



Universidad del Desarrollo
Facultad de Ingeniería

METODOLOGÍAS DE ESTIMACIÓN DEL VALOR DE JUGADORES EN EL FÚTBOL
PROFESIONAL: UN ENFOQUE DE DATA SCIENCE

POR: MIGUEL RUIZ BENAVIDES

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dra. Loreto Bravo

Diciembre 2024

SANTIAGO

Dedico este trabajo a mi familia, cuyo apoyo y confianza han sido fundamentales para alcanzar este objetivo. Su acompañamiento ha sido mi mayor motivación.

AGRADECIMIENTO

Agradezco a mi familia por su apoyo y confianza incondicional, que han sido fundamentales para alcanzar este objetivo. Su acompañamiento ha sido mi mayor inspiración.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	4
3. HIPÓTESIS Y OBJETIVOS.....	10
4. DATOS Y METODOLOGÍA	12
4.1. DATOS	12
4.2. METODOLOGÍA	13
5. RESULTADOS	19
5.1. ANÁLISIS EXPLORATORIO DE DATOS:	19
5.2. EXTRACCIÓN DE DATOS	26
5.3. ELECCIÓN DE VARIABLES.....	28
<i>Delanteros.....</i>	<i>31</i>
<i>Mediocampistas:.....</i>	<i>31</i>
<i>Defensas.....</i>	<i>32</i>
<i>Arquero</i>	<i>32</i>
5.4. PROBLEMÁTICAS	33
5.5. REGRESIONES	34
<i>Delanteros.....</i>	<i>35</i>
<i>Mediocampistas.....</i>	<i>36</i>
<i>Defensas.....</i>	<i>37</i>
<i>Arquero</i>	<i>38</i>
<i>Resumen.....</i>	<i>39</i>
5.6. BASE DE DATOS	40
5.7. EXPORTACIÓN DE MODELO	42

5.8.	VISUALIZACIÓN	43
5.9.	GITHUB	52
6.	CONCLUSIONES	54
	BIBLIOGRAFÍA	56

Resumen

Este proyecto aborda la estimación del valor de mercado de los futbolistas mediante el uso de metodologías de machine learning, redes neuronales y análisis de datos. Se desarrollaron modelos predictivos para calcular los valores de los jugadores, integrando variables deportivas, económicas y sociales. Como parte del análisis, se emplearon herramientas como Transfermarkt y técnicas de scraping en redes sociales para incrementar la precisión de las valoraciones.

Para garantizar la utilidad y replicabilidad del trabajo, se creó una aplicación web utilizando Flask, permitiendo a los usuarios visualizar las predicciones de manera interactiva. Además, el código completo del proyecto, incluyendo el aplicativo Flask, se publicó en un repositorio de GitHub para facilitar su acceso y reutilización.

1. Introducción

Durante más de una década, los equipos de la Premier League han buscado formas de determinar el valor real de un jugador, considerando factores tangibles como el rendimiento, la edad, la posición y las condiciones del contrato, así como otros aspectos intangibles. Aunque algunos equipos han desarrollado modelos de valoración complejos, estos no ofrecen una cifra exacta, sino una estimación aproximada del valor económico del futbolista. A pesar de estos esfuerzos, los precios en el mercado de transferencias suelen ser impredecibles y no siempre reflejan el valor real de los jugadores. El entrenador del Chelsea, Antonio Conte, describió este fenómeno como "una locura", lo que evidencia la volatilidad del mercado.

La ausencia de un patrón claro en la fijación de precios ha llevado a que muchos clubes y directores deportivos confíen en sus instintos y conocimientos del mercado, generando un entorno opaco. Si bien, algunos equipos aplican metodologías más formales, el contexto del mercado y las circunstancias específicas influyen notablemente en el precio final de un jugador. En última instancia, el valor de un futbolista dependerá de lo que un club esté dispuesto a pagar, lo que añade una capa de complejidad al proceso de valoración.

Casos recientes en el fútbol, como el de la Juventus, han puesto de manifiesto cómo algunos clubes manipulan fraudulentamente los precios de los jugadores para evitar riesgos contables. Un ejemplo emblemático fue el intercambio de jugadores entre la Juventus y el Barcelona en 2020, donde los precios inflados de Miralem Pjanic y Arthur Melo revelaron discrepancias difíciles de justificar. Este tipo de prácticas sugiere la

necesidad de contar con mecanismos más transparentes y precisos que permitan verificar las transacciones y detectar posibles infracciones legales en el mercado de transferencias.

El mercado de transferencias de futbolistas sigue siendo una industria multimillonaria marcada por la influencia de factores deportivos y económicos. A pesar de los esfuerzos por lograr mayor transparencia en la valoración de jugadores, la variabilidad e incertidumbre en los precios continúa siendo significativa. Este estudio ha propuesto la implementación de modelos predictivos basados en machine learning y redes neuronales como una solución para disminuir la subjetividad en este proceso, aprovechando datos de plataformas especializadas como Transfermarkt y redes sociales. A través de estas herramientas de análisis de datos, se busca proporcionar a los clubes una mayor precisión y fundamento en sus decisiones de compra y venta de jugadores, contribuyendo así a un mercado de transferencias más eficiente y justo.

2. Trabajo Relacionado

Las habilidades en data science son esenciales para realizar estudios basados en datos, ya que la falta de conocimientos en programación, estadística o análisis puede obstaculizar la toma de decisiones fundamentadas y generar conclusiones imprecisas. Un estudio de (Phatak) resalta cómo los macrodatos impactan la toma de decisiones en el deporte, pero señala que sólo una pequeña parte de estos datos se utiliza debido a la falta de habilidades en programación, mencionando herramientas como OPTA y bibliotecas de Python y R para facilitar el análisis.

Por otro lado, los datos de rendimiento deportivo son cruciales para el análisis de aspectos como el desempeño individual y colectivo en el fútbol. En el trabajo de (Mohammad Mahdi Nasiri, Mojtaba Ranjbar, Madjid Tavana, Francisco J. Santos Arteaga, & Reza Yazdanparast, 2018), se destaca la importancia de la calidad de los jugadores para el éxito de un equipo. (Daniel Dinsdale & Joe Gallagher, 2022) abordan el reto de predecir el rendimiento de jugadores transferidos a otros clubes. Finalmente, (Moreno, 2022): desarrolla una red neuronal para predecir el rendimiento de los jugadores, utilizando Machine Learning para identificar patrones y predecir errores con alta precisión.

La aplicación de enfoques de machine learning y data science en el fútbol ha sido ampliamente documentada en la literatura, demostrando su efectividad en varios contextos.

Algunos estudios clave incluyen:

- (Frutos, 2020) Analiza la relación entre ingresos por taquilla, derechos televisivos y marketing con el valor de mercado de los clubes, utilizando un modelo de regresión lineal que establece que los ingresos comerciales aumentan con el valor de mercado de los equipos.
- (Sylvio Barbon Junior, Felipe Arruda Moura , & Ricardo da Silva Torres , 2024) : Exploran cómo los métodos basados en datos pueden analizar el fútbol, un deporte con complejas interacciones y abundantes datos. Se enfocan en automatizar la recopilación, transformación y análisis de datos para tomar decisiones inteligentes, destacando el impacto de la inteligencia artificial y la ciencia del deporte en la optimización del rendimiento de atletas y equipos.
- (Miohk Ryoo, Namjung Kim , & Kyoungju Park , 2017) Proponen un sistema innovador para analizar y visualizar el rendimiento de los equipos de fútbol, centrándose en las relaciones causa-efecto entre jugadores y equipos a partir de datos de transferencias.
- (Vineet M. Payyappalli & Jun Zhuang, 2019) Presentan un modelo matemático que ofrece recomendaciones óptimas para la transferencia de jugadores entre clubes, utilizando datos públicos sobre jugadores y equipos.
- (Uribe) : Destaca la necesidad de modernizar los métodos de valoración de futbolistas en un contexto de globalización y cambios financieros. Propone el uso de modelos como el M.V.A. (Market Value Added) y el EVA (Economic Value Added) para tratar a los jugadores como activos intangibles, mejorando la toma de decisiones y optimizando el valor agregado al club.

En los últimos años, el precio de los jugadores ha experimentado un aumento sostenido e inexplicable, lo que requiere la implementación de metodologías actualizables anualmente para la estimación del valor de los futbolistas. Por ejemplo, (Cenalmor) propone el modelo CIES, que considera variables de los clubes (desempeño deportivo y financiero), de los jugadores (contrato, edad, rendimiento) y una variable contextual (inflación del mercado). Este modelo ha demostrado ser preciso en las grandes ligas europeas, con un R^2 del 85%, reflejando de manera adecuada las dinámicas del mercado. No obstante, no incluye factores como tensiones internas, urgencias financieras o el potencial comercial de los jugadores, aspectos que podrían mejorar con la incorporación de métricas adicionales, como el número de seguidores en redes sociales.

Lo anterior sugiere que el mercado deportivo sigue siendo impredecible, por lo tanto, se hace necesario el uso de metodologías avanzadas para predecir los precios de transferencias, como se puede observar en diversos estudios.

- (Aaron Santiago Pedraza Cárdenas, 2023) desarrollar un modelo predictivo basado en datos deportivos para estandarizar la valoración de jugadores, accesible a clubes, entrenadores, agentes y sponsors.
- (Alejandro Guzman & Emilio Catapano, 2024): Este trabajo presenta un modelo basado en el Proceso Analítico Jerárquico (AHP) para estimar el valor de mercado de futbolistas delanteros. El modelo considera tres criterios principales: deportivo, personal y profesional, desglosados en subcriterios. La propuesta reúne técnicas

de valoración inmobiliaria adaptadas al fútbol, incorporando el valor del contrato del jugador en el análisis. Esto permite evaluar si un jugador está subvalorado o sobrevalorado al comparar su desempeño integral con el de jugadores comparables y su contrato. La metodología fue aplicada a Lionel Messi y Cristiano Ronaldo, analizando su desempeño durante los nueve años que coincidieron en la liga española, proporcionando una herramienta para estimar el valor real de mercado de los jugadores.

- (Markel Rico-González, José Pino-Ortega, Amaia Méndez, Filipe Manuel Clemente, & Arnold Baca, 2023) : La naturaleza caótica del fútbol presenta desafíos para los modelos estadísticos predictivos en la toma de decisiones basadas en evidencia. Este estudio identificó sistemáticamente investigaciones que aplicaron el aprendizaje automático (ML)
- (BOSMAN): Evalúa modelos de regresión para predecir tarifas de transferencia de jugadores, concluyendo que la regresión lineal es la más efectiva en esta tarea.
- (Pedraza Cárdenas & Sánchez Castañeda, 2023) Tiene como objetivo identificar las variables deportivas que más impactan el valor de los futbolistas y desarrollar un modelo de Machine Learning para estimar sus precios en ligas europeas. Utilizando datos de Kaggle y Transfermarkt, se busca estandarizar el cálculo de precios para facilitar estimaciones basadas en datos y apoyar negociaciones entre clubes. El proyecto crea un dataframe por posición y utiliza varios modelos para estimar el valor de los jugadores:

- Regresión Lineal: Predice el valor basándose en variables como altura, nacionalidad, y posición.
- Support Vector Regression (SVR): Establece relaciones no lineales entre variables de entrada y el precio estimado.
- Gradient Boosting Regressor: Combina varios modelos de bajo nivel para mejorar la precisión a través de iteraciones.
- Random Forest: Utiliza múltiples árboles de decisión para generar una estimación combinada.
- Red Neuronal: Emplea múltiples capas de decisión para predecir el precio basándose en las variables de los jugadores.

