



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

**Modelo predictivo para determinar resultados de partidos de fútbol  
en la Liga EA Sports 2024**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:**

Ingeniero de Sistemas

**AUTORES:**

Siancas Huaman, Daniel Alberto ([orcid.org/0000-0002-5930-6341](https://orcid.org/0000-0002-5930-6341))

Valencia Yovera, Franklin Paul ([orcid.org/0000-0002-1344-1919](https://orcid.org/0000-0002-1344-1919))

**ASESOR:**

M. Sc. Tavera Ramos, Anthony Paul ([orcid.org/0000-0002-4159-930X](https://orcid.org/0000-0002-4159-930X))

**LÍNEA DE INVESTIGACIÓN:**

Sistema de Información y Comunicaciones

**LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:**

Apoyo a la reducción de brechas y carencias en la educación en todos sus niveles

PIURA – PERÚ

2024

## Declaratoria de autenticidad del asesor



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

### Declaratoria de Autenticidad del Asesor

Yo, TAVARA RAMOS ANTHONY PAUL, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - PIURA, asesor de Tesis titulada: "Modelo Predictivo para determinar resultados de partidos de Fútbol en La Liga EA Sports 2024", cuyos autores son SIANCAS HUAMAN DANIEL ALBERTO, VALENCIA YOVERA FRANKLIN PAUL, constato que la investigación tiene un índice de similitud de 10%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

PIURA, 01 de Junio del 2024

Apellidos y Nombres del Asesor:	Firma
TAVARA RAMOS ANTHONY PAUL DNI: 40784283 ORCID: 0000-0002-4159-930X	Firmado electrónicamente por: ATAVARAR el 08-06- 2024 10:10:11

Código documento Trilce: TRI - 0755322



# Declaratoria de originalidad del autor/ autores



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

## Declaratoria de Originalidad de los Autores

Nosotros, SIANCAS HUAMAN DANIEL ALBERTO, VALENCIA YOVERA FRANKLIN PAUL estudiantes de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - PIURA, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Modelo Predictivo para determinar resultados de partidos de Fútbol en La Liga EA Sports 2024", es de nuestra autoría, por lo tanto, declaramos que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. Hemos mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumimos la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual nos sometemos a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

<b>Nombres y Apellidos</b>	<b>Firma</b>
DANIEL ALBERTO SIANCAS HUAMAN <b>DNI:</b> 75689387 <b>ORCID:</b> 0000-0002-5930-6341	Firmado electrónicamente por: DASIANCASS el 02-06-2024 10:22:22
FRANKLIN PAUL VALENCIA YOVERA <b>DNI:</b> 77430362 <b>ORCID:</b> 0000-0002-1344-1919	Firmado electrónicamente por: FPVALENCIAY el 02-06-2024 10:27:09

Código documento Trilce: TRI - 0755399

### **Dedicatoria**

A Dios sobre todas las cosas, porque ha sido quien me ha ayudado y me ha guiado a concluir esta etapa de mi vida. A mi familia, en primer lugar, a mis padres A todas las personas que de alguna manera influyeron positivamente en mi continuo aprendizaje en toda esta época universitaria con sus consejos en momentos oportunos, sus oraciones, sus desvelos y palabras de aliento.

### **Agradecimiento**

Quiero expresar mi gratitud al Ing. Anthony Tavera por su invaluable apoyo y conocimientos durante la realización de este proyecto. También agradezco profundamente a mi familia por ser una parte integral de mi formación profesional y por creer siempre en mí.

