

# Machine Learning Project

Autora:  
Inma Jiménez



# Índice

1. Contexto
2. Datos
3. Análisis de datos
4. Optimización
5. Conclusión



# 1. Contexto

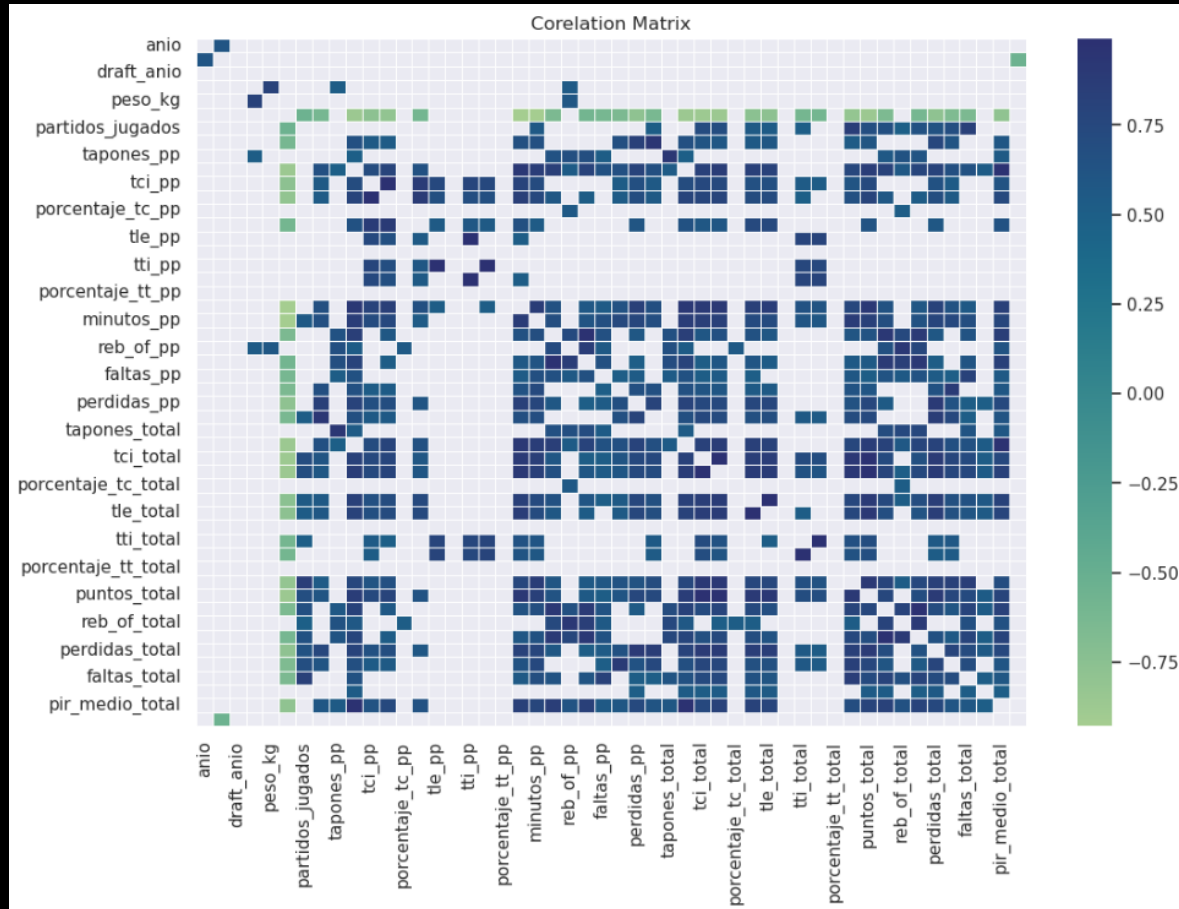
- PIR como indicador del rendimiento y potencial de los jugadores de la NBA
- Predicción del PIR medio para mejorar la selección de jugadores
- Modelo de Regresión

## 2. Datos

- Datos obtenidos directamente de la API de la web oficial de la NBA:
  - Información general de la temporada y año de juego.
  - Información relativa a los datos personales de todos los jugadores de la NBA por temporada.
  - Información específica de sobre los estadísticos de todos los jugadores de la NBA por temporada.



