



# Tomo 06

## *Ingenierías*

# Algoritmo de Machine Learning para la Predicción de un Partido de Fútbol

Julio Cesar Galván Bellman<sup>1</sup>, Dr. Luis Carlos Méndez González<sup>2</sup>,  
Dr. Iván Juan Carlos Pérez Olguín<sup>3</sup>, MC. Abel Eduardo Quezada Carreón<sup>4</sup>

**Resumen**—Se ha desarrollado un algoritmo de machine learning para predecir los resultados de la Premier League, con el objetivo de agilizar el análisis de estadísticas para aquellos que realizan apuestas deportivas. Este algoritmo examina múltiples variables en cuestión de segundos y genera predicciones basadas en dichas estadísticas. Su propósito es proporcionar una herramienta eficiente que ahorre tiempo a los apostadores, eliminando la necesidad de analizar manualmente una gran cantidad de datos. Al hacerlo, ofrece una ventaja competitiva al permitir decisiones informadas y rápidas sobre a qué resultados apostar.

**Palabras clave**—Machine learning, estadísticas, predicción, scraping, algoritmo.

## Introducción

Las apuestas deportivas siempre han estado presentes en el mundo, gracias a la innovación tecnológica, los casinos y las casas de apuestas han experimentado un gran cambio en la forma en que operaban [1]. Ahora, los usuarios pueden disfrutar de las apuestas deportivas desde cualquier lugar del mundo, ya que pueden ingresar desde un dispositivo electrónico a los casinos y plataformas de apuestas de su elección, lo que les brinda una experiencia más cómoda sin tener que salir de su hogar.

En la actualidad, las personas que se toman en serio el tema de realizar apuestas deportivas suelen dedicar una cantidad significativa de tiempo y esfuerzo a analizar las estadísticas y el desempeño de su equipo favorito antes de que se juegue un partido. Este análisis incluye varios aspectos clave, tales como si la plantilla titular es la que va a jugar, cuántos goles ha marcado o recibido el equipo en los últimos encuentros, el historial de enfrentamientos que tienen contra el equipo contrario, dónde se jugará el partido, entre muchos más. Por lo que realizar este tipo de estudio puede llevar varias horas, ya que implica la recopilación y análisis de cada uno de los datos.

Para agilizar la recopilación y análisis de los encuentros, algunas personas han optado por utilizar algoritmos de machine learning, ya que dichos algoritmos pueden realizar predicciones en un lapso más corto de tiempo y más precisas sobre los resultados de los partidos, debido a que pueden analizar una gran cantidad de datos y así lograr brindar una predicción más acertada sobre cómo podría finalizar el encuentro [2][3].

## Metodología

Para que el algoritmo de machine learning funcione adecuadamente para la predicción de partidos de fútbol tenga éxito, primero se debe desarrollar un código de Scraping. El Scraping es una técnica que permite extraer información de sitios web de manera automatizada, enviando solicitudes a las páginas web, analizan su contenido HTML y extraen datos específicos [4]. Ver figura 1.

Después de extraer los datos y almacenarlos en un archivo CSV (valores separados por comas), se procede a cargar esta información en nuestro algoritmo predictor de resultados. Este algoritmo utiliza la técnica de Random Forest para procesar y analizar todos los datos disponibles. Entrenamos el algoritmo con datos históricos, donde cada árbol de decisión en el bosque examina distintos aspectos y subconjuntos de la información [5]. Al combinar las predicciones generadas por todos estos árboles de decisión, obtenemos una predicción más fiable y precisa del resultado del partido. Esta metodología no solo mejora significativamente la precisión de nuestras predicciones, sino que también asegura que se tengan en cuenta múltiples factores y perspectivas, ofreciendo una visión más completa y sólida del posible desenlace del encuentro. Principalmente, los datos que se toman para entrenar el algoritmo son con partidos anteriores al 2023 y para realizar predicciones se basa en partidos que surgieron en el 2023. Ver figura 2.

