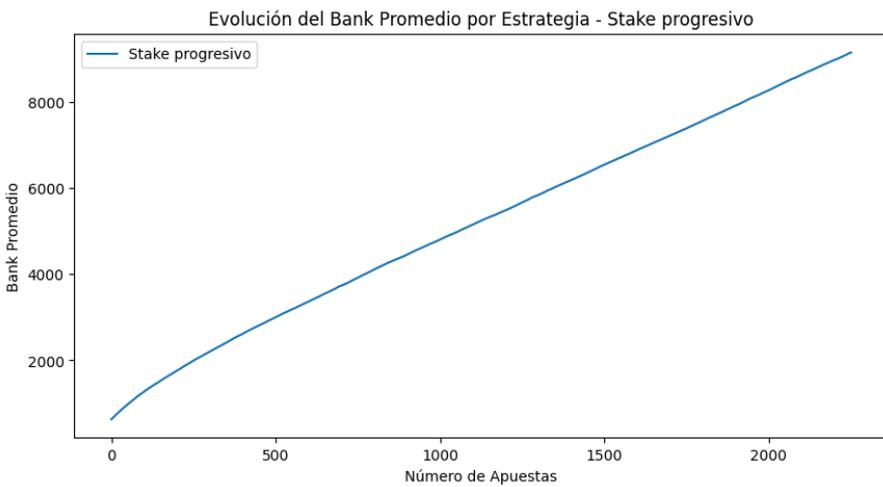


Figura 5.6: Evolución del bank con stake progresivo (simulación)

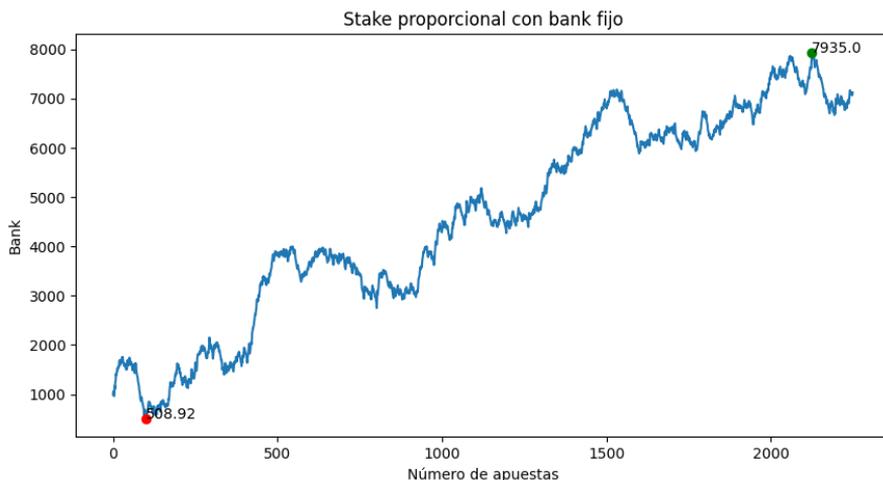


Figura 5.7: Evolución del bank con stake proporcional y bank fijo

(22.9%), lo cual es lógico al ser una estrategia más arriesgada, donde tanto el beneficio como la cantidad total apostada son mayores.

### 5.3. Stake proporcional

La estrategia de stake proporcional no requiere del cálculo de un stake óptimo, ya que este está fijado por el valor de la apuesta. Esto facilita la obtención de los resultados, los cuales se muestran a continuación para cada una de las variantes ya mencionadas.

#### 5.3.1. Bank fijo

La variante que emplea bank fijo representa una estrategia más conservadora en comparación al resto. Como se ve en la tabla 5.4, para el dataset, el bank final obtenido es de **7098.46€**, con un beneficio del **609.8%**, siendo el más bajo hasta el momento. Las simulaciones muestran esto mismo, siendo el beneficio medio más bajo con **7563.15€**. Sin embargo, también es la estrategia que menor cantidad total apostada requiere (87676.64€), lo cual indica que se está sacrificando parte del beneficio a cambio de asumir un menor riesgo. Esto también se ve reflejado en la probabilidad de bancarrota, que desciende hasta el **15.1%**.