Lo anterior resalta la importancia de utilizar modelos de regresión como herramienta clave para estimar el precio de los futbolistas. Sin embargo, basarse exclusivamente en el rendimiento deportivo puede ser un enfoque limitado o erróneo al valorar a un jugador. Es fundamental considerar también factores como su popularidad y el impacto emocional que genera en los aficionados. Algunos estudios relevantes en este ámbito incluyen:

- (González Baigorri & Marí Roselló, 2021) Investiga los factores que influyen en la valoración de mercado de los jugadores de fútbol, incorporando un índice de sentimiento basado en la opinión de los aficionados y utilizando datos del videojuego FIFA desde 2014/15 hasta 2021/22. Mediante regresión lineal múltiple

- (Agustina Margaritis, y otros) : Este estudio investiga los factores que influyen en el valor de mercado de los jugadores de fútbol, así como el impacto de las redes sociales en dicho valor

Los estudios anteriores mencionan datos de rendimiento como la popularidad, pero también hay un factor señalado por (Geurts, 2022), quien estudia lesiones, factores de riesgo, anatomía de los jugadores y análisis táctico mediante data-tracking, su objetivo es evaluar tanto el rendimiento individual como colectivo, transformando los resultados en información valiosa para mejorar el desempeño del equipo y agilizar la toma de decisiones.

3. Hipótesis y Objetivos

- **Hipótesis**

Los modelos predictivos basados en machine learning y redes neuronales pueden mejorar significativamente la precisión en la estimación del valor de mercado de los jugadores de fútbol, integrando tanto datos deportivos como sociales (como seguidores en redes sociales).

- **Objetivos generales**

Desarrollar un modelo predictivo basado en machine learning y redes neuronales que permita estimar con precisión el valor de mercado de los jugadores de fútbol, integrando tanto datos de rendimiento deportivo como variables externas, incluyendo la popularidad en redes sociales.

- **Objetivos Específicos**

- Recopilar y unificar datos de diversas fuentes (Transfermarkt, redes sociales, rendimiento en partidos) para formar un conjunto de datos adecuado para el análisis.
- Implementar varios modelos de machine learning (Regresión Lineal, Support Vector Regression, Random Forest, y Red Neuronal) para predecir el valor de mercado de los jugadores de fútbol.

- Evaluar el impacto de variables no deportivas, como el número de seguidores en redes sociales, en la valoración económica de los jugadores, a través de técnicas de análisis de datos y machine learning.
- Comparar la precisión de los modelos predictivos propuestos utilizando métricas como R^2 , y RMSE con el fin de seleccionar el modelo que ofrezca las mejores predicciones.
- Desarrollar una plataforma interactiva en Flask que permita a los usuarios visualizar los resultados de las predicciones y los factores que más influyen en la valoración de los jugadores.

4. Datos y Metodología

4.1. Datos

Los datos corresponden a rendimiento de jugadores obtenidos de Wyscout

E	F	G	H
Vencimiento contrato ▾	Partidos jugados ▾	Minutos jugados ▾	Goles ▾
2023-06-30 00:00:00	51	4534	25
2024-06-30 00:00:00	45	3575	1
2024-12-31 00:00:00	47	4276	2
2023-06-30 00:00:00	29	2013	2
2024-12-31 00:00:00	46	3698	14
	42	3828	1
2024-12-31 00:00:00	46	3574	13
2024-12-31 00:00:00	40	3537	0
2024-12-31 00:00:00	28	2097	11
2024-12-31 00:00:00	31	2577	1
2023-12-31 00:00:00	27	2248	1
2025-06-30 00:00:00	30	1846	1
2024-06-30 00:00:00	26	1443	0
2023-12-31 00:00:00	43	3738	1

Se utilizarán además datos de transfermarkt:

- Transfermarkt: Rumores
- Transfermarkt: Valor de transferencia
- Transfermarkt: Lesiones

También se utilizarán datos de seguidores de redes sociales de Instagram

- Web scrapping a Bing: Número de seguidores en Instagram.
- Además, ocupando la biblioteca de Python pytrends podemos obtener datos de popularidad de un jugador en Google Trends durante la última semana.

- Finalmente obtenemos las publicaciones relacionadas con un jugador en el subreddit de fútbol (r/soccer) con la cantidad de publicaciones encontradas y el total de (votos positivos).

4.2. Metodología

Según el sitio web transfermarkt evalúan el precio de los jugadores según:

- Perspectivas de futuro
- Edad
- Rendimiento en el club y la selección nacional
- Nivel y estatus de la liga, tanto en términos deportivos como económicos
- Reputación/prestigio
- Potencial de desarrollo
- Características específicas de la liga
- Valor de marketing
- Número y reputación de clubes interesados
- Potencial de rendimiento
- Nivel de experiencia
- Propensión a las lesiones
- Diferentes condiciones financieras de clubes y ligas
- Demanda general y "tendencias" en el mercado
- Desarrollo general del precio de los fichajes

- Factores externos como la pandemia de coronavirus y sus consecuencias

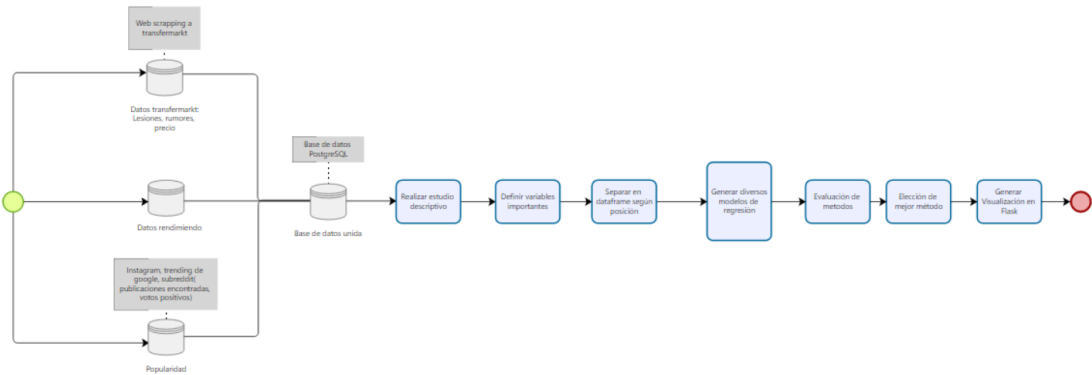
Ante esto, lo que se contempla es determinar el precio de los jugadores teniendo como base los datos de rendimiento, y se complementarán con los obtenidos de transfermarkt realizando un scraping para su obtención:

- Precio
- Rumores
- Lesiones

Se analizará el estudio de (Geurts, 2022) para ver el impacto que tienen las lesiones en el valor, no obstante, en la construcción se analizará si este dato formará parte del modelo.

Además, se debe obtener un scraping para obtener los seguidores que tienen los futbolistas en redes sociales siguiendo los estudios de (González Baigorri & Marí Roselló, 2021) y (Agustina Margaritis, y otros)

Dicho procedimiento se visualiza a continuación:



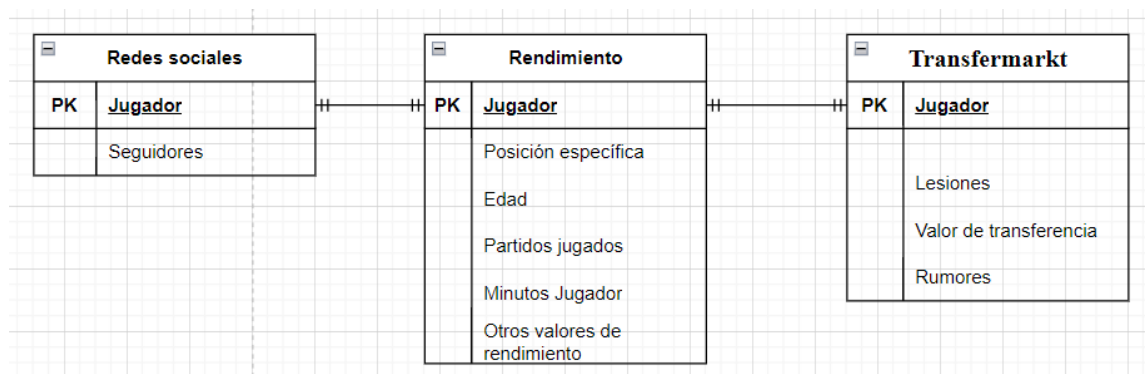
Los tres sets de datos (rendimiento de jugadores, transfermarkt y los seguidores de redes sociales se unirán en una sola base de datos PostgreSQL

Los campos por considerar son:

- Posición específica
- Edad
- Valor de mercado
- Vencimiento contrato
- Partidos jugados
- Minutos jugados
- Otros valores de rendimiento a considerar
- Nueva variable (Jugador juega en medio local)
- Transfermarkt: Rumores

- Transfermarkt : Valor de transferencia
- Transfermarkt: Lesiones
- Web scrapping a Bing: Número de seguidores en Instagram

Un ejemplo de modelo de datos propuestos es:



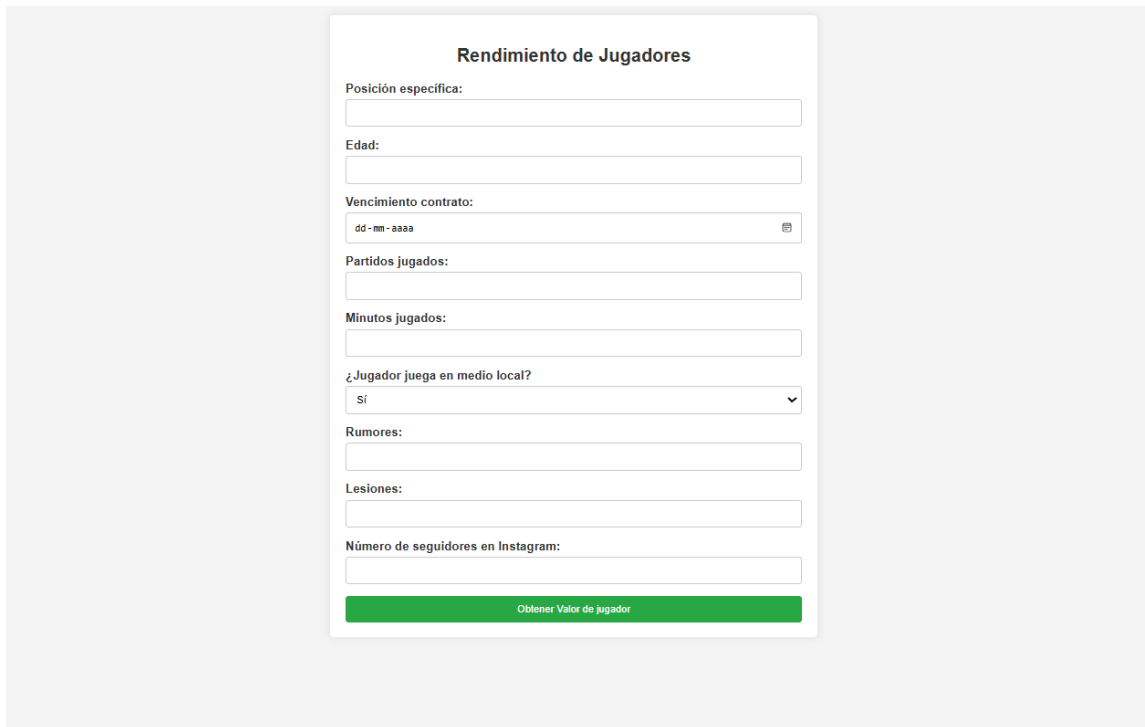
Se generarán diferentes modelos, siguiendo lo realizado por (Pedraza Cárdenas & Sánchez Castañeda, 2023)

- Regresión Lineal
- Support Vector Regression (SVR)
- Gradient Boosting Regressor
- Random Forest
- Red Neuronal
- Regresión polinómica

Tal como menciona (Pedraza Cárdenas & Sánchez Castañeda, 2023) en vez de generar un dataframe por posición, se crea una variable numérica que indique la posición.

