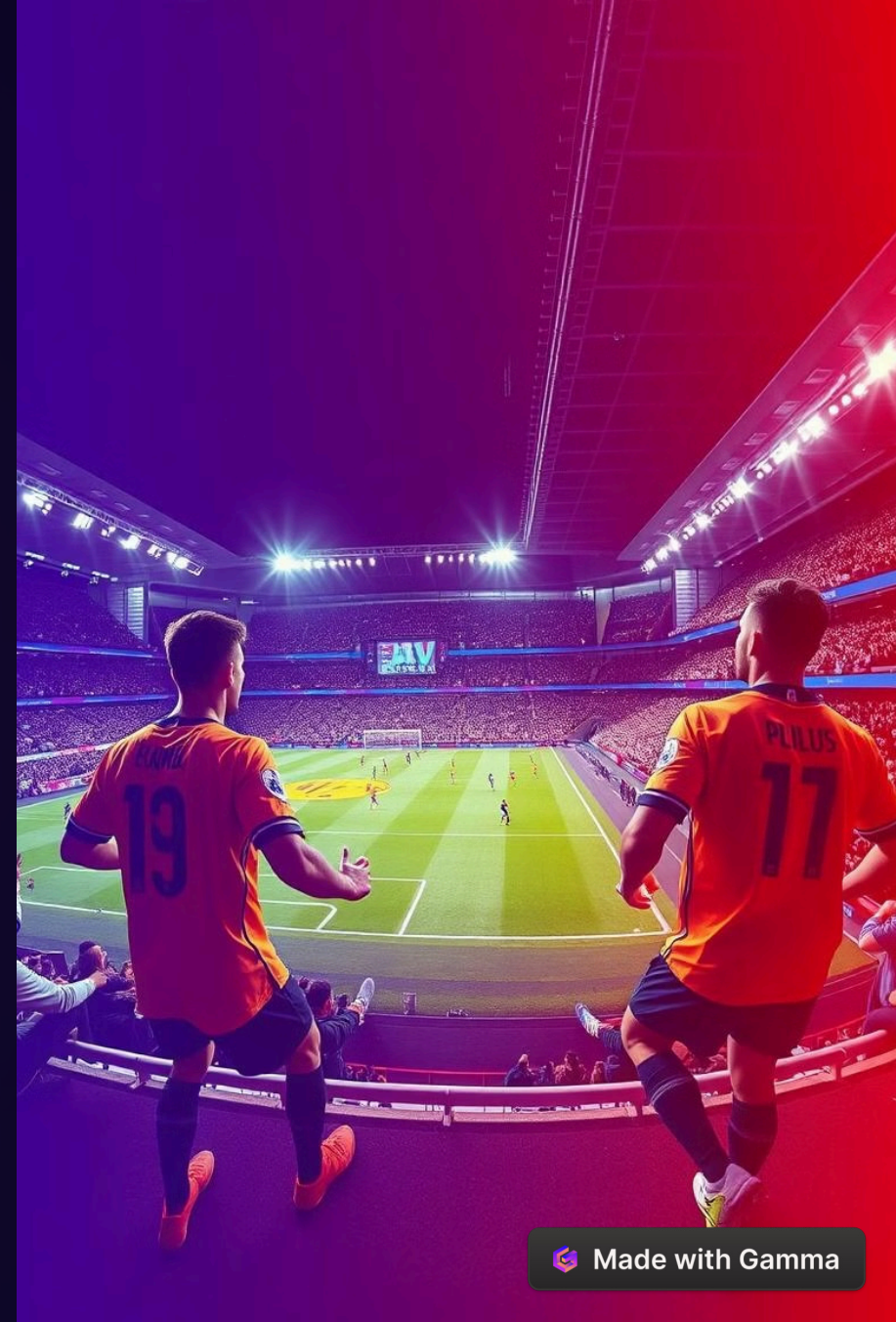


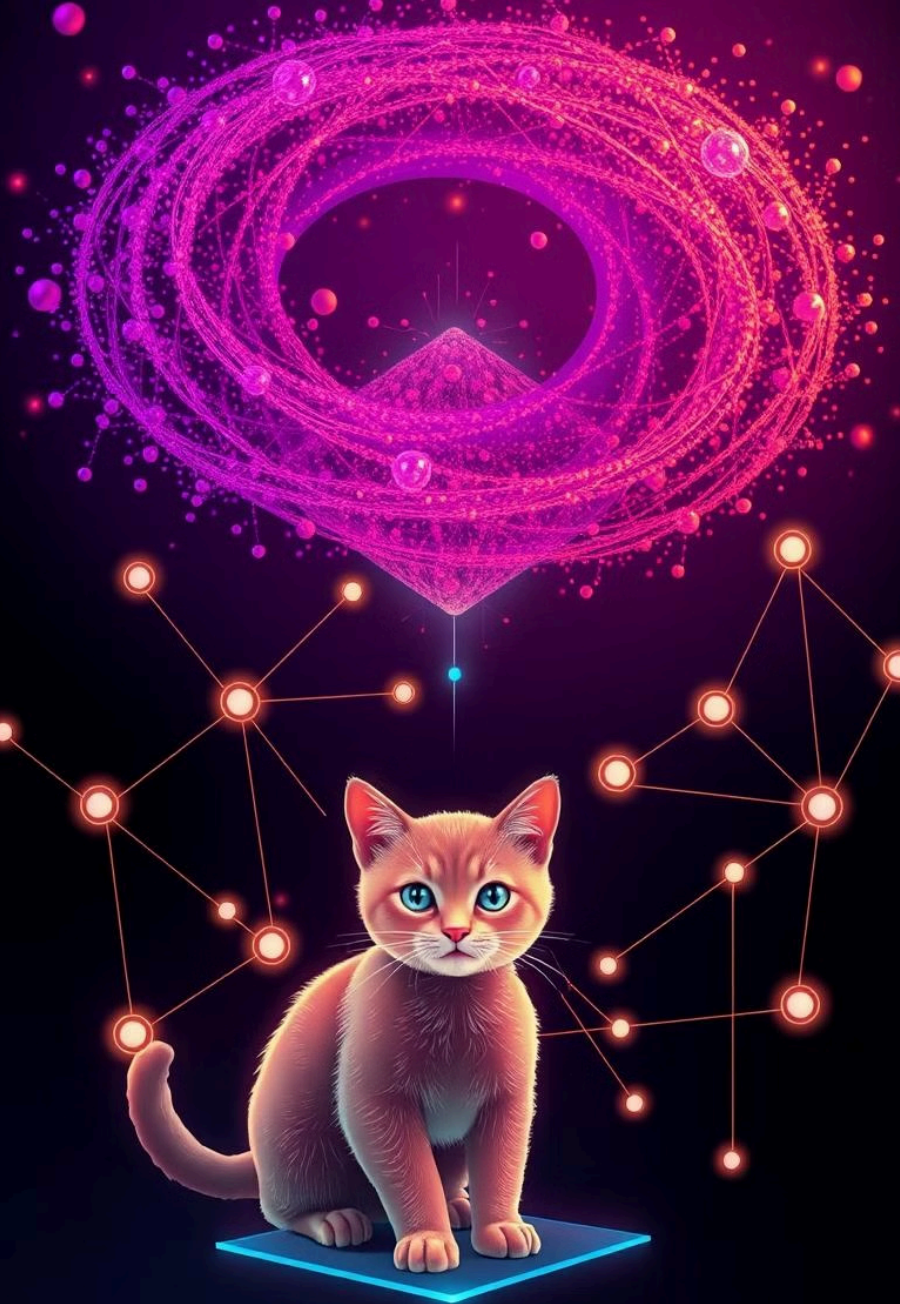
Predicción de Partidos de Fútbol en la Premier League

En esta presentación, exploraremos un emocionante proyecto de ciencia de datos que busca predecir los resultados de partidos de la Premier League utilizando el poderoso algoritmo de Random Forest.