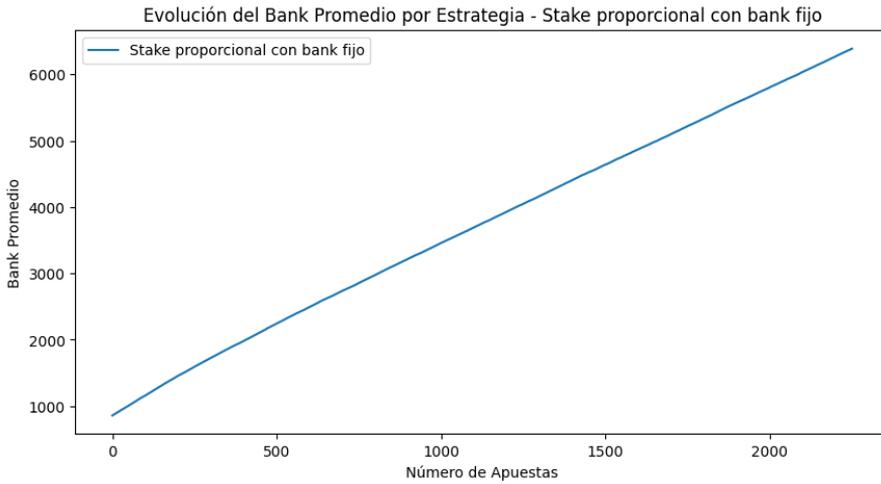


Figura 5.8: Evolución del bank con stake proporcional y bank fijo (simulación)

	Dataset	Simulación
Bank final	7098.46	7563.15
% Beneficio	609.8	656.31
Bank máximo	7935	8072.19
Bank mínimo	508.92	669.99
Total apostado	87676.92	87676.64

Tabla 5.4: Comparación simulación y dataset con stake proporcional y bank fijo

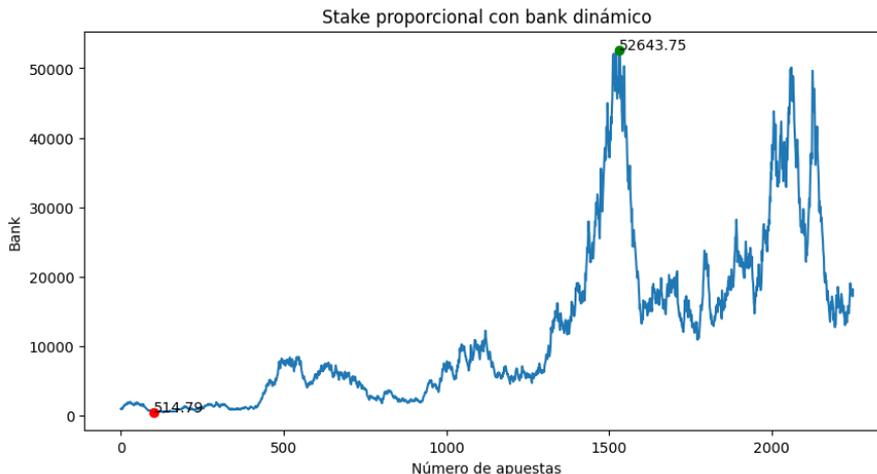


Figura 5.9: Evolución del bank con stake proporcional y bank dinámico

### 5.3.2. Bank dinámico

En el caso de esta estrategia no es posible la bancarrota, ya que siempre se va a apostar un porcentaje del bank actual y nunca va a llegar a 0. Sin embargo, podemos calcular la tasa de simulaciones que terminan en pérdidas para hacernos una idea del riesgo que conlleva esta estrategia.