## Índice de contenidos

Carátula .....	i
Declaratoria de autenticidad del asesor .....	ii
Declaratoria de originalidad del autor/ autores .....	iii
Dedicatoria.....	iv
Agradecimiento .....	v
Índice de contenidos.....	vi
Índice de tablas.....	vii
Índice de figuras .....	viii
Resumen .....	ix
Abstract .....	x
I. INTRODUCCIÓN .....	1
II. MÉTODOLÓGÍA.....	9
III. RESULTADOS.....	14
IV. DISCUSIÓN.....	26
V. CONCLUSIONES .....	30
VI. RECOMENDACIONES .....	31
REFERENCIAS .....	32
ANEXOS.....	37

## Índice de tablas

Tabla 1:Frecuencias y Porcentajes de Resultados de Partidos.....	17
Tabla 2: Goles .....	17
Tabla 3: Tiros.....	18
Tabla 4:Tarjetas.....	19
Tabla 5:Corners .....	20
Tabla 6:Prueba normalidad goles .....	21
Tabla 7:ANOVA goles.....	21
Tabla 8:Prueba normalidad tiros .....	22
Tabla 9: ANOVA tiros.....	22
Tabla 10: Prueba normalidad tarjetas .....	23
Tabla 11: ANOVA tarjetas.....	23
Tabla 12: Prueba normalidad corners .....	24
Tabla 13:ANOVA corners.....	24

## Índice de figuras

Figura 1:Columnas del conjunto de datos .....	14
Figura 2:Numero de temporadas de la liga EASports .....	15
Figura 3:Lista de equipos en la liga EASports.....	15
Figura 4:Total de partidos en la liga EASports .....	16
Figura 5:Frecuencia de partidos.....	16
Figura 6: Matriz de confusión .....	25

## Resumen

El presente estudio tuvo como objetivo general desarrollar e implementar un modelo predictivo para determinar el resultado de un partido en La Liga EA Sports 2024. Esta investigación es de tipo aplicada con un enfoque cuantitativo y tuvo un diseño no experimental. Se tomo como población todos los datos relevantes para la construcción del modelo predictivo. La recolección de datos se obtuvo mediante las fichas de registro y la observación, el análisis de datos se realizó mediante el software estadístico Jamovi. Entre los principales resultados se llegó a demostrar que uno de los factores más influyentes en el resultado de un partido es el de marcar goles esto de acuerdo a la prueba de ANOVA. Se logro un porcentaje de precisión del 73.02% el cual se logró mediante la matriz de confusión que se desarrolló en el mismo modelo. Por lo tanto, se logra alcanzar los objetivos de investigación. Como conclusión se logró desarrollar e implementar un modelo predictivo para determinar el resultado de un en La Liga EA Sports 2024.

**Palabras clave:** Modelo predictivo, resultado de un partido de futbol, factores influyentes, redes neuronales, machine learning

## **Abstract**

The general objective of this study was to develop and implement a predictive model to determine the outcome of a match in the EA Sports 2024 League. This is an applied research with a quantitative approach and a non-experimental design. The population was taken as all relevant data for the construction of the predictive model. The data collection was obtained through registration forms and observation, the data analysis was performed using Jamovi statistical software. Among the main results, it was demonstrated that one of the most influential factors in the outcome of a match is scoring goals, according to the ANOVA test. A percentage of accuracy of 73.02% was achieved through the confusion matrix that was developed in the same model. Therefore, the research objectives were achieved. In conclusion, it was possible to develop and implement a predictive model to determine the outcome of an EA Sports 2024 League.

**Keywords:** predictive model, soccer match outcome, influencing factors, neural networks, machine learning

## I. INTRODUCCIÓN

El fútbol es probablemente el deporte de equipo más popular a nivel mundial, capturando la atención de millones de aficionados y analistas deportivos. La fascinación por este deporte no solo reside en su dinamismo y emoción, sino también en la complejidad de predecir los resultados de los partidos. Este desafío ha sido un área de interés significativo para investigadores, dado que una mezcla compleja de factores y eventos aleatorios influyen en el resultado final de un partido (Berrar et al., 2019). El aprendizaje automático ha emergido como una herramienta crucial en el análisis deportivo, permitiendo identificar variables predictivas clave para el rendimiento en el fútbol profesional. (Geurkink et al. 2021). Asimismo, el uso del aprendizaje automático ha revolucionado la predicción de resultados en el fútbol, permitiendo un análisis más preciso y detallado de los factores que influyen en el desempeño de los equipos. (Azhari et al. 2018).