<sup>1</sup> Julio Cesar Galván Bellman alumno de la carrera de Ingeniería en Mecatrónica de la Universidad Autónoma de Ciudad Juárez [al192780@alumnos.uacj.mx](mailto:al192780@alumnos.uacj.mx)

<sup>2</sup> Dr. Luis Carlos Méndez González es profesor investigador del departamento de ingeniería industrial y manufactura en la Universidad Autónoma de Ciudad Juárez. [luis.mendez@uacj.mx](mailto:luis.mendez@uacj.mx)

<sup>3</sup> Dr. Iván Juan Carlos Pérez Olguín, es profesor investigador del departamento de ingeniería eléctrica y computación en la Universidad Autónoma de Ciudad Juárez. [ivan.perez@uacj.mx](mailto:ivan.perez@uacj.mx)

<sup>4</sup> MC. Abel Eduardo Quezada Carreón es del departamento de eléctrica y computación en la Universidad Autónoma de Ciudad Juárez. [abel.quezada@uacj.com](mailto:abel.quezada@uacj.com)



Figura 1. Mapa conceptual del pseudocódigo del Scraping.

	date	time	round	day	venue	result	gf	ga	opponent	xg	...	fk	pk	pkatt	season	team	target	venue_code	opp_code	hour	day_code
1	2023-08-11	20:00	Matchweek 1	Fri	Away	W	3.0	0.0	Burnley	1.9	...	0.0	0	0	2024	Manchester City	1	0	5	20	4
3	2023-08-19	20:00	Matchweek 2	Sat	Home	W	1.0	0.0	Newcastle Utd	1.0	...	0.0	0	0	2024	Manchester City	1	1	16	20	5
4	2023-08-27	14:00	Matchweek 3	Sun	Away	W	2.0	1.0	Sheffield Utd	3.5	...	2.0	0	1	2024	Manchester City	1	0	18	14	6
5	2023-09-02	15:00	Matchweek 4	Sat	Home	W	5.0	1.0	Fulham	2.2	...	0.0	1	1	2024	Manchester City	1	1	9	15	5
6	2023-09-16	15:00	Matchweek 5	Sat	Away	W	3.0	1.0	West Ham	3.6	...	1.0	0	0	2024	Manchester City	1	0	21	15	5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
42	2023-04-30	14:00	Matchweek 34	Sun	Away	L	1.0	3.0	Newcastle Utd	0.8	...	0.0	0	0	2023	Southampton	0	0	16	14	6
43	2023-05-08	20:00	Matchweek 35	Mon	Away	L	3.0	4.0	Nottingham Forest	3.1	...	0.0	1	1	2023	Southampton	0	0	17	20	0
44	2023-05-13	15:00	Matchweek 36	Sat	Home	L	0.0	2.0	Fulham	0.4	...	0.0	0	0	2023	Southampton	0	1	9	15	5
45	2023-05-21	14:00	Matchweek 37	Sun	Away	L	1.0	3.0	Brighton	0.6	...	1.0	0	0	2023	Southampton	0	0	4	14	6
46	2023-05-28	16:30	Matchweek 38	Sun	Home	D	4.0	4.0	Liverpool	1.8	...	1.0	0	0	2023	Southampton	0	1	12	16	6

1496 rows x 30 columns

Figura 2. Datos con los que trabaja el Random Forest.

La precisión en la predicción de victorias se refiere a cuántas veces nuestro modelo acertó al predecir una victoria. De manera similar, la precisión en la predicción de derrotas indica cuántas veces nuestro modelo acertó al predecir una derrota. Para tener una idea más exacta, combinamos estas dos métricas, considerando la frecuencia de cada resultado en nuestros datos. Así, la precisión combinada es una especie de promedio ponderado, donde las precisiones de victorias y derrotas se multiplican por sus respectivas frecuencias y se dividen por la frecuencia total de partidos. Esto nos da una visión más completa de cómo nuestro algoritmo se desempeña en la predicción de resultados de partidos de fútbol, dando como 0 las veces que el algoritmo acertó y 1 si no acertó al resultado, obteniendo incluso un porcentaje alto de precisión. Ver figura 3 y 4.

```

pd.crosstab(index=combined["actual"], columns=combined["predicted"])

```

predicted	0	1
actual		
0	567	146
1	303	148

Figura 3. Predicciones realizadas.

```
[ ] error = accuracy_score(test["target"], preds)
error
0.6142611683848798
```

Figura 4. Porcentaje de efectividad.

### Resultados

Utilizando el algoritmo, los apostadores pueden analizar una amplia cantidad de datos, en cuestión de minutos. Esta capacidad de procesamiento de grandes cantidades de información permite identificar patrones que de otro modo serían difíciles o tardadas de detectar. Así que, como resultado los apostadores pueden tomar decisiones más informadas y precisas, gracias al algoritmo de machine learning para la predicción de partidos de futbol aumentando sus posibilidades de éxito.

