

**VCF Forecasting Scouting Model**

Aplicación para la predicción y visualización del valor de mercado y el rendimiento de los jugadores.

Autor: Juan Cornejo, Andrés Cervera, Jesús Jornet, Cristian Marty.

Tutor: Pedro Nieto

2023-2024



TFM – Máster en Data Analytics

EDEM – Escuela de Empresarios

AGRADECIMIENTOS:

Este Trabajo de Fin de Máster ha sido un desafío tanto personal como académico, y no habría sido posible sin el apoyo de muchas personas a las que nos gustaría expresar nuestro más sincero agradecimiento.

Desde nuestros tutores y mentores Pedro, Victoria, Javi y Miguel, hasta todos y cada uno de los docentes que han pasado por nuestra aula en cada uno de los módulos. Gracias a cada una de las piedras que han puesto en nuestra formación hemos sido capaces de desarrollar con éxito este proyecto, aplicando todo lo aprendido en este maravilloso año lleno de experiencias.

**TABLA DE CONTENIDO**

[1. AGRADECIMIENTOS 3](#_Toc176255264)

[2. AGRADECIMIENTOS 3](#_Toc176255265)

[3. INTRODUCCIÓN GENERAL 5](#_Toc176255266)

[3.1. EVOLUCIÓN TECNOLÓGICA 7](#_Toc176255267)

[3.2. RELEVANCIA DEL ASPECTO PSICOLÓGICO 11](#_Toc176255268)

[4. OBJETIVOS Y CONTEXTO 12](#_Toc176255269)

[4.1. OBJETIVO GENERAL DEL TRABAJO 12](#_Toc176255270)

[4.2. ANÁLISIS DE JUEGO 12](#_Toc176255271)

[CONTEXTO 12](#_Toc176255272)

[OBJETIVOS 14](#_Toc176255273)

[4.3. DIRECCIÓN DE CANTERAS 15](#_Toc176255274)

[CONTEXTO 15](#_Toc176255275)

[OBJETIVOS 16](#_Toc176255276)

[4.4. GESTIÓN DE EQUIPOS 17](#_Toc176255277)

[OBJETIVO 17](#_Toc176255278)

[4.5. DIRECCIÓN DEPORTIVA 18](#_Toc176255279)

[OBJETIVO 18](#_Toc176255280)

[5. METODOLOGÍA 19](#_Toc176255281)

[5.1. METODOLOGÍA - ANÁLISIS DE JUEGO 19](#_Toc176255282)

[DISEÑO DEL ESTUDIO 19](#_Toc176255283)

[MUESTRA 19](#_Toc176255284)

[PROCEDIMIENTO 20](#_Toc176255285)

[5.2. METODOLOGÍA - DIRECCIÓN DE CANTERAS 24](#_Toc176255286)

[DISEÑO DEL ESTUDIO 24](#_Toc176255287)

[MUESTRA 24](#_Toc176255288)

[PROCEDIMIENTO 25](#_Toc176255289)

[6. MARCO TEÓRICO 26](#_Toc176255290)

[6.1. GENERAL 26](#_Toc176255291)

[6.2. MARCO TEÓRICO - GESTIÓN DE EQUIPOS 28](#_Toc176255292)

[INTRODUCCIÓN 28](#_Toc176255293)

[EVOLUCIÓN DEL CONCEPTO INTELIGENCIA EMOCIONAL 29](#_Toc176255294)

[PRINCIPIOS DE LA TEORÍA 31](#_Toc176255295)

[REPERCUSIÓN E INVESTIGACIONES POSTERIORES 34](#_Toc176255296)

[CONCLUSIONES 37](#_Toc176255297)

[6.3. MARCO TEÓRICO – DIRECCIÓN DEPORTIVA 38](#_Toc176255298)

# ABSTRACT Y RESUMEN

## VERSIÓN EN CASTELLANO

En un contexto de máxima exigencia como el fútbol de élite, el cuál actualmente se está viendo afectado por algunas constricciones y limitaciones financieras, cualquier pequeño detalle puede contribuir a marcar la diferencia.