# 3. Análisis de Datos



Matriz de correlación  
entre las variables

# 3. Análisis de Datos

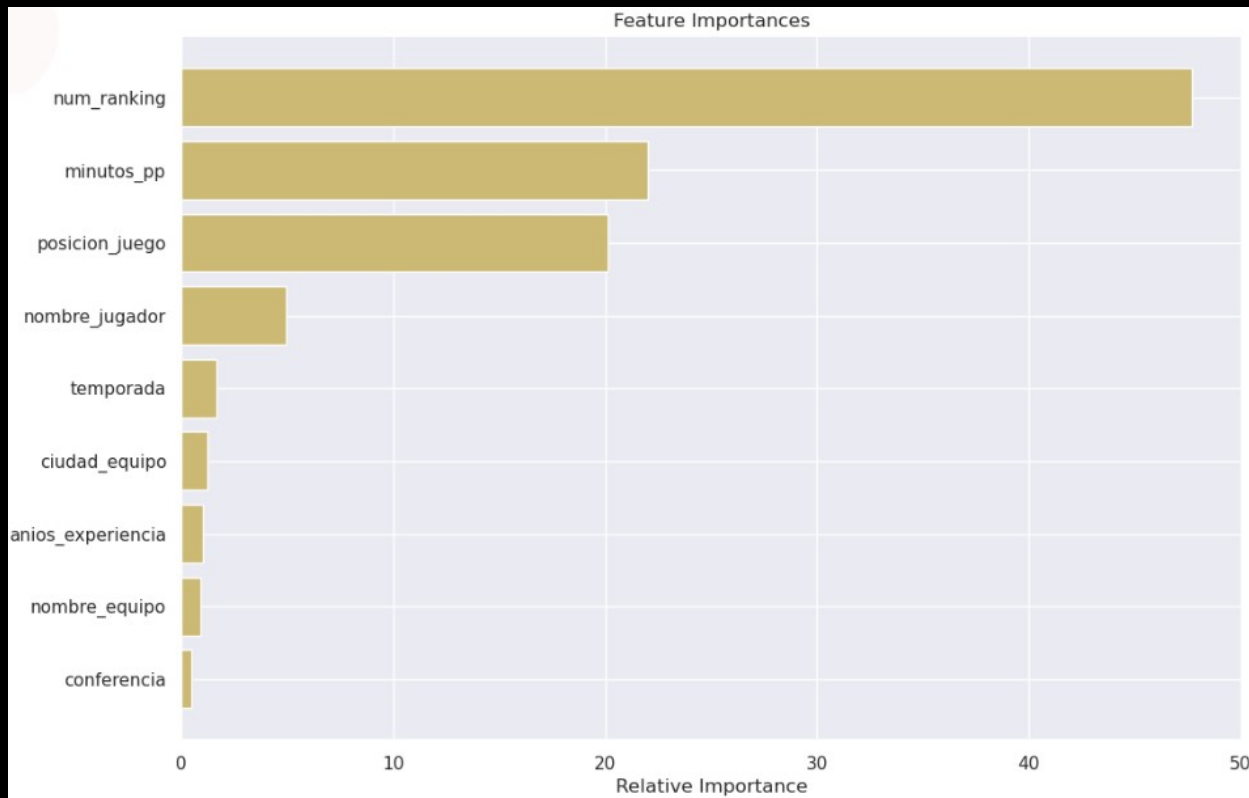
	anio	id_jugador	draft_anio	altura	peso_kg	num_ranking	minutos_pp	minutos_total	pir_medio_total	anios_experiencia
anio	1.000000	0.601412	0.068534	-0.080765	-0.097126	0.045623	-0.002116	-0.050493	0.048809	-0.064420
id_jugador	0.601412	1.000000	0.031585	-0.062909	-0.129554	0.217256	-0.192029	-0.216445	-0.146980	-0.553470
draft_anio	0.068534	0.031585	1.000000	0.007366	0.010462	-0.062580	0.065352	0.078463	0.052372	0.075711
altura	-0.080765	-0.062909	0.007366	1.000000	0.801804	0.059000	-0.083685	-0.036673	0.172631	0.037317
peso_kg	-0.097126	-0.129554	0.010462	0.801804	1.000000	0.007595	-0.039653	-0.007425	0.230086	0.142467
num_ranking	0.045623	0.217256	-0.062580	0.059000	0.007595	1.000000	-0.920958	-0.814319	-0.782897	-0.194066
minutos_pp	-0.002116	-0.192029	0.065352	-0.083685	-0.039653	-0.920958	1.000000	0.874195	0.768343	0.210756
minutos_total	-0.050493	-0.216445	0.078463	-0.036673	-0.007425	-0.814319	0.874195	1.000000	0.694588	0.186668
pir_medio_total	0.048809	-0.146980	0.052372	0.172631	0.230086	-0.782897	0.768343	0.694588	1.000000	0.197224
anios_experiencia	-0.064420	-0.553470	0.075711	0.037317	0.142467	-0.194066	0.210756	0.186668	0.197224	1.000000

Matriz de  
correlación en el  
conjunto train

```
pir_medio_total    1.000000
num_ranking        0.782897
minutos_pp         0.768343
anios_experiencia  0.197224
```

Features numéricas que seleccionadas

# 4. Optimización



Modelo Catboost –  
Feature Importance

# 4. Optimización

- Modelos:
  - *Modelo GradientBoostingRegressor* → “posicion\_juego” → *OrdinalEncoding*
    - R2 en test = 0.79
    - R2 en conjunto test reservado = 0.82
  - *Modelo GradientBoostingRegressor + RandomSearch* → “posicion\_juego” → *OrdinalEncoding*
    - R2 en test = 0.99764
  - *Modelo GradientBoostingRegressor + GridSearch* → “posicion\_juego” → *OrdinalEncoding*
    - R2 en test = 0.9999



# 5. Conclusiones

- 1- Modelo final: GradientBoostingRegressor + GridSearch
- 2- El R2 y la precisión del modelo es alta porque resulta un modelo simple que trabaja sobre 3 variables principalmente: “num\_ranking”, “posicion\_juego” y “minutos\_pp”.
- 3- Limitaciones y mejoras:
  - Simplicidad del modelo
  - No tener acceso a datos anteriores

¡Gracias por  
vuestra atención!

