

Scouting y predicción de resultados en ligas semiprofesionales de baloncesto

Autor: Álvaro Simón del Corral

Tutor: Alejandro Bellogín Kouki

Índice:

1.- Introducción

1.1.- Motivación

1.2.- Objetivos

2.- Diseño e Implementación

2.1.- Diseño

2.2.- Implementación

3.- Pruebas y Resultados

3.1.- Sistema

3.2.- Predicción de resultados

4.- Conclusiones y Trabajo Futuro

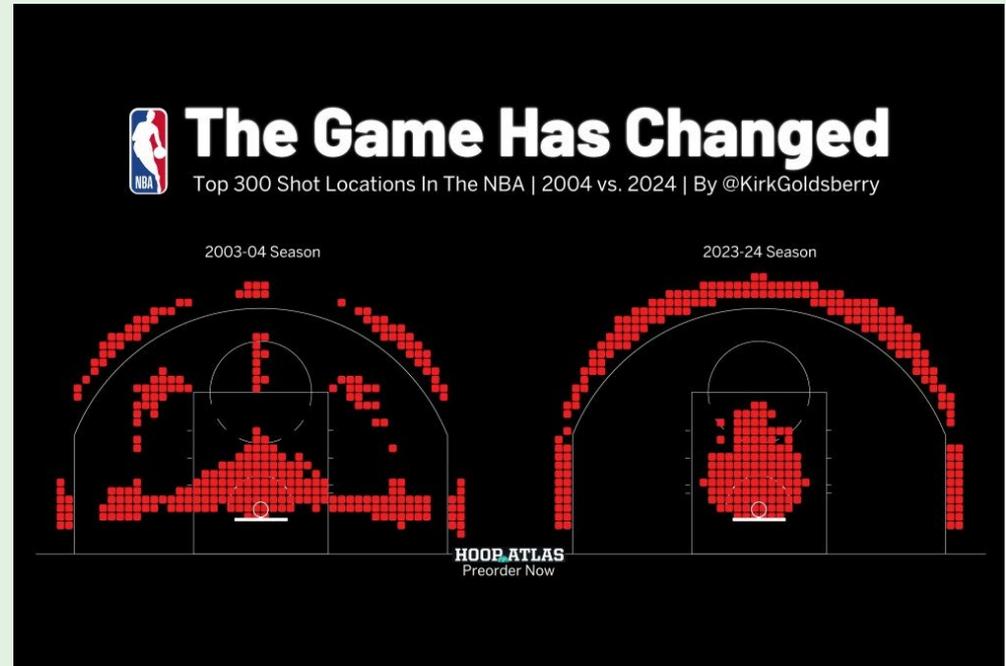
4.1.- Conclusiones

4.2.- Trabajo futuro

1- Introducción

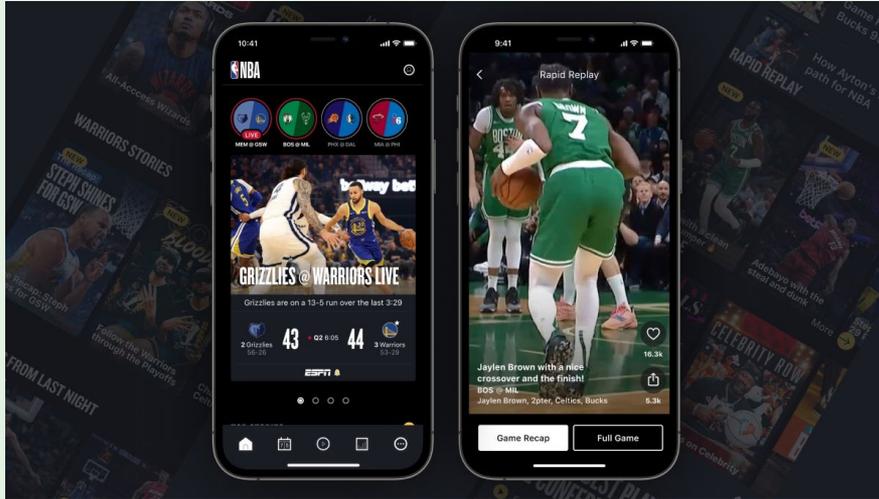
1.1. Motivación (I)

- ➔ Rendimiento individual
- ➔ Rendimiento colectivo
- ➔ Negociación de contratos
- ➔ Planificaciones estratégicas
- ➔ Reducción de lesiones
- ➔ Entre otros ámbitos



