

Proyecto de Integración y Automatización de Datos para IA para el proyecto de Modelos Predictivos de la Premier League

Andrei Ricardo Saldarriaga Gutiérrez¹, Ariel Cifuentes Osorio²,
John Isidro Roa Reina³, Juan Pablo Mogollón Lozano⁴

Facultad de Ingeniería y Ciencias Básicas
Universidad Central
Maestría en Análisis de Datos
Automatización e Integración de Datos para IA
Bogotá, Colombia

April 6, 2024

Contents

| | | |
|----------|---|----------|
| 1 | Introducción | 2 |
| 2 | Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA | 2 |
| 2.1 | Objetivo general | 2 |
| 2.1.1 | Objetivos específicos | 2 |
| 2.2 | Alcance | 3 |
| 2.3 | Pregunta de investigación | 3 |
| 2.4 | Hipotesis | 3 |
| 3 | Reflexiones sobre el origen de datos e información | 4 |
| 3.1 | ¿Cual es el origen de los datos e información? | 4 |
| 3.2 | ¿Cuales son las consideraciones legales o éticas del uso de la información? | 4 |
| 3.3 | ¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? | 4 |
| 3.4 | ¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? | 4 |
| 4 | Diseño de integración y Automatización de Datos para IA (Diagrama) | 5 |

1 Introducción

El mercado de las apuestas deportivas está en auge, con ocho millones de cuentas activas en Colombia y un movimiento de más de 26 billones de pesos en 2022, siendo el fútbol, el principal motor de ingresos para las casas de apuestas. Sin embargo, solo el 1% de los apostadores obtiene beneficios económicos, y la mayoría son jóvenes que subestiman el riesgo de pérdida.

Usualmente al apostar las personas se basan en conocimientos y sensaciones, pero esto suele resultar en pérdidas económicas. Algunos recurren a canales de redes sociales para obtener asesoramiento gratuito o pago, que prometen tasas de éxito superiores al 80%. Sin embargo, estos asesoramientos carecen de fundamentos sólidos y no siempre cumplen con las expectativas.

Para tener éxito en las apuestas en la Premier League, donde el promedio de goles es alto (2.43 por partido), es necesario desarrollar un modelo predictivo utilizando análisis estadísticos y modelos como Random Forest o Modelo Bayesiano para aprovechar oportunidades en mercados más específicos, como lo son el número de goles, tiros de esquina, tarjetas, entre otros, de un equipo en particular.

A pesar de la emoción y el entretenimiento que ofrece el mundo de las apuestas deportivas, es crucial comprender los riesgos involucrados y adoptar estrategias informadas. La combinación de la intuición humana con el análisis estadístico y los modelos predictivos puede ser la clave para obtener éxito en este campo desafiante.

2 Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA

Integración y Automatización de Datos para Modelos Predictivos Basados en Datos que Maximicen las Oportunidades en el Mercado de Apuestas de la Premier League

2.1 Objetivo general

Desarrollar un sistema de integración y automatización de datos para aplicaciones de Inteligencia Artificial que permita el análisis predictivo de resultados en partidos de la Premier League, con el fin de mejorar la precisión en las apuestas deportivas.

2.1.1 Objetivos específicos

- Recopilar datos históricos de las últimas cinco temporadas de la Premier League, incluyendo resultados de partidos, estadísticas de equipos y jugadores,

y cualquier otra información relevante.

- Integrar los datos recopilados en una base de datos y diseñar un esquema de almacenamiento eficiente para facilitar su acceso y consulta.
- Desarrollar procesos de ETL (Extracción, Transformación y Carga) para automatizar la recopilación, limpieza y carga de datos en la base de datos.
- Explorar y evaluar diferentes algoritmos de aprendizaje automático, como Random Forest, Redes Neuronales, o Regresión Logística, para construir modelos predictivos.
- Validar los modelos mediante pruebas con datos históricos y comparar su desempeño con otros métodos de predicción.

2.2 Alcance

El alcance de este proyecto comprende predicciones de diferentes mercados de apuestas de la Premier League, entre los que se encuentran, el número de goles marcados en un partido, el número de tiros de esquina, tiros a puerta, entre otros mercados ya sean del equipo local y/o visitante.

2.3 Pregunta de investigación

¿Cómo pueden integrarse y automatizarse datos históricos de la Premier League para construir modelos de aprendizaje automático precisos que permitan predecir resultados de partidos y mejorar la efectividad en aciertos de apuestas deportivas?

