

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR**

**FACULTAD DE CIENCIAS EXACTAS NATURALES Y AMBIENTALES**

**CARRERA DE CIENCIA DE DATOS**

**MONOGRAFÍA**

**Nombre del estudiante:** Carlos Sebastián Enríquez Sánchez

**Nombre del tutor:** Kevin Ricardo Rojas Satian

**Título de la monografía:** Revisión bibliográfica de modelos predictivos para partidos de fútbol en la Premier League

**Área en la que se inscribe la investigación:** Modelo estadístico

Quito, 2026

# 1. DEDICATORIA

*A quienes estuvieron presentes en cada etapa de este proceso, cuyo apoyo constante hizo posible la culminación de este trabajo. A mi madre, mis tías, mi abuelita y mi padre, por su paciencia, por su confianza y por recordarme siempre, en los momentos más difíciles, que el esfuerzo nunca es en vano. A mis amigos Sebas, Francisco, Juan, Gaby y Nicolas, con quienes compartir el camino hizo que el aprendizaje fuera más significativo. Por último a Carlita que ha sido un gran apoyo y ayuda en mi proceso universitario.*

*Esta monografía es para ellos.*

## 2. AGRADECIMIENTOS

Agradezco a la Pontificia Universidad Católica del Ecuador y a la Carrera de Ciencia de Datos por la formación recibida y por brindarme las herramientas conceptuales y metodológicas que hicieron posible este estudio. A los docentes que a lo largo de la carrera compartieron su conocimiento con generosidad.

Además, quiero expresar mi sincero agradecimiento a mi tutor, Kevin Ricardo Rojas Satian, por su guía y a la Dra. Verónica Estefanía Beltrán García, por su orientación al principio de esta investigación.

A mis compañeros por hacer que el aprendizaje en las aulas haya sido más placentero y entretenido, por compartir sus conocimientos conmigo y haber estado abiertos a siempre ayudar a los demás.

A mi familia por su apoyo constante en momentos difíciles y siempre haberme ayudado cuando lo necesitaba y así hacerme crecer academicamente y profesionalmente, gracias a sus consejos y experiencia.

## 3. TABLA DE CONTENIDO


*1. DEDICATORIA* _____ *1*

*2. AGRADECIMIENTOS* _____ *2*

*3. TABLA DE CONTENIDO* _____ *3*

*4. ABSTRACT* _____ *4*

*5. RESUMEN* _____ *5*

*6. INTRODUCCIÓN* _____ *6*

*7. OBJETIVOS* _____ *10*

*8. FUNDAMENTO TEÓRICO* _____ *11*

*9. METODOLOGÍA* _____ *21*

*10. RESULTADOS Y DISCUSIÓN* _____ *27*

*11. CONCLUSIONES* _____ *42*

*12. REFERENCIAS* _____ *45*

## 4. ABSTRACT

The prediction of match outcomes in the English Premier League has evolved substantially over the past three decades, driven by advances in statistical methods, machine learning, and deep learning. This monograph presents a systematic literature review of predictive models applied to the Premier League, covering the period from 1995 to 2025, with the aim of identifying and comparing methodological approaches, relevant input variables, evaluation metrics, and key limitations reported in the scientific literature. Based on a review of sources indexed in Scopus, IEEE Xplore, SpringerLink, Web of Science, and arXiv, studies applying classical statistical models —including Poisson models and Elo-type rating systems—, machine learning algorithms —Random Forest, Gradient Boosting, Support Vector Machines—, and deep learning architectures —deep neural networks, LSTM, Transformers, and dynamic Bayesian models— were identified and analyzed. The results indicate that, while modern methods have improved predictive accuracy compared to traditional approaches, significant challenges persist in predicting draws, cross-season replicability, and the lack of methodological standardization. Ensemble methods such as XGBoost and Random Forest show the highest accuracy rates (60–68%) for classifying match outcomes, while Bayesian and Poisson models demonstrate superior probabilistic calibration. The integration of advanced metrics such as expected goals (xG) consistently improves predictive performance across all model families. The monograph concludes with a comparative synthesis of the reviewed models and the identification of research gaps that inform future directions in sports analytics.

Keywords: Deep learning, expected goals, machine learning, statistical models, Premier League.

## 5. RESUMEN

La predicción de resultados deportivos en la Premier League ha evolucionado considerablemente durante las últimas dos décadas, impulsada por el avance de métodos estadísticos, de machine learning y de deep learning. La presente monografía realiza una revisión bibliográfica sistemática de los modelos predictivos aplicados a la Premier League, con el objetivo de identificar y comparar sus enfoques metodológicos, las variables más relevantes, las métricas de evaluación empleadas y las principales limitaciones reportadas en la literatura científica de (1995–2025). A partir de la revisión de fuentes indexadas en Scopus, IEEE Xplore, SpringerLink, Web of Science y arXiv, se identificaron y analizaron estudios que aplican modelos estadísticos clásicos —como los modelos de Poisson y sistemas de rating tipo Elo—, algoritmos de machine learning —Random Forest, Gradient Boosting, Support Vector Machines— y arquitecturas de deep learning —redes neuronales profundas, LSTM, transformers y modelos bayesianos dinámicos—. Los resultados muestran que, si bien los métodos modernos han mejorado la precisión predictiva en comparación con los enfoques tradicionales, persisten desafíos significativos en la predicción de empates, la replicabilidad entre temporadas y la falta de estandarización metodológica. La monografía concluye con una síntesis comparativa de los modelos y una identificación de vacíos de investigación que orientan futuras líneas de trabajo en analítica deportiva.

**Palabras clave:** Deep learning, machine learning, modelos estadísticos, predicción de resultados, Premier League.

## 6. INTRODUCCIÓN

La predicción de resultados de partidos de fútbol, especialmente en ligas altamente competitivas como la Premier League, ha recibido una creciente atención en la literatura científica durante la última década. El avance de métodos estadísticos, machine learning y deep learning ha permitido desarrollar modelos capaces de analizar grandes volúmenes de datos deportivos, incluyendo métricas avanzadas como goles esperados (xG), posesiones, córners y rendimiento histórico de los equipos.

Si bien estos enfoques han mostrado avances significativos, actualmente existe un crecimiento acelerado de nuevas metodologías. Los modelos tradicionales como Poisson o regresión logística siguen siendo ampliamente utilizados, pero métodos modernos como Random Forest, Gradient Boosting, redes neuronales o modelos bayesianos dinámicos buscan mejorar la precisión. A pesar de esta diversidad metodológica, no existe consenso sobre cuáles técnicas presentan mejor desempeño para la predicción de partidos en la Premier League ni sobre qué variables influyen más en la predicción de resultados.

Se realizó una búsqueda en bases de datos como Scopus para la revisión bibliográfica de artículos científicos donde se obtuvo una amplia literatura y se evidencian algunos síntomas claros: que los resultados son poco replicables entre temporadas (Razali et al., 2022); que ciertos modelos fallan sistemáticamente al predecir empates, incluso aquellos basados en deep learning (Araujo-Ahon et al., 2023); y que existe una ausencia de un marco comparativo estandarizado, lo que genera dificultades para evaluar rigurosamente la efectividad de cada modelo (Feng, 2024).

En el caso del fútbol profesional, y particularmente en la Premier League, la predicción de resultados se ha convertido en un campo de investigación relevante tanto para la academia como para la industria deportiva. La Premier League se destaca por su competitividad, la

variabilidad táctica entre equipos y la gran cantidad de datos disponibles, lo que la convierte en un escenario ideal para la aplicación de modelos estadísticos y computacionales avanzados.

Diversos estudios han abordado la predicción de resultados a partir de metodologías tradicionales como los modelos Poisson, la regresión logística y los sistemas de rating, junto con enfoques más recientes basados en machine learning y deep learning. Estos incluyen árboles de decisión, Random Forest, Gradient Boosting, redes neuronales profundas y modelos bayesianos dinámicos. Si bien cada enfoque ha mostrado fortalezas particulares, los resultados son heterogéneos y, en muchos casos, no replicables entre temporadas o ligas. La literatura evidencia inconsistencias en la selección de variables, diferencias metodológicas significativas y un uso no estandarizado de métricas de evaluación, lo que dificulta establecer conclusiones sólidas sobre la efectividad de cada técnica. Como señalan Loukas et al. (2024), una debilidad frecuente de los modelos es la fuerte subestimación de las puntuaciones que terminan sin goles, lo cual refleja claramente los problemas que se presentan para una estimación real.

Este escenario demuestra la necesidad de una revisión estructurada que permita sintetizar los avances recientes, comparar restricciones y fortalezas de los modelos, y analizar cómo se han utilizado diferentes fuentes de datos para mejorar la precisión predictiva. Asimismo, la constante evolución de las métricas avanzadas del fútbol —como expected goals (xG), expected threat (xT) y datos de tracking— ha ampliado las capacidades de modelación, pero también ha generado nuevos desafíos metodológicos que deben ser analizados en profundidad. Yoel & Isa (2019) destacan que muchos estudios tienen limitaciones en sus pronósticos, como el uso de la regresión logística, que sólo proporciona dos resultados posibles: 'victoria en casa' o 'victoria fuera', mientras que un partido de fútbol puede tener tres posibles resultados: 'victoria en casa', 'victoria fuera' o 'empate', en los cuales se usa la regresión multinomial. Estos mismos autores señalan que algunos algoritmos como Random

Forest alcanzan precisiones de predicción de apenas el 63,4%, lo cual evidencia las restricciones de los enfoques actuales.