El algoritmo de Scraping solo está diseñado para tomar datos de una temporada anterior a la que se juega y de la que se esté cursando, por lo que, durante las pruebas realizadas sobre el algoritmo de machine learning para la predicción de partidos de futbol ha demostrado ser viable para poder predecir los resultados de los partidos ya que tiene una fiabilidad superior al 50%. Ver figura 3 y 4.

Después de recopilar y procesar los datos, el siguiente paso crucial es visualizar los resultados reales de los partidos y las predicciones generadas por el algoritmo. Esta visualización permite evaluar el desempeño del modelo y entender mejor cómo se comporta en la práctica. En la Figura 5, se puede observar claramente esta comparación, donde los resultados reales del encuentro se representan junto con las predicciones del algoritmo. Además, para una comprensión más detallada se muestra en una tabla, como se ilustra en la Figura 3, que indica las veces que el algoritmo acertó o cuando no lo hizo.

	actual_x	predicted_x	date	team_x	opponent_x	result_x	new_team_x	actual_y	predicted_y	team_y	opponent_y	result_y	new_team_y
0	0	1	2023-01-03	Arsenal	Newcastle Utd	D	Arsenal	0	0	Newcastle United	Arsenal	D	Newcastle Utd
1	1	1	2023-01-15	Arsenal	Tottenham	W	Arsenal	0	0	Tottenham Hotspur	Arsenal	L	Tottenham
2	1	1	2023-01-22	Arsenal	Manchester Utd	W	Arsenal	0	0	Manchester United	Arsenal	L	Manchester Utd
3	0	0	2023-02-04	Arsenal	Everton	L	Arsenal	1	0	Everton	Arsenal	W	Everton
4	0	1	2023-02-11	Arsenal	Brentford	D	Arsenal	0	0	Brentford	Arsenal	D	Brentford
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1050	0	0	2024-04-20	Wolverhampton Wanderers	Arsenal	L	Wolves	1	1	Arsenal	Wolves	W	Arsenal
1051	0	0	2024-04-24	Wolverhampton Wanderers	Bournemouth	L	Wolves	1	0	Bournemouth	Wolves	W	Bournemouth
1052	1	0	2024-04-27	Wolverhampton Wanderers	Luton Town	W	Wolves	0	0	Luton Town	Wolves	L	Luton Town
1053	0	0	2024-05-04	Wolverhampton Wanderers	Manchester City	L	Wolves	1	1	Manchester City	Wolves	W	Manchester City
1054	0	0	2024-05-11	Wolverhampton Wanderers	Crystal Palace	L	Wolves	1	1	Crystal Palace	Wolves	W	Crystal Palace

1055 rows x 13 columns

Figura 5. Resultados reales y predicciones del partido.

### **Conclusiones**

Aunque predecir resultados deportivos es todo un reto, con el algoritmo de machine learning para la predicción de partidos de futbol, los apostadores pueden contar con una herramienta que ofrece una ventaja que puede ser significativa sobre los casinos o plataformas de apuestas, ya que, el algoritmo ha demostrado una efectividad superior al 50% lo cual lo convierte en una herramienta confiable y rentable para todos aquellos que estén interesados en realizar apuestas deportivas.

El proyecto del algoritmo de machine learning para la predicción de partidos de futbol aún tiene mucho camino por recorrer. Se seguirá realizando pruebas y mejorando el algoritmo para hacerlo más efectivo y útil para todas aquellas personas que quieran realizar apuestas deportivas. En resumen, consideramos que hasta ahora el proyecto ha sido un éxito.

### **Referencias**

- [1] Schwartz, D. G. (2019). Futures of gaming: how casinos and gambling might evolve in the near future. *Gaming Law Review*, 23(5), 306-318.
- [2] Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F., & Baca, A. (2023). Machine learning application in soccer: a systematic review. *Biology of sport*, 40(1), 249-263.
- [3] Capobianco, G., Di Giacomo, U., Mercaldo, F., Nardone, V., & Santone, A. (2019). Can machine learning predict soccer match results?. In *ICAART* (2) (pp. 458-465).
- [4] Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3).
- [5] Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 13(1), 1063-1095.