Es por ello, que este trabajo se centra principalmente en el diseño de una solución que permita al Valencia Club de Fútbol disponer de conocimientos de mercado y, en particular, poder disponer de un método de estimación de precios de transferencia de los derechos federativos y económicos en las operaciones de traspaso mediante algoritmos de machine learning, con el objetivo de poder alcanzar una ventaja competitiva frente al resto de equipos.

A tal efecto, la solución propuesta está basada en todas las metodologías, herramientas y recursos adquiridos en el Máster de Data Analytics con el fin de poder demostrar el buen aprendizaje y el conocimiento de qué herramienta es más adecuada para cada caso de uso.

Adicionalmente, el equipo ha prestado especial énfasis a la simplificación de la experiencia de usuario de la herramienta. Creemos que una buena interfaz y un proceso fácil y limpio pueden ayudar al uso de la misma, y más en un sector que habitualmente suele estar algo más alejado del ámbito tecnológico. Esta búsqueda se puede ver plasmada en pequeños detalles que serán analizados y desglosados más adelante.

**Palabras clave**: Transferencia, Fútbol, Machine Learning, Experiencia de Usuario (UX)

## VERSIÓN EN INGLÉS

In a highly demanding context such as elite soccer, which is currently being affected by some financial constraints and limitations, any small detail can make a difference.

That is why this work is mainly focused on the design of a solution that allows Valencia Club de Fútbol to have market knowledge and, in particular, to have a method for estimating transfer prices of federative and economic rights in transfer operations using machine learning algorithms, in order to achieve a competitive advantage over other teams.

To this end, the proposed solution is based on all the methodologies, tools and resources acquired in the Master of Data Analytics in order to demonstrate good learning and knowledge of which tool is best suited for each use case.

Additionally, the team has placed special emphasis on simplifying the user experience of the tool. We believe that a good interface and an easy and clean process can help the use of the tool, especially in a sector that usually tends to be somewhat more distant from the technological field. This search can be seen in small details that will be analyzed and broken down later.

**Keywords**: Transfer, Football, Machine Learning, User Experience (UX)

# INTRODUCCIÓN

El objetivo de este TFM es poder ofrecer al Valencia C.F. una solución que le permita estimar precios de traspaso y poder obtener información que le pueda ayudar a tener una ventaja competitiva en el mundo del fútbol.

Sin embargo, para ello, antes es necesario conocer un poco más la idiosincrasia del propio sector, a fin de entender que rol jugará este TFM en la cobertura de las necesidades del club.

## INTRODUCCIÓN AL SECTOR

El fútbol es uno de los deportes más populares en todo el mundo. Este deporte, tan seguido y practicado por tanta gente, presenta una de las paradojas más importantes del deporte. A pesar de que sus reglas básicas son muy sencillas, una vez comienza el partido, el desarrollo del juego es muy complejo.

Esta complejidad del fútbol hace que muchos autores hayan intentado desgranar o clasificar todos los aspectos que tienen relevancia en un partido (táctico, técnico, físico, etc.) a fin de entender cada uno de ellos no sólo de forma individual sino también la correlación entre ellos. De hecho, esta división en tres partes se puede apreciar en la mayoría de literatura de este sector, y en incluso en los cursos de formación de entrenadores, la mayor parte del tiempo se dedica únicamente a estas tres ramas.

Sin embargo, uno de los grandes cambios que ha experimentado el fútbol en los últimos años ha sido la **evolución de la tecnología** y su **implantación** a lo largo de las estructuras de los clubes. Esta evolución se desglosa en el siguiente epígrafe.

## EVOLUCIÓN TECNOLÓGICA EN EL SECTOR

Vivimos en un entorno en el cual muchas empresas y organizaciones, gracias a la ayuda de la tecnología, tienen la posibilidad de recabar una inmensa cantidad de datos y procesarlos, con el fin de extraer conclusiones que les permitan tomar las mejores decisiones posibles para alcanzar sus objetivos.

