



Predicción Ganador UEFA Champions League

Lidia Velicia Ruiz
Ana García Saiz

Inteligencia Artificial
Grado Ingeniería Matemática
Universidad Alfonso X el Sabio

Índice

1. Introducción	2
2. Metodología	2
2.1. Recopilación de datos	2
2.1.1. Extracción	2
2.1.2. Estadísticas equipos	3
2.2. Preprocesamiento	4
2.2.1. Ponderaciones Ligas Nacionales	4
2.2.2. Ponderaciones UCL	4
2.2.3. Dataset partidos	5
3. Análisis y Resultados	5
3.1. Modelo Regresión Lineal	5
3.2. Modelo Red Neuronal	5
3.3. Modelos Orange	6
4. Conclusiones	6

1. Introducción

El objetivo de este proyecto era simple, haciendo uso de técnicas de Análisis de Datos y Machine Learning, predecir el equipo ganador del torneo de la UEFA Champions League (UCL) 2023/24. Con total libertad para elegir el planteamiento del análisis, los datos, los modelos a utilizar...

2. Metodología

2.1. Recopilación de datos

El vasto mundo del fútbol y las páginas de estadísticas futbolísticas es complicado. Hay muchas webs con mucha información y recopilar datos en grandes cantidades no es tarea fácil.

2.1.1. Extracción

Por suerte, dimos con la librería ScraperFC de **@ouseymour**, para el scrappeo de datos futbolísticos. Entre las diferentes webs de las que permite obtener datos, elegimos FBRef, ya que nos pareció la más completa en términos de las estadísticas que ofrece y la gran variedad de ligas internacionales que maneja.

Con el fin de predecir el ganador del torneo de la UCL 23/24, hemos tomado los siguientes datos:

- Estadísticas de la UCL 2022/2023
- Estadísticas de la fase de grupos de la UCL 2022/2023
- Estadísticas de las ligas nacionales 2022/2023
- Estadísticas de las ligas nacionales 2023/2024 (hasta la fecha)

Decidimos coger los datos de las 2 últimas temporadas porque, en general, los equipos suelen mantener su plantilla relativamente constante durante este períodos.

2.1.2. Estadísticas equipos

Para obtener una visión completa de cada equipo, seleccionamos variables que cubren tanto aspectos de ataque como de defensa, así como otras variables relevantes. Las variables recopiladas para el análisis y su significado son las siguientes:

Ataque

- **Gls90**: Goles por 90 minutos.
- **Ast90**: Asistencias por 90 minutos.
- **G+A90**: Goles y asistencias combinados por 90 minutos.
- **G-PK90**: Goles no provenientes de penaltis por 90 minutos.
- **G+A-PK90**: Goles y asistencias no provenientes de penaltis por 90 minutos.
- **SCA90**: Acciones que conducen a una oportunidad de gol por 90 minutos.
- **GCA90**: Acciones que conducen a un gol por 90 minutos.
- **Sh90**: Disparos por 90 minutos.
- **SoT90**: Disparos a puerta por 90 minutos.
- **SoT %**: Porcentaje de disparos a puerta.
- **G/Sh**: Goles por disparo.
- **G/SoT**: Goles por disparo a puerta.

Defensa

- **GA90**: Goles concedidos por 90 minutos.
- **Save %**: Porcentaje de paradas realizadas por el portero.
- **SoTA90**: Disparos a puerta enfrentados por 90 minutos.

Otras

- **Probabilidad_SemiF**: Probabilidad de alcanzar las semifinales.
- **Probabilidad_Won**: Probabilidad de ganar el torneo.

2.2. Preprocesamiento

2.2.1. Ponderaciones Ligas Nacionales

Dado que cada liga europea presenta un nivel competitivo diferente, creímos necesario ponderar las estadísticas de los equipos en sus ligas nacionales, según el "nivel futbolístico" de la liga de procedencia:

- Las cinco principales ligas europeas Premiere League (Inglaterra), La Liga (España), Ligue 1 (Francia), Serie A (Italia) y Bundesliga (Alemania) se ponderaron con un factor de 1.
- Las ligas "de 2^o categoría" serían la Eredivisie (Países Bajos), Superliga (Dinamarca) y Primeira Liga (Portugal), consideradas de un nivel ligeramente inferior, se ponderaron con un factor de 0.67.

2.2.2. Ponderaciones UCL

Para reflejar de manera precisa las diferencias de nivel entre los partidos de fase de grupos de la UCL y los enfrentamientos en fases más avanzadas del torneo, es esencial ponderar las estadísticas de los equipos. La mayoría de los partidos en las fases iniciales de la UCL son contra equipos significativamente inferiores, lo cual puede inflar las estadísticas de los equipos más fuertes. Para contrarrestar este efecto y lograr una representación más equilibrada, es necesario ajustar las estadísticas de la UCL añadiendo las estadísticas de las ligas nacionales de cada equipo.

La fórmula de ponderación utilizada para calcular las estadísticas combinadas es la siguiente:

$$\text{Ponderación final} = \frac{(\text{Est}_{\text{UCL}} + 0,5 \cdot \text{Est}_{\text{NAC}})}{2}$$

Esta fórmula permite integrar las estadísticas nacionales, dándoles un peso moderado al multiplicarlas por 0.5, antes de combinarlas con las estadísticas de la UCL. Finalmente, el resultado se divide por 2 para obtener

una ponderación equilibrada entre ambos conjuntos de datos. Esta metodología asegura que las estadísticas utilizadas en el análisis sean representativas del rendimiento real de los equipos, considerando tanto su desempeño en la UCL como en sus respectivas ligas nacionales.