Se evaluará y guardará el mejor modelo, con ello se generará un aplicativo de visualización en Flask

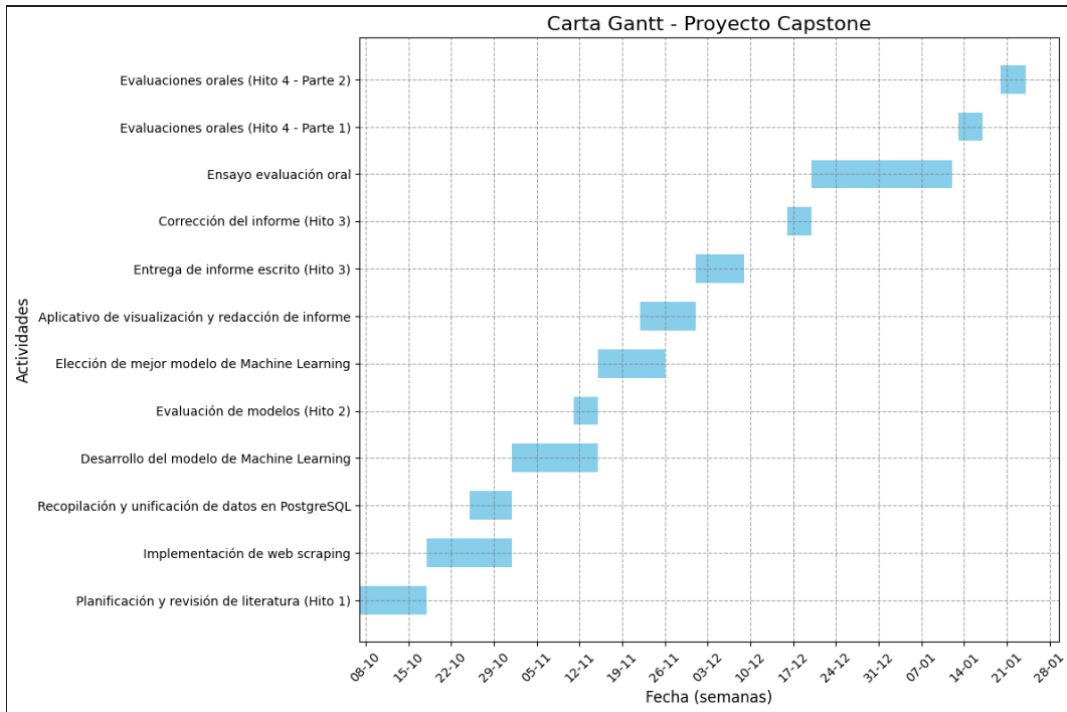
El modelo propuesto es:



The image shows a web application interface titled "Rendimiento de Jugadores". It contains several input fields for user data: "Posición específica:", "Edad:", "Vencimiento contrato:" (with a date format "dd-mm-aaaa" and a calendar icon), "Partidos jugados:", "Minutos jugados:", "¿Jugador juega en medio local?" (a dropdown menu with "Sí" selected), "Rumores:", "Lesiones:", and "Número de seguidores en Instagram:". At the bottom, there is a green button labeled "Obtener Valor de jugador".

La carta Gantt del proceso se visualiza a continuación:

Actividad	Inicio	Fin
Planificación y revisión de literatura (Hito 1)	2024-10-07 00:00:00	2024-10-18 00:00:00
Implementación de web scraping	2024-10-18 00:00:00	2024-11-01 00:00:00
Recopilación y unificación de datos en PostgreSQL	2024-10-25 00:00:00	2024-11-01 00:00:00
Desarrollo del modelo de Machine Learning	2024-11-01 00:00:00	2024-11-15 00:00:00
Evaluación de modelos (Hito 2)	2024-11-11 00:00:00	2024-11-15 00:00:00
Elección de mejor modelo de Machine Learning	2024-11-15 00:00:00	2024-11-26 00:00:00
Aplicativo de visualización y redacción de informe	2024-11-22 00:00:00	2024-12-01 00:00:00
Entrega de informe escrito (Hito 3)	2024-12-09 00:00:00	2024-12-01 00:00:00
Corrección del informe (Hito 3)	2024-12-16 00:00:00	2024-12-20 00:00:00
Ensayo evaluación oral	2024-12-20 00:00:00	2025-01-12 00:00:00
Evaluaciones orales (Hito 4 - Parte 1)	2025-01-13 00:00:00	2025-01-17 00:00:00
Evaluaciones orales (Hito 4 - Parte 2)	2025-01-20 00:00:00	2025-01-24 00:00:00



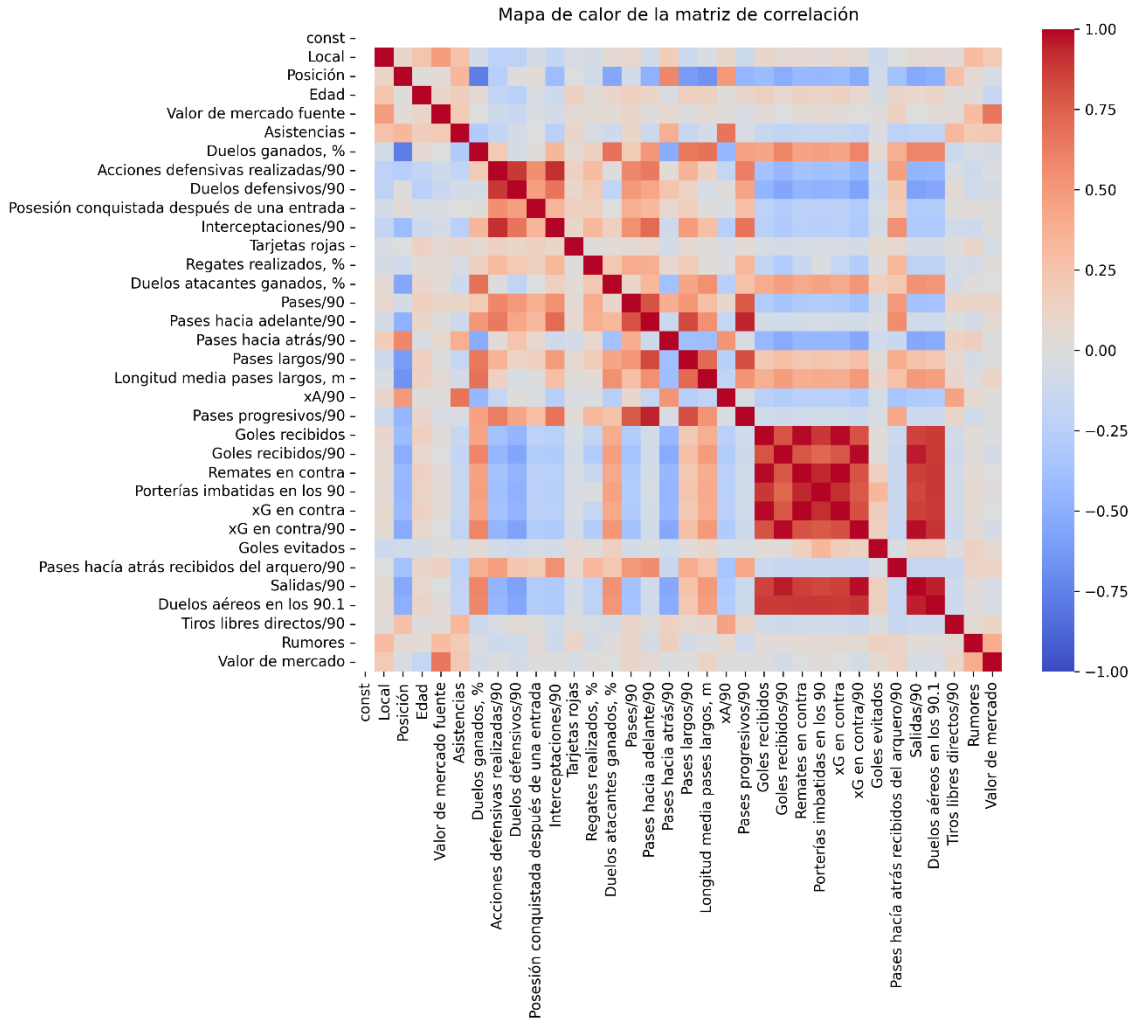
5. Resultados

5.1. Análisis exploratorio de datos:

Se comienza el análisis exploratorio con las correlaciones más importantes:

Valor de mercado	1.000000
Valor de mercado fuente	0.657457
Rumores	0.403402
Asistencias	0.219730
Local	0.209413
Pases hacía atrás recibidos del arquero/90	0.127326
Longitud media pases largos, m	0.123936
Pases/90	0.115512
Tiros libres directos/90	0.101913
Goles evitados	0.058824
Pases hacia adelante/90	0.057219
Porterías imbatidas en los 90	0.026509
Regates realizados, %	0.018045
Pases largos/90	0.014653
Posesión conquistada después de una entrada xA/90	0.010868 0.009109
Interceptaciones/90	0.008778
Pases progresivos/90	0.001383
Remates en contra	-0.001287
Duelos atacantes ganados, %	-0.003789
xG en contra	-0.003893
Goles recibidos	-0.012039
Pases hacia atrás/90	-0.012823
Acciones defensivas realizadas/90	-0.018263
Duelos aéreos en los 90.1	-0.038269

El mapa de calor con las correlaciones se visualiza en:



Si utilizamos otro método para calcular correlaciones como spearman:

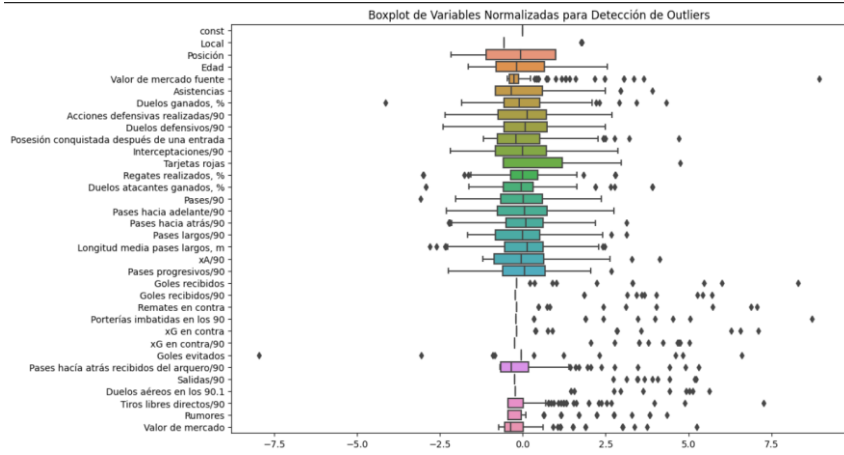
Valor de mercado	1.000000
Rumores	0.482201
Valor de mercado fuente	0.350119
Pases/90	0.228877
Pases hacía atrás recibidos del arquero/90	0.225862
Pases hacia adelante/90	0.224285
Longitud media pases largos, m	0.213405
Interceptaciones/90	0.209422
Duelos atacantes ganados, %	0.204451
Duelos ganados, %	0.183587
Pases progresivos/90	0.166054
Acciones defensivas realizadas/90	0.153862
Regates realizados, %	0.152714
Pases largos/90	0.141206
Local	0.137542
Goles evitados	0.126466
Asistencias	0.114953
Posesión conquistada después de una entrada	0.077473
Tiros libres directos/90	0.068190
Tarjetas rojas	0.042392
Duelos defensivos/90	0.038063
Porterías imbatidas en los 90	-0.014826
Remates en contra	-0.018505
Duelos aéreos en los 90.1	-0.018564
Goles recibidos	-0.018668