Dentro de este entorno encontramos el mundo del deporte un sector que genera miles de millones en todo el mundo y en el cual los propios deportistas buscan alcanzar el éxito mediante disciplina, sacrificio y entrenamiento.

El fútbol, en particular, siempre ha sido uno de los deportes que más ha tardado en abrirse a nuevas tecnologías y ha ido evolucionando tanto sus reglas como la forma de practicarlo siempre de forma paulatina.

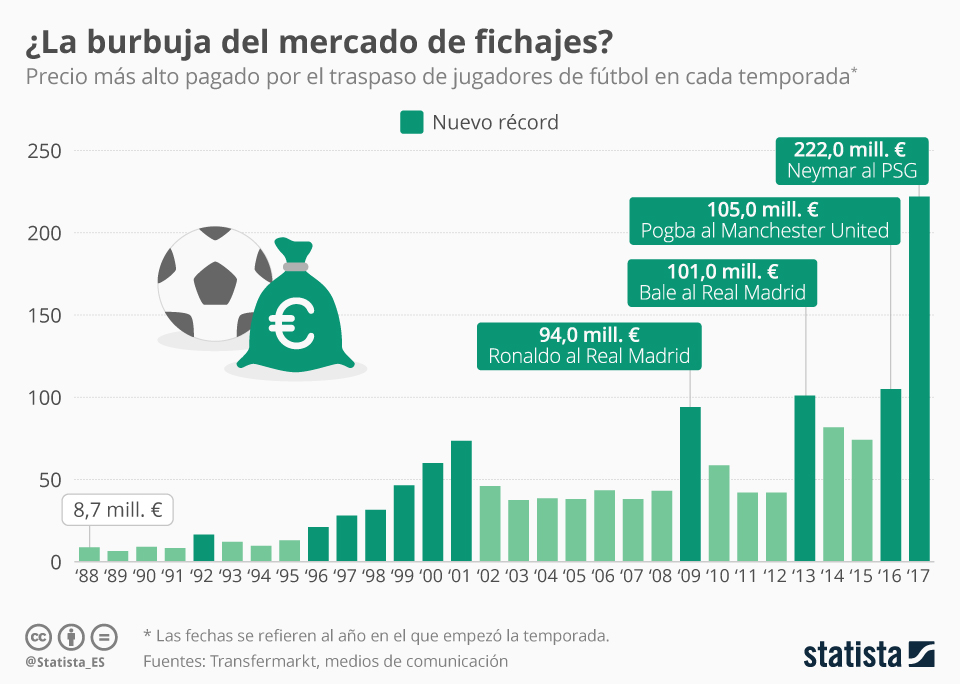
A pesar de ello, la esencia del fútbol a priori sigue siendo bastante básica: se enfrentan dos equipos de once jugadores cada uno y el objetivo es marcar más goles que el rival. No obstante, una vez comienza el partido, este juego de reglas sencillas se transforma en un juego de desarrollo altamente complejo en el cual pueden influir múltiples variables. Esta característica recuerda un poco al experimento de John Conway y su juego “Game of Life” publicado en el artículo “*Mathematical Games”* (Gardner M. , 1970) en el cual se reflejaba cómo un juego con reglas sencillas puede tener desarrollos muy complejos.

Esta percepción de aparente simpleza podría ser el origen de que mucha gente piense que es fácil entender cómo funciona este deporte (recordándonos al famoso efecto Dunning-Kruger (Dunning, 2011)) y, por tanto, no ven la necesidad de aplicar de estas nuevas tendencias y tecnologías que incluso ya se han empezado a probar en otros deportes de gran prestigio como el baloncesto, el baseball o el fútbol americano.