2.4 Hipotesis

La integración y automatización de datos de la Premier League, para crear modelos predictivos mejorará la precisión en apuestas deportivas. Utilizando estadísticas de equipos, jugadores, condiciones de partidos y tendencias históricas, se espera alcanzar al menos un 70% de acierto en pronósticos.

Este modelo ofrecerá a los apostadores una herramienta basada en IA para tomar decisiones acertadas, identificando patrones y tendencias de los partidos.

3 Reflexiones sobre el origen de datos e información

3.1 ¿Cual es el origen de los datos e información?

Se utilizará una base histórica de las últimas cinco temporadas de la Premier League, la cual, contiene variables como fecha del partido, equipo local, equipo visitante, goles anotados por el equipo local, goles anotados por el equipo visitante, entre otras variables.

Adicionalmente, mediante técnicas de web scraping utilizando Python, actualizaremos semanalmente los resultados de la liga en curso, para efectos de comparar las predicciones realizadas.

3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información?

Los datos utilizados para este proyecto corresponden a datos disponibles en la web, los cuales no están restringidos para su consulta, por lo que no estimamos la existencia de asuntos legales asociados al uso de estos.

Por otra parte, los datos resultantes de las predicciones no estarán destinados para consulta de un público menor de 18 años. Adicionalmente, se recomendará a los usuarios de la información el uso responsable de la misma, teniendo en cuenta que las apuestas pueden generar adicción.

3.3 ¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA?

El reto de este proyecto es contar con información completa y actualizada que permita predecir de manera acertada diferentes eventos deportivos de la Premier League y adicionalmente que pueda usarse en tiempo real.

3.4 ¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto?

Con la ayuda de la IA esperamos crear un modelo de predicción de eventos deportivos de Premier League que esté en un nivel de automatización de procesos, en el cual, podamos monitorear los resultados en tiempo real.

4 Diseño de integración y Automatización de Datos para IA (Diagrama)

A continuación se puede observar el diagrama para la integración y automatización de modelos predictivos para la Premier League:

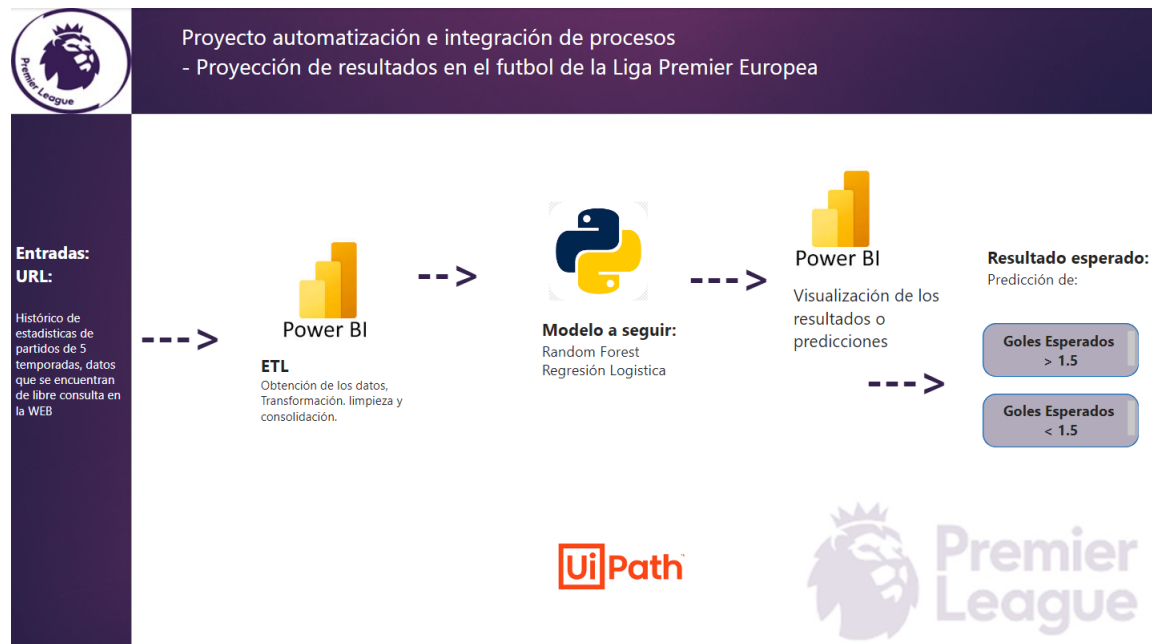


Figure 1: Diagrama del modelo