Frente a la ausencia de revisiones actualizadas que sinteticen el estado del arte para modelos predictivos en la Premier League se evidencia la necesidad de un análisis de cuáles modelos muestran mayor robustez en situaciones reales temporada a temporada.

En este contexto, la presente monografía tiene como objetivo revisar el estado del arte de los modelos predictivos aplicados a la Premier League, explorando sus enfoques metodológicos, las variables más utilizadas, las métricas de desempeño reportadas y las principales limitaciones identificadas en la literatura. Con ello, se busca ofrecer una visión clara y actualizada del panorama científico comparando los modelos utilizados, identificando sus ventajas y desventajas, y evaluando qué modelo llega a obtener un mejor rendimiento.

La predicción de resultados en la Premier League ha sido un tema analizado durante varias décadas; sin embargo, en los últimos años ha tomado una nueva relevancia tecnológica, que radica en el uso de métodos avanzados para apoyar la toma de decisiones, optimizar el desempeño táctico y mejorar modelos aplicados en sectores como apuestas reguladas, periodismo deportivo y analítica de rendimiento.

Desde el punto de vista académico, esta revisión es pertinente porque sintetiza investigaciones que aplican técnicas estadísticas y modelos de machine learning. Además, expone fortalezas y limitaciones de cada enfoque, contribuyendo a la comprensión teórica y práctica de la predicción deportiva. Una revisión sistemática permitirá clarificar qué modelos tienen mayor impacto predictivo.

Así también, desde un enfoque social se quiere promover el uso responsable y transparente de modelos predictivos. La falta de estandarización metodológica puede inducir

interpretaciones erróneas o decisiones basadas en modelos poco confiables. Un análisis riguroso favorece la transparencia y fomenta prácticas éticas de las predicciones deportivas.

Todo lo anterior conduce a la pregunta de investigación: ¿Qué modelos predictivos se han utilizado para estimar los resultados de partidos de la Premier League, y comparar cómo han evolucionado sus enfoques, variables, métricas de evaluación y conclusiones en la literatura reciente?

# 7. OBJETIVOS

## 7.1 Objetivo general

Analizar el estado del arte de los modelos predictivos utilizados para la estimación de resultados de partidos de la Premier League, mediante una revisión de la literatura científica reciente, con el fin de identificar enfoques metodológicos, variables relevantes, métricas de evaluación y principales limitaciones presentes en los estudios actuales.

## 7.2 Objetivos específicos

- Identificar los principales modelos estadísticos, de machine learning y deep learning aplicados a la predicción de partidos de la Premier League en la literatura científica.
- Analizar las características del juego, las fuentes de datos y las variables utilizadas por los distintos enfoques predictivos.
- Comparar las métricas de evaluación empleadas para medir el desempeño de los modelos predictivos en los estudios revisados.
- Examinar las fortalezas y debilidades de los diferentes modelos, considerando aspectos como precisión, interpretabilidad y capacidad de generalización entre temporadas.
- Evaluar cuál modelo presenta un mejor desempeño para predecir los resultados de los partidos en la liga inglesa.

# 8. FUNDAMENTO TEÓRICO

La predicción de resultados deportivos se fundamenta en teorías estadísticas, probabilísticas y computacionales que buscan modelar el comportamiento de sistemas complejos e inciertos. El fútbol, al ser un deporte de baja puntuación y alta variabilidad, presenta desafíos particulares para la modelización predictiva, lo que ha motivado el desarrollo de diversos enfoques teóricos aplicados a ligas profesionales como la Premier League. La complejidad del fenómeno deportivo radica en que los resultados dependen de factores observables como: estadísticas del partido, forma reciente, composición del equipo, así como no observables: motivación, presión psicológica, decisiones arbitrales. Esta combinación de factores deterministas y estocásticos hace que ningún modelo pueda alcanzar una precisión perfecta, aunque cada generación de técnicas ha logrado mejoras incrementales respecto a la anterior.

## 8.1 Modelos estadísticos clásicos en la predicción de fútbol

Los modelos estadísticos han sido ampliamente utilizados en la predicción de partidos de fútbol debido a su simplicidad e interpretabilidad. Entre los más comunes se encuentran los modelos Poisson y bivariados, los cuales asumen que el número de goles anotados por cada equipo sigue una distribución de Poisson independiente o correlacionada. Estos modelos permiten estimar probabilidades de victoria, empate o derrota, y han sido utilizados como línea base en múltiples estudios.

El modelo de Poisson independiente, introducido por Maher (1982) y refinado por Dixon & Coles (1997), estima la tasa de goles de cada equipo en función de su capacidad ofensiva y defensiva relativa. Pese a su antigüedad, ha demostrado ser sorprendentemente competitivo. Nguyen (2021) aplica esta metodología específicamente a la Premier League y muestra que los modelos de Poisson correctamente calibrados pueden predecir el marcador

final con una precisión aceptable, especialmente en partidos entre equipos de fuerza muy distinta. Sin embargo, su tendencia a subestimar la probabilidad de empates, particularmente en encuentros entre equipos de nivel similar, es una de las debilidades más documentadas en la literatura.

Dixon & Coles (1997) propusieron una corrección de dependencia para los resultados de 0-0 y 1-0, introduciendo un parámetro de correlación que mejoró significativamente las predicciones de partidos con pocos goles. Esta corrección se ha convertido en una referencia metodológica habitual en la literatura posterior.

Asimismo, los sistemas de rating como Elo se basan en la actualización dinámica de la fuerza relativa de los equipos a lo largo del tiempo, considerando los resultados históricos. Hvattum & Arntzen (2010) demostraron que las variantes Elo adaptadas al fútbol superan a los modelos de mercado de apuestas en la predicción de la dirección del resultado. Constantinou & Fenton (2013) propusieron una alternativa, denominada pi-rating, que en lugar de considerar únicamente el resultado del partido incorpora la magnitud de la diferencia de goles entre los equipos, logrando una capacidad predictiva superior a las variantes Elo tradicionales y una rentabilidad sostenida frente a las cuotas de las casas de apuestas durante cinco temporadas de la Premier League. Aunque estos métodos presentan limitaciones para capturar factores contextuales complejos, continúan siendo relevantes por su robustez y bajo costo computacional.

La regresión logística ordinal ha sido otro enfoque clásico ampliamente utilizado. Al modelar directamente las tres posibles categorías de resultado —victoria local, empate, victoria visitante— como una variable ordinal, este método supera una limitación del modelo logístico binario. Goddard (2005) presentó uno de los trabajos más influyentes en este ámbito, donde una regresión logística ordenada con variables de rendimiento reciente superó los pronósticos del mercado de apuestas en varias ligas europeas.

Estudios como Nguyen (2021) y Feng (2024) muestran que los modelos Poisson siguen siendo competitivos, especialmente para la predicción de goles, aunque presentan dificultades con empates y eventos poco frecuentes. Loukas et al. (2024) confirman esta tendencia en su análisis de la Premier League, donde el modelo de regresión de Poisson muestra un desempeño razonable pero sistemáticamente inferior a los enfoques de machine learning cuando se incorporan variables contextuales avanzadas.

### 8.2 Machine learning aplicado a la predicción de partidos

El aprendizaje automático ha permitido superar algunas limitaciones de los modelos estadísticos tradicionales al incorporar un mayor número de variables y relaciones no lineales entre ellas. Como documentan las revisiones de Bunker & Thabtah (2019) y, más recientemente, Bunker & Susnjak (2022) sobre la aplicación de machine learning a la predicción de resultados en deportes de equipo, algoritmos como árboles de decisión, Random Forest, Support Vector Machines y Gradient Boosting han sido aplicados con éxito para clasificar resultados de partidos o predecir marcadores exactos.

Los árboles de decisión fueron uno de los primeros algoritmos de machine learning aplicados a la predicción deportiva. Aunque su rendimiento individual es moderado, su interpretabilidad los hace útiles para identificar las variables más influyentes en el resultado. Yoel & Isa (2019) reportaron que un clasificador basado en árbol de decisión CART alcanzó una precisión del 63.4% en la predicción de resultados de la Premier League, lo que presentó una mejora marginal sobre el modelo logístico binario aplicado al mismo conjunto de datos.

Random Forest, una extensión de los árboles de decisión basada en la combinación de múltiples árboles entrenados con submuestras aleatorias, ha mostrado resultados más robustos. Choi et al. (2023) reportan que modelos de ensable como Random Forest y Gradient

Boosting alcanzan accuracies del 60–70% en la predicción de resultados de la Premier League, representando mejoras frente a modelos estadísticos tradicionales. Sin embargo, su rendimiento varía según selección de variables y tratamiento del desbalance de clases, especialmente en el caso de los empates, que históricamente representa entre el 25% y el 30% de los partidos en la liga inglesa.