Fuente: ChessBase, 2024

1.1. Motivación (II)



Fuente: Microsoft, 2022



Fuente: NBN23, 2021

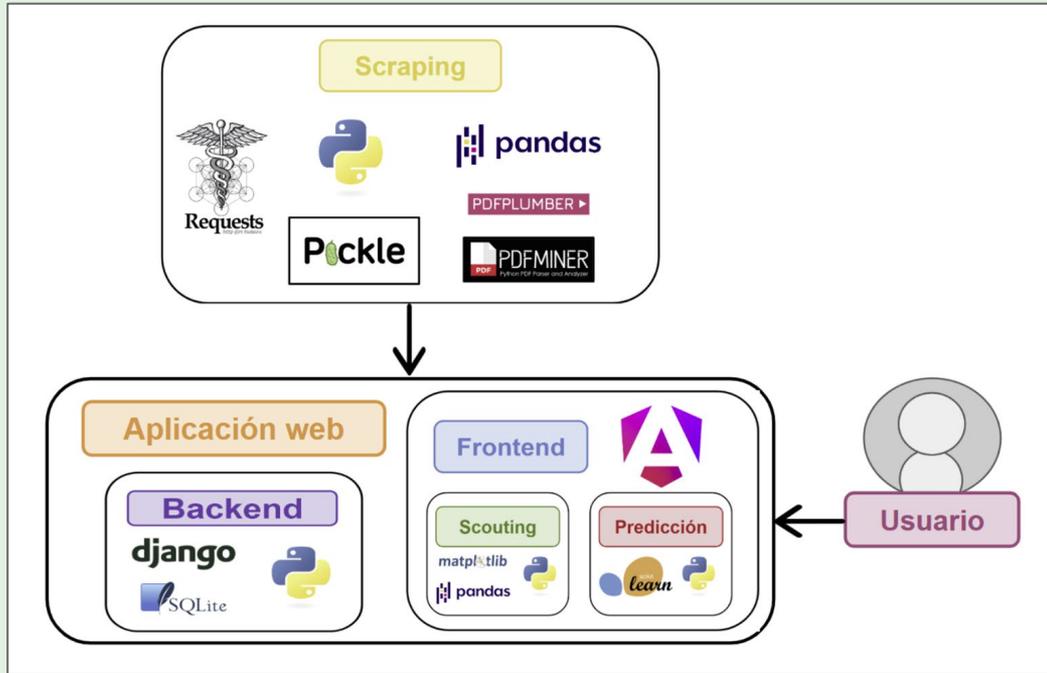
1.2: Objetivos



- Gráficos
- Estadísticas
- Rankings en base a estadísticas
- Comparaciones de jugadores/equipos
- Búsqueda de perfiles
- Clasificación de la liga
- Información completa de partidos disputados
- Pronósticos de futuros enfrentamientos

2.- Diseño e Implementación

2.1- Diseño



Scraping:

- Extracción automatizada de datos.
- Suministra dicha información a la aplicación web.

Aplicación web:

- Núcleo central del sistema.
- 2 funcionalidades principales: **scouting** y **predicción de resultados**.

Usuario: Interactúa de forma directa con la aplicación web.

2.2.- Implementación

2.2.1.- Scraping

2.2.2.- Aplicación web

2.2.3.- Scouting

2.2.4.- Predicción de resultados

LEYENDA:

Funcionalidad existente en la FEB	
Funcionalidad existente en la FEB pero ampliada	
Nueva funcionalidad no existente en la FEB	

2.2.1. Scraping (I)



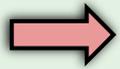
TERCERA
FEB

Fuente: Federación Española de Baloncesto, 2024

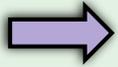


Fuente: Federación Española de Baloncesto, 2024

2.2.1. Scraping (II)



A través de su API se obtienen respuestas que contienen:



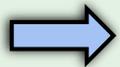
Box Score de cada equipo



Play by play



Shot Chart



Desde su web se descargan los archivos PDF que contienen las estadísticas completas de cada cuarto.

2.2.1. Scraping (III)

➡ A través de su API se obtienen respuestas que contienen:

➡ **Box Score de cada equipo**

➡ **Play by play**

➡ **Shot Chart**

I	D	JUGADOR	MIN	PT	T2	T3	TC	TL	REBOTES			TAPONES			FALTAS			VA	+/-		
									RO	RD	RT	AS	BR	BP	TF	TC	HT			FC	FR
3		J. ATIENZA PEREA	06:50	0	0/0	0/2	0/2	0/0	1	1	2	0	0	1	0	0	0	1	0	-2	-3
5		D. FERNANDEZ CAÑAS	11:17	4	2/3	0/0	2/3	0/0	0	1	1	0	0	1	0	0	0	2	1	2	-15
*	6	J. DOMINGUEZ LARRE	29:24	11	4/5	1/10	5/15	0/0	0	4	4	1	1	1	0	0	0	3	1	4	12
*	9	F. GOMEZ DE ENTERRIA LOPEZ	22:17	6	0/2	2/5	2/7	0/0	1	3	4	0	1	2	0	0	0	0	1	5	16
*	10	J. DE DOMINGO GARAY	12:46	5	1/2	1/3	2/5	0/0	0	1	1	1	0	0	0	0	0	1	3	6	2
14		G. DIAZ MONTERO	25:57	16	2/2	3/9	5/11	3/3	2	2	4	6	2	1	0	0	0	3	2	20	-1
15		A. SANCHO PEREZ	19:40	5	1/4	0/1	1/5	3/6	0	2	2	0	0	0	0	0	0	2	7	5	-18
*	25	G. LARDIES GUILLEN	22:01	11	4/9	0/0	4/9	3/3	5	4	9	2	1	1	0	1	0	4	2	15	27
*	27	A. SIMON DEL CORRAL	32:11	8	4/6	0/0	4/6	0/0	1	5	6	4	4	1	0	0	0	0	0	19	8
29		J. MOLINA MELENDEZ	08:38	8	0/0	2/3	2/3	2/2	0	0	0	0	0	0	0	0	0	1	1	7	-5
34		H. MAIGA	08:59	0	0/1	0/0	0/1	0/0	0	1	1	0	0	0	0	0	0	1	1	0	-13
99		C. BARADJI	00:00	0	0/0	0/0	0/0	0/0	0	0	0	0	0	0	0	0	0	0	0	0	0
			200:00	74	18/34	9/33	27/67	11/14	10	24	34	14	9	8	0	1	0	18	19	81	

Fuente: Federación Española de Baloncesto, 2025

➡ Desde su web se descargan los archivos PDF que contienen las estadísticas completas de cada cuarto.