En los resultados, se observa un enorme aumento del beneficio respecto a la variante de bank fijo, tanto en el dataset como en las simulaciones. El beneficio mediano para esta estrategia es de **26286.5€** según las simulaciones, siendo el mayor entre las 4 estrategias. Además, este valor es notablemente superior al obtenido en el dataset, ya que, como se puede ver en el gráfico 5.9, al final se produce una caída provocada por una racha negativa de apuestas.

La mayor diferencia respecto a las demás estrategias se encuentra en el bank máximo, que es casi 6 veces superior al de su variante de bank fijo. Esto se debe a que el crecimiento en este caso es exponencial, mientras que en el resto de estrategias es lineal. También es importante destacar la cantidad total apostada, que es 4 veces superior a la de su perseguidora, la estrategia de stake progresivo. Por último, la probabilidad de que se incurra en pérdidas tras 2250 apuestas es del **13.1%**, que no solo es la más baja, sino que además no implica que en esos casos se llegue a la bancarrota, simplemente que el retorno de la inversión hasta la apuesta número 2250 es negativo.

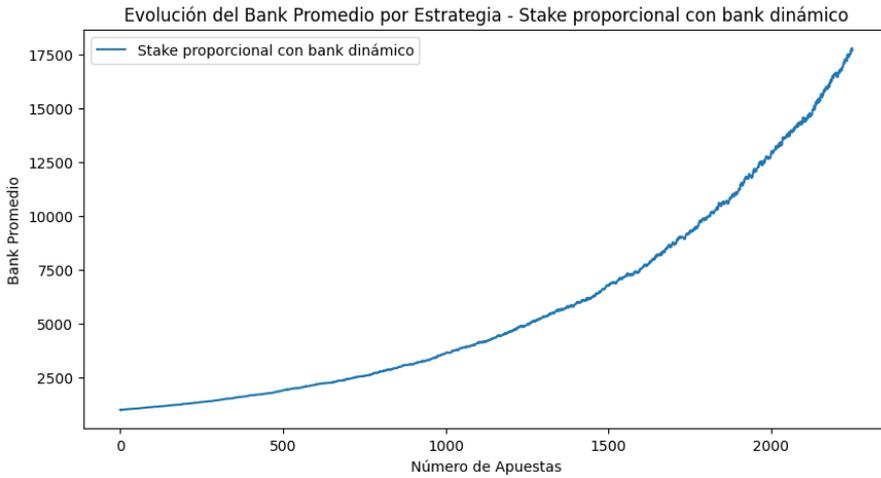


Figura 5.10: Evolución del bank con stake proporcional y bank dinámico (simulación)

	Dataset	Simulación
Bank final	17759.41	27286.5
% Beneficio	1675.9	2628.6
Bank máximo	52643.75	54865.02
Bank mínimo	514.79	565.36
Total apostado	1050381	672626.66

Tabla 5.5: Comparación simulación y dataset con stake proporcional y bank dinámico

## 5.4. Comparación de resultados

Una vez evaluadas las estrategias individualmente, es necesario compararlas entre sí para dimensionar los resultados. En los gráficos 5.11 y 5.12 se puede ver la evolución del bank, en el dataset y en las simulaciones, respectivamente, para cada una de las 4 estrategias. Es fácil observar que la estrategia que mejores beneficios alcanza es la representada en color rojo, correspondiente al stake proporcional con bank dinámico. Sin embargo, esta estrategia también posee una enorme volatilidad, como se puede ver en los grandes picos que se forman en el gráfico del dataset.

El gráfico correspondiente a las simulaciones muestra el comportamiento del beneficio para las distintas estrategias. Hay una clara diferencia en los casos donde se emplea bank dinámico y en los que se emplea bank fijo, ya que en el primer caso el crecimiento es exponencial, aumentando la volatilidad y el riesgo, mientras que en el segundo caso el crecimiento es lineal, adoptando así un enfoque más conservador.