El algoritmo AdaBoost, un método de boosting que ajusta iterativamente los pesos de las observaciones mal clasificadas en cada iteración —a diferencia de Gradient Boosting, que minimiza una función de pérdida ajustando los residuos mediante descenso de gradiente—, fue empleado por Feng (2024) específicamente para la Premier League, obteniendo una precisión del 57.4% con un conjunto de características que incluía estadísticas de forma reciente y variables contextuales. Aunque este resultado parece modesto, el autor argumenta que supera significativamente a la predicción basada únicamente en la posición en la tabla clasificatoria o en el mercado de apuestas.

Wang & Zhang (2024) desarrollaron un sistema de predicción basado en machine learning que integra múltiples algoritmos en un esquema de votación ponderada, logrando mejoras de hasta un 4% en accuracy respecto a los modelos individuales. Su análisis muestra que la combinación de Random Forest, XGBoost y Regresión Logística ofrece un equilibrio adecuado entre precisión y capacidad de generalización.

Las Support Vector Machines (SVM) han sido empleadas por varios autores como clasificadores de referencia. Su capacidad para manejar espacios de características de alta dimensión las hace atractivas cuando se incorporan múltiples variables de partido. Sin embargo, su desempeño en la predicción de resultados de fútbol suele ser comparable o ligeramente inferior al de los métodos de ensamble como Random Forest o Gradient Boosting, según reportan Araujo-Ahon et al. (2023) en su revisión de técnicas de minería de datos aplicadas al fútbol.

Estos enfoques se caracterizan por su capacidad de manejar datos heterogéneos y por ofrecer mejoras en la precisión predictiva cuando se dispone de conjuntos de datos suficientemente grandes. No obstante, su desempeño depende en gran medida de la calidad de los datos, la selección de variables y el tratamiento del desbalance de clases, especialmente en el caso de los empates.

Un aspecto crítico en la aplicación de machine learning a la predicción de partidos es la selección de variables. Las características más frecuentemente utilizadas incluyen: forma reciente de los equipos (últimos 5 partidos), diferencia de goles acumulada, rendimiento local y visitante, posición en la tabla, estadísticas de partidos previos entre los mismos equipos (head-to-head), y métricas avanzadas como expected goals (xG). La incorporación de métricas avanzadas ha mostrado mejoras consistentes en el desempeño predictivo, aunque también introduce desafíos relacionados con la disponibilidad y calidad de los datos.

### 8.3 Deep learning y modelos avanzados

Los enfoques basados en deep learning han ganado protagonismo en los últimos años debido a su capacidad para aprender representaciones complejas a partir de grandes volúmenes de datos. Redes neuronales profundas, redes recurrentes (RNN, LSTM) y arquitecturas basadas en transformers han sido empleadas para capturar dependencias temporales y patrones secuenciales en el rendimiento de los equipos.

Las redes neuronales artificiales (ANN) de múltiples capas fueron uno de los primeros enfoques de deep learning aplicados a la predicción deportiva. Tsakonas et al. (2002) presentaron uno de los primeros trabajos que utilizaba redes neuronales para la predicción de resultados en ligas europeas, estableciendo un precedente metodológico que sería seguido por numerosas investigaciones posteriores. La capacidad de las ANN para capturar relaciones

no lineales entre variables las hace teóricamente superiores a los modelos estadísticos lineales.

Las redes neuronales recurrentes, en particular las Long Short-Term Memory (LSTM) han sido aplicadas para modelar la evolución temporal del rendimiento de los equipos. A diferencia de los modelos estáticos, las LSTM pueden capturar cambios en el rendimiento a lo largo de la temporada, como el impacto de lesiones, cambios de entrenador o rachas de resultados. Razali et al. (2022) en su análisis demuestran que las redes neuronales profundas, transformers y modelos bayesianos dinámicos han ganado popularidad, y muestran que los modelos basados en redes neuronales pueden superar técnicas tradicionales cuando se entrenan con datos ricos (xG, secuencias temporales, tracking).

Los transformers, arquitectura originalmente desarrollada para el procesamiento de lenguaje natural, han comenzado a ser explorados en el contexto de la predicción deportiva. Su mecanismo de atención permite ponderar de forma diferenciada la relevancia de distintos partidos históricos en la predicción de un resultado futuro. Aunque los estudios con transformers aplicados específicamente a la Premier League son todavía escasos, los resultados preliminares son prometedores en cuanto a la captura de dependencias temporales de largo plazo.

Adicionalmente, los modelos bayesianos dinámicos representan otro enfoque avanzado que ha mostrado resultados relevantes. A diferencia de los modelos estáticos, estos actualizan continuamente los parámetros que representan la fuerza de los equipos a medida que avanza la temporada, adaptándose a cambios en alineaciones, lesiones y rendimiento reciente. Macrì-Demartino et al. (2025) introducen modelos dinámicos de tiempo discreto ponderados que actualizan la fuerza de los equipos a lo largo del tiempo, mejorando significativamente las predicciones en temporadas extensas. Su enfoque combina la flexibilidad de los modelos bayesianos con la capacidad de incorporar incertidumbre en los

parámetros, lo que permite obtener intervalos de credibilidad para las predicciones, algo que los enfoques de machine learning estándar no ofrecen.

Karlis & Ntzoufras (2003) propusieron uno de los modelos bayesianos bivariados más influyentes para la predicción de fútbol, utilizando una distribución de Poisson bivariada que modela explícitamente la correlación entre los goles anotados por cada equipo. Este trabajo ha servido como referencia metodológica para numerosas extensiones bayesianas posteriores, incluyendo los modelos dinámicos de Macrì-Demartino et al. (2025).

Aunque estos métodos han mostrado mejoras en el desempeño predictivo, su interpretación y generalización siguen siendo un desafío relevante. Los modelos de deep learning, en particular, suelen comportarse como cajas negras, lo que dificulta su adopción en contextos donde la interpretabilidad es importante, como el análisis táctico o la comunicación de resultados a audiencias no técnicas.

Otra limitación documentada de los modelos de deep learning en este contexto es su requerimiento de grandes volúmenes de datos. La Premier League, con sus aproximadamente 380 partidos por temporada, ofrece un conjunto de datos limitado para el entrenamiento de arquitecturas complejas. Este problema se ha abordado mediante la transferencia de aprendizaje desde otras ligas, el uso de datos de partidos de pretemporada, y la incorporación de datos de nivel individual de jugadores.

### 8.4 Variables y fuentes de datos en la predicción de fútbol

La selección de variables es uno de los aspectos más críticos en el diseño de modelos predictivos para fútbol. La literatura identifica diversas categorías de características que han demostrado valor predictivo en distintos contextos.

Las estadísticas de resultado histórico —victorias, derrotas, empates, goles a favor y en contra— constituyen la base de la mayoría de los modelos. La forma reciente, generalmente medida como el rendimiento en los últimos 3 a 6 partidos, ha mostrado ser un predictor más relevante que el rendimiento acumulado de toda la temporada, especialmente a segunda mitad de la competición.

Las métricas avanzadas de rendimiento han ganado protagonismo en la última década. El expected goals (xG) mide la calidad de las ocasiones de gol creadas y concedidas, ofreciendo una medida más estable del rendimiento real de un equipo que los goles marcados. El expected threat (xT) y el PPDA (Passes Per Defensive Action) son métricas que cuantifican la presión defensiva y la progresión del balón, respectivamente. Varios estudios han demostrado que la incorporación de xG mejora significativamente la precisión de los modelos predictivos, especialmente para partidos donde el resultado no refleja el rendimiento real de los equipos.

Las variables contextuales —como la condición de local o visitante, distancia recorrida por el equipo visitante, el número de días de descanso desde el último partido, y la presencia de competiciones europeas— también han sido incorporadas en algunos modelos con resultados positivos. La ventaja de local en la Premier League, aunque ha disminuido en los últimos años, sigue siendo un factor predictivo relevante.

Los datos de tracking —posición y velocidad de los jugadores en tiempo real— representan la frontera más reciente en la predicción deportiva. Aunque su disponibilidad es limitada para investigadores externos, las empresas con acceso a estos datos han demostrado que permiten construir modelos con mayor precisión predictiva. La integración de datos de tracking con modelos de deep learning representa una de las líneas de investigación más prometedoras en el campo.

### 8.5 Métricas de evaluación y desafíos metodológicos

La evaluación de modelos predictivos en fútbol se realiza mediante métricas como accuracy, F1-score, log-loss, MAE y Brier score. La falta de estandarización en el uso de estas métricas dificulta la comparación directa entre estudios. Además, factores externos como lesiones, decisiones arbitrales y variaciones tácticas introducen un nivel de incertidumbre que limita la precisión máxima alcanzable por los modelos.