Hay múltiples ejemplos de cómo estos deportes (con gran repercusión en EEUU) han mostrado una apertura mayor a nuevas tendencias. Uno de ellos podemos encontrarlo en el baseball, y más concretamente en los Oakland Athletics. Gracias al uso de datos de los jugadores, fueron capaces de confeccionar una plantilla de bajo coste que a la postre acabó dando un gran rendimiento. Dicha hazaña, que fue recogida en el famoso libro de Moneyball (Lewis, 2004), tuvo tanta repercusión que la NBA y la NFL no tardaron en empezar a utilizar datos de los jugadores para confeccionar sus plantillas.

El uso de herramientas de recogida y análisis de grandes volúmenes de datos es utilizado por la gran mayoría de las empresas más potentes del mercado con múltiples objetivos: reducir costes, detectar anomalías en los procesos, identificar patrones de consumo de los clientes etc. De la misma forma, un club de fútbol podría utilizar dicha tecnología no sólo para su área directiva o de gestión, sino también para el ámbito deportivo, pudiendo ser una herramienta muy efectiva no sólo para la mejora de entrenamientos y toma de decisiones, **sino también para la identificación de jugadores con potencial que todavía no hayan sido descubiertos**.

Importancia especial cobra esta última posibilidad. Durante los últimos años, el mercado de traspasos de futbolistas ha experimentado una inflación sin precedentes tal y como muestra la Figura 1, debido principalmente al aumento del valor de venta de los derechos televisivos de los clubs y en algunos casos a la entrada de capital extranjero y fondos de inversión en los clubs.



**Figura 1*. Evolución del importe del traspaso más caro de cada año 1988-2017***

*Gráfico que muestra la evolución del máximo importe pagado por el traspaso de un jugador durante esa temporada* (Moreno, 2017)*. Recuperado de la web:* [*https://es.statista.com/grafico/10537/neymar-lleva-al-precio-por-traspaso-a-un-record-historico/*](https://es.statista.com/grafico/10537/neymar-lleva-al-precio-por-traspaso-a-un-record-historico/)

En consecuencia, el hecho de utilizar inteligentemente la información podría ayudar a los equipos a ser más precisos en la sección de un futbolista y en la predicción de su potencial y su valor. De esta forma, podrían anticiparse al mercado y por lo tanto obtener un mejor precio por el futbolista, con dos principales ventajas como consecuencia:

1. Mayor precisión en la captación de talento
2. Mejor rendimiento del capital disponible para traspasos.

Por estos motivos, el uso de la tecnología de recogida y proceso de datos puede jugar un papel fundamental no sólo en la toma de decisiones en el partido o la temporada, sino también durante el periodo de traspasos.

Sin embargo, en el fútbol siempre ha habido cierto escepticismo en el uso de los datos como una herramienta más. A pesar de que en los últimos años ese enfoque ha cambiado y cada vez más clubs apuestan por sistemas y especialistas de datos, todavía hay muchos profesionales que se muestran reluctantes a esta nueva herramienta.

Algunos de estos casos en los que se muestra la desconfianza en los datos podemos encontrarlos en el libro “*The numbers game: why everything you know about soccer is wrong”* (Anderson & Sally, 2013).

Hasta hace tan sólo una década atrás la recogida de información estadística y visual se realizaba de forma manual. Esto hacía que, en muchos casos, el volumen de datos a recabar fuese limitado y susceptible de errores manuales (errores de conteo, criterios heterogéneos según cada observador, etc.). En consecuencia, la relevancia que se le daba al análisis integral de la información era generalmente bajo.

A pesar de ello, en los últimos años, múltiples clubes de la elite profesional han decidido apostar por sistemas que permitan la fácil captura de información de partidos y jugadores, así como por profesionales que sean capaces de interpretar los resultados obtenidos. Gracias a ello, el cuerpo técnico puede disponer de información en tiempo real acerca de la actuación de los jugadores en el partido y la temporada, además de conocer detalles de jugadores que puedan ser considerados fichajes potenciales que permitan al club anticiparse al mercado, pudiendo suponer un ahorro en costes.