	Tipo de Stake			
	Plano	Progresivo	Proporcional	
			Bank fijo	Bank dinámico
<b>Beneficio</b>	9259.55	13590.87	6563.15	26286.5
<b>% Bancarrota</b>	22.9	36.7	15.1	13.1*
<b>Relación</b>	404.34	370.32	458.96	2037.71

Tabla 5.6: Relación beneficio - riesgo de cada estrategia (datos de las simulaciones)

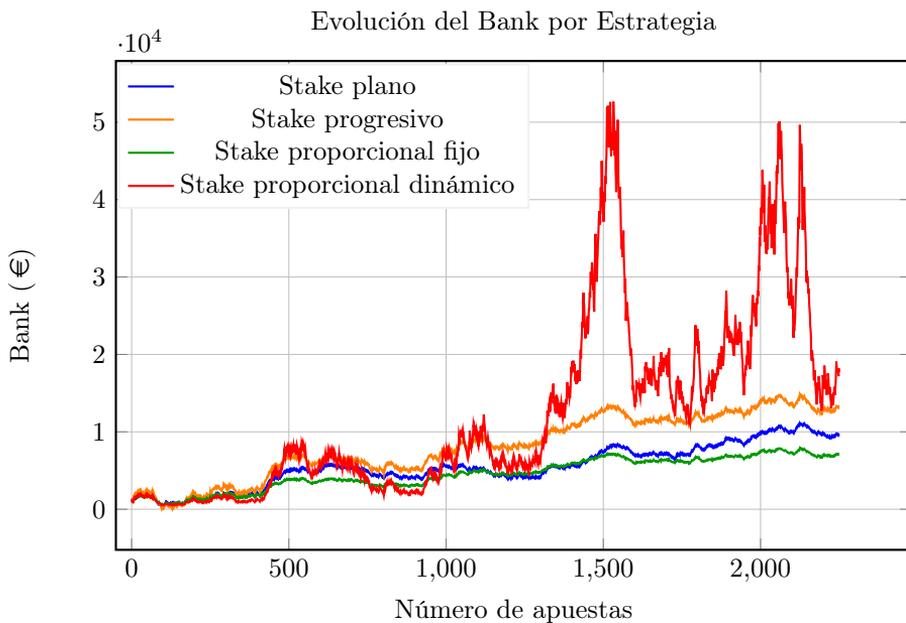


Figura 5.11: Comparativa de la evolución del bank por estrategia.