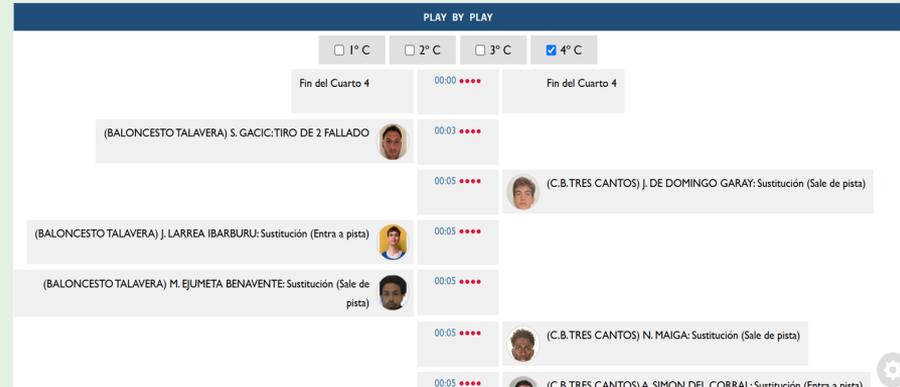
2.2.1. Scraping (IV)

➡ A través de su API se obtienen respuestas que contienen:

➡ Box Score de cada equipo

➡ Play by play

➡ Shot Chart



Fuente: Federación Española de Baloncesto, 2025

➡ Desde su web se descargan los archivos PDF que contienen las estadísticas completas de cada cuarto.

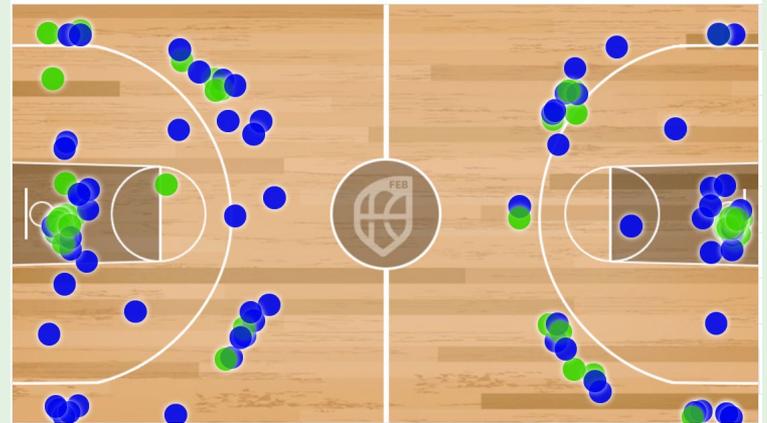
2.2.1. Scraping (V)

➡ A través de su API se obtienen respuestas que contienen:

➡ Box Score de cada equipo

➡ Play by play

➡ Shot Chart



Fuente: Federación Española de Baloncesto, 2025

➡ Desde su web se descargan los archivos PDF que contienen las estadísticas completas de cada cuarto.

2.2.1. Scraping (VI)

➔ A través de su API se obtienen respuestas que contienen:

➔ Box Score de cada equipo

➔ Play by play

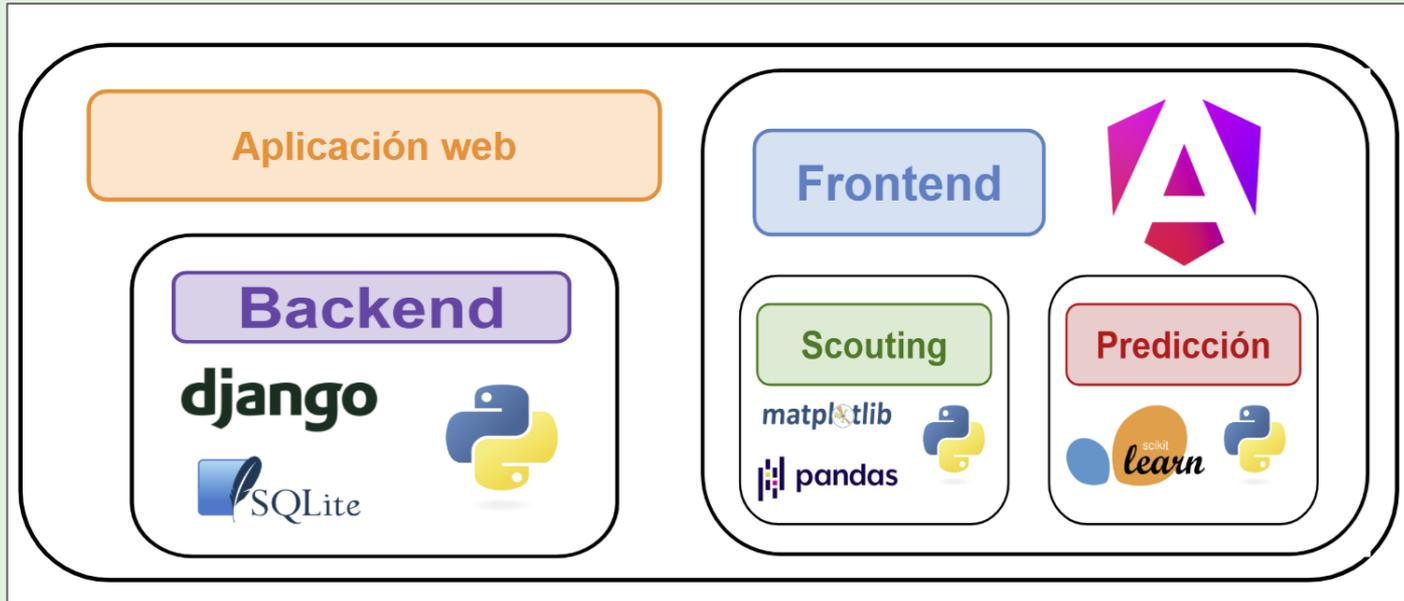
➔ Shot Chart

➔ Desde su web se descargan los archivos PDF que contienen las estadísticas completas de cada cuarto.