Estudios previos han explorado diversos enfoques para predecir resultados de fútbol, incluyendo redes neuronales artificiales (Anfilets et al., 2020), árboles de decisión (Azeman et al., 2020) y modelos lineales generalizados mixtos (Badiella et al., 2022), asimismo otro enfoque como Artificial Neural Networks (ANN) se utilizaron para la predicción de goles en un partido de futbol (Jones et al. 2023). Por otro lado, exploraron métodos híbridos que combinan algoritmos de agrupamiento y clasificación para predecir resultados de partidos (Kinalioğlu & Kuş 2023). Estos estudios han identificado variables clave, como goles marcados, tiros a puerta y tarjetas, como predictores significativos del rendimiento de los equipos (Geurkink et al., 2021; Capobianco et al., 2019). Asimismo, se identificó que la información sobre los jugadores es una variable clave para la predicción de partidos (Kinalioğlu & Kuş 2023).

El presente estudio tiene como objetivo desarrollar y evaluar un modelo predictivo basado en redes neuronales artificiales para pronosticar los resultados de partidos de fútbol en La Liga EA Sports 2024. Específicamente, nos proponemos: (1) Identificar las variables clave que influyen en el resultado de los partidos y (2) Determinar la eficacia del modelo predictivo en términos de precisión y capacidad para predecir victorias, empates y derrotas."

En la presente revisión de literatura según, Anfilets, S, et al (2020) en su artículo de investigación científica se centra en la predicción del ganador de un partido de futbol en la Premier league inglesa donde proponen una red neuronal multicapa basada en la regularización de la red elástica. Como método utilizaron datos importantes que pueden determinar el ganador de un partido siendo estos los tiros a puerta como también los corners y las faltas cometidas por los equipos entre otros parámetros importantes que ayudarían en esta investigación. Los resultados de este sistema dieron resultados muy buenos en cuanto a la predicción de partidos. También nos dice que su sistema puede mejorarse desarrollando así una red neuronal más compleja. (Anfilets et al., 2020).

Según, Manack y Van Zyl (2020). Su artículo de investigación consiste en utilizar redes siamesas y los modelos de pares ranknet en conjunto con el aprendizaje por transferencia. Se implementaron estos modelos junto a las clasificaciones de los partidos de futbol como también así PageRank que está basado en gráficos estadísticos todo esto para obtener clasificaciones en la premier league de Inglaterra desde la temporada 2006 hasta la temporada 2018. Los conjuntos de datos utilizados para esta investigación fueron las estadísticas estacionales de los equipos y aumentaron estas estadísticas mediante datos externos como lo son las transferencias y los datos monetarios. Su modelo llega a una puntuación menor a la de una red neuronal, pero a la hora de identificar empates y clasificaciones de la temporada tiene un mejor comportamiento. En cuanto a resultados el aprendizaje por transferencia dio mejores resultados obteniendo mejores resultados en las clasificaciones por partidos, mientras que los datos aumentados proporcionan un mejor resultado en clasificaciones por temporada. (Manack & Van Zyl, 2020).

Asimismo Konefał Marek, et all (2020) en su artículo la cual su objetivo era examinar si las actividades físicas y técnicas de los jugadores de fútbol y las ubicaciones de los partidos pueden relacionarse con un mayor o menor probabilidad para ganar partidos que cuyo resultados pueden calificarse como reñido, por ende los autores observaron 7972 jugadores individuales de la Bundesliga alemana durante las temporadas, 2014, 2015, 2016 y 2017, la cual utilizaron el modelo estadísticos para determinar la carga del entrenamiento individual relacionada a la ubicación del jugador en el campo, donde concluyeron

que el modelo propuesto puede ser utilizado por los entrenadores para la identificación de actividades físicas y técnicas de los jugadores para determinar los resultados del partido de fútbol. (Konefał et al., 2020).