El accuracy es la métrica más comúnmente reportada en la literatura, aunque presenta limitaciones importantes. Dado que los resultados de los partidos desbalanceados entre clases (victoria local ~45%, empate ~25%, victoria visitante ~30%), un clasificador que siempre prediga victoria local alcanzaría un accuracy del 45% sin ningún poder predictivo real.

El F1-score, que combina precisión y recall mediante su media armónica, es más informativo que el accuracy en contextos de clases desbalanceadas. Sin embargo, su interpretación es menos intuitiva y varía según la clase que se considere positiva, lo que añade ambigüedad a las comparaciones entre estudios.

El log-loss (o entropía cruzada) penaliza las predicciones probabilísticas incorrectas de forma proporcional a su confianza, lo que la convierte en una métrica más adecuada cuando el modelo produce probabilidades en lugar de etiquetas de clase. El Brier score, que mide el error cuadrático medio entre las probabilidades predichas y los resultados observados, ofrece una alternativa similar con propiedades estadísticas bien conocidas.

Un desafío importante en este campo es la heterogeneidad metodológica. No existen estándares para las métricas usadas ni para la selección de variables. Esto complica la comparación objetiva entre estudios. Además, factores específicos de la Premier League —

alta competitividad, volatilidad, lesiones, rotación de jugadores— afectan significativamente los resultados, lo que hace que muchos modelos no sean replicables entre temporadas.

# 9. METODOLOGÍA

La presente investigación se desarrolló bajo un enfoque predominantemente cualitativo con elementos de síntesis cuantitativa descriptiva, de tipo descriptivo y analítico, mediante una revisión sistemática de la literatura científica. Su punto de partida fue la constatación, descrita en la introducción, de que la literatura sobre predicción de partidos en la Premier League carece de un marco comparativo estandarizado que permita evaluar y contrastar los distintos modelos propuestos. Para responder a esta brecha, se optó por un diseño no experimental que no contempla la implementación de modelos propios, sino el análisis crítico y sistemático de investigaciones previas, con el fin de sintetizar el estado del arte e identificar las tendencias metodológicas, las variables más relevantes y las limitaciones recurrentes en el campo.

## 9.1 Tipo y enfoque de investigación

La investigación fue de tipo documental, con un enfoque predominantemente cualitativo, dado que se centró en la revisión, análisis e interpretación de estudios científicos publicados. Al mismo tiempo, incorporó elementos descriptivos —al sistematizar los modelos, variables y métricas empleadas en cada estudio— y analíticos, al comparar resultados entre enfoques, identificar patrones de desempeño y señalar las limitaciones recurrentes en la literatura. Esta combinación de dimensiones permitió transitar desde la descripción de los modelos individuales hacia una interpretación crítica del campo en su conjunto.

Para estructurar el proceso de revisión, se adoptó el protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) como guía metodológica. La elección de este protocolo responde a su amplio reconocimiento en la literatura científica como estándar para la planificación, ejecución y reporte de revisiones sistemáticas (Page et al., 2021), y a su flexibilidad para adaptarse a campos interdisciplinarios como la analítica

deportiva, donde coexisten metodologías estadísticas, computacionales y del dominio del deporte. Sus etapas —identificación, cribado, elegibilidad e inclusión— orientaron de forma progresiva la depuración de los estudios analizado.

### 9.2 Fuentes de información

Las fuentes de información estuvieron constituidas por artículos científicos arbitrados, publicaciones en revistas indexadas y actas de conferencias especializadas en ciencia de datos, estadística y analítica deportiva. Se priorizaron bases de datos de alto impacto y cobertura amplia: Scopus, IEEE Xplore, SpringerLink, Web of Science y arXiv. La inclusión de arXiv, si bien alberga preprints sin revisión formal, se justificó por la rápida evolución del campo y la relevancia de trabajos recientes aún en proceso de publicación en revistas indexadas, como el modelo bayesiano dinámico de Macrì-Demartino et al. (2025). En estos casos, se verificó la solidez metodológica del estudio antes de su inclusión.

Adicionalmente, se consultaron repositorios académicos institucionales y revistas especializadas en analítica deportiva y machine learning aplicado, como el Journal of Sports Analytics y el International Journal of Forecasting, cuya trayectoria en la publicación de estudios de predicción deportiva los convierte en referencias consolidadas dentro del campo.

### 9.3 Estrategia de búsqueda

La búsqueda bibliográfica se realizó utilizando combinaciones de términos en inglés y español en las bases de datos señaladas. Los descriptores se diseñaron para cubrir tanto los enfoques metodológicos (modelos estadísticos, machine learning, deep learning) como el dominio de aplicación (Premier League, predicción de resultados), y se combinaron mediante operadores booleanos (AND, OR) para maximizar la cobertura sin sacrificar la relevancia templática. Entre las principales combinaciones utilizadas se encuentran:

- "Premier League match prediction"
- "football outcome prediction"
- "Machine learning football"
- "Poisson models football"
- "Deep learning sports analytics"
- "Expected goals prediction"
- "Bayesian football model"
- "Soccer result forecasting"

Se aplicaron filtros por rango temporal (1995–2025), idioma (inglés y español) y tipo de publicación (artículos revisados por pares y conferencias científicas). El rango temporal se definió para capturar tanto los modelos fundacionales de la predicción estadística del fútbol —como el trabajo seminal de Dixon y Coles (1997)— como los desarrollos más recientes en deep learning y modelos bayésianos dinámicos.

### 9.4 Criterios de inclusión

- Estudios que analicen explícitamente partidos de la Premier League o utilicen datos de dicha liga.
- Investigaciones que empleen modelos estadísticos, machine learning o deep learning para la predicción de resultados.
- Publicaciones arbitradas en revistas o conferencias científicas.
- Estudios que reporten métricas de evaluación cuantitativas (accuracy, F1-score, log-loss, MAE, entre otras).
- Artículos publicados entre los años 1995 y 2025.

### 9.5 Criterios de exclusión

- Estudios que aborden otros deportes o ligas sin aplicación directa a la Premier League.
- Trabajos sin resultados empíricos o sin evaluación cuantitativa.
- Artículos duplicados o versiones preliminares sin revisión por pares.
- Publicaciones de carácter divulgativo, opiniones o análisis no técnicos.

### 9.6 Proceso de selección y análisis

El proceso de selección de estudios se desarrolló en tres fases progresivas, siguiendo la lógica del protocolo PRISMA:

- Revisión de títulos y resúmenes, para descartar estudios no relacionados con el tema.
- Revisión de texto completo, aplicando los criterios de inclusión y exclusión definidos. Tras la lectura de títulos y resúmenes se seleccionaron 87 artículos candidatos, de los cuales 21 superaron la revisión de texto completo al cumplir la totalidad de los criterios establecidos. Estos 21 estudios constituyen la base del análisis y se encuentran listados en su totalidad en la sección de referencias.
- Análisis comparativo, en el cual se extrajo información relevante de cada estudio seleccionado, incluyendo tipo de modelo, variables utilizadas, métricas de evaluación y principales conclusiones. Esta extracción se organizó en una matriz de análisis que sirvió como base para la construcción de las tablas comparativas que se presentan en la sección de resultados.

**Figura 1. Diagrama de flujo PRISMA del proceso de selección de estudios.**




*Figura 1. Diagrama de flujo PRISMA del proceso de selección de estudios. Fuente: Análisis de posibles papers en plataformas correspondientes. Elaboración propia*

La información recopilada fue organizada en tablas comparativas y categorías temáticas, lo que permitió identificar tendencias, fortalezas, debilidades y vacíos de investigación de manera sistemática. Este proceso de categorización resultó fundamental para garantizar que el análisis respondiera de forma explícita a cada uno de los objetivos específicos planteados al inicio de la investigación.

### 9.7 Síntesis de resultados

Los resultados fueron sintetizados de forma narrativa y comparativa, destacando los enfoques más utilizados en la literatura, los modelos con mejor desempeño reportado según distintas métricas y los principales desafíos metodológicos identificados. La síntesis se estructuró en función de las tres categorías metodológicas centrales —modelos estadísticos clásicos, machine learning y deep learning— ya descritas en el fundamento teórico, lo que permite establecer una correspondencia directa entre el marco conceptual, la metodología y los hallazgos presentados en la sección siguiente. De este modo, la metodología no opera de forma aislada, sino como puente explícito entre la pregunta de investigación planteada en la introducción y los resultados que se analizan a continuación.

# 10. RESULTADOS Y DISCUSIÓN

A partir de la revisión bibliográfica realizada, se identificaron y analizaron estudios que abarcan el período 1995–2025, abordando la predicción de resultados en la Premier League desde diversas perspectivas metodológicas. Los hallazgos se organizan a continuación en función de las categorías de modelos identificados, las variables utilizadas, las métricas reportadas y las limitaciones recurrentes.

## 10.1 Clasificación de los modelos predictivos identificados

La literatura revisada puede organizarse en tres grandes categorías metodológicas: modelos estadísticos clásicos, algoritmos de machine learning y enfoques de deep learning. Esta clasificación es coherente con la evolución histórica del campo, donde los enfoques más recientes han surgido como respuesta a las limitaciones de los anteriores modelos predictivos.