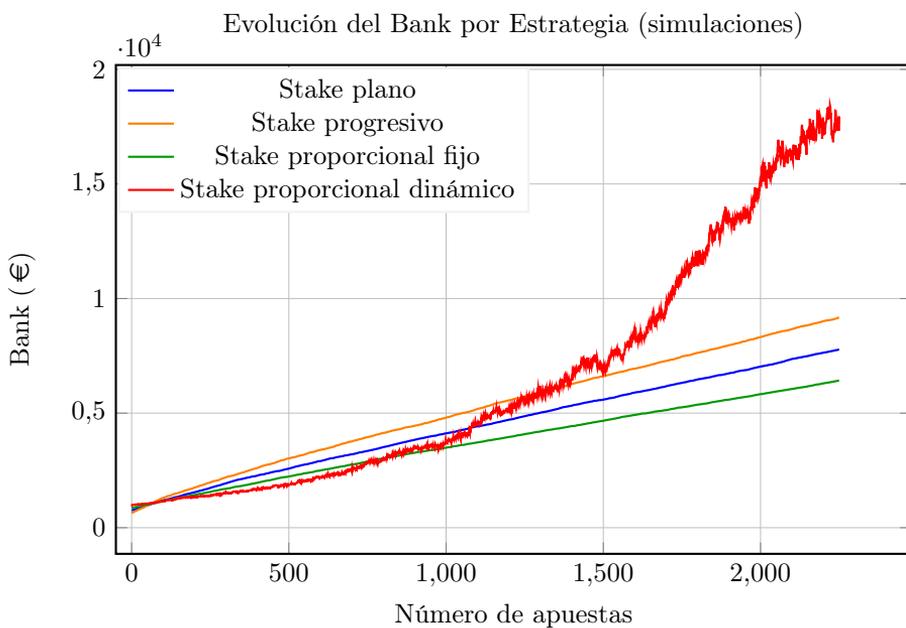


Figura 5.12: Comparativa de la evolución del bank para cada estrategia (simulaciones).

Para verificar que las estrategias generan valor, se han comparado sus resultados con los de una estrategia con stake aleatorio, demostrando así que los resultados no son fruto de la aleatoriedad. Para ello, se ha calculado la evolución del bank para cada una de las estrategias sobre una muestra aleatoria de 10.000 apuestas. La comparativa se observa en la figura 5.13, donde se observa que la estrategia aleatoria es la peor de las 5. En la tabla 5.7 se muestran los banks finales tras 10.000 apuestas, y se puede ver que hay una diferencia significativa de la aleatoria respecto al resto. A pesar de ello, la estrategia aleatoria sigue reportando beneficios debido al alto porcentaje de acierto en las apuestas del dataset, las cuales deben tomarse como algo meramente teórico.

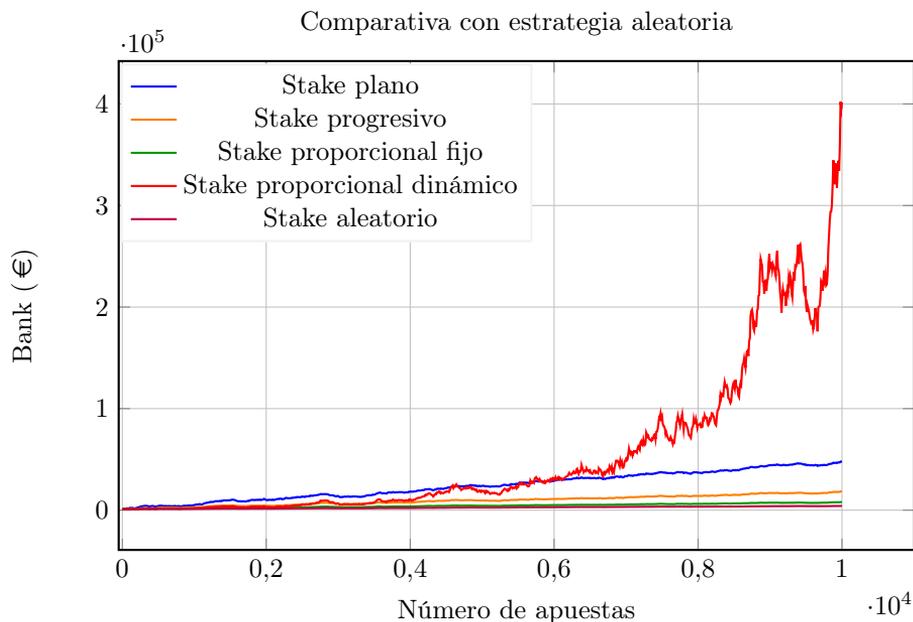


Figura 5.13: Evolución del bank para cada estrategia.

Estrategia	Bank final (€)
Stake plano	47,917.60
Stake progresivo	18,422.55
Stake proporcional fijo	7,969.02
Stake proporcional dinámico	401,451.12
Stake aleatorio	4,173.97

Tabla 5.7: Resumen del bank final tras 10.000 apuestas por estrategia.