Según, Cho, Y, et al. (2018) En su artículo utiliza el análisis de las redes sociales junto con el potenciador de gradiente para ayudar al supervisor de campo a tener una mejor estrategia para poder ganar los partidos. El aprendizaje automático es la principal técnica para poder predecir la victorias y derrotas en los partidos de futbol, como también utiliza otras técnicas como la regresión logística, el árbol de decisión entre otros. Llegaron a la conclusión que mediante un análisis se puede representar el rendimiento de los equipos y que con el aprendizaje automático se puede tener un sistema más preciso de predicción. (Cho et al., 2018).

Pablo Caballero, et all. (2017) Desarrollaron un estudio empírico para determinar el resultado de un partido, los cuales utilizaron como muestra una temporada del 2014 y 2015 de la liga regional sub'18 y la que contaba con 12 equipos, donde sus variables fueron analizar los goles de la victoria, empate o derrota. Para comprobar la influencia de las variables situacionales, tanto como el resultado de los partidos desarrollaron una regresión lineal. (Caballero et al., 2017)

Según, Diniz, M. et al. (2018) En su artículo de revisión científica propone 2 modelos bayesianos utilizando la distribución multinomial de Dirichlet los cuales los compara con 3 modelos predictivos estándar. Estos modelos fueron utilizados en la predicción de partidos en la liga brasileña en concreto en 1710 partidos dando resultados en tiempo real. Sus resultados determinan que los modelos Dirichlet no tienen nada que envidiarles a los modelos estándar con los que se comparó y que son muy competitivos. (Diniz et al., 2019)

Como también, Manderson, et al (2018) En su artículo de revisión científica desarrolla un modelo bayesiano dinámico para determinar los resultados de los partidos de futbol de los equipos de la Liga de futbol Australiana basado en las diferencias de distribución de Poisson .(Manderson et al., 2018)

Asimismo, Malantinos et al., (2018) en su artículo de revisión científica evalúa los métodos de predicción de resultados de partidos de futbol en la Superliga de Grecia donde utiliza varios métodos de ingeniería para realizar análisis de datos como también para hacer la limpieza de estos mismos asimismo hacer un

aumento cuando se requiera. Donde también utiliza cinco algoritmos conocidos para realizar comparativas en dos ligas más como la Premier League y la Eredivise. (Malamatinos et al., 2022)

Connor Mark, et al (2022) su artículo de investigación consiste en cuantificar y modelar las demandas de velocidad de los jugadores de fútbol profesional que participan en la liga del campeonato inglés, para así comparar el efecto de la ubicación del partido y analizar el efecto de la posición en los resultados de la velocidad del jugador a través de la temporada donde se aplicaron cálculos de medias móviles a los datos brutos de velocidad en la duración de los partidos donde concluyeron de que la velocidad de juego se ajustaba a un modelo de ley de potencia .(Connor et al., 2022)

Pablo Medina, et al.(2021) En su artículo el cual su objetivo era un modelo de regresión para determinar las principales pautas de un partido de fútbol, a través de las probabilidades de ganar, empatar o perder, como visitante o como equipo local, el modelo propuesto está basado en el rendimiento histórico como también mejora al agregar descriptores de red de los dos equipos en un partido de fútbol, la cual estas medidas de red brindan información útil para determinar los resultados de los partidos, donde validaron su enfoque utilizando los datos de la liga española, donde concluyeron que la centralidad de la interrelación parece brindar información útil y relevante relacionada con el rendimiento de un equipo de fútbol durante el torneo de la liga española. (Medina et al., 2021)

Por otra parte, Tsokos et al., 2019 Realizan una comparación de dos modelos, tales como Bradley-Terry y un modelo jerárquico de Poisson para determinar su rendimiento en cuanto a la predicción de resultados de partidos de fútbol lo cual se concluye que los resultados de los dos modelos tienen un rendimiento similar en cuanto a la predicción de resultados. (Tsokos et al., 2019)