Y su gráfico de calor:

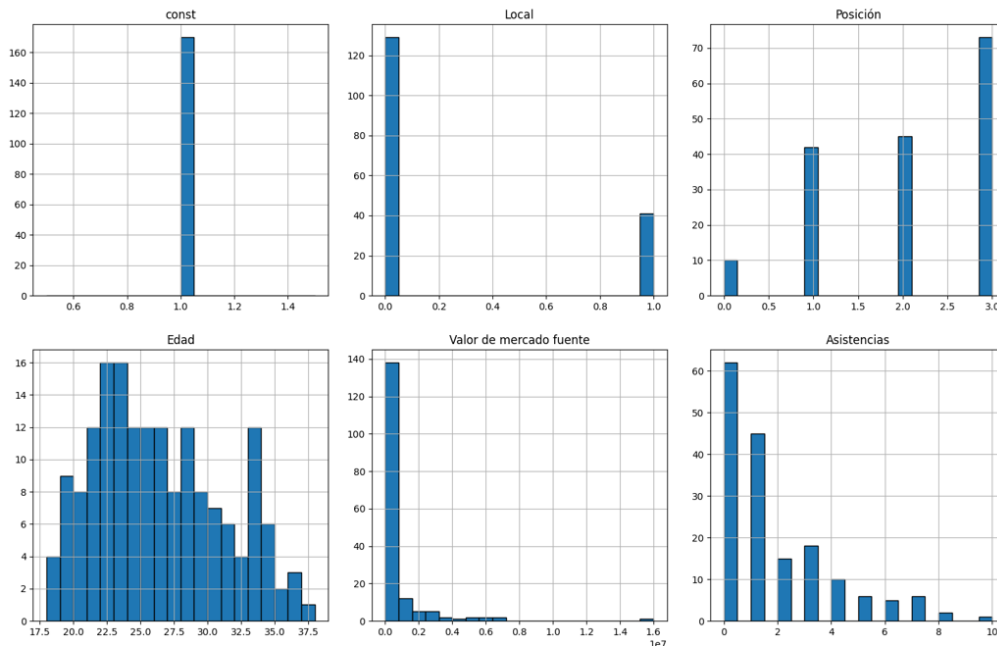


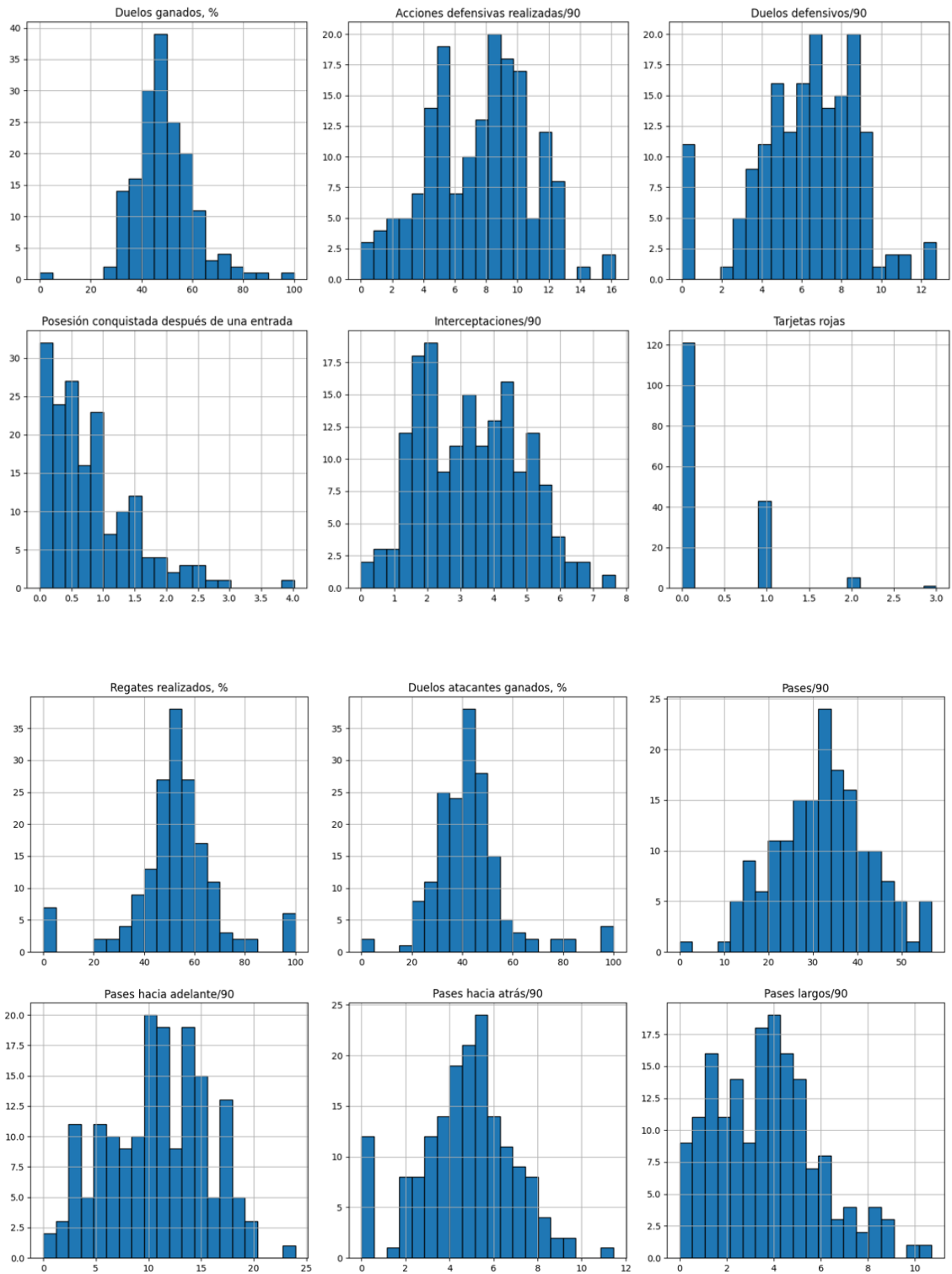
De lo anterior, podemos inferir que existen outliers en algunas columnas que podrían estar afectando el rendimiento del modelo. Además, es posible que no haya una relación lineal entre las variables.

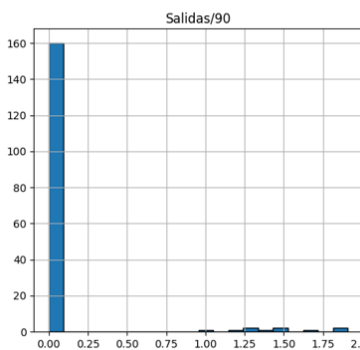
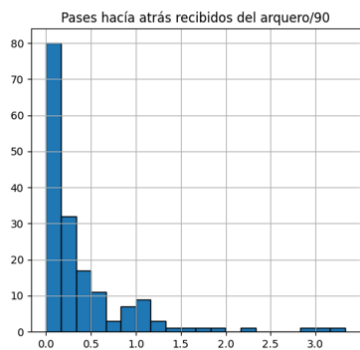
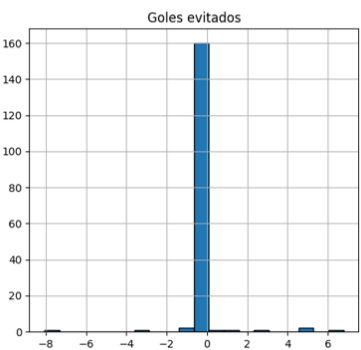
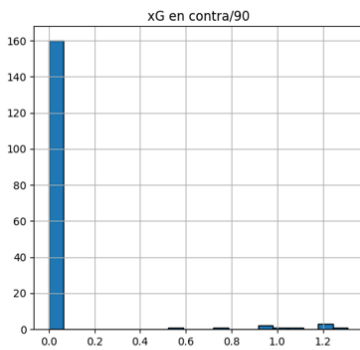
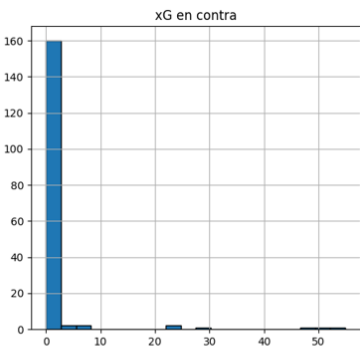
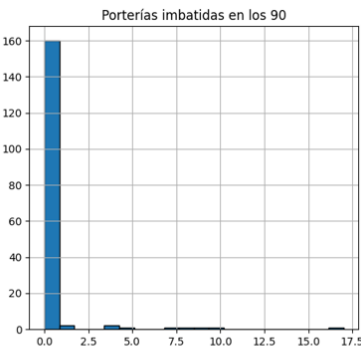
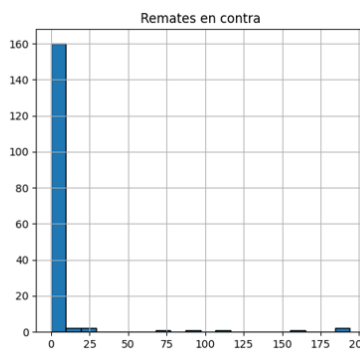
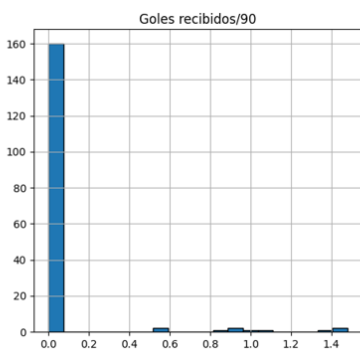
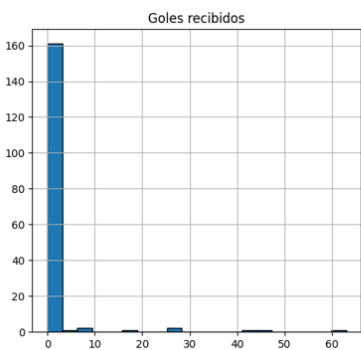
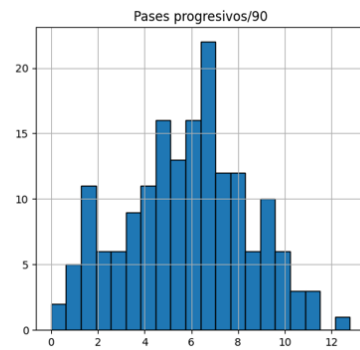
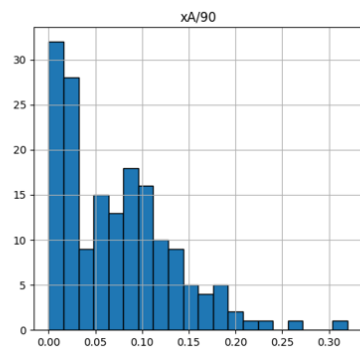
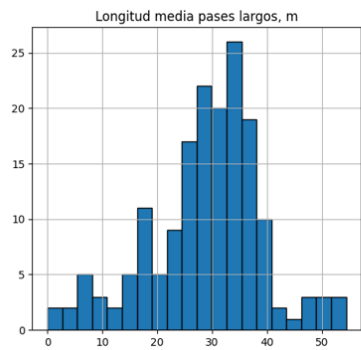
Los valores atípicos se visualizan:

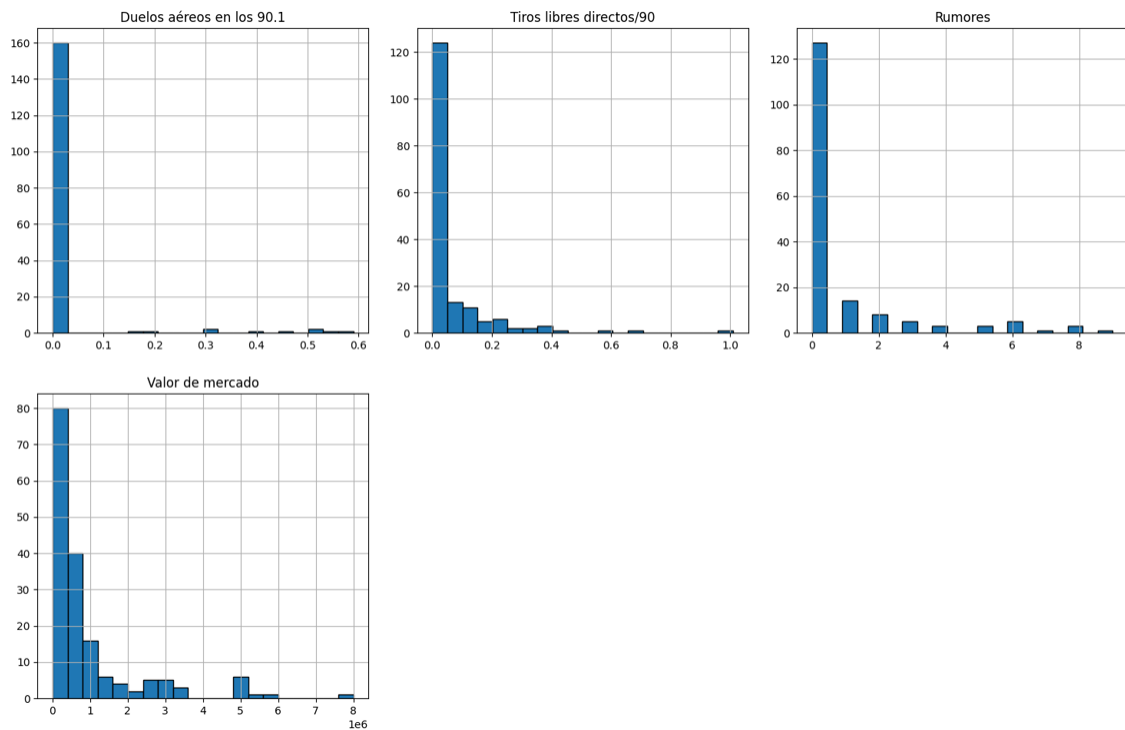


Los gráficos de distribución se visualizan a continuación:









5.2. Extracción de datos

Primero se realiza la extracción de información extra para complementar los datos de rendimiento:

- Se realiza un scrapping a transfermarkt utilizando BeautifulSoup para obtener los de transferencias, rumores, lesiones.
- Luego se realiza utilizando selenium para obtener los seguidores de Instagram de cada futbolista,
- Además, ocupando la biblioteca de Python pytrends obtenemos para obtener datos de popularidad de un jugador en Google Trends durante la última semana

- d. Finalmente obtenemos las publicaciones relacionadas con un jugador en el subreddit de fútbol con la cantidad de publicaciones encontradas y el total de votos positivo.

Un ejemplo de lo obtenido es:

Followers	Following	Rumores	Lesiones
1000000	915	2	1
251000	282	0	0
828000	300	6	15
13000	145	6	15
537000	1267	3	2
153000	1261	1	0
179000	598	1	4
149000	779	0	2
0	0	0	8
0	0	0	0
1000000	992	1	3
52713	639	0	1
196000	693	2	2
76000	912	3	6
1135	79	3	1
0	0	4	1
0	0	2	3
0	0	1	0
138000	672	1	2
64000	791	0	0
134000	696	0	0
253000	547	0	0
0	0	0	3
253000	547	0	0
0	0	0	4

Luego se crean variables extras al modelo:

- Se crea una variable “local” con valores 0 jugadores que jueguen en medio local y 1 aquellos que jueguen en medio extranjero
- Posteriormente se agrega una variable posición con 0:Aquero,1: Defensa,2: Mediocampista y 3: Delanteros

Local	Posición
1	3
1	1
1	1
1	2
1	3
1	1
1	3
1	2
1	3
1	1
1	1
1	3
1	2
0	2
0	0
0	3
1	3
1	3
0	2
1	3
1	1
1	3
1	3
1	1

5.3. Elección de variables

Determinación de variables importantes, se tiene un dataset con 115 columnas, no es factible realizar un modelo de predicción con esta cantidad de variables, por ello, se eliminan las variables que tienen una correlación menor a 27% con la variable objetivo.

Lo anterior es para obtener una metodología automatizada en la elección de variables por posición, no obstante, se requieren otras metodologías complementarias para eliminar variables, primero se analiza la significancia estadística eliminando aquellas variables, que obtienen valores p mayores al 5%, que indican una baja significancia estadística.

```

# Iterar hasta que todas las variables sean significativas
while True:
    # Ajustar el modelo OLS
    modelo = sm.OLS(y, X).fit()

    # Identificar los p-valores excepto el término constante
    p_values = modelo.pvalues.drop('const', errors='ignore')

    # Identificar la variable con el p-valor más alto, excluyendo las columnas protegidas
    max_p_variable = None
    for variable, p_value in p_values.sort_values(ascending=False).items():
        if variable not in columns_to_keep:
            max_p_variable = variable
            max_p_value = p_value
            break

    # Si no se encontró ninguna variable eliminable o todas son significativas, termina el ciclo
    if max_p_variable is None or max_p_value < 0.05:
        print("Todas las variables restantes son significativas o no hay más variables para eliminar.")
        break

    # Eliminar la variable con el p-valor más alto
    print(f"Eliminando variable '{max_p_variable}' con p-valor: {max_p_value}")
    X = X.drop(columns=[max_p_variable])

```

Luego analizando algunas variables se observan que la información se repite en algunos casos un ejemplo de esto son las variables:

- Acciones defensivas realizadas/90
- Duelos defensivos/90
- Duelos defensivos ganados, %

Lo anterior se visualiza en:

Q	R	S
Acciones defensivas realizadas/90	Duelos defensivos/90	Duelos defensivos ganados, %
1,69	0,11	16,67
1,58	0,09	50
0,75	0,05	100
1,91	0	0
1,86	0,12	80
1,79	0,07	0
1,78	0,13	57,14
2,05	0,56	33,33
1,52	0	0
1,84	0	0
1,18	0	0
2	0,11	50
1,96	0,11	100
2,17	0,08	100
1,43	0,13	25
2,25	0,17	0

Lo anterior sugiere un evidente problema de colinealidad lo cual se corregirá eliminando aquellas variables con el índice VIF (Factor de Inflación de la Varianza), las columnas a considerar son aquellas con un $VIF > 10$

```
# Función para eliminar columnas con VIF alto directamente en el DataFrame
def eliminar_vif_alto(dataframe, umbral_vif=10):
    # Crear un bucle para iterar hasta que todas las columnas tengan un VIF aceptable
    while True:
        # Calcular VIF para cada característica en el DataFrame
        vif = pd.DataFrame()
        vif["feature"] = dataframe.columns
        vif["VIF"] = [variance_inflation_factor(dataframe.values, i)
                    | for i in range(dataframe.shape[1])]

        # Encontrar la columna con el VIF más alto
        max_vif = vif["VIF"].max()

        # Si el VIF más alto está por debajo del umbral, terminar el bucle
        if max_vif < umbral_vif:
            break

        # Eliminar la columna con el VIF más alto
        feature_to_drop = vif.sort_values("VIF", ascending=False).iloc[0]["feature"]
        print(f'Eliminando "{feature_to_drop}" con VIF de {max_vif}')
        dataframe.drop(columns=[feature_to_drop], inplace=True)
```

También se menciona que se separa el conjunto de datos en:

- Atacantes
- Mediocampistas
- Defensas
- Arqueros

En la selección de variables, solo los "followers" de Instagram resultaron significativos, descartándose los valores provenientes de Pytrends y Reddit.

Delanteros

Para el caso de los delanteros las columnas a considerar son:

- const
- Valor de mercado fuente
- Minutos jugados
- Faltas/90
- Goles, excepto los penaltis/90
- Penaltis realizados, %
- Followers
- Rumores

Mediocampistas:

- const
- Valor de mercado fuente
- Asistencias
- xA
- Goles, excepto los penaltis/90
- Longitud media pases largos, m
- Tiros libres directos/90
- Penaltis a favor
- Followers
- Rumores

- vencimiento_contrato_dias

Defensas

- const
- Local
- Valor de mercado fuente
- Goles de cabeza
- Carreras en progresión/90
- Longitud media pases, m
- Centros desde el último tercio/90
- Followers
- Rumores
- Lesiones

Arquero

- const
- Edad
- Porterías imbatidas en los 90
- Followers
- Rumores

5.4. Problemáticas

Se intenta corregir la nomenclatura de los jugadores realizando un scraping a Bing, pero no se resuelve en todos los casos:

```
from bs4 import BeautifulSoup
def normalizar_nombre(nombre):
    nombre=nombre.replace(".", "")
    nombre_Q=nombre.replace(" ", "+")
    # URL de búsqueda
    url = f'https://www.bing.com/search?q={nombre_Q}+chile+nombre+jugador'

    # Hacer la solicitud HTTP
    response = requests.get(url)
    headers = {
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/129.0.0.0 Safari/537.36"
    }
    response = requests.get(url, headers=headers)
    #response.raise_for_status() # Verifica si la solicitud fue exitosa

    # Crear el objeto BeautifulSoup
    soup = BeautifulSoup(response.text, 'html.parser')

    # Encontrar el elemento deseado
    element = soup.find_all('div', class_='b_focusTextMedium')
    i=0
    valor=nombre
    e=None
    while i < len(element):
        e=element[i]
        try:
            valor=e.text
            break
        except:pass
        i+=1
    print(nombre_Q,e,valor)
    print(url)
    return valor
```

Debido a la falta de una nomenclatura definida en los nombres de los jugadores, la obtención de resultados se ve afectada tanto en el scraping de datos desde Transfermarkt como en la recopilación de información sobre la popularidad de los jugadores en diversas fuentes.