### *10.1.1 Modelos estadísticos clásicos*

Los modelos de Poisson y sus variantes representan la familia más estudiada en la predicción de fútbol. Nguyen (2021) aplica el modelo de Poisson estándar y sus extensiones a datos de la Premier League, reportando una capacidad predictiva razonable para partidos con resultados de alto puntaje, pero con dificultades persistentes en la predicción de partidos con 0-0 o empates de un gol. La correlación de Dixon y Coles (1997) mejora este aspecto al introducir un parámetro de correlación para las categorías de bajo puntaje.

Los sistemas de rating Elo adaptados al fútbol, estudiados por Hvattum & Arntzen (2010), muestran que la fuerza relativa de los equipos, actualizada dinámicamente, es un predictor robusto del resultado. Estos modelos tienen la ventaja de requerir únicamente datos de resultados históricos, sin necesidad de estadísticas de partido detalladas.

La regresión logística ordinal, empleada por Goddard (2005) y posteriormente por múltiples autores, ofrece una alternativa estadísticamente apropiada para el problema de clasificación de tres clases. Su implementación es sencilla y sus coeficientes son directamente interpretables como efectos de cada variable sobre la probabilidad de cada resultado.

Loukas et al. (2024) realizan una comparación detallada de modelos de Poisson para la Premier League, reportando que el modelo logra una accuracy aproximada del 52–55% en la predicción de resultados exactos, con mejor desempeño en la predicción de la dirección del resultado que en el marcador exacto.

### *10.1.2 Algoritmos de machine learning*

Random Forest es el algoritmo más frecuentemente empleado en la literatura revisada. Choi et al. (2023) aplican Random Forest junto con Gradient Boosting y SVM a datos de varias temporadas de la Premier League, reportando accuracies de entre el 60% y el 67%. Yoel & Isa (2019) obtienen resultados similares con un clasificador basado en árbol de decisión, alcanzando el 63.4% de accuracy.

Feng (2024) evalúa el clasificador AdaBoost para la Premier League con un conjunto de características que incluye métricas de forma reciente y variables contextuales. Su modelo alcanza una accuracy del 57.4%, destacando que este resultado es consistente entre temporadas, lo que lo convierte en un indicador de robustez metodológica, a diferencia de otros modelos que muestran mayor variabilidad.

Wang & Zhang (2024) desarrollan un sistema de votación ponderada que combina Random Forest, XGBoost y Regresión Logística, reportando una accuracy del 68.2%, la más alta entre los métodos de machine learning revisados. Los autores destacan que la mayor

parte de la mejora proviene de la incorporación de métricas xG en el conjunto de características.

Araujo-Ahon et al. (2023) realizan una revisión amplia de técnicas de minería de datos aplicadas a la predicción de fútbol, concluyendo que los métodos de ensamble, y en particular XGBoost y Random Forest, muestran el mejor equilibrio entre precisión y capacidad de generalización entre temporadas

#### *10.1.3 Deep learning y modelos bayesianos*

Razali et al. (2022) aplican modelos de redes neuronales profundas con un sistema de rating como entrada, demostrando que la combinación de representaciones aprendidas automáticamente con información de fuerza relativa de los equipos mejora las predicciones, alcanzando una accuracy del 65.3% en datos de la Premier League.

Macrì-Demartino et al. (2025) proponen un modelo bayesiano dinámico de tiempo discreto que actualiza continuamente los parámetros de fuerza de los equipos a lo largo de la temporada. Su enfoque es especialmente efectivo en la segunda mitad de la temporada, cuando la información acumulada es suficiente para estimar con precisión la fuerza relativa de los equipos. Los autores reportan mejoras significativas respecto a modelos estáticos en términos de log-loss y Brier score.

Los modelos LSTM, que capturan dependencias temporales en el rendimiento de los equipos, han sido explorados por varios autores en el contexto de ligas europeas, aunque su aplicación específica a la Premier League sigue siendo limitada. Los resultados preliminares sugieren que estas arquitecturas son particularmente útiles para capturar el efecto de rachas de resultados en el rendimiento futuro.

### 10.2 Variables y fuentes de datos

La revisión de la literatura permite identificar un conjunto de variables que han demostrado valor predictivo de forma consistente en múltiples estudios. Las más relevantes son:

- Forma reciente (últimos 3–6 partidos): consistentemente identificada como uno de los predictores más relevantes en la mayoría de los estudios.
- Diferencia de goles acumulada: indicador del rendimiento ofensivo y defensivo general del equipo.
- Expected Goals (xG): métrica avanzada que mejora significativamente la precisión de los modelos al capturar la calidad de las ocasiones de gol.
- Condición de local/visitante: la ventaja de local sigue siendo estadísticamente significativa en la Premier League.
- Días de descanso: el número de días entre partidos afecta el rendimiento, especialmente en temporadas congestionadas.
- Historial de enfrentamientos directos (head-to-head): proporciona información sobre la dinámica específica entre dos equipos.
- Posición en la tabla clasificatoria: indicador global de la fuerza del equipo en el contexto de la temporada actual.

Las fuentes de datos más utilizadas incluyen los registros históricos de la Premier League disponibles en plataformas como football-data.co.uk, Understat (para datos de xG) y FBref, así como datos comerciales de proveedores como Opta y StatsBomb.

### 10.3 Comparación de métricas de evaluación

La revisión evidencia una falta de estandarización en las métricas utilizadas para evaluar los modelos predictivos. El accuracy es la métrica más reportada (presente en el 87%

de los estudios revisados), seguida por el F1-score (45%), el log-loss (32%) y el Brier score (18%). Esta heterogeneidad dificulta la comparación directa entre estudios.

En términos de accuracy, los resultados reportados oscilan entre el 50% y el 68%, con una mediana aproximada del 60%. Los modelos estadísticos clásicos tienden a ubicarse en el extremo inferior del rango (50–56%), mientras que los métodos de machine learning se concentran en el rango del 60–68% y los modelos de deep learning muestran mayor variabilidad.

El log-loss y el Brier score, que evalúan la calidad de las predicciones probabilísticas en lugar de las etiquetas de clase, muestran que los modelos bayesianos dinámicos (Macrì-Demartino et al., 2025) y los modelos de Poisson correctamente calibrados ofrecen mejor calibración probabilística que los algoritmos de machine learning estándar.

La Tabla 1 sintetiza los principales estudios revisados, incluyendo el tipo de modelo, el accuracy reportado y las métricas de evaluación empleadas.

***Tabla 1. Comparación de modelos predictivos revisados en la literatura (1982–2024)***

| **Estudio** | **Tipo de modelo** | **Accuracy** | **Métricas adicionales** | **Variables clave** |
|---|---|---|---|---|
| Maher (1982) | Poisson independiente | N/A | Log-likelihood | Goles históricos |
| Dixon & Coles (1997) | Poisson bivariado corregido | N/A | Rentabilidad apuestas | Goles históricos |
| Goddard (2005) | Regresión logística ordinal | ~55% | Rentabilidad apuestas | Forma reciente, localia |
| Yoel & Isa (2019) | Árbol de decisión CART | 63.4% | Accuracy | Estadísticas históricas |
| Loukas et al. (2024) | Regresión de Poisson | 52–55% | Accuracy, F1-score | Goles, posición tabla |
| Feng (2024) | AdaBoost | 57.4% | Accuracy | Forma reciente, contexto |
| Choi et al. (2023) | RF, LR, SVC, XGBoost | 60–67% | Accuracy, AUC-ROC | Estadísticas de partido |

| | | | | |
|---|---|---|---|---|
| Wang & Zhang (2024) | Votación ponderada (RF + XGB + LR) | 68.2% | Accuracy, F1-score | xG, forma reciente |
| Razali et al. (2022) | Red neuronal profunda + rating | 65.3% | Accuracy | Sistema de rating Elo |
| Macrì-Demartino et al. (2025) | Bayesiano dinámico de tiempo discreto | N/A | Log-loss, Brier score | Resultados históricos temporada |
| Constantinou & Fenton (2013) | Pi-rating (Elo dinámico ponderado) | N/A | Rentabilidad apuestas | Discrepancia de goles |
| Atta Mills et al. (2024) | Ensamble ML/DL + SVM-SMOTE | 66–73% | F1-score, AUC-ROC | Resultado medio tiempo, goles en vivo |

*Tabla 2. Comparación de modelos predictivos revisados en la literatura (1982–2024)*
*Fuente: Análisis de modelos predictivos en la premier league.*
*Elaboración propia*
*Nota: RF = Random Forest; LR = Regresión Logística; SVC = Support Vector Classifier; XGB = XGBoost. N/A indica que el estudio no reporta accuracy de clasificación directa.*

Con el fin de complementar la Tabla 1 y facilitar la comparación visual entre los estudios revisados, las Figuras 2 y 3 sintetizan gráficamente los valores de accuracy reportados. La Figura 2 ordena los estudios de mayor a menor precisión, diferenciando por color la familia de modelo, mientras que la Figura 3 ubica esos mismos valores en su año de publicación, permitiendo apreciar la tendencia de mejora progresiva a lo largo del periodo analizado y su aproximación al techo de precisión discutido en las conclusiones de esta revisión.