Fuente: Federación Española de Baloncesto, 2025

2.2.2.- Aplicación web (I)



2.2.2.- Aplicación web (II)

Bienvenido a la API

Accesos:

- `/api/teams/` : Lista de equipos de la liga
- `/api/players/` : Lista de jugadores de la liga
- `/api/games/` : Lista de partidos jugados de la liga
- `/api/compare_players/?player1=jugador1&player2=jugador2` : Obtiene la comparación entre 'jugador1' y 'jugador2'
- `/api/games_predictions/` : Obtiene las predicciones de los partidos

2.2.2.- Aplicación web (III)



2.2.3- Scouting (I)

- 1.- Tablas convertidas a imágenes
- 2.- Tablas filtrables
- 3.- Comparaciones
- 4.- Gráficos de puntuación de los partidos
- 5.- Gráficos de tiro de los partidos
- 6.- Mapas de calor

2.2.3- Scouting (II)

1.- Tablas convertidas a imágenes

 Clasificación de la liga 

 Rankings de partidos 

 Cuadros de puntuación de partidos 

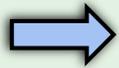
 Estadísticas generales equipos/jugadores 

 Rankings por estadística equipos/jugadores 

PO	Equipo	PJ	PG	PP	PF	PC	PT	R
1	MOVISTAR ESTUDIANTES	13	10	3	1075	953	23	+6
2	CB ARIDANE	13	9	4	1017	884	22	+4
3	REAL CANOE N.C.	13	9	4	1053	982	22	+1
4	C.B. TRES CANTOS	13	9	4	915	852	22	-2
5	LUJISA GUADALAJARA BASKET	13	9	4	1035	973	22	-3
6	SUN CHLORELLA DRAGONS	13	8	5	1014	935	21	+3
7	CB LA MATANZA	13	8	5	1011	973	21	-1
8	BALONCESTO TALAVERA	13	6	7	968	938	19	-1
9	BALONCESTO ALCALA	13	5	8	924	939	18	-1
10	TOSCONES CORRALEJO	13	5	8	826	937	18	-5
11	CABEZUELO CB SOCUÉLLAMOS	13	4	9	891	965	17	-2
12	CESUR DISTRITO OLIMPICO	13	4	9	883	963	17	+2
13	C.B. VALSEQUILLO	13	3	10	834	1006	16	+1
14	PABLO LASO ACADEMY	13	2	11	898	1044	15	+1

2.2.3- Scouting (III)

2.- Tablas filtrables



Búsqueda por perfil de jugadores



Play by play de partidos



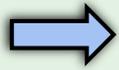
¿Cómo usar los filtros?



- En todas las columnas puedes filtrar por coincidencia: "11", "C.B.", "05:32".
- En las columnas de texto no se distingue por mayúsculas o minúsculas.
- En las columnas de números y tiempos puedes usar operadores para filtrar: ">11", ">=11.3", "<05:32".
- Los operadores disponibles son: "=", "<=", ">=", "<", ">".

2.2.3- Scouting (IV)