Para evitar errores en los resultados, se filtran aquellos valores con precio de transferencia =0

```
df = df[df["Valor de mercado"] != 0]
```

df

✓ 0.0s Python

	Local	Edad	Valor de mercado fuente	Partidos jugados	Minutos jugados	Goles	xG	Asistencias	xA	Duelos/90	...	Penaltis realizados, %	post_upvotes	Popularidad google	Followers	Following	Rumor
0	1.00	23	16000000	51	4534	25	20.33	4	3.19	21.64	...	100.00	10	1213	0	1000000	915
4	1.00	28	6000000	46	3698	14	11.59	10	8.72	20.59	...	100.00	10	118	100	537000	1275
6	1.00	24	5000000	46	3574	13	11.40	7	3.89	15.18	...	100.00	10	195	0	51000	1002
8	1.00	28	4000000	28	2097	11	9.25	3	2.74	17.42	...	100.00	10	4099	0	2073	447
11	1.00	23	3000000	30	1846	1	1.69	1	1.57	14.04	...	0.00	10	2097	0	52000	690
...
712	0.00	36	50000	35	3188	11	9.91	4	4.64	25.37	...	100.00	10	5768	182	32000	1477
715	0.00	33	50000	12	564	0	0.63	0	0.33	11.60	...	0.00	6	357	0	0	0
725	0.00	38	0	21	1020	1	2.24	3	2.68	17.82	...	0.00	10	2303	0	1000000	633
733	0.00	36	0	28	1861	1	2.37	5	2.95	21.92	...	0.00	9	299	0	524000	7552
739	0.00	41	0	30	2414	6	7.56	7	4.00	12.08	...	0.00	10	350	0	0	0

193 rows × 17 columns

Otro aspecto para considerar antes de realizar las predicciones es la diferencia temporal entre los datos: los datos de rendimiento de los jugadores son antiguos, mientras que la información obtenida de Transfermarkt es actual. Por ejemplo, Ben Brereton Díaz actualmente juega en el Southampton, pero en los datos de rendimiento su club registrado es el Blackburn Rovers.

También se menciona que jugadores que aparecen actualmente en la base de datos ya están retirados lo que justifica el filtro mencionado de precio de transferencia =0

5.5. Regresiones

Después de definir las variables que se incluirán en el modelo, procederemos a realizar las regresiones utilizando los modelos seleccionados, evaluando para cada uno el coeficiente de determinación (R^2) y el error cuadrático medio (RMSE), se evaluarán las

regresiones comparando los valores con y sin datos de seguidores verificando sus resultados y su impacto en cada modelo:

Delanteros

Para el caso de los delanteros el resultado de la regresión es:

Followers

	Modelo	R ²	RMSE
0	Regresión OLS	0.67	622477.16
1	Regresión Ridge	0.67	622480.33
2	Regresión Lasso	0.67	622477.16
3	Regresión ElasticNet	0.66	629038.75
4	Regresión MLP	0.47	788121.58
5	Regresión redes neuronales	0.39	846332.92
6	Regresión pol 2	0.86	404504.07

Sin Followers

	Modelo	R ²	RMSE
0	Regresión OLS	0.68	612106.56
1	Regresión Ridge	0.68	612195.38
2	Regresión Lasso	0.68	612106.56
3	Regresión ElasticNet	0.65	638693.66
4	Regresión MLP	0.41	827973.40
5	Regresión redes neuronales	0.41	827897.05
6	Regresión pol 2	0.75	541636.49

Mediocampistas

Para el caso de los mediocampistas el resultado de la regresión es:

Followers

	Modelo	R²	RMSE
0	Regresión OLS	0.60	673028.10
1	Regresión Ridge	0.60	673936.58
2	Regresión Lasso	0.60	673028.10
3	Regresión ElasticNet	0.53	725764.64
4	Regresión MLP	0.31	877493.09
5	Regresión redes neuronales	0.18	960665.06
6	Regresión pol 2	0.97	190078.85

Sin Followers

	Modelo	R²	RMSE
0	Regresión OLS	0.60	673729.62
1	Regresión Ridge	0.59	674656.29
2	Regresión Lasso	0.60	673729.62
3	Regresión ElasticNet	0.53	726843.13
4	Regresión MLP	0.16	970339.48
5	Regresión redes neuronales	0.16	970132.89
6	Regresión pol 2	0.83	442126.93

Defensas

Followers

	Modelo	R²	RMSE
0	Regresión OLS	0.71	438085.11
1	Regresión Ridge	0.71	438085.97
2	Regresión Lasso	0.71	438085.11
3	Regresión ElasticNet	0.68	457026.32
4	Regresión MLP	0.61	505836.65
5	Regresión redes neuronales	0.58	527753.10
6	Regresión pol 2	0.89	269094.83

Sin Followers

	Modelo	R²	RMSE
0	Regresión OLS	0.71	438202.14
1	Regresión Ridge	0.71	438202.99
2	Regresión Lasso	0.71	438202.14
3	Regresión ElasticNet	0.68	457039.45
4	Regresión MLP	0.55	547836.55
5	Regresión redes neuronales	0.55	547813.48
6	Regresión pol 2	0.84	324691.28

Arquero

Followers

	Modelo	R²	RMSE
0	Regresión OLS	0.86	193357.53
1	Regresión Ridge	0.86	193359.27
2	Regresión Lasso	0.86	193357.53
3	Regresión ElasticNet	0.82	217609.62
4	Regresión MLP	-0.16	548161.61
5	Regresión redes neuronales	-0.26	570040.84
6	Regresión pol 2	0.98	78338.02

Sin Followers


	Modelo	R²	RMSE
0	Regresión OLS	0.85	195445.98
1	Regresión Ridge	0.85	195447.69
2	Regresión Lasso	0.85	195445.98
3	Regresión ElasticNet	0.81	219302.81
4	Regresión MLP	-0.23	564812.50
5	Regresión redes neuronales	-0.32	583148.45
6	Regresión pol 2	0.96	97788.03


Resumen


Por posición se procesaron los siguientes elementos:


- Arquero: Regresión polinómica de orden 2
- Defensas: Regresión polinómica de orden 2
- Mediocampistas: Regresión polinómica de orden 2
- Delanteros: Regresión polinómica de orden 2

Se guarda el “.joblib” de cada modelo

 Arquero.joblib

 Defensa.joblib

 Mediocampista.joblib

 Delantero.joblib

5.6. Base de datos

Finalmente, almacenamos los datos originales en una base de datos en PostgreSQL.

```
import pandas as pd
from sqlalchemy import create_engine, text

# Definir las rutas y nombres
excel_file_path = r"C:\Users\miguel\OneDrive\Escritorio\capstone\avance_29\jugadores_nombre_normalizado_redit_google_trending_v2.xlsx"
table_name = 'rendimiento'

# Leer el archivo de Excel
df = pd.read_excel(excel_file_path)

try:
    # Conexión a PostgreSQL (ajusta las credenciales)
    engine = create_engine('postgresql://user:password@localhost:5432/dbname')

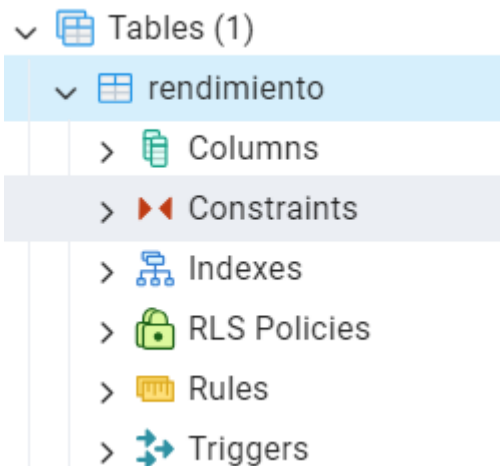
    # Crear la base de datos
    with engine.connect() as connection:
        connection.execute(text("COMMIT")) # Se asegura que no haya transacciones pendientes
        connection.execute(text("CREATE DATABASE futbol"))

    # Crear nueva conexión a la nueva base de datos
    engine = create_engine('postgresql://user:password@localhost:5432/futbol')

    # Guardar el DataFrame en la base de datos
    df.to_sql(table_name, engine, if_exists='replace', index=False)
    print("Datos cargados exitosamente en la base de datos.")

except Exception as e:
    print(f"Ocurrió un error: {e}")
```

La tabla generada en PostgreSQL se visualiza a continuación:



se genera un cursor para acceder a los datos:

```
try:
    # Crea un cursor
    cursor = conn.cursor()

    # Ejecuta la consulta
    cursor.execute("SELECT * FROM rendimiento")

    # Recupera los datos de la consulta
    rows = cursor.fetchall()

    # Obtén los nombres de las columnas
    column_names = [desc[0] for desc in cursor.description]

    # Cierra el cursor
    cursor.close()

    # Crea un DataFrame con los datos
    rendimiento_df = pd.DataFrame(rows, columns=column_names)

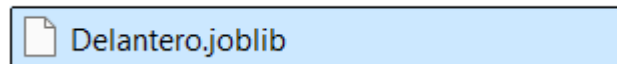
    # Muestra las primeras filas del DataFrame
    print(rendimiento_df.head())

except Exception as e:
    print(f"Ocurrió un error al leer la tabla: {e}")

finally:
    # Cierra la conexión
    conn.close()
```

5.7. Exportación de modelo

A continuación, se guarda el archivo joblib correspondiente al mejor modelo de regresión generado.



Se visualizan algunas predicciones:

Valor Real	Valor Predicho Poisson
8000000.00	8462430.85
5000000.00	4712448.86
5000000.00	5192314.76
3500000.00	2547168.03
5000000.00	4546808.99
...	...
125000.00	740104.17
100000.00	100000.00
125000.00	541999.97
200000.00	431220.49
25000.00	911128.71

5.8. Visualización

Finalmente, obtenemos nuestra aplicación de visualización desarrollada en Flask.

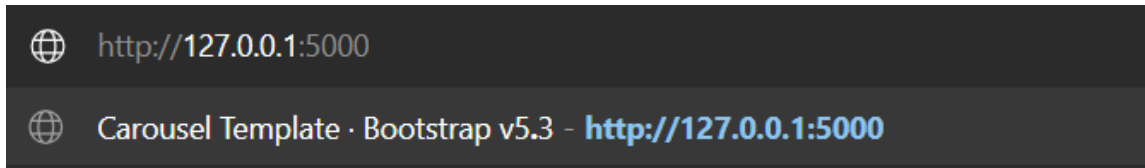
Primero se instalan las bibliotecas y entornos necesarios para ejecutar el aplicativo flask

- `conda create --name mi_entorno python=3.12`
- `conda info --env`
- `conda activate mi_entorno`
- `pip install Flask.`
- `pip install pandas.`
- `pip install scikit-learn.`
- `pip install unidecode.`
- `pip install openpyxl.`
- `pip install statsmodels.`

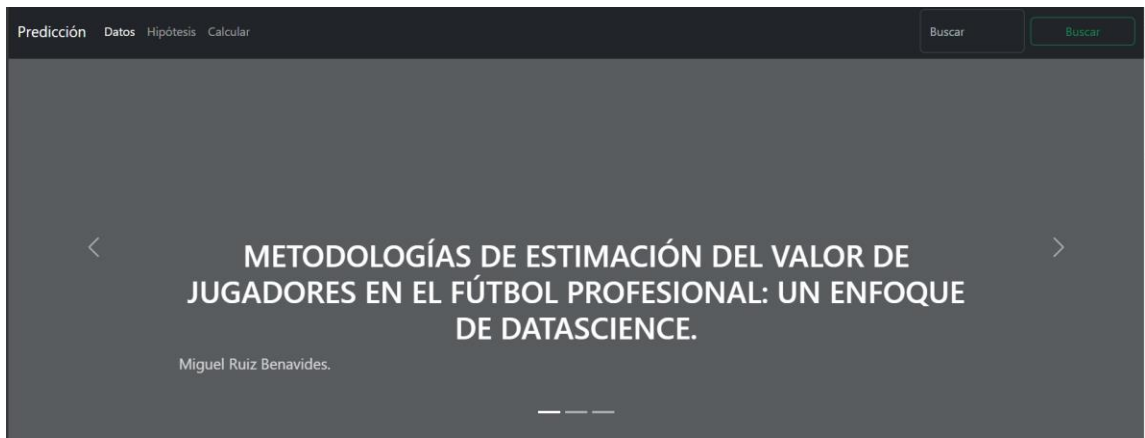
Luego se ejecuta `{ mi_entorno } \python.exe app.py runserver` en la línea de comandos del cmd

```
C:\Users\miguel\OneDrive\Documentos\flask>C:\Users\miguel\anaconda3\envs\capstone_futbol\python.exe app.py
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 938-851-214
```

Se visualiza la opción “Running on” copiando la url <http://127.0.0.1:5000> en un navegador



Desplegando la interfaz web:



EJEMPLO.