Dependiendo del criterio utilizado, hay diferentes estrategias que se pueden considerar ganadoras frente al resto:

- **Mayor beneficio.** Si lo que se busca es maximizar el beneficio, la estrategia óptima es el **stake proporcional con bank dinámico**, pues es la que alcanza máximos más altos, y la que logrará un crecimiento más rápido del bank.
- **Menor riesgo.** Si se busca una estrategia conservadora, la opción más adecuada es el **stake proporcional con bank fijo**. Esta estrategia requiere el menor importe total apostado y presenta un porcentaje de bancarrotas más bajo, lo que proporciona mayor estabilidad a cambio de un beneficio menor.
- **Mejor relación beneficio-riesgo.** Este criterio suele ser el más adecuado a la hora de escoger una estrategia. En la tabla 5.6, podemos ver que la estrategia óptima en este aspecto es el **stake proporcional con bank dinámico**. Sin embargo, su alta volatilidad hace difícil estimar con precisión el beneficio esperado y se pueden obtener resultados muy diferentes a los predichos.
- **Facilidad de aplicación.** En caso de que el jugador se decante por la estrategia más simple, la mejor opción es el **stake plano**, ya que no requiere cálculos y mantiene una apuesta fija en cada jugada. Aunque no es la estrategia más rentable en términos de retorno, sí puede ser la más conveniente en cuanto a tiempo y esfuerzo.

Para finalizar, se resumen los resultados obtenidos sobre el dataset y sobre las simulaciones en las tablas 5.8 y 5.9.

	Tipo de Stake			
	Plano	Progresivo	Proporcional	
			Bank fijo	Bank dinámico
<b>Bank Final</b>	9559.0	13196.92	7098.46	17759.41
<b>% Beneficio</b>	855.9	1219.69	609.85	1675.94
<b>Bank Máximo</b>	11134.0	14870.0	7935.0	52643.75
<b>Bank Mínimo</b>	260.2	17.85	508.92	514.79
<b>Total apostado</b>	135000.0	175353.85	87676.92	1050381.0

Tabla 5.8: Resultados de las estrategias en el dataset

	Tipo de Stake			
	Plano	Progresivo	Proporcional	
			Bank fijo	Bank dinámico
<b>Bank Final</b>	10259.55	14590.87	7563.15	27286.5
<b>% Beneficio</b>	925.95	1359	656.31	2628.65
<b>Bank Máximo</b>	10922.9	15499.73	8072.19	54865.02
<b>Bank Mínimo</b>	632.76	606.77	669.99	565.36
<b>Total apostado</b>	135000	175289.81	87676.64	672626.66

Tabla 5.9: Resultados de las estrategias en las simulaciones

## Capítulo 6

# Conclusiones y trabajo futuro

### 6.1. Conclusiones

- I. Las *value bets* generan beneficios a largo plazo únicamente si la esperanza matemática del beneficio es positiva. Esta depende de la cuota media en los aciertos y de las probabilidades de acierto y fallo.
- II. No existe una estrategia de gestión del *bank* óptima en términos absolutos. Algunas priorizan la rentabilidad asumiendo mayor riesgo, mientras que otras son más conservadoras pero ofrecen menor retorno.
- III. Las discrepancias entre las cuotas de diferentes casas de apuestas sirven para identificar *value bets*. Como se ha demostrado en este estudio, este método es capaz de generar beneficios a largo plazo.
- IV. Es posible obtener beneficios sostenidos en apuestas deportivas, pero se requiere un método fiable para detectar *value bets* y una aplicación rigurosa de la estrategia elegida.
- V. No es posible obtener un beneficio asegurado con las *value bets*. Aunque la esperanza matemática indique rentabilidad a largo plazo, restricciones como las limitaciones impuestas por las casas de apuestas o la falta de capital suficiente, son factores que imposibilitan este hecho.

### 6.2. Limitaciones del estudio

A pesar de los llamativos resultados obtenidos, este estudio presenta ciertas limitaciones que es importante considerar. En primer lugar, el análisis se ha centrado

únicamente en un subconjunto específico de eventos deportivos y casas de apuestas, lo que puede limitar la generalización de los resultados a otros contextos. Aunque se han realizado simulaciones para ver el comportamiento a largo plazo, no deja de ser una mera estimación.