3.- Comparaciones

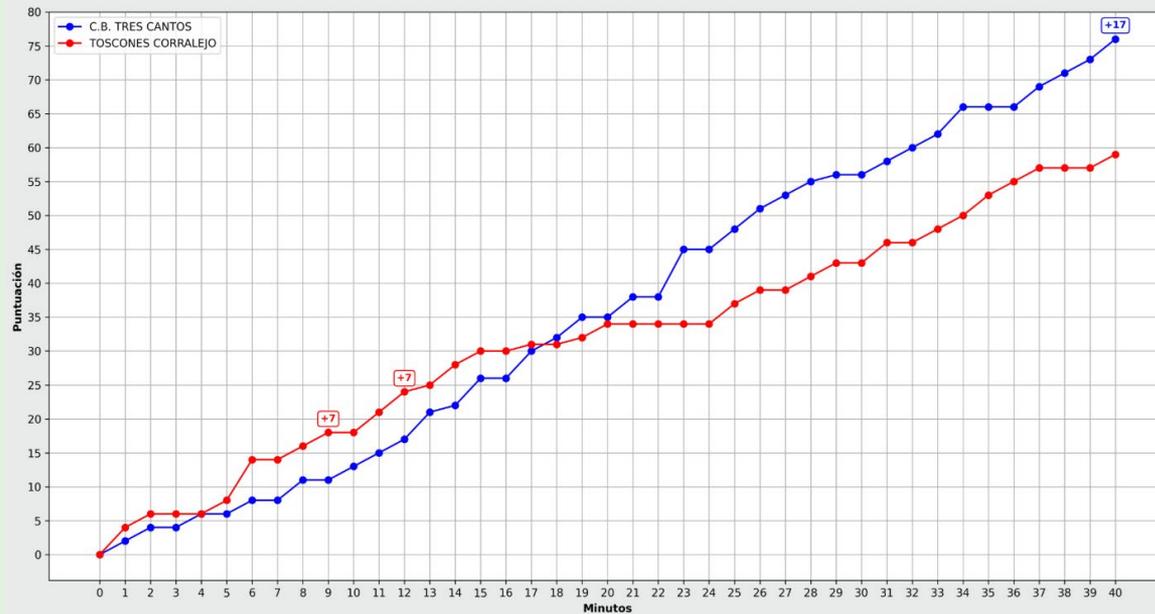


En todos los apartados de comparativas



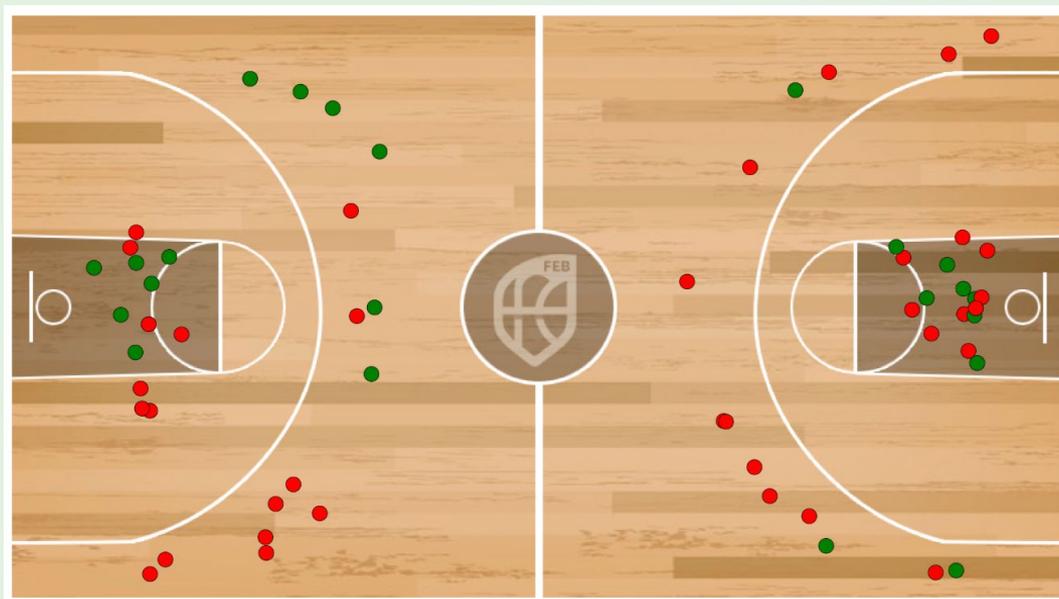
2.2.3- Scouting (V)

4.- Gráficos de puntuación de los partidos



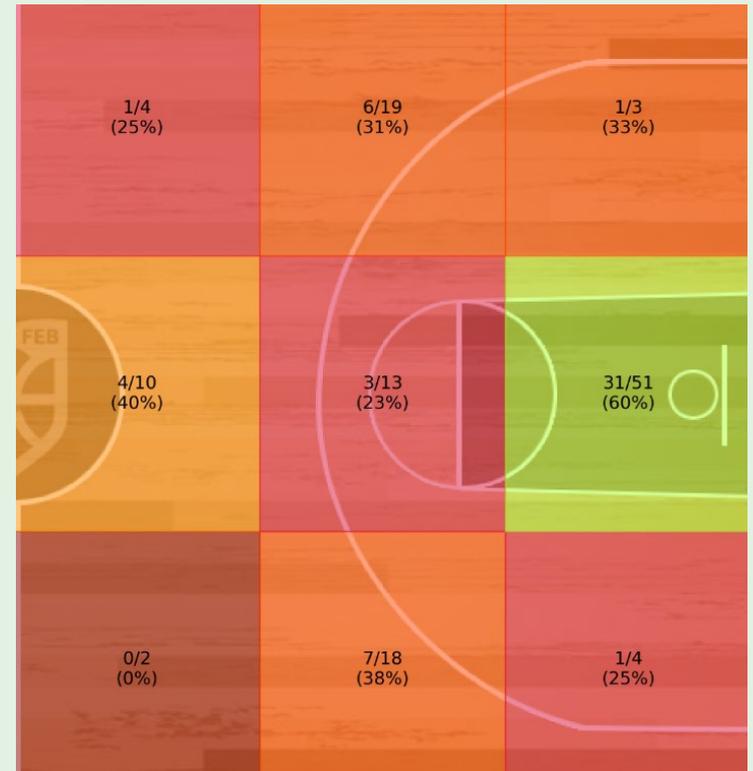
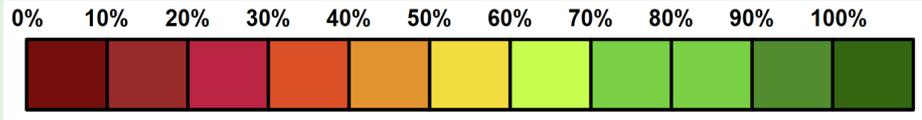
2.2.3- Scouting (VI)

5.- Gráficos de tiro de los partidos



2.2.3- Scouting (VII)

6.- Mapas de calor



2.2.4- Predicción de resultados (I)★

Baremos
principales

$$P = B + C = \sum_{i=1}^n w_i \cdot b_i + C$$

Coefficiente
de ajuste

Donde:

- w_i y b_i son el peso y el valor del i -ésimo baremo, respectivamente.
- n es el número total de baremos considerados.