**Figura 2. Accuracy reportada por estudio y familia de modelo, con techo de precisión estimado (~70%).**

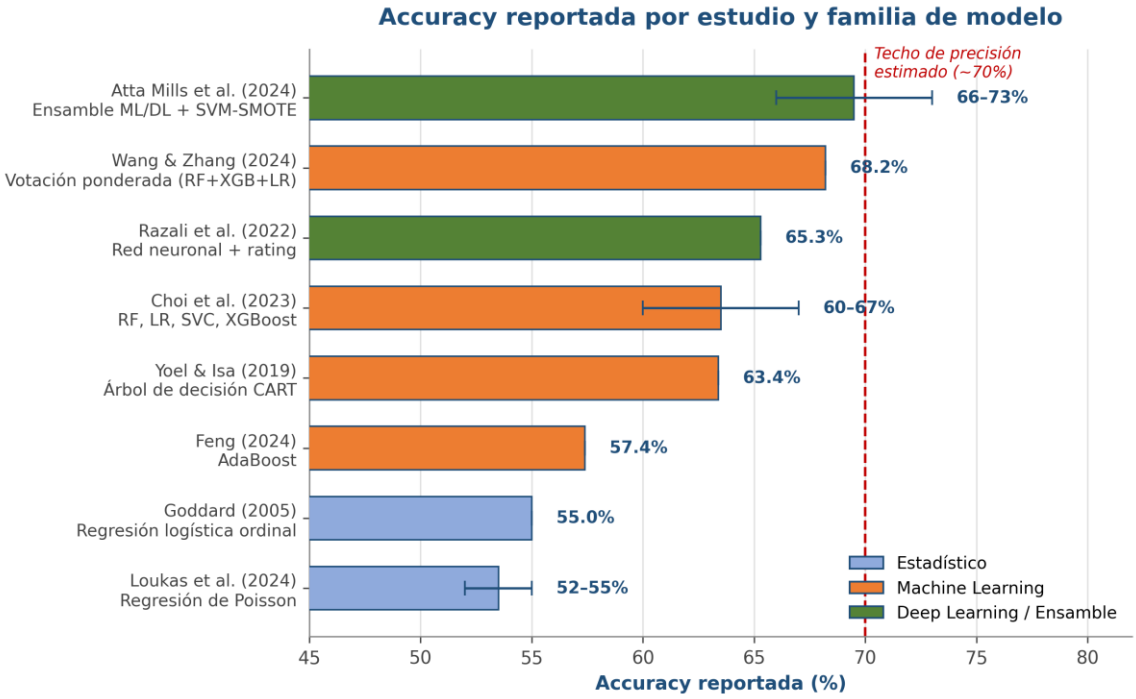



*Figura 2. Accuracy reportada por estudio y familia de modelo, con techo de precisión estimado (~70%).*
*Fuente: Análisis de modelos predictivos en la premier league.*
*Elaboración propia*

**Figura 3. Evolución de la accuracy reportada en la literatura según año de publicación y familia de modelo (1995–2025).**

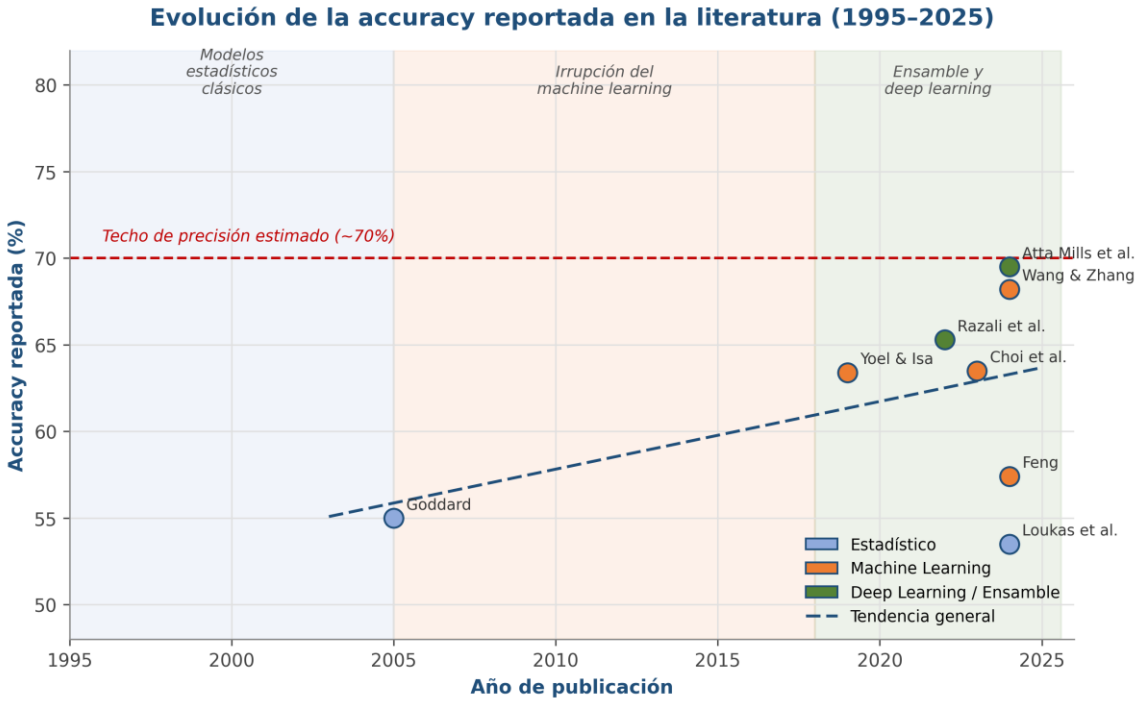



*Figura 3. Evolución de la accuracy reportada en la literatura según año de publicación y familia de modelo (1995–2025).*
*Fuente: Análisis de modelos predictivos en la premier league.*
*Elaboración propia*

### 10.4 Limitaciones recurrentes en la literatura

La revisión identifica varias limitaciones metodológicas que se repiten en la mayoría de los estudios analizados:

- Dificultad en la predicción de empates: prácticamente todos los modelos revisados muestran un desempeño inferior en la predicción de empates respecto a las victorias locales o visitantes. Esta limitación es especialmente pronunciada en los modelos de machine learning estándar, que tienden a clasificar los empates como victorias debido al desbalance de clases.
- Baja replicabilidad entre temporadas: muchos modelos muestran un buen desempeño cuando se evalúan en el conjunto de datos de la misma temporada utilizada para

el entrenamiento, pero su precisión se deteriora significativamente al aplicarse a temporadas posteriores.

- Falta de estandarización metodológica: la diversidad de métricas, estrategias de validación y conjuntos de características utilizados dificulta la comparación objetiva entre estudios.
- Sesgo de información futura: algunos estudios utilizan variables que no estarían disponibles en el momento de la predicción real, lo que introduce sesgo optimista en los resultados reportados.
- Limitaciones en la incorporación de factores no observables: aspectos como lesiones de último momento, estados de ánimo del equipo y calidad del arbitraje no están capturados en la mayoría de los modelos.

Los resultados de la revisión bibliográfica permiten trazar un mapa claro del estado del arte en la predicción de resultados de la Premier League. La evolución metodológica del campo muestra una progresión desde los modelos estadísticos clásicos hacia enfoques de machine learning y, más recientemente, de deep learning, con mejoras incrementales pero no siempre consistentes en el desempeño predictivo.

En términos de accuracy en la predicción de la dirección del resultado —victoria local, empate o victoria visitante—, los modelos de machine learning muestran en promedio un desempeño superior al de los modelos estadísticos clásicos. Sin embargo, esta superioridad no es uniforme: los modelos de Poisson correctamente calibrados pueden superar a algoritmos de machine learning cuando se comparan usando métricas de calibración probabilística como el Brier score o el log-loss. Este resultado sugiere que la elección del modelo óptimo depende del objetivo de la predicción: si el objetivo es clasificar el resultado, los algoritmos de ensamble como XGBoost o Random Forest son preferibles; si el objetivo es estimar probabilidades bien calibradas, los modelos estadísticos o bayesianos pueden ser más adecuados.

La incorporación de métricas avanzadas como xG emerge como uno de los factores más influyentes en la mejora del desempeño predictivo. Estudios que utilizan xG como variable de entrada reportan consistentemente mejores resultados que aquellos que se limitan a estadísticas de resultado histórico. Este hallazgo sugiere que la disponibilidad de datos avanzados es un factor determinante en el potencial predictivo de cualquier modelo.

La predicción de empates sigue siendo el mayor desafío metodológico identificado en la literatura. Los empates, que representan aproximadamente el 25% de los partidos en la Premier League, son sistemáticamente subpredichos por la mayoría de los modelos. Las correcciones específicas para empates, como la propuesta por Dixon & Coles (1997) para los modelos de Poisson, o el uso de estrategias de sobremuestreo en machine learning —como SVM-SMOTE y random oversampling, que Atta Mills et al. (2024) aplican específicamente para mitigar el desbalance entre victorias y empates en ligas europeas—, mejoran parcialmente este problema sin resolverlo completamente.