Por otro lado, el estudio se ha realizado en base a los datos proporcionados directamente por el software, lo que indica que las cuotas de las que disponemos son las que había en el momento en que el bot las ha detectado. Sin embargo, los valores de las cuotas varían en el tiempo, por lo que, si el usuario no es lo suficientemente rápido, es posible que las cuotas bajen y ya no exista valor en esa apuesta. Esto hace necesaria una gran dedicación por parte del usuario, que debe ser veloz a la hora de realizar las apuestas para que la cuota no cambie.

Por último, es importante mencionar que los resultados obtenidos son meramente teóricos. Las apuestas contenidas en el dataset no han sido realizadas por mí en la vida real, por lo que las posibles limitaciones mencionadas anteriormente son producto de especulaciones e información extraída de vídeos y artículos consultados para este trabajo, y no de experiencias personales.

## 6.3. Líneas futuras

A lo largo de este estudio, han ido surgiendo ideas para la mejora y ampliación del trabajo. En primer lugar, una mejora sustancial que aportaría gran valor a este trabajo sería el desarrollo de un software que realice las apuestas de forma automática. Esto, en el caso de la casa de apuestas de Winamax, se antoja complicado, pues no dispone de una API pública. Sin embargo, hay algunas casas que proporcionan este servicio, lo cual facilita mucho el desarrollo de esta herramienta. Esto tendría enormes beneficios, reduciendo la carga de trabajo para el usuario y evitando sufrir cambios en las cuotas que provoquen una pérdida de valor.

Otra posible ampliación reside en incluir *sure bets* en el estudio. Estas apuestas garantizan beneficios inmediatos, y podría ser de interés comparar el rendimiento de *value bets* y *sure bets* a corto y largo plazo, y aplicar distintas estrategias para evaluar los pros y contras de cada tipo de apuestas.

Por último, para tener una mejor visión de las métricas calculadas, sería de utilidad crear una interfaz estilo *dashboard* que muestre el estado de cada una de las estrategias y permita al usuario establecer filtros y modificar parámetros, como puede ser el lapso de tiempo o el tipo de apuestas que se visualizan. Herramientas como Power BI de Microsoft permiten realizar este tipo de visualizaciones y conectarlas en tiempo real con fuentes de datos como archivos Excel.

# Bibliografía

- [1] Betburger. <https://www.betburger.com/>.
- [2] Andrew Grant, Anastasios Oikonomidis, Alistair C. Bruce and Johnnie E. V. Johnson. New entry, strategic diversity and efficiency in soccer betting markets: The creation and suppression of arbitrage opportunities (2018). [https://irep.ntu.ac.uk/id/eprint/35180/1/12737\\_Bruce..pdf](https://irep.ntu.ac.uk/id/eprint/35180/1/12737_Bruce..pdf).
- [3] Carlos Gómez-González and Julio del Corral. The betting market over time: Overround and surebets in european football (2018).
- [4] casasdeapuestas360. Opiniones bet365. <https://www.casasdeapuestas360.es/>.
- [5] Constantinou, A.C.; Fenton, N.E. Profiting from arbitrage and odds biases of the european football gambling market. the journal of gambling business and economics (2013). <http://constantinou.info/downloads/papers/evidenceofinefficiency.pdf>.
- [6] Data Bridge Market Research. Global sports betting market. <https://www.databridgemarketresearch.com/es/reports/global-sports-betting-market>.
- [7] Egon Franck, Erwin Verbeek, Stephan Nuesch. Inter-market arbitrage in betting (2013). [https://www.wiwi.uni-muenster.de/uf/sites/uf/files/PublikationenNuuesch/2013franck\\_verbeek\\_nuesch\\_2013.pdf](https://www.wiwi.uni-muenster.de/uf/sites/uf/files/PublikationenNuuesch/2013franck_verbeek_nuesch_2013.pdf).
- [8] Egon Franck, Erwin Verbeek, Stephan Nuesch. Prediction accuracy of different market structures — bookmakers versus a betting exchange (2010). [https://www.sciencedirect.com/science/article/abs/pii/S0169207010000105?fr=RR-2&ref=pdf\\_download&rr=91e8eed238cdcc02](https://www.sciencedirect.com/science/article/abs/pii/S0169207010000105?fr=RR-2&ref=pdf_download&rr=91e8eed238cdcc02).
- [9] Giovanni Angelini, Vincenzo Candila, Luca De Angelis. Weighted elo rating for tennis match predictions (2022). <https://www.sciencedirect.com/science/article/abs/pii/S0377221721003234>.
- [10] Hubacek, O. Exploiting betting market inefficiencies with machine learning, master's thesis, czech technical university, faculty of electrical engineering (2017).

