

INFORME PREVIO

De Soto Manilla, David
Llorián González, Pablo
Martínez Jiménez, Yaiza
Menéndez Fernández, Cristina
Pérez Navarro, David

30 de abril de 2025

1. Datos

La base de datos empleada para este proyecto está compuesta por la unión de dos bases de datos obtenidas en <https://www.basketball-reference.com/>. Estas son:

- https://www.basketball-reference.com/leagues/NBA_2023_totals.html, donde se reflejan las estadísticas de los jugadores NBA durante la temporada 2022/2023.
- <https://hoopshype.com/salaries/players/2022-2023/>¹, donde se refleja el salario de los jugadores esa misma temporada.

Además, se ha creado una base con la misma información pero de la temporada 2023/2024 que se usará como datos de prueba en diferentes modelos.

2. Objetivos

El presente trabajo tiene como finalidad realizar un análisis estadístico y predictivo de los salarios de los jugadores de la NBA durante la temporada 2022-2023. Para ello, se parte de una base de datos que recopila múltiples variables de rendimiento y características individuales de los jugadores. Los objetivos específicos del estudio son los siguientes:

1. **Preprocesamiento y depuración de datos:** Eliminar registros con información incompleta o irrelevante (por ejemplo, jugadores con escasa participación en la liga) y transformar variables claves como el salario, expresándolo en millones de dólares para una mejor interpretación.
2. **Análisis exploratorio de datos:** Estudiar la distribución de variables fundamentales como la edad, posición en el campo, y estadísticas individuales por partido (puntos, asistencias, rebotes, robos, tapones, etc.), así como su relación con la variable salarial. Se incluyen visualizaciones gráficas (histogramas, diagramas de caja, gráficos de dispersión) que permiten detectar patrones y posibles asociaciones.
3. **Clasificación de jugadores:** Implementar técnicas de agrupamiento (*clustering*), especialmente mediante el algoritmo *k-means*, para identificar perfiles defensivos de jugadores (por ejemplo, grandes taponadores o ladrones de balón) y analizar la relación entre estos perfiles y las posiciones en el campo.
4. **Cálculo de indicadores avanzados:** Incorporar métricas como la eficiencia del jugador (combinación ponderada de sus estadísticas positivas y negativas), y construir nuevas variables como los promedios por partido, a fin de enriquecer el análisis y mejorar las predicciones.
5. **Modelización y predicción salarial:** Aplicar técnicas de aprendizaje automático como árboles de regresión y redes neuronales artificiales, con el objetivo de predecir el salario de los jugadores en función de sus características y rendimiento. Se evaluará la precisión de los modelos y se interpretarán las variables con mayor importancia en la predicción.

¹Pese a ser una página web ajena a **Basketball Reference**, la información ha sido obtenida de la misma.

6. Interpretación de resultados y validación: Identificar los jugadores más eficientes y mejor remunerados, analizar posibles discrepancias, y validar el rendimiento de los modelos utilizando particiones de entrenamiento y prueba. Además, se exploran relaciones estadísticas significativas entre edad y salario a través de tests de hipótesis.

En conjunto, el objetivo del trabajo busca describir los datos de los que disponemos, además de generar modelos predictivos que nos permitan estimar de forma razonable el salario de un jugador, en función de su rendimiento y características personales. Esto es muy útil para equipos, agentes y analistas del deporte.

3. Reparto

El reparto del trabajo será de la siguiente forma:

- 1. Búsqueda de los datos:** David Pérez Navarro
- 2. Realización del código:** David Pérez Navarro, Cristina Menéndez Fernández
- 3. Redacción de los resultados obtenidos a partir del código:** Pablo Llorián González
- 4. Redacción de las conclusiones:** David De Soto Manilla
- 5. Desarrollo de la presentación:** Cristina Menéndez Fernández, Yaiza Martínez Jiménez
- 6. Revisión del informe:** Yaiza Martínez Jiménez
- 7. Revisión de la presentación:** David De Soto Manilla