## HISTORIA Y CONTEXTO DEL CLUB

Valencia Club de Fútbol es un club de fútbol histórico de la Liga Española. Este club, que ya es centenario (creado un 1 de marzo de 1919), se ha ganado a pulso ser uno de los referentes en el panorama futbolístico tanto nacional como internacional.

En el ámbito nacional, desde su fundación, salvo en dos breves etapas en las que jugó en segunda división (1928-31 y 1986-87), el equipo siempre se ha mantenido en la división de oro del fútbol español. De hecho, hasta en 6 ocasiones ha levantado el título de campeón de liga y en 8 ocasiones conquistó el título de campeón de Copa, siendo la más reciente esta pasada temporada 2018-19 tal y como se puede ver en la propia web del equipo. (Valencia C.F., 2019)

En el ámbito internacional también ha demostrado su poderío y casta alzando tres títulos europeos como son las Copas de la UEFA cosechadas en las temporadas 1961-62,62-63 y 2003-04. Adicionalmente, también fue dos veces finalista de la UEFA Champions League de forma consecutiva en las temporadas 1999-00 y 2000-01, quedando patente la importancia de este equipo en el panorama futbolístico internacional.

Sin embargo, en las últimas temporadas el club se ha visto afectado por la crisis financiera del COVID. Esto justo con un claro decrecimiento general en los beneficios de los clubes deportivos de la primera división española, ha afectado a la capacidad de maniobra del propio club.

# OBJETIVOS Y CONTEXTO

Una vez entendido el contexto del sector y la historia del propio club, procederemos a detallar los objetivos de este trabajo de fin de máster (en adelante, “tfm”).

## OBJETIVO GENERAL DEL TFM

Poder proporcionar una solución al Valencia C.F. que le permita hacer una estimación de los precios de transferencia de los jugadores de forma simple y sencilla, a fin de poder obtener una ventaja competitiva en su sector.

## CONTEXTO DEL DESAFÍO Y MOTIVACIÓN

Como se ha analizado en las secciones previas, el mundo del fútbol se ha vuelto cada vez más competitivo. La llegada (e implantación paulatina) de nuevas de tecnologías ha permitido a muchos clubes comenzar a desarrollar nuevas metodologías de trabajo, así como poder invertir más en I+D tanto en los procesos de captación de jugadores, como en los de análisis de rendimiento.

De la misma forma, la viabilidad y sostenibilidad económica se ha vuelto un aspecto fundamental de los clubes de fútbol. Para entender la relevancia de este tfm, es importante entender que, actualmente, aunque los derechos de retransmisión son la principal fuente de ingresos de los clubes de fútbol profesional, los beneficios por traspasos juegan un papel sumamente importante en el balance económico de un club.

Es por eso que, contar con buenas estimaciones que permitan establecer un correcto precio de transferencia, es clave tanto para las compras, como para las ventas. Un indicador como este puede tener un impacto directo tanto en el aumento de ingresos (uso de la estimación para ajustar el precio de venta), como para la reducción de costes (uso de la estimación para ajustar el precio de compra). Por tanto, este tfm puede potencialmente ayudar al club en su gestión y tener un impacto directo en su economía.

## DESAFÍOS PRINCIPALES

Como se ha indicado en el apartado de objetivo general, se busca proveer de una estimación. Es por ello que, aunque el resto de elementos del proyecto puedan jugar un papel relevante, el protagonismo se encuentra en la propia estimación.

Parece lógico pensar entonces que el gran desafío se encontrará en lo relativo a la creación y optimización del proceso de estimación.

A pesar de contar con un proveedor de datos de calidad, es aquí donde podemos encontrar uno de los mayores desafíos: la calidad y cantidad de la información para generar la estimación.

