

# Predicción de Resultados en Partidos de la Premier League mediante Ciencia de Datos

## Objetivo

Desarrollar modelos predictivos basados en algoritmos de machine learning para estimar las probabilidades de los resultados (local, empate, visitante) en partidos de la Premier League. Además, reflexionar sobre las implicaciones éticas de la aplicación de estos modelos en contextos como las apuestas deportivas.

### 1 Preprocesamiento de datos

Base de datos con estadísticas históricas de partidos. Limpieza y normalización de datos para garantizar su calidad y relevancia. División en conjuntos de entrenamiento (70%) y prueba (30%).

## Metodología

Se desarrollaron modelos predictivos, utilizando principalmente 2 tipos: Modelo de árbol de decisión y Modelo de red neuronal para estimar resultados en partidos de fútbol. Los datos históricos fueron preprocesados y ajustados a cada modelo. Se emplearon técnicas como búsqueda de hiperparámetros y validación cruzada para optimizar el rendimiento, asegurando la confiabilidad de las predicciones en escenarios reales.

### 2 Desarrollo de modelos

Random Forest: elegido por su capacidad de interpretar características relevantes y calcular probabilidades.

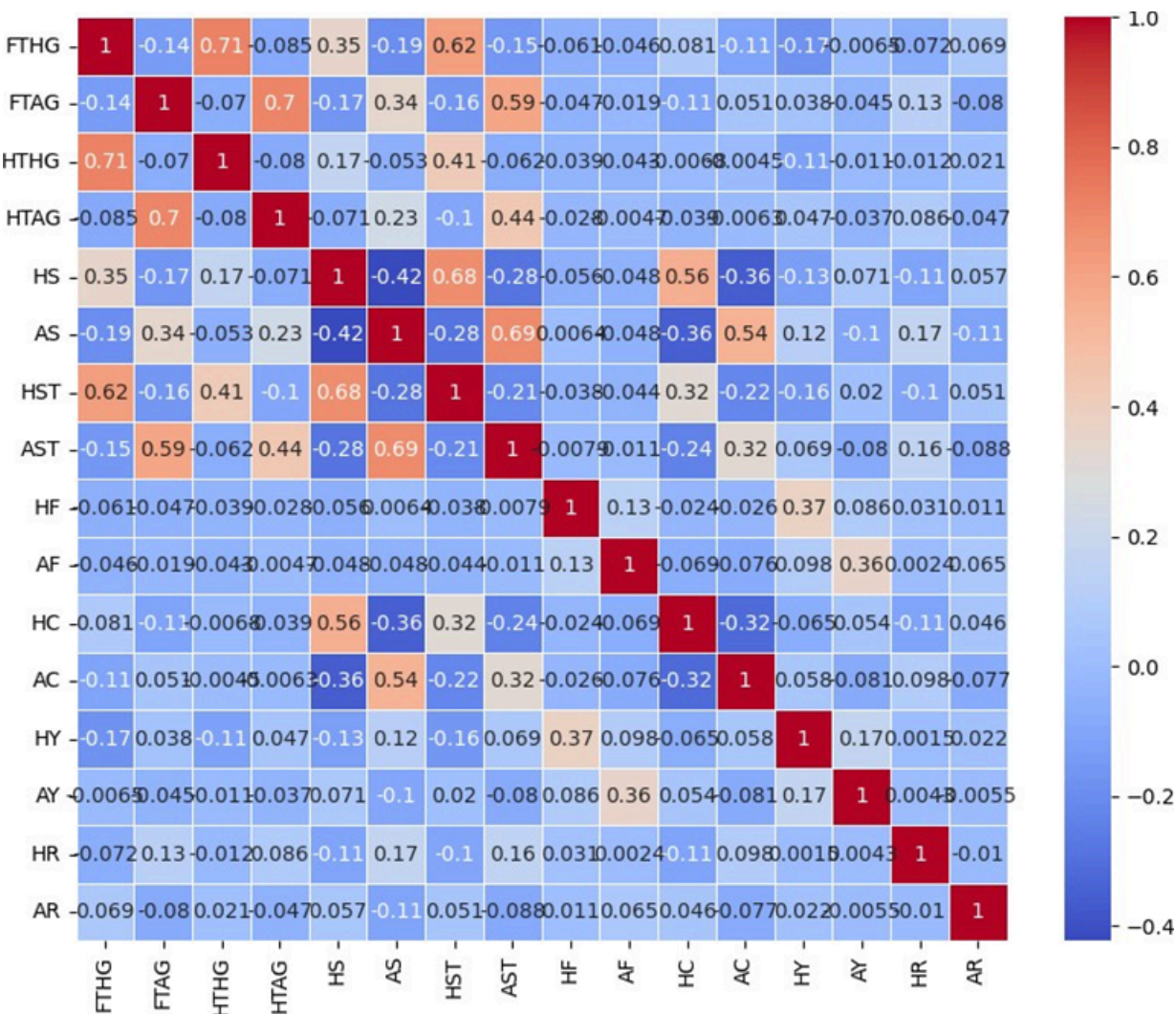
Red Neuronal: utilizado para capturar relaciones no lineales y manejar problemas complejos de clasificación.