Casos recientes en el fútbol, como el de la Juventus, han puesto de manifiesto cómo algunos clubes manipulan artificialmente los precios de los jugadores para evitar riesgos contables. Un ejemplo emblemático fue el intercambio de jugadores entre la Juventus y el Barcelona en 2020, donde los precios inflados de Miralem Pjanic y Arthur Melo revelaron discrepancias difíciles de justificar. Este tipo de prácticas subraya la necesidad de contar con mecanismos más transparentes y precisos que permitan verificar las transacciones y detectar posibles infracciones legales en el mercado de transferencias. .



Wyscout

Contiene datos de rendimientos de jugadores chilenos.



Transfermarkt

Contiene datos de precio de transferencia, rumores y lesiones de jugadores.



Popularidad

Scraping realizado a Bing para obtener seguidores de los jugadores de fútbol chilenos, además de la cantidad de visualizaciones de video en Reddit y los trending de Google.

• Hipótesis.

Los modelos predictivos basados en machine learning y redes neuronales pueden mejorar significativamente la precisión en la estimación del valor de mercado de los jugadores de fútbol, integrando tanto datos deportivos como sociales (como seguidores en redes sociales)



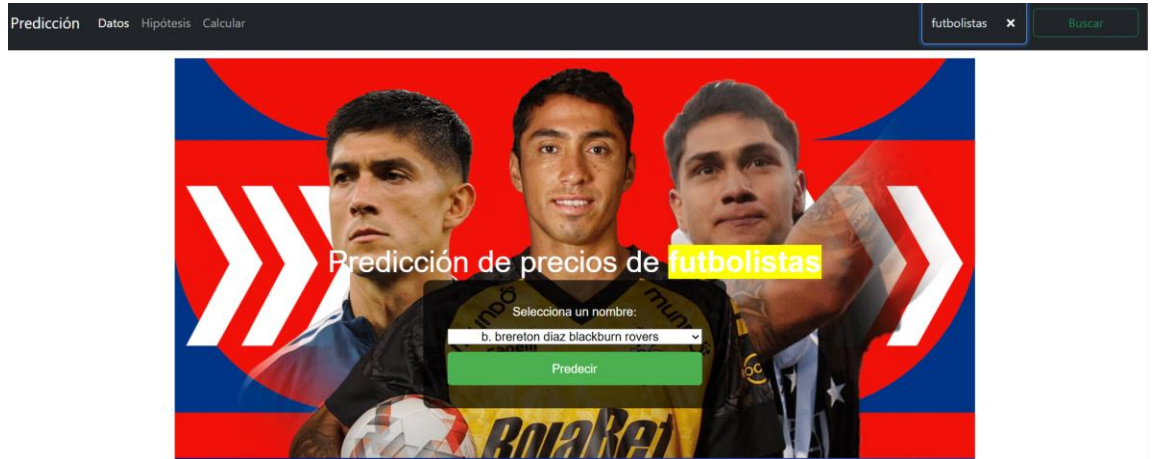


La interfaz web cuenta con:

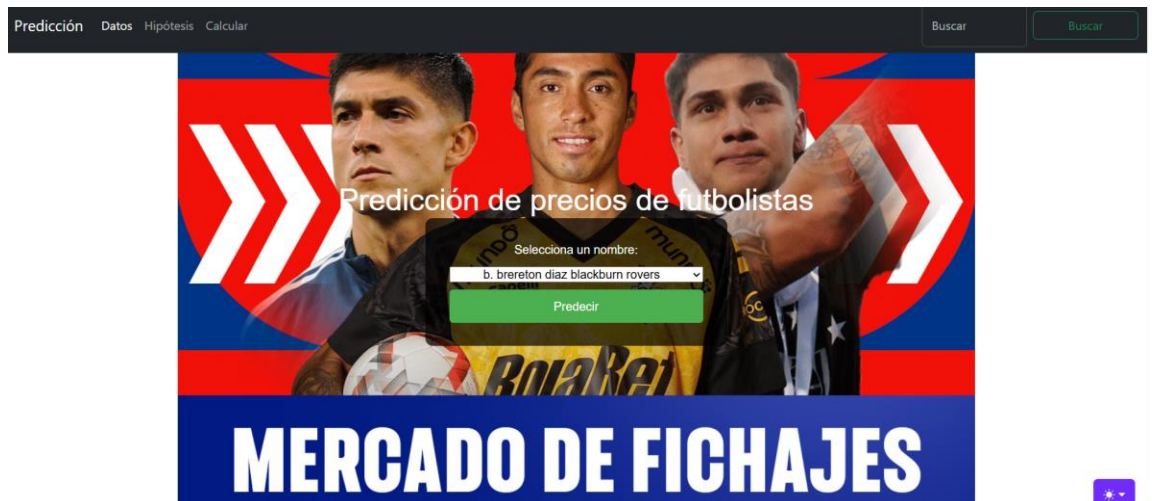
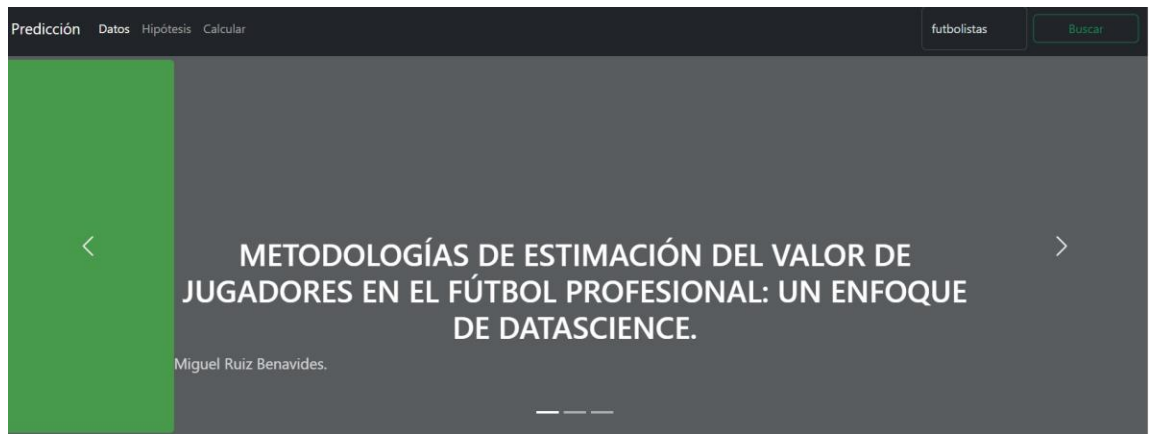
1. Buscador: Contiene una función javascript que permite buscar dentro del aplicativo flask:



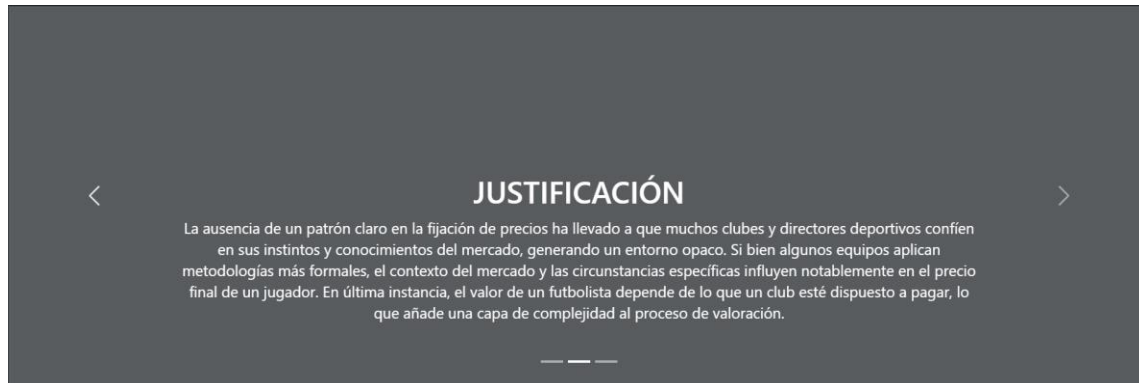
También tiene una función javascript que permite bajar automáticamente a la posición en la página donde está la búsqueda



2. Navbar: Tiene una función javascript que al seleccionar el elemento me permite ir a la zona de navegación:



3. Carruzel: Opción que permite desplegar información contextual del proyecto:



4. Datos: contiene un resumen de los datos utilizados en el proyecto:



5. Hipótesis: Sección que contiene la hipótesis

• Hipótesis.