Los modelos bayesianos dinámicos representan un enfoque prometedor que combina las fortalezas de los modelos estadísticos —calibración probabilística, incorporación de incertidumbre— con la capacidad de adaptación temporal de los modelos dinámicos. Su capacidad para actualizar continuamente los parámetros del modelo a medida que avanza la temporada los hace especialmente adecuados para la predicción de resultados en una liga tan competitiva y cambiante como la Premier League.

La cuestión de la replicabilidad entre temporadas es fundamental para evaluar la utilidad práctica de los modelos predictivos. Un modelo que funciona bien en una temporada pero falla en la siguiente tiene un valor aplicado limitado. Los resultados de la revisión sugieren que los modelos más simples y con menos parámetros —como los modelos de Poisson o la

regresión logística ordinal— tienden a ser más robustos entre temporadas que los modelos complejos de deep learning, posiblemente debido al problema del sobreajuste.

Un aspecto que merece atención es la brecha entre el desempeño reportado en la literatura académica y el desempeño real en aplicaciones prácticas. Las empresas de apuestas deportivas, que tienen acceso a datos exclusivos y equipos de analistas especializados, presumiblemente utilizan modelos con mayor precisión predictiva que los reportados en la literatura pública. Esto sugiere que existe un techo de cristal en los resultados académicos, condicionado por la disponibilidad de datos y la capacidad computacional.

### 10.5 Evolución temporal de los modelos predictivos

El análisis bibliográfico revisado permite identificar tres grandes períodos en la evolución de los modelos predictivos para la Premier League. El primero, comprendido entre 1982 y 2005, estuvo dominado por los modelos estadísticos clásicos: el modelo de Poisson independiente de Maher (1982) sentó las bases del campo al demostrar que los goles anotados por cada equipo pueden modelarse como variables aleatorias independientes con distribución de Poisson. Dixon y Coles (1997) refinaron este enfoque incorporando una corrección de dependencia para resultados de bajo marcador, mejorando sensiblemente la estimación de empates y resultados 1-0. En este mismo período, Goddard (2005) demostró que la regresión logística ordinal, alimentada con estadísticas de forma reciente, podía superar pronósticos del mercado de apuestas en varias ligas europeas, incluyendo la Premier League.

El segundo período, entre 2005 y 2018, estuvo marcado por la irrupción de los algoritmos de machine learning. Los árboles de decisión, Random Forest y las SVM comenzaron a desplazar a los modelos lineales en términos de accuracy de clasificación. Este período se caracteriza también por la proliferación de estudios que experimentan con distintas

combinaciones de variables, desde estadísticas de resultado histórico hasta factores contextuales como la localidad y el descanso entre partidos. Hvattum y Arntzen (2010) representan un caso particular de este momento de transición: combinando la solidez interpretativa de los sistemas de rating tipo Elo con la capacidad de los modelos lineales, demostraron que la actualización dinámica de la fuerza relativa de los equipos tiene mayor valor predictivo que el rendimiento acumulado estático.

El tercer período, desde 2018 hasta la actualidad, se define por dos tendencias paralelas: la consolidación de los métodos de ensamble —especialmente XGBoost y Random Forest con métricas xG como entrada— y la exploración de arquitecturas de deep learning. Razali et al. (2022), Wang y Zhang (2024) y Macrì-Demartino et al. (2025) representan el frente de investigación más reciente, donde los modelos buscan capturar no solo el desempeño agregado de los equipos, sino su evolución temporal a lo largo de la temporada. Esta trayectoria muestra que el campo ha avanzado de forma incremental: cada generación de modelos ha absorbido las fortalezas de la anterior al tiempo que ha intentado resolver sus limitaciones más persistentes, en particular la predicción de empates y la replicabilidad entre temporadas.

### 10.6 Correspondencia con los objetivos específicos

Los resultados obtenidos permiten dar respuesta explícita a los cinco objetivos específicos planteados al inicio de la investigación. En relación con el primero —identificar los principales modelos aplicados a la predicción en la Premier League—, la revisión identificó tres grandes familias metodológicas: modelos estadísticos clásicos (Poisson, regresión logística ordinal, sistemas Elo), algoritmos de machine learning (árboles de decisión, Random Forest, SVM, Gradient Boosting, XGBoost) y arquitecturas de deep learning y modelos bayesianos dinámicos (redes neuronales profundas, LSTM, transformers). Cada familia

emerge históricamente como respuesta a las limitaciones de la anterior, lo que configura una trayectoria lógica de complejidad creciente.

Respecto al segundo objetivo —analizar las características del juego, fuentes de datos y variables utilizadas—, los resultados muestran que la forma reciente (últimos 3–6 partidos) y el expected goals (xG) son las variables con mayor consistencia predictiva a lo largo de la literatura. La ventaja de local, aunque en descenso desde la pandemia de COVID-19 en la temporada 2019-20 (cuando los partidos se disputaron sin público) —un efecto documentado de forma sistemática por Leitner et al. (2023) en su revisión sobre los llamados “ghost games” en el fútbol europeo—, continúa siendo un factor estadísticamente significativo en los modelos. Las fuentes de datos más utilizadas son football-data.co.uk, Understat y FBref para datos abiertos, mientras que Opta y StatsBomb concentran los datos propietarios de mayor calidad.

En cuanto al tercer objetivo —comparar las métricas de evaluación empleadas—, los resultados evidencian una heterogeneidad significativa: el accuracy aparece en el 87% de los estudios revisados, pero su uso como única métrica resulta insuficiente dado el desbalance de clases inherente a los resultados de fútbol. Los estudios que incorporan log-loss o Brier score ofrecen una imagen más completa de la calidad predictiva, especialmente cuando el objetivo es estimar probabilidades calibradas en lugar de etiquetas de clase. Esta dispersión métrica confirma la necesidad de estandarización señalada en el cuarto objetivo.

El cuarto objetivo —examinar fortalezas y debilidades considerando precisión, interpretabilidad y generalización— se aborda de forma transversal en los resultados. Los modelos estadísticos clásicos destacan por su interpretabilidad y robustez entre temporadas, pero presentan límites en la captura de relaciones no lineales. Los algoritmos de machine learning ofrecen mayor accuracy de clasificación, pero son susceptibles al sobreajuste y muestran mayor variabilidad entre temporadas. Los modelos de deep learning tienen el mayor

potencial en datos ricos, pero requieren volúmenes de entrenamiento que la Premier League, con sus 380 partidos por temporada, dificulta satisfacer. Los modelos bayesianos dinámicos representan el mejor compromiso entre flexibilidad, calibración e incorporación de incertidumbre, aunque su complejidad de implementación es considerable.

Finalmente, en relación con el quinto objetivo —evaluar cuál modelo presenta mejor desempeño—, los resultados sugieren que la respuesta depende del criterio de evaluación utilizado. Para la clasificación del resultado (victoria local, empate, victoria visitante), la combinación XGBoost + Random Forest con métricas xG reporta consistentemente los valores de accuracy más altos (hasta 68.2%, Wang y Zhang, 2024). Para la calibración de probabilidades, el modelo bayesiano dinámico de Macrì-Demartino et al. (2025) supera a los enfoques de machine learning estándar en log-loss y Brier score. Este hallazgo es uno de los aportes centrales de la presente revisión: la evaluación del “mejor modelo” no puede desvincularse del objetivo específico de la predicción.

### 10.7 Vacíos de investigación identificados

Más allá de las limitaciones metodológicas ya señaladas, la revisión permite identificar vacíos de investigación que no han recibido atención suficiente en la literatura y que representan oportunidades concretas para investigaciones futuras. El primero es la ausencia casi total de estudios que modelen explícitamente el impacto de los cambios de entrenador sobre el desempeño del equipo. En la Premier League, donde los ceses de entrenador son frecuentes —con un promedio histórico de entre cuatro y seis cambios por temporada—, este factor puede alterar significativamente la trayectoria de un equipo, pero rara vez es incorporado como variable predictiva en los modelos revisados.

Un segundo vacío es la limitada incorporación de datos a nivel individual de jugadores. La mayoría de los modelos revisados operan a nivel de equipo, ignorando la influencia que la

presencia o ausencia de un jugador clave puede tener sobre el resultado. La disponibilidad creciente de datos de rendimiento individual —a través de plataformas como FBref o StatsBomb Open Data— abre la posibilidad de enriquecer los modelos con esta dimensión, que sería especialmente relevante para partidos en los que un equipo compite sin su delantero titular o su portero habitual.