### 3 Evaluación de modelos

Métricas como precisión, recall y F1-score para evaluar el desempeño de cada modelo.

Comparación con probabilidades implícitas de casas de apuestas.

## Matriz de correlación



## Resultados

Random Forest: Precisión del 92.7%, con un F1-score de 0.925. Ventaja principal: mayor interpretabilidad.

Red Neuronal: Precisión del 95.3%, con un F1-score de 0.952. Ventaja principal: mejor capacidad para capturar patrones complejos.

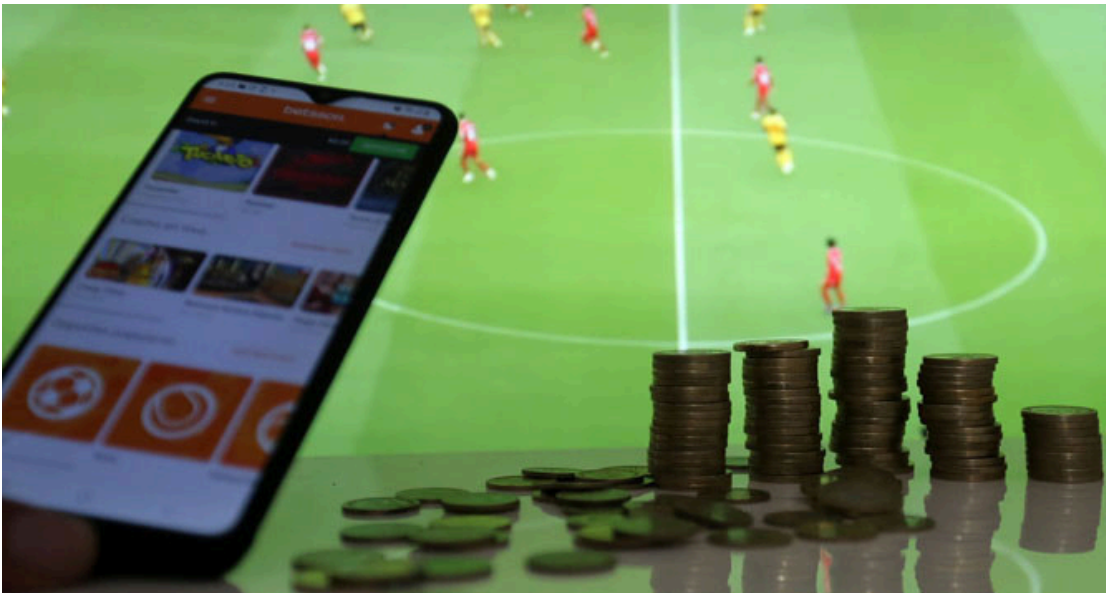
Comparación con apuestas deportivas: el modelo Random Forest predijo una mayor probabilidad de empate frente a las estimaciones más optimistas de las casas de apuestas.

A continuación se presentan resultados de probabilidades obtenidas para cada posible evento dentro de los resultados de un partido (Local, Empate, Visita)

Para el ejemplo se seleccionó el partido que se disputará entre el Manchester City y Tottenham Hotspurs, obteniendo mayor probabilidad de victoria el local con 54,7% vs un 39,8% de empate y tan sólo 5,5% de victoria visitante

## Discusión

El uso de modelos predictivos en contextos deportivos plantea desafíos éticos, especialmente en relación con las apuestas. La falta de regulación en mercados como el chileno puede derivar en riesgos para los consumidores, como expectativas irreales o comportamientos adictivos. Además, es fundamental garantizar que estos modelos no comprometan la integridad del deporte ni se utilicen para manipular resultados. Esto subraya la necesidad de una regulación clara y un enfoque responsable en el desarrollo y aplicación de estas tecnologías.



## Prueba de predicción