- [11] James Surowiecki. The wisdom of crowds (2004).
- [12] John L Kelly. A new interpretation of information rate (1956).
- [13] Kollár, A. Betting models using ai: A review on ann, svm, and markov chain (2020). [https://mpra.ub.uni-muenchen.de/106821/1/MPRA\\_paper\\_106821.pdf](https://mpra.ub.uni-muenchen.de/106821/1/MPRA_paper_106821.pdf).
- [14] L. Breiman et al. Optimal gambling systems for favorable games, 1961.
- [15] Lawrence Clegg, John Cartlidge. Not feeling the buzz: Correction study of mispricing and inefficiency in online sportsbooks (2024). <https://arxiv.org/abs/2306.01740v4>.
- [16] H. Markowitz. Portfolio selection (1952).
- [17] Matej Uhrín, Gustav Šourek, Ondřej Hubáček, Filip Železný. Optimal sports betting strategies in practice: an experimental review (2021). <https://arxiv.org/abs/2107.08827>.
- [18] Michal Sipko. Machine learning for the prediction of professional tennis matches (2015). <http://www.doc.ic.ac.uk/teaching/distinguished-projects/2015/m.sipko.pdf>.
- [19] Mladen Josivljevic, Daniel Hlubinka. Estimators of probability density function level-sets and their volumes (2024). <https://dspace.cuni.cz/bitstream/handle/20.500.11956/192913/120485881.pdf>.
- [20] NIKOLAOS VLASTAKIS, GEORGE DOTSI AND RAPHAEL N. MARKELOS. How efficient is the european football betting market? evidence from arbitrage and trading strategies (2009). <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/for.1085>.
- [21] Oktsa Dwika Rahmashari, Wuttichai Srisodaphol. Advanced outlier detection methods for enhancing beta regression robustness (2025). <https://www.sciencedirect.com/science/article/pii/S277266222500013X>.
- [22] Philip Ramirez, J. James Reade, Carl Singleton. Betting on a buzz: Mispricing and inefficiency in online sportsbooks (2023). [https://www.sciencedirect.com/science/article/pii/S0169207022001091?ref=pdf\\_download&fr=RR-2&r=91024ac8b9c7f777](https://www.sciencedirect.com/science/article/pii/S0169207022001091?ref=pdf_download&fr=RR-2&r=91024ac8b9c7f777).
- [23] Ping-Feng Pai, Lan-Hung Chang, Liao Kuo-Ping Lin. Analyzing basketball games by a support vector machines with decision tree model (2016). <https://link.springer.com/article/10.1007/s00521-016-2321-9>.
- [24] Rory P. Bunker, Fadi Thabtah. A machine learning framework for sport result prediction (2019). <https://www.sciencedirect.com/science/article/pii/S2210832717301485#b0030>.

- [25] Thomas P. Hayes. How not to win a million dollars: A counterexample to a conjecture of l. breiman (2011). <https://arxiv.org/abs/1112.0829>.