2.2.4- Predicción de resultados (II)

1- Baremos principales

➡ **win_probability:** probabilidad de victoria del equipo

➡ **site_probability:** probabilidad de victoria como local/visitante (según corresponda)

➡ **points_differential:** diferencia entre los puntos totales anotados y recibidos del equipo

➡ **average_pir:** media de valoración del equipo

➡ **remainder:** racha del equipo

➡ **site:** si el equipo juega como local/visitante/en campo neutral

$$B = (win_probability \times 3 + remainder) \\ + (site_probability \times 2,5 + site) \\ + (points_differential + average_pir)$$

2.2.4- Predicción de resultados (III)

2.- Coeficiente de ajuste

Situación	Valor
Si gana por más de 19 como local	1.5
Si gana por más de 9 como visitante	1.5
Si gana por [10, 19] como local	1
Si gana por [1, 9] como visitante	1
Si gana por [6, 9] como local	0.5
Si gana por [1, 5] como local	0.25
Si no han disputado ningún partido previo	0

Situación	Valor
Si pierde por [1, 5] como visitante	-0.25
Si pierde por [6, 9] como visitante	-0.5
Si pierde por [1, 9] como local	-1
Si pierde por [10, 19] como visitante	-1
Si pierde por más de 9 como local	-1.5
Si pierde por más de 19 como visitante	-1.5

2.2.4- Predicción de resultados (IV)

Seleccione un equipo
BALONCESTO ALC... ▾

Seleccione otro equipo
C.B. TRES CANTOS ▾

Equipos	Predicciones
BALONCESTO ALCALA	3.85
C.B. TRES CANTOS	5.41

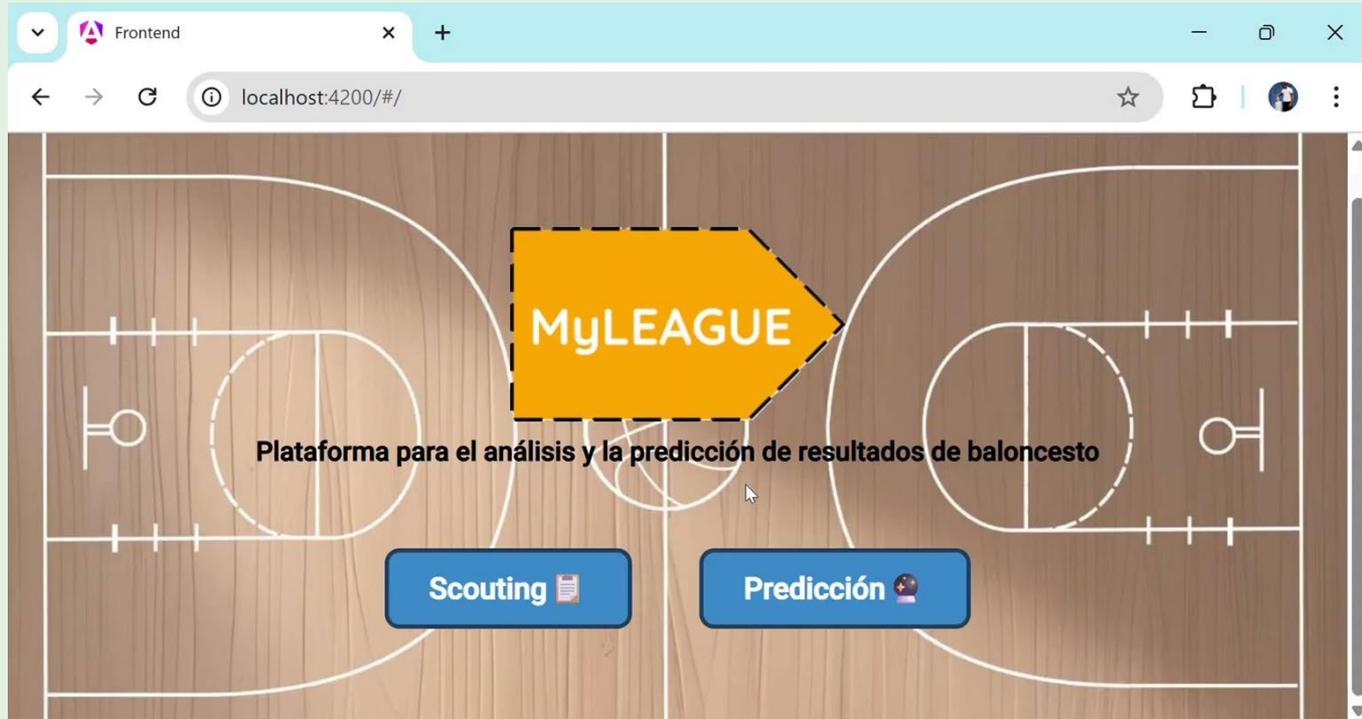
Ganador de la predicción:
C.B. TRES CANTOS



Probabilidad de acierto:
BAJA

3.- Pruebas y Resultados

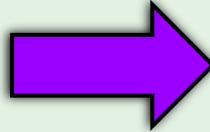
3.1- Sistema



3.2.- Predicción de resultados (I)

Punto de partida:

13 jornadas de
partidos



Objetivo de predicción:

7 jornadas de
partidos



49 partidos

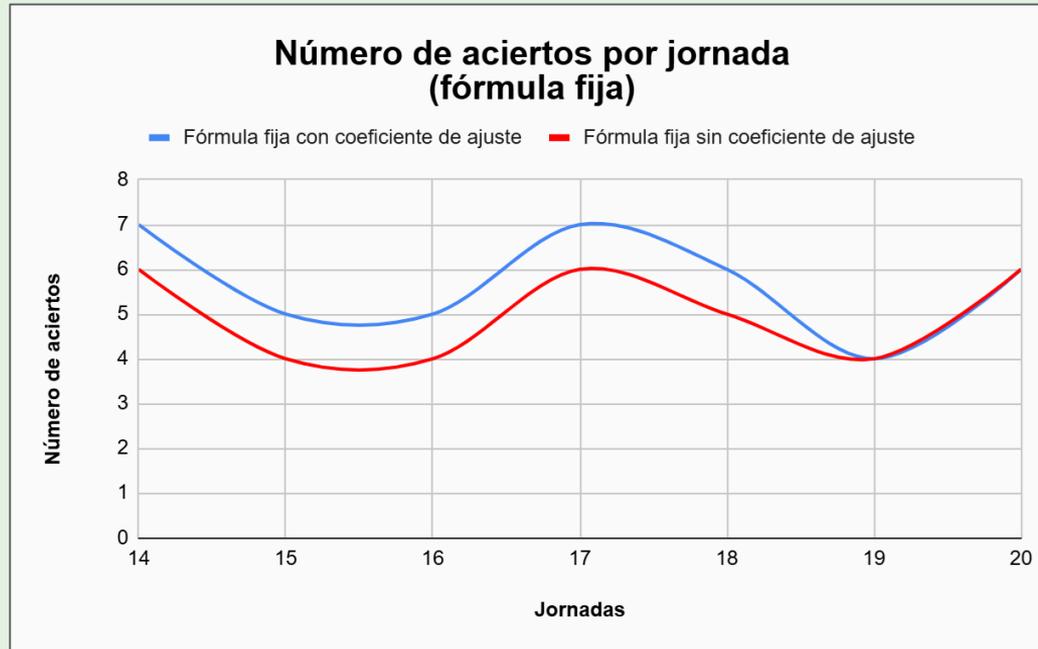
3.2.- Predicción de resultados (II)