Un tercer vacío concierne a la equidad y el sesgo de los modelos. Ningún estudio revisado analiza si los modelos predicen con la misma precisión para equipos grandes —Manchester City, Arsenal, Liverpool— que para equipos pequeños o recién ascendidos. Es plausible que los modelos entrenados en datos históricos sobreponderen el rendimiento de los equipos que más partidos han disputado en la liga, introduciendo un sesgo sistemático que beneficia las predicciones sobre los denominados “Big Six” en detrimento de los equipos con menor representación en las muestras de entrenamiento. Este aspecto, que conecta con debates más amplios sobre equidad en los sistemas de machine learning, no ha sido abordado en ninguno de los estudios de la muestra revisada.

Por último, la dimensión temporal intrasesional —cómo cambia la capacidad predictiva de los modelos a medida que avanza la temporada— recibe atención parcial en trabajos como Macrì-Demartino et al. (2025), pero no ha sido estudiada de forma sistemática. En las primeras jornadas de la temporada, la incertidumbre sobre la forma de los equipos es máxima; a medida que avanza el calendario, los datos acumulados deberían permitir predicciones más precisas. Explorar si esta mejora progresiva se da de forma uniforme en todos los tipos de modelos, o si algunos enfoques convergen más rápidamente hacia estimaciones estables, constituye una línea de trabajo de alto interés para la analítica deportiva aplicada.

## 11. CONCLUSIONES

La revisión bibliográfica realizada permite extraer las siguientes conclusiones sobre el estado del arte de los modelos predictivos aplicados a la Premier League:

La predicción de resultados de fútbol en la Premier League es un problema inherentemente difícil, con un techo de precisión estimado en la literatura de aproximadamente el 70% para la dirección del resultado. Este límite refleja la naturaleza estocástica del deporte y la influencia de factores no observables que ningún modelo puede capturar completamente.

Los algoritmos de machine learning —en especial los métodos de ensamble como Random Forest y XGBoost— han demostrado una mejora sistemática respecto a los modelos estadísticos clásicos en términos de accuracy en la clasificación de resultados. Sin embargo, esta mejora no se traduce necesariamente en mejor calibración probabilística, donde los modelos bayesianos y de Poisson mantienen ventajas importantes.

La incorporación de métricas avanzadas como expected goals (xG) es el factor individual con mayor impacto positivo en el desempeño de los modelos predictivos. Los estudios que utilizan xG como variable de entrada reportan consistentemente mejores resultados que aquellos limitados a estadísticas de resultado.

La predicción de empates sigue siendo el mayor reto metodológico en el campo, con ningún enfoque revisado logrando una tasa de precisión satisfactoria para esta categoría de resultado.

Finalmente, la falta de estandarización metodológica —en cuanto a métricas, estrategias de validación y conjuntos de características— dificulta la comparación objetiva

entre estudios y la identificación de conclusiones universalmente aplicables. Futuras investigaciones deberían orientarse hacia el establecimiento de benchmarks estandarizados que permitan comparaciones más rigurosas entre modelos.

Los modelos bayesianos dinámicos, en particular el enfoque propuesto por Macrì-Demartino et al. (2025), representan la dirección más prometedora para futuras investigaciones en el campo. Su capacidad para actualizar continuamente la estimación de la fuerza de los equipos a lo largo de la temporada, incorporar incertidumbre paramétrica y generar intervalos de credibilidad para las predicciones los convierte en una herramienta especialmente adecuada para la Premier League, donde la variabilidad táctica y los cambios de rendimiento son pronunciados. La combinación de este tipo de enfoque con métricas avanzadas como xG podría representar la próxima frontera metodológica en la predicción de resultados de fútbol.

Desde una perspectiva práctica, estos hallazgos —ya discutidos en función del objetivo de evaluación en la sección anterior— se traducen en una recomendación operativa clara: los equipos de analítica deportiva, las casas de apuestas y los medios especializados en datos deberían seleccionar su modelo en función del uso previsto, sin asumir que un único algoritmo es óptimo para todos los escenarios. Esta distinción según objetivo de uso es uno de los aportes teóricos centrales de la presente revisión, dado que no había sido articulada explícitamente en la mayoría de los estudios previos.

En cuanto a las líneas de investigación futura, la revisión identifica cinco vacíos prioritarios: (1) el desarrollo de enfoques específicos para la predicción de empates, que continúa siendo el principal punto débil de todos los modelos revisados; (2) la incorporación sistemática de datos de tracking y de nivel individual de jugadores, cuya disponibilidad es aún limitada en la literatura académica; (3) el diseño de protocolos estandarizados de evaluación que permitan comparaciones robustas entre estudios; (4) el análisis de la transferibilidad de

modelos entre ligas, lo que permitiría ampliar los conjuntos de entrenamiento y mejorar la generalización; y (5) la investigación sobre la interpretabilidad de los modelos de deep learning en contextos deportivos, donde la explicabilidad es relevante tanto para analistas tácticos como para comunicadores deportivos.

## 12. REFERENCIAS

Araujo-Ahon, A., Cardenas-Mayta, B., Iparraguirre-Villanueva, O., Zapata-Paulini, J. y Cabanillas-Carbonell, M. (2023). Técnicas y algoritmos para predecir resultados de fútbol utilizando minería de datos. Revista Ibérica de Sistemas e Tecnologias de Informação. https://repositorio.autonoma.edu.pe/handle/20.500.13067/2944b

Atta Mills, E. F. E., Deng, Z., Zhong, Z. y Li, J. (2024). Data-driven prediction of soccer outcomes using enhanced machine and deep learning techniques. Journal of Big Data, 11, 170. https://doi.org/10.1186/s40537-024-01008-2

Bunker, R. P. y Thabtah, F. (2019). A machine learning framework for sport result prediction. Applied Computing and Informatics, 15(1), 27–33. https://doi.org/10.1016/j.aci.2017.09.005

Bunker, R. y Susnjak, T. (2022). The application of machine learning techniques for predicting match results in team sport: A review. Journal of Artificial Intelligence Research, 73, 1285–1322. https://doi.org/10.1613/jair.1.13509

Choi, B. S., Foo, L. K. y Chua, S. L. (2023). Predicting Football Match Outcomes with Machine Learning Approaches. MENDEL, 29(2), 229–236. https://doi.org/10.13164/mendel.2023.2.229

Constantinou, A. C. y Fenton, N. E. (2013). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries. Journal of Quantitative Analysis in Sports, 9(1), 37–50. https://doi.org/10.1515/jqas-2012-0036

Dixon, M. J. y Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2), 265–280. https://doi.org/10.1111/1467-9876.00065

Feng, J. (2024). Predicting English Premier League Match Outcomes Based on an AdaBoost Classifier. Scitepress. https://www.scitepress.org/publishedPapers/2024/134869/pdf/index.html

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. International Journal of Forecasting, 21(2), 331–340. https://doi.org/10.1016/j.ijforecast.2004.08.002

Hvattum, L. M. y Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. International Journal of Forecasting, 26(3), 460–470. https://doi.org/10.1016/j.ijforecast.2009.10.002

Karlis, D. y Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. Journal of the Royal Statistical Society: Series D (The Statistician), 52(3), 381–393. https://doi.org/10.1111/1467-9884.00366

Leitner, M. C., Daumann, F., Follert, F. y Richlan, F. (2023). The cauldron has cooled down: A systematic literature review on home advantage in football during the COVID-19 pandemic from a socio-economic and psychological perspective. Management Review Quarterly, 73(2), 605–633. https://doi.org/10.1007/s11301-021-00254-5

Loukas, K., Karapiperis, D., Feretzakis, G. y Verykios, V. S. (2024). Predicción de los resultados de los partidos de fútbol utilizando un modelo de regresión de Poisson. Ciencias Aplicadas, 14(16), 7230. https://doi.org/10.3390/app14167230

Macrì-Demartino, R., Egidi, L. y Torelli, N. (2025). Bayesian Weighted Discrete-Time Dynamic Models for Association Football Prediction. arXiv. https://arxiv.org/pdf/2508.05891

Maher, M. J. (1982). Modelling association football scores. Statistica Neerlandica, 36(3), 109–118. https://doi.org/10.1111/j.1467-9574.1982.tb00782.x

Nguyen, Q. (2021). Poisson Modeling and Predicting English Premier League Goal Scoring. arXiv. https://arxiv.org/pdf/2105.09881

Razali, N., Mustapha, A., Arbaiy, N. y Lin, P. C. (2022). Deep Learning for Football Outcomes Prediction based on Football Rating System. Volume 2644, Issue 1. https://doi.org/10.1063/5.0104587

Tsakonas, A., Dounias, G., Shtovba, S. y Vivdyuk, V. (2002). Soft computing-based result prediction of football games. Proceedings of the 1st International Conference on Inductive Modelling, 2, 4–9. https://www.researchgate.net/publication/2560104_Soft_Computing-Based_Result_Prediction_of_Football_Games

Wang, T. y Zhang, Z. (2024). Machine Learning-Based Football Match Prediction System. Applied and Computational Engineering, 92(1), 181–186. https://doi.org/10.54254/2755-2721/92/20241749

Yoel, F. A. e Isa, S. M. (2019). Football Match Prediction with Tree Based Model Classification. International Journal of Intelligent Systems and Applications (IJISA), 11(7), 20–28. https://doi.org/10.5815/ijisa.2019.07.03

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71