Los modelos predictivos basados en machine learning y redes neuronales pueden mejorar significativamente la precisión en la estimación del valor de mercado de los jugadores de fútbol, integrando tanto datos deportivos como sociales (como seguidores en redes sociales)



6. Sección de cálculo de precio: En esta sección mediante un panel desplegado de los futbolistas obtenemos su cálculo en base a los datos de rendimiento, popularidad y transfermarkt, elegimos, por ejemplo:

b. brereton diaz blackburn rovers

g. maripan monaco

p. diaz river plate

e. pulgar galatasaray

d. valdes america

s. vegas monterrey

v. davila leon

c. baeza toluca

i. lichnovsky tigres uanl

t. alarcon cadiz

p. galdames genoa

i. saavedra universidad catolica

b. cortes colo colo

m. nunez universidad catolica

m. allende torque

f. sierralta watford

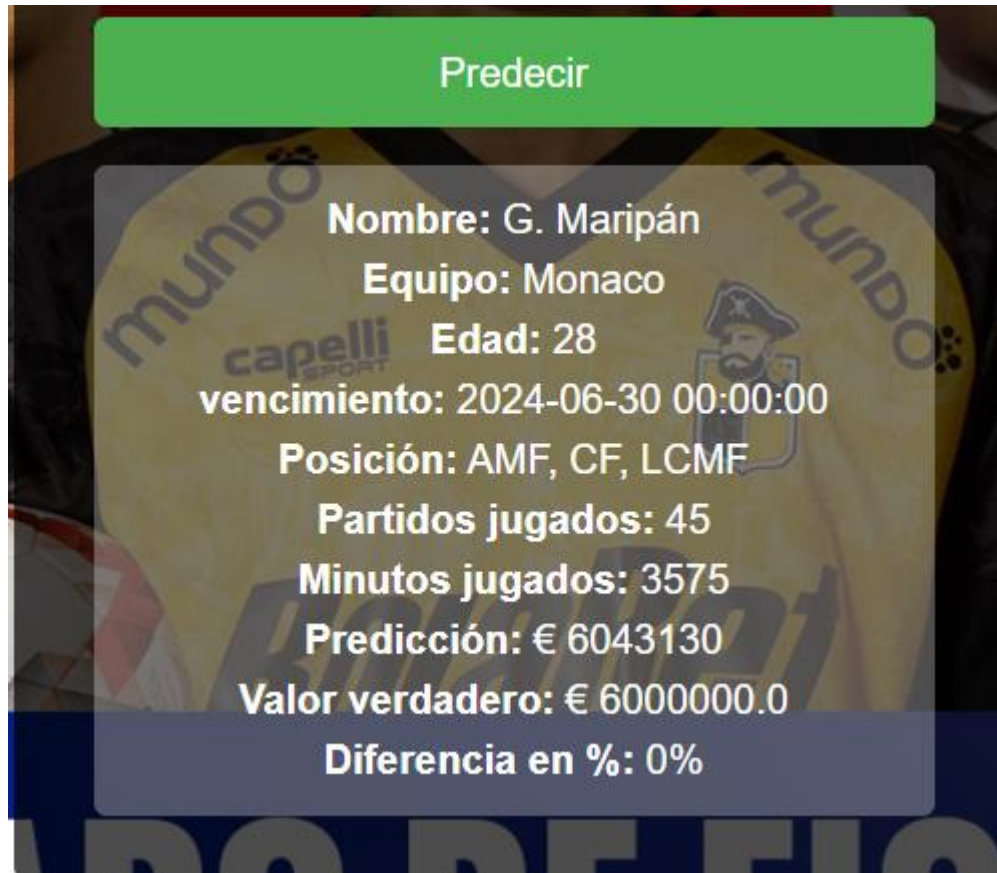
p. ruiz real salt lake

b. kuscevic palmeiras

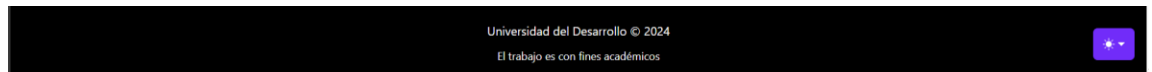
b. brereton diaz blackburn rovers

Predecir

Seleccionamos “predecir”, y nos devuelve los datos con la predicción realizada:



7. Footer: Contiene información del proyecto:



Es importante mencionar que se utiliza Bootstrap en el diseño del aplicativo de visualización.

5.9. Github

Se crea un repositorio en github con información del proyecto

[https://github.com/migueldvdsws/Capstone_futbol/tree/master:](https://github.com/migueldvdsws/Capstone_futbol/tree/master)

Name	Last commit message	Last commit date
models	Primer commit: Inicialización del proyecto	10 minutos ago
src/data	Primer commit: Inicialización del proyecto	10 minutos ago
static	Primer commit: Inicialización del proyecto	10 minutos ago
templates	Primer commit: Inicialización del proyecto	10 minutos ago
README.md	Create README.md	now
app.py	Primer commit: Inicialización del proyecto	10 minutos ago

```
README.md
```

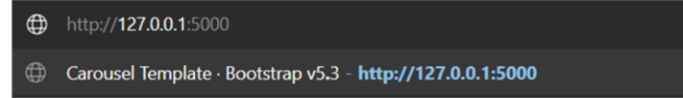
Proyecto que con machine learning permite obtener los precios de los jugadores de futbol

Primero se instalan las bibliotecas y entornos necesarios para ejecutar el aplicativo flask

- conda create --name mi_entorno python=3.12
- conda info --env • conda activate mi_entorno • pip install Flask • pip install pandas • pip install scikit-learn • pip install unicode • pip install openpyxl • pip install statsmodels Luego se ejecuta [mi_entorno]\python.exe app.py runserver en la línea de comandos del cmd

```
C:\Users\migue\OneDrive\Documentos\Futbol>C:\Users\migue\anaconda3\envs\capstone_futbol\python.exe app.py
* Serving Flask app "app"
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 928-851-214
```

Se visualiza la opción "Running on" copiando la url <http://127.0.0.1:5000> en un navegador



Desplegando la interfaz web:



6. Conclusiones

En conclusión, los objetivos planteados han sido cumplidos al desarrollar un modelo predictivo basado en machine learning y redes neuronales que integra datos deportivos y sociales para estimar con precisión el valor de mercado de los jugadores de fútbol. Se logró recopilar y unificar datos de diversas fuentes, implementar y comparar múltiples modelos predictivos, evaluar el impacto de variables no deportivas, como la popularidad en redes sociales, y seleccionar el modelo más preciso según métricas como R^2 y RMSE. Además, se desarrolló una plataforma interactiva en Flask que facilita la visualización de los resultados y de los factores clave en la valoración económica de los jugadores.

El trabajo realizado ha propuesto una metodología innovadora que aplica técnicas de machine learning y redes neuronales para la valoración de futbolistas, estableciendo un marco robusto para futuros estudios en este campo. Esta metodología enfatiza la importancia de recopilar y unificar datos relevantes, como el rendimiento deportivo, las lesiones y la popularidad en redes sociales, lo que puede enriquecer significativamente la base de datos mediante herramientas de scraping. Además, sugiere desarrollar modelos predictivos que permitan estimar el valor de mercado de los jugadores, ofreciendo un proceso de valoración más estandarizado y accesible para clubes y agentes del sector.

A pesar de las limitaciones del estudio, como la escasa cantidad de datos disponibles (sólo 740 futbolistas) y la antigüedad de la información, que incluye jugadores actualmente en equipos diferentes, los resultados obtenidos son interesantes y replicables en futuros

estudios. La falta de normalización en los nombres también es un desafío que se debe abordar para mejorar la precisión de los modelos.

Entre los hallazgos más relevantes se destaca que la popularidad de los jugadores, medida a través de su presencia en redes sociales, es una variable importante que influye en la determinación de su valor de mercado. Los modelos de regresión, particularmente la regresión polinómica de orden 2, demostraron ser efectivos para estimar los precios, siendo este el modelo de estudio más prometedor.

En resumen, la metodología propuesta tiene el potencial de mejorar la precisión de las estimaciones del valor de mercado de los futbolistas y ofrecer una comprensión más profunda de los factores que influyen en su valoración, abriendo nuevas oportunidades para la automatización y estandarización en la industria del fútbol.

Bibliografía

- Aaron Santiago Pedraza Cárdenas, O. (17 de 6 de 2023). *urosario*. (E. d. Escuela de Administración, Ed.) Obtenido de <https://repository.urosario.edu.co/server/api/core/bitstreams/605b2699-da56-4946-b598-80f85e8b64e2/content>
- Agustina Margaritis, Natalia Macchi, Román Tanducci, Santiago García, Javier Bussi, & Lucía Andreozzi. (s.f.). *revistaingenieria*. (U. N. Rosario, Ed.) Obtenido de <https://revistaingenieria.uda.cl/Publicaciones/360006.pdf>
- Alejandro Guzman, & Emilio Catapano. (24 de 01 de 2024). *revistasenlinea*. (C. V. Universidad Metropolitana, Ed.) Obtenido de <https://revistasenlinea.saber.ucab.edu.ve/index.php/tekhne/article/view/6322>
- BOSMAN, D. (s.f.). *arno*. (T. UNIVERSITY, Ed.) Obtenido de <https://arno.uvt.nl/show.cgi?fid=172614>
- Cenalmor, M. C. (s.f.). *comillas*. (comillas, Ed.) Obtenido de <https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/46797/TFG%20-%20Cortes%20Cenalmor%2c%20Miguel.pdf?sequence=2&isAllowed=y>
- Daniel Dinsdale, & Joe Gallagher. (27 de Jan de 2022). *arxiv*. (C. University, Ed.) Obtenido de <https://arxiv.org/abs/2201.11533>
- Frutos, S. B. (Abril de 2020). (C. U. PONTIFICIA, Ed.) Obtenido de <https://repositorio.comillas.edu/rest/bitstreams/408629/retrieve>

Geurts, N. R. (8 de 7 de 2022). (L. Laguna, Ed.) Obtenido de

<https://riull.ull.es/xmlui/bitstream/handle/915/29416/Ciencia%20de%20datos%20con%20python%20en%20su%20aplicacion%20a%20la%20practica%20del%20futbol.pdf?sequence=1&isAllowed=y>

González Baigorri, A., & Marí Roselló, V. (2021). *Estimación y evolución de los determinantes del valor de mercado de los jugadores de fútbol profesional*. (U. P. Barcelona, Ed.) Obtenido de

<https://repositori.upf.edu/bitstream/handle/10230/54572/TFGECO22MARIGONZEsti.pdf?sequence=1&isAllowed=y><https://docta.ucm.es/entities/publication/1d2d680a-9ca7-40bc-9752-43cecac45258>

Markel Rico-González, José Pino-Ortega, Amaia Méndez, Filipe Manuel Clemente, & Arnold Baca. (2023). *Biology of Sport*. (B. o. Sport, Ed.) Obtenido de

<https://www.termedia.pl/Machine-learning-application-in-soccer-a-systematic-review,78,46287,0,1.html>

Miohk Ryoo, Namjung Kim , & Kyoungju Park . (11 de September de 2017).

SpringerLink. (SpringerLink, Ed.) Obtenido de

<https://link.springer.com/article/10.1007/s11042-017-5137-4>

Mohammad Mahdi Nasiri, Mojtaba Ranjbar, Madjid Tavana, Francisco J. Santos

Arteaga, & Reza Yazdanparast. (22 de 10 de 2018). *onlinelibrary*. Obtenido de

<https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12342>

Moreno, Z. (2022). <https://oa.upm.es/71425/>. (U. P. Madrid, Ed.) Obtenido de

<https://oa.upm.es/71425/>

Pedraza Cárdenas, A., & Sánchez Castañeda, O. (17 de junio de 2023). *urosario*. (U. d. Rosario, Ed.) Obtenido de <https://repository.urosario.edu.co/server/api/core/bitstreams/605b2699-da56-4946-b598-80f85e8b64e2/content>

Phatak, A. (s.f.). <https://fis.dshs-koeln.de/>. (D. S. Köln., Ed.) Obtenido de <https://fis.dshs-koeln.de/de/publications/data-driven-performance-analysis-in-soccer-a-compilation-of-data->

Sylvio Barbon Junior, Felipe Arruda Moura , & Ricardo da Silva Torres . (03 de September de 2024). *springer*. (springer, Ed.) Obtenido de https://link.springer.com/chapter/10.1007/978-3-031-67256-9_14

Uribe, A. d. (s.f.). *bibliotecadigital*. (I. U. Envigado, Ed.) Obtenido de <http://bibliotecadigital.iue.edu.co/bitstream/20.500.12717/898/1/CD0974A.pdf>

Vineet M. Payyappalli , & Jun Zhuang. (01 de 03 de 2019). *link.springer*. Obtenido de <https://link.springer.com/article/10.1007/s10669-019-09721-7>