Algunas de las limitaciones identificadas:

* **Información sesgada**: Información ya refinada, es decir, no viene en crudo (“raw”) sino que puede haber sido previamente preprocesada por el proveedor de datos, lo que puede otorgar algo de confusión al no saber cómo se ha construido esa ratio. Ejemplo: ELO de un equipo.
* **Volumen de datos limitado**: No disponemos de históricos de muchos años, tan sólo de algunas temporadas.
* **Alcance de datos limitado**: No se dispone de algunos datos más complejos, o de datos relativos a factores psicológicos o de toma de decisiones que pueden influir en el rendimiento de un jugador y, por tanto, en su posible valor de transferencia.
* **Acceso a la información:** El proveedor de datos no dispone de una API para conexión directa con los datos necesitados. La descarga de datos se realiza de forma manual y en caso de necesitar nuevas temporadas, se realiza una petición al proveedor.
* **Distribución de los precios de transferencia:** Parece lógico comprender que la mayoría de traspasos en el mundo del fútbol se realizan por importes pequeños. Esto muestra que no siguen una distribución “Normal”, y por tanto el número de casos según el importe del traspaso puede mostrar desequilibrios (habiendo muchos traspasos bajos, y pocos traspasos altos). Esto puede tener un impacto a la hora de entrenar el algoritmo, ya que tiene menos casos de los que aprender y, por tanto, puede presentar más dificultades para predecir de forma correcta.

Debido a estas limitaciones, la creación de un proceso de estimación que ofrezca unas garantías mínimas se presenta como uno de los principales desafíos del tfm.

# METODOLOGÍA

En este apartado procederemos a definir la metodología propuesta para alcanzar el objetivo principal identificado.

## INTRODUCCIÓN

El presente estudio tiene como objetivo principal la predicción del precio de traspaso actual de jugadores de fútbol en base a sus estadísticas históricas y su valor de mercado. Para ello, se utilizarán datos de un listado de jugadores que incluyen diferentes atributos (edad, posición, equipo de origen y destino, goles, asistencias, entre otros). Este análisis se lleva a cabo mediante un enfoque observacional retrospectivo, dado que se parte de información histórica ya registrada y se procesa para realizar predicciones basadas en modelos de machine learning.

### DISEÑO DEL ESTUDIO

#### El estudio se enfoca en la predicción del precio de traspaso de los jugadores de fútbol mediante modelos predictivos que consideran múltiples variables que afectan su valoración en el mercado. Los datos serán tratados desde dos niveles de análisis:

#### ANÁLISIS MACRO

#### En este nivel, se trabajará con información de un amplio conjunto de jugadores de fútbol provenientes de distintas ligas y temporadas. Este análisis busca identificar patrones generales en la valoración de mercado de los jugadores y su rendimiento en el campo.

#### ANÁLISIS MICRO

El análisis micro se centrará en un subconjunto específico de jugadores, realizando un análisis detallado de cómo sus estadísticas individuales, rendimiento, y otros factores impactan en su precio de traspaso. El objetivo es evaluar si los modelos pueden predecir de manera precisa el precio de traspaso de jugadores individuales.

Posteriormente, se integrará en una hoja de Excel para realizar el tratamiento y depuración de la información y proceder al análisis de los datos.

### MUESTRA

#### La muestra está compuesta por un conjunto de jugadores de diversas ligas y temporadas, incluyendo estadísticas clave como la cantidad de goles, asistencias, partidos jugados, edad, posición, y sus valores de mercado en diferentes momentos. El listado de jugadores y sus datos se cargan a través de una aplicación de Streamlit, lo que facilita la interacción con los usuarios. Los datos son procesados y almacenados en una base de datos de BigQuery para su análisis.

### PROCEDIMIENTO

#### ANÁLISIS MACRO

Se ha diseñado un sistema automatizado en el que los usuarios cargan un archivo **Excel** con las estadísticas de los jugadores a través de una interfaz web construida con **Streamlit**. Estos datos son almacenados en **BigQuery** y se utiliza un modelo previamente entrenado para hacer las predicciones de los precios de traspaso. El modelo emplea técnicas de machine learning para evaluar múltiples variables, como el rendimiento de los jugadores, su edad, su nacionalidad, entre otros.