Resultados fórmula fija:

Métrica	Sin coeficiente de ajuste	Con coeficiente de ajuste
Aciertos	35/49	40/49
Fallos	14/49	9/49
Accuracy	71.43%	81.63%
Precision (0)	73.53%	81.82%
Recall (0)	83.33%	90.00%
F1 Score (0)	78.12%	85.71%
Precision (1)	66.67%	81.25%
Recall (1)	52.63%	68.42%
F1 Score (1)	58.82%	74.29%

3.2.- Predicción de resultados (III)

Comparativa resultados fórmula fija:



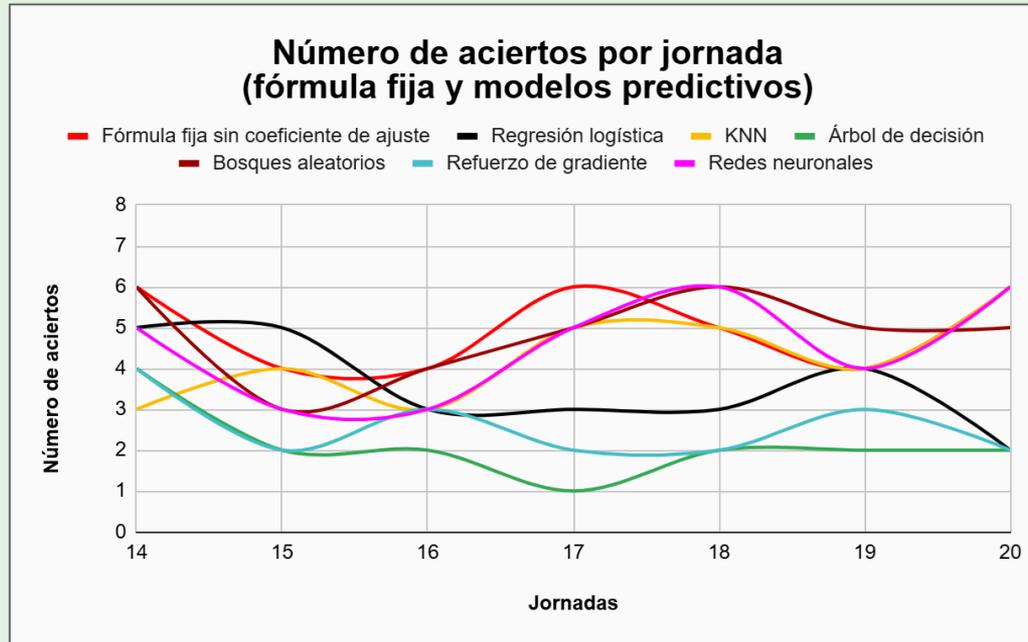
3.2.- Predicción de resultados (IV)

Resultados modelos de aprendizaje automático:

Métrica	Regresión logística	KNN	Árbol de decisión	Bosques aleatorios	Refuerzo de gradiente	Redes neuronales
Aciertos	25/49	30/49	15/49	34/49	18/49	32/49
Fallos	24/49	19/49	34/49	15/49	31/49	17/49
Accuracy	51.02%	61.22%	30.61%	69.39%	36.73%	65.31%
Precision (0)	87.50%	61.22%	30.00%	74.19%	42.86%	70.97%
Recall (0)	23.33%	100.00%	10.00%	76.67%	10.00%	73.33%
F1 Score (0)	36.84%	75.95%	15.00%	75.41%	16.22%	72.13%
Precision (1)	43.90%	0.00%	30.77%	61.11%	35.71%	55.56%
Recall (1)	94.74%	0.00%	63.16%	57.89%	78.95%	52.63%
F1 Score (1)	60.00%	0.00%	41.38%	59.46%	49.18%	54.05%

