

SISTEMA DE PRECICCIÓN DEL RESULTADO DE PARTIDOS DE FÚTBOL

SISTEMAS DE INFORMACIÓN DE GESTIÓN Y
BUSSINESS INTELLIGENCE

Julio César Ruiz Calle

ÍNDICE

1. Introducción
2. Objetivos del trabajo
3. Desarrollo del trabajo
 1. Obtención de los datos
 2. Limpieza de los datos
 3. Creación de la base de datos de grafos
 4. Desarrollo del algoritmo de aprendizaje automático
 5. Creación de la aplicación web
4. Líneas de futuro

INTRODUCCIÓN

- El trabajo consiste en la predicción de los resultados de partidos de fútbol usando Python.
- El usuario introducirá los datos del partido
- Podrá elegir una liga de las siguientes: Liga alemana, liga española, liga inglesa, liga francesa y liga italiana.
- Podrá elegir el equipo local y el equipo visitante, así como las cuotas que tenga el partido en cuestión.
- El sistema le devolverá el resultado calculado y las probabilidades.

OBJETIVOS DEL TRABAJO

- Terminar el proyecto dentro de los plazos estipulados y que tenga un funcionamiento aceptable
- Utilizar en el proyecto las herramientas que nos ha recomendado el profesor (Neo4J, Python, machine learning...)
- Indagar en el tema de la predicción de resultados mediante técnicas de aprendizaje automático
- Indagar en el tema de los distintos algoritmos de aprendizaje automático y ver las ventajas y desventajas de ellos

OBJETIVOS DEL TRABAJO (2)

- No es un objetivo mejorar los sistemas de predicción que ya existen
- Aprender a utilizar las distintas herramientas que ofrece Python para trabajar con grandes cantidades de datos como NumPy, Pandas, Matplot...

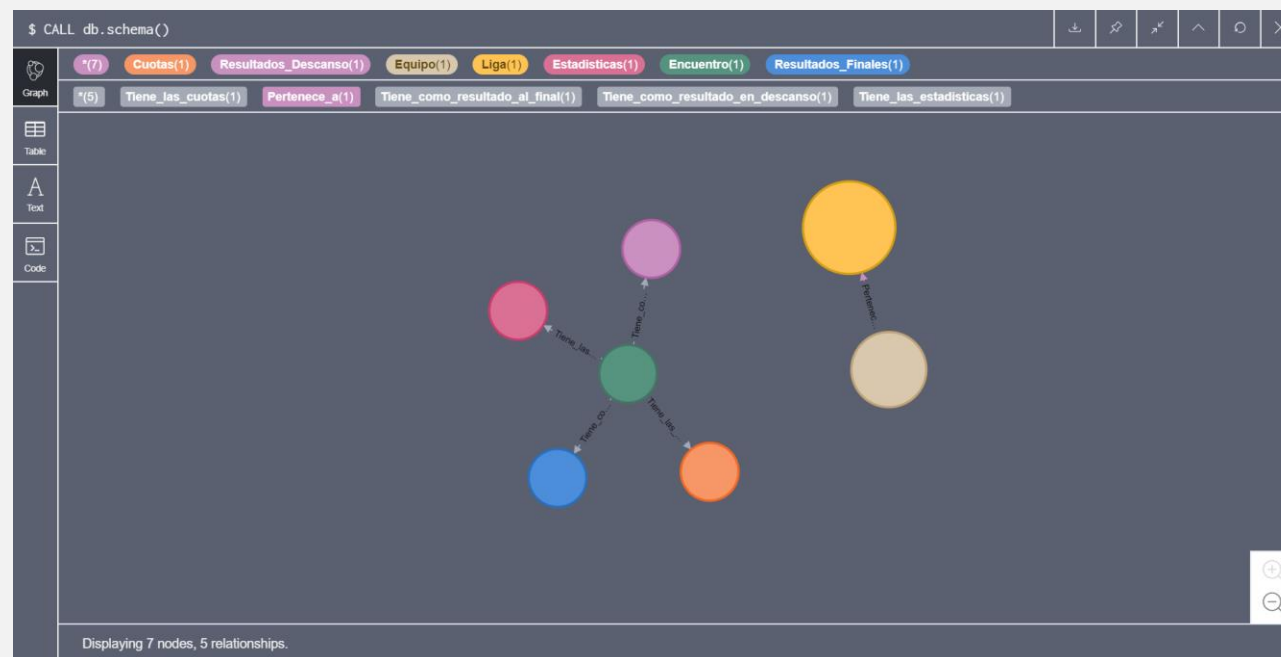
DESARROLLO DEL TRABAJO : OBTENCIÓN DE LOS DATOS

- Necesitaba sets de datos con un gran número de estadísticas registradas, que cada partido fuese único y que se incluyesen las cuotas para cada partido
- Encontré sets de datos que me convencían en la página <https://www.football-data.co.uk/>
- Están separados por ligas y por temporadas

DESARROLLO DEL TRABAJO : LIMPIEZA DE LOS DATOS

- La limpieza de los datos era necesaria porque había estadísticas de mis archivos csv que no iba a utilizar
- Además existían estadísticas vacías (null) en algunas filas, al pasarle dichos valores al modelo de entrenamiento crearían conflicto
- Solucioné ambos problemas creando dos pequeños scripts de Python: uno para quitar las columnas que no iba a utilizar y otro para quitar las filas que contenían valores vacíos

DESARROLLO DEL TRABAJO : CREACIÓN DE LA BASE DE DATOS



DESARROLLO DEL TRABAJO : CREACIÓN DEL ALGORITMO DE APRENDIZAJE AUTOMÁTICO

- En primer lugar, accedí desde Python a la base de datos de Neo4J para poder crear una estructura de datos con la que poder trabajar. Para ello usé las librerías de Python GraphDatabase y Pandas
- Decidí usar un algoritmo de regresión logística porque se usa para resolver problemas de clasificación, usa una variable dependiente además de variables independientes y es un algoritmo simple y eficiente

DESARROLLO DEL TRABAJO : CREACIÓN DEL ALGORITMO DE APRENDIZAJE AUTOMÁTICO (2)

- A la hora de entrenar el modelo se usan todas las estadísticas pero a la hora de predecir un nuevo resultado sólo se usan unas pocas
- Debido a esto, a la hora de predecir decidí utilizar una media de los valores en cada caso para solventar el problema

DESARROLLO DEL TRABAJO : CONSTRUCCIÓN DE LA APP WEB

- Decidí crearla para que el usuario pudiese interactuar más cómodamente con el sistema
- El frontend escucha en el puerto 8000 y está desarrollado en lenguaje Javascript con Vuejs además de Vuetify
- El backend escucha en el puerto 5000 y está desarrollado en lenguaje Python con Flask
- Se comunican mediante peticiones GET/POST

LÍNEAS DE FUTURO

1. Alimentar la base de datos
2. Implementar la funcionalidad de que se pueda elegir una casa de apuestas
3. Añadir más equipos por liga
4. Añadir más ligas
5. Mejorar el algoritmo de aprendizaje automático
6. Añadir aún más características al modelo
7. Abrir la puerta a otros deportes