2.2.3. Dataset partidos

Nuestro enfoque para predecir quién ganará el campeonato será simular los enfrentamientos, prediciendo los goles marcados por cada equipo, con ello, según quien gane, llegaremos al ganador de la UCL 23/24.

Para ello, entrenaremos los modelos con los partidos de la temporada completa 22/23 y la fase de grupos de la 23/24. Junto con las variables relativas al enfrentamiento, incluiremos las estadísticas de cada equipo, para que el modelo pueda aprender de ellas para predecir los goles.

De esta forma el dataset nos queda:

3. Análisis y Resultados

Nuestro objetivo es predecir el número de goles que marca cada equipo en cada partido en función de sus estadísticas y las del adversario. Por lo tanto, al estar prediciendo una variable numérica, nuestros modelos serán de regresión.

Los modelos que hemos utilizado son:

3.1. Modelo Regresión Lineal

3.2. Modelo Red Neuronal

En este caso, hemos creado dos tipos de modelo, el primero de ellos, más simple, solo con las variables de ataque y defensa. El segundo, añadiendo además las variables de probabilidad de alcanzar semifinales y probabilidad de ganar la champions.

Veamos los resultados del primero:

Fase	Equipo 1	Gol Acum. Eq 1	Equipo 2	Gol Acum. Eq 2
Octavos	PSG	2.37	Real Sociedad	1.064
Octavos	Copenhague	1.272	Manchester City	4.482
Octavos	Barcelona	3.024	Napoli	0.932
Octavos	Atlético Madrid	2.276	Inter	1.621
Octavos	Dortmund	2.046	PSV	1.318
Octavos	Bayern Múnich	2.669	Lazio	0.899
Octavos	Arsenal	2.531	Porto	2.351
Octavos	Real Madrid	2.929	Leipzig	1.417
Cuartos	Dortmund	1.772	Atlético Madrid	2.118
Cuartos	Barcelona	2.303	PSG	1.859
Cuartos	Bayern Múnich	1.577	Arsenal	2.522
Cuartos	Real Madrid	1.694	Manchester City	3.12
Semis	Barcelona	2.402	Atlético Madrid	1.845
Semis	Arsenal	1.682	Manchester City	3.359
Final	Barcelona	1.402	Manchester City	3.212

Fase	Equipo 1	Gol Acum. Eq 1	Equipo 2	Gol Acum. Eq 2
Octavos	PSG	1.969	Real Sociedad	1.055
Octavos	Copenhague	1.534	Manchester City	3.468
Octavos	Barcelona	3.061	Napoli	0.762
Octavos	Atlético Madrid	1.912	Inter	1.452
Octavos	Dortmund	2.379	PSV	1.124
Octavos	Bayern Múnich	2.417	Lazio	0.747
Octavos	Arsenal	1.991	Porto	1.791
Octavos	Real Madrid	2.608	Leipzig	0.982
Cuartos	Dortmund	1.538	Atlético Madrid	1.668
Cuartos	Barcelona	1.967	PSG	1.698
Cuartos	Bayern Múnich	1.586	Arsenal	1.862
Cuartos	Real Madrid	1.958	Manchester City	2.608
Semis	Barcelona	2.104	Atlético Madrid	1.457
Semis	Arsenal	1.245	Manchester City	2.741
Final	Barcelona	1.256	Manchester City	2.742

En cuanto al otro modelo:

Fase	Equipo 1	Gol Acum. Eq 1	Equipo 2	Gol Acum. Eq 2
Octavos	PSG	3.581	Real Sociedad	2.767
Octavos	Copenhague	3.159	Manchester City	4.446
Octavos	Barcelona	3.622	Napoli	2.892
Octavos	Atlético Madrid	3.312	Inter	3.011
Octavos	Dortmund	3.282	PSV	3.117
Octavos	Bayern Múnich	3.006	Lazio	2.402
Octavos	Arsenal	3.294	Porto	3.456
Octavos	Real Madrid	3.954	Leipzig	3.053
Cuartos	Dortmund	2.992	Atlético Madrid	3.397
Cuartos	Barcelona	3.667	PSG	3.765
Cuartos	Bayern Múnich	3.042	Arsenal	3.288
Cuartos	Real Madrid	4.07	Manchester City	4.462
Semis	PSG	3.688	Atlético Madrid	3.502
Semis	Arsenal	3.42	Manchester City	4.433
Final	PSG	3.83	Manchester City	4.62

Fase	Equipo 1	Gol Acum. Eq 1	Equipo 2	Gol Acum. Eq 2
Octavos	PSG	2.223	Real Sociedad	1.539
Octavos	Copenhague	1.775	Manchester City	3.293
Octavos	Barcelona	2.398	Napoli	1.544
Octavos	Atlético Madrid	2.184	Inter	1.854
Octavos	Dortmund	1.995	PSV	1.76
Octavos	Bayern Múnich	2.052	Lazio	1.34
Octavos	Arsenal	2.396	Porto	2.351
Octavos	Real Madrid	2.542	Leipzig	1.939
Cuartos	Dortmund	1.948	Atlético Madrid	2.141
Cuartos	PSG	2.2	Barcelona	2.328
Cuartos	Arsenal	2.376	Bayern Múnich	1.911
Cuartos	Real Madrid	2.442	Manchester City	2.98
Semis	Barcelona	2.295	Atlético Madrid	2.169
Semis	Arsenal	2.301	Manchester City	3.049
Final	Barcelona	2.226	Manchester City	2.956

3.3. Modelos Orange

4. Conclusiones

Referencias

[1]

[2]

[3]

[4]

[5]