Integrantes:

- Benjamin De La Torre Rojas
- Mauricio David Correa Hernandez
- Vladlen Shatunov





Introducción a Machine Learning

1 Aprendizaje Automático

Machine Learning permite a las computadoras aprender de los datos sin ser programadas explícitamente.

2 Patrones y Predicciones

Los algoritmos de Machine Learning identifican patrones en los datos para hacer predicciones y tomar decisiones.

3 Aplicaciones Amplias

El Machine Learning tiene aplicaciones en una variedad de campos, desde el reconocimiento de imágenes hasta la detección de fraudes.

Modelos de Machine Learning

Regresión

Modelos de regresión predicen valores continuos, como el precio de una casa.

Clasificación

Modelos de clasificación predicen categorías discretas, como spam o no spam.



El Algoritmo de Random Forest

Múltiples Árboles de Decisión

Random Forest utiliza varios árboles de decisión para tomar una decisión final.

Muestras Aleatorias

Cada árbol se entrena con una muestra aleatoria de los datos, evitando el sobreajuste.

Votación Mayoritaria

La predicción final se obtiene mediante una votación entre todos los árboles.



Definición del Problema

El objetivo del proyecto es predecir si un equipo de la Premier League ganará o no un partido utilizando datos históricos.

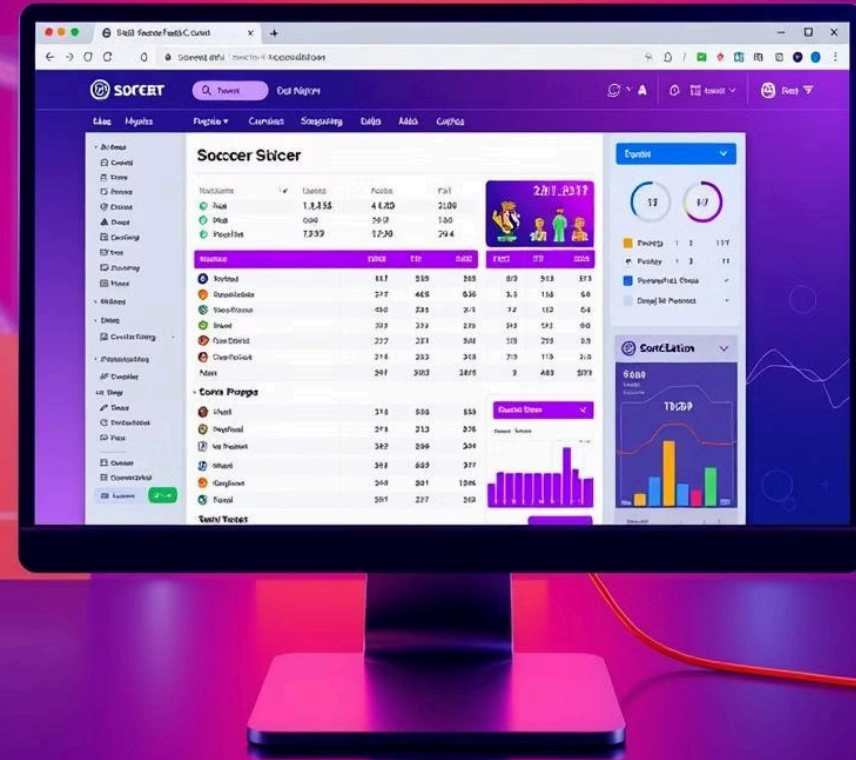


Objetivo del Modelo