#### ANÁLISIS MICRO

Este nivel de análisis se centra en jugadores específicos. A partir de las estadísticas individuales de cada jugador, el sistema predice cuál debería ser su precio de traspaso en el mercado actual. La predicción es procesada a través de un modelo **PKL** alojado en un servidor, utilizando una API basada en **Flask**. El resultado final se muestra nuevamente en **Streamlit**, donde se presenta al usuario la predicción del precio de traspaso del jugador.

# DESARROLLO

El desarrollo del presente proyecto se ha estructurado en varias fases claramente definidas, cada una con objetivos y tareas específicas. A continuación, se presenta la composición final del desarrollo del proyecto, dividido en las distintas partes que lo conforman:

**1. Fase de Preprocesamiento de Datos**

En esta primera etapa se llevó a cabo la preparación del conjunto de datos necesario para la ejecución del modelo. El proceso incluyó la recolección de la información inicial en formato Excel, que contenía un listado de jugadores de fútbol con sus estadísticas y valores de mercado.

Posteriormente, se procedió a realizar la carga del archivo en la aplicación desarrollada en **Streamlit**, desde donde se automatizó la conversión de los datos en formato adecuado. A lo largo de esta fase se abordaron tareas de limpieza de datos, codificación de variables categóricas y normalización de características numéricas, empleando técnicas como el **Label Encoding** y el uso de embeddings para campos categóricos (como la posición de los jugadores y nacionalidad).

**2. Carga de Datos en BigQuery**

Tras el preprocesamiento, los datos fueron cargados en un bucket de **Google Cloud Storage**. Desde allí, mediante un servicio de **Cloud Run**, se ejecutó un proceso automatizado para almacenar la información en **BigQuery**, la base de datos de almacenamiento en la nube. Este proceso aseguró la persistencia de los datos y su disponibilidad para las siguientes fases del proyecto.

**3. Entrenamiento y Aplicación del Modelo de Predicción**

En esta fase, se desarrolló un servicio en **Cloud Run** para el entrenamiento del modelo de predicción de precios de traspaso de jugadores. El modelo se entrena utilizando datos históricos de jugadores, aplicando técnicas de **machine learning** para aprender de las características como edad, posición, equipo de origen, rendimiento en el campo (goles, asistencias, etc.) y valor de mercado en transferencias anteriores.

Una vez entrenado el modelo, se generaron archivos **.pkl** con los modelos entrenados, que fueron almacenados para su posterior uso. Además, se estableció una rutina para cargar el modelo más reciente y aplicar las predicciones sobre los datos procesados.

**4. Desarrollo de la API de Comunicación**

Para intermediar entre la interfaz de **Streamlit** y el modelo entrenado, se implementó una **API REST** utilizando **Flask**. La API permite a Streamlit enviar los datos preprocesados, recibir las predicciones generadas por el modelo y devolver los resultados para ser mostrados al usuario de manera interactiva en la interfaz de la aplicación.

Esta API sirve como puente entre el modelo de machine learning y la visualización en Streamlit, gestionando de manera eficiente las peticiones y respuestas en tiempo real.

**5. Despliegue y Visualización en Streamlit**

Finalmente, la fase de visualización en **Streamlit** permite que el usuario cargue nuevos archivos de datos en formato Excel y vea, tras el procesamiento, el precio de traspaso estimado para cada jugador. Los resultados de las predicciones son mostrados de forma clara e interactiva, permitiendo al usuario evaluar las estimaciones en función de las estadísticas de cada jugador.

Además, la interfaz incluye funcionalidades para filtrar y ordenar los jugadores por diversos parámetros como el equipo, la posición o la nacionalidad.

Interfaz de usuario gráfica, Aplicación, Teams

Descripción generada automáticamente

Arquitectura del proyecto