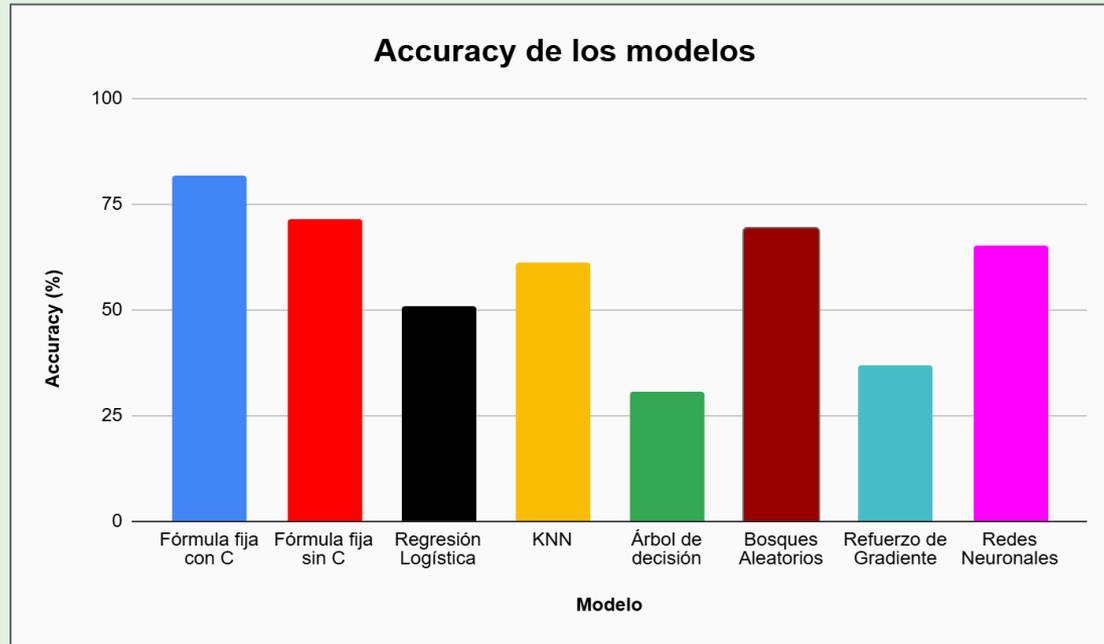
3.2.- Predicción de resultados (V)

Comparativa resultados fórmula fija sin C y modelos de aprendizaje automático:



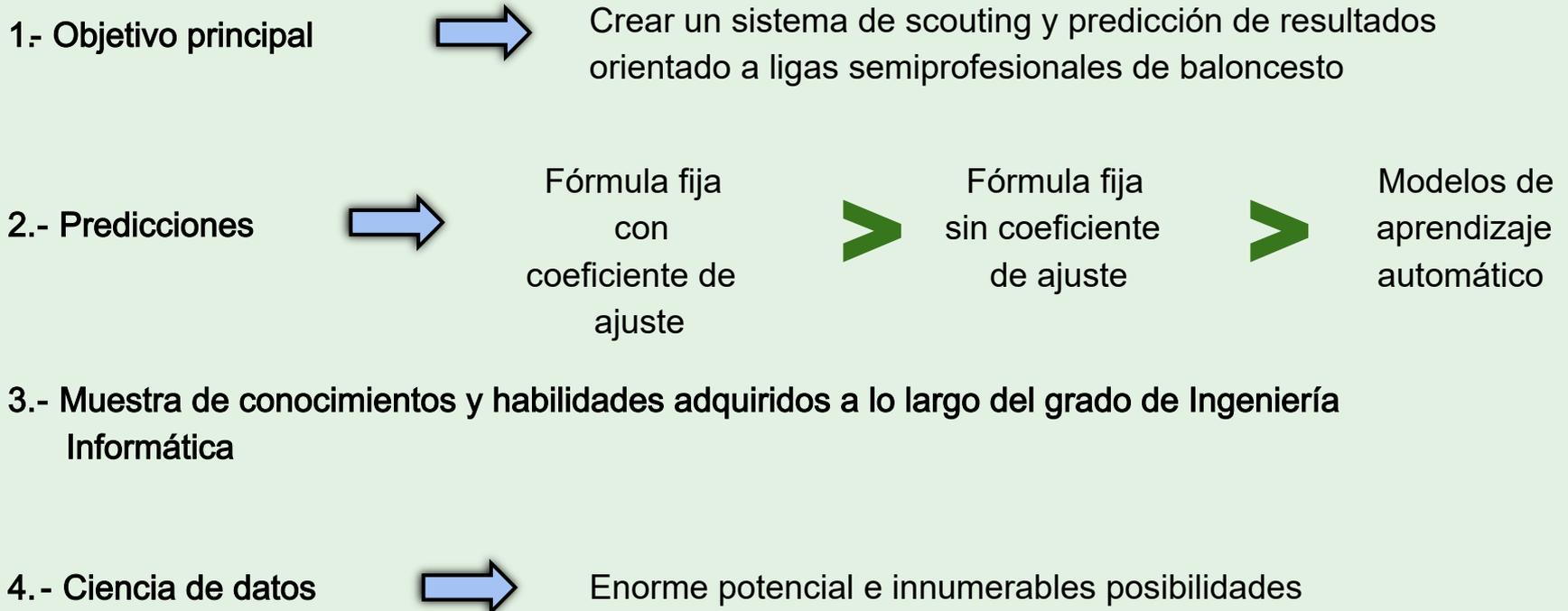
3.2.- Predicción de resultados (VI)

Accuracy de los modelos:



4.- Conclusiones y Trabajo Futuro

4.1- Conclusiones



4.2.- Trabajo futuro

COMPLETO

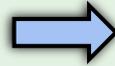
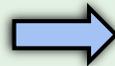
1.- Automatización completa del proceso de representación de datos

2.- Mejora del rendimiento de las predicciones de resultados

3.- Incorporar nuevas estadísticas

4.- Incluir nuevas funcionalidades

ROBUSTO



AUTOMATIZADO

Reduce tiempos de mantenimiento y minimiza posibles riesgos en la introducción de datos

Sistema híbrido que combine la fórmula actual y los modelos de aprendizaje automático

Número de posesiones por equipo en cada encuentro

Generación de informes personalizados o sistema de notificaciones

Gracias por su atención

Álvaro Simón del Corral
Ingeniería Informática
Universidad Autónoma de Madrid