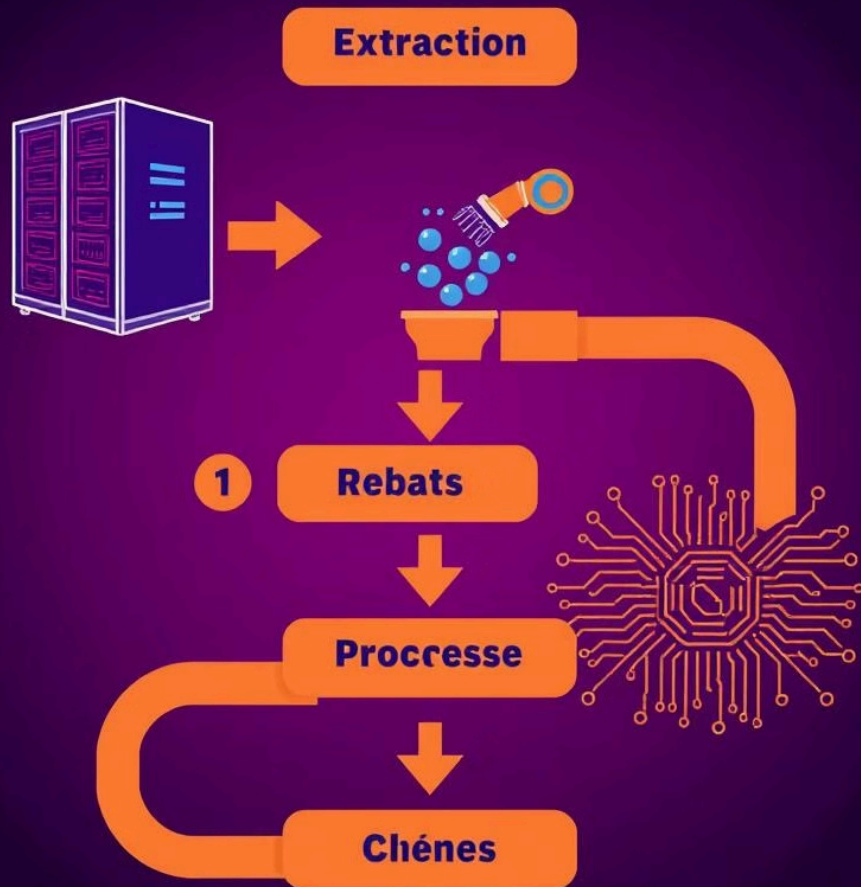
El modelo es de clasificación binaria, prediciendo si un equipo ganará (1) o perderá/empatará (0) un partido.

Fuente de Datos

La fuente de datos es fbref.com, que ofrece estadísticas detalladas sobre partidos de fútbol de la Premier League.



Data Pipeline



Obtención y Procesamiento de Datos

1

Descarga de Datos

Se utiliza la biblioteca requests de Python para descargar datos de partidos de fbref.com.

2

Análisis y Limpieza

Se utiliza BeautifulSoup para analizar el HTML y pandas para manipular y estructurar los datos.

3

Integración de Datos

Se crea un DataFrame de pandas que contiene todos los partidos para el entrenamiento del modelo.

Características del Conjunto de Datos

Variable	Tipo	Descripción
date	Categórica (ordinal)	Fecha del partido
venue	Categórica	Local/Visitante
gf	Numérica	Goles a favor
ga	Numérica	Goles en contra
xg	Numérica	Goles esperados
xga	Numérica	Goles esperados contra
poss	Numérica	Posesión
formation	Categórica	Formación del equipo
opp formation	Categórica	Formación del oponente
sh	Numérica	Tiros
sot	Numérica	Tiros a puerta
dist	Numérica	Distancia promedio de los tiros
fk	Numérica	Tiros libres
pk	Numérica	Penaltis marcados
pkatt	Numérica	Penaltis intentados
attendance	Numérica	Asistencia
result	Categórica	Resultado (W/L/D)
season	Numérica	Temporada
team	Categórica	Nombre del equipo