Como también Arntzen & Hvattum, (2021) Utilizan dos modelos estadísticos para predecir los resultados de fútbol los cuales son el modelo de regresión logarítmica y el segundo que se basa en la modelización del riesgo, donde demuestra que no hay diferencias en cuanto a las predicciones antes de los partidos. (Arntzen & Hvattum, 2021)

Mientras que, Hassan et al., (2020) Proponen un enfoque diferente para predecir victorias y derrotas y la sensibilidad de los atributos en el mundial de Rusia 2018, para esto hace uso de una red neuronal para determinar la eficacia de todos los atributos de los partidos de futbol. (Hassan et al., 2020)

Asimismo, en el presente año se realiza un estudio en La liga, la Premier League y la Major League Soccer donde se hace uso de modelos de aprendizaje automático. (Bai et al., 2022)

Razali, N ,et al (2018) En su artículo de investigación la cual tuvo como objetivo estudiar las redes bayesianas como método para poder predecir resultados de partidos de futbol, su objetivo de estudio fue revisar la aplicación de las redes bayesianas en cuanto a la predicción de partidos de futbol, también nos dice que este método se puede utilizar en futuras investigaciones para poder desarrollar un nuevo método de predicción utilizando las redes bayesianas en conjunto con características para ayudar a aumentar el rendimiento de este modelo de predicción. (Razali et al., 2018)

Talko B. Dijkhuis et al, (2021) en su artículo con respecto a su documentación de sus resultados para predecir el rendimiento físico individual, la cual se utilizaron algoritmos basados en arboles llamados Random Forest y Decision Tree y que mediante los datos de seguimiento de los jugadores en los que se basaron en 302 partidos de fútbol de elite en la temporada holandesa, donde las variables de los jugadores se expresan por la distancia recorrida, velocidad y gasto energético donde obtuvieron como resultados que si es posible detectar a los futbolistas de bajo rendimiento mediante una aplicación de aprendizaje automático y combinación de monitoreo del gasto emergentico. (Dijkhuis et al., 2021)

Lee,J et al. (2022)En su artículo de investigación científica nos propone un método utilizando redes bayesianas de poisson para poder evaluar cual parámetros que ayudan a tener ventaja jugando de local y dicho cambio de esta ventaja, estos parámetros utilizados los utilizan para mejorar el rendimiento de los modelos de aprendizaje automático en los partidos de futbol en el marco del covid-19. (Lee et al., 2022)

Hubaček, Šourek y Zálezny (2019) En su artículo de investigación científica proponen a investigar sobre los diferentes modelos de predicción de resultados

de partidos de fútbol probándolos así para dar un comparativas. Los modelos que probaron fueron: El modelo de Poisson doble, las clasificaciones ELO, las clasificaciones Pi, y PageRank. Donde concluyen que el modelo de Poisson doble fue el que mejor se comportó, pero las clasificaciones Pi fueron las que superaron a los otros modelos. (Hubaček et al., 2019)

Zhou et al. (2021) en su artículo Donde su propósito fue determinar el papel de cada indicador sobre el rendimiento de la determinación de los resultados de los partidos donde ha variado a través de un análisis después de un periodo a largo tiempo, donde utilizaron el modelo de regresión logística binaria para poder medir el nivel de relación de estos factores, donde obtuvieron como resultados que la influencia de varios factores influye sobre el cambio de los resultados de los partidos a largo plazo. (Zhou et al., 2021)

A continuación, se mostrarán las bases teóricas que sustentan la investigación con respecto a modelo predictivo para determinar los resultados de partidos de fútbol.

El estudio tiene relación con múltiples teóricas que fortalecen las razones por las cuales es de importancia un modelo predicativo en base a determinar los resultados de partidos de fútbol.

Según, Yash Ajgaonkar et al (2021) un modelo predictivo es una de las técnicas inteligentes para demostrar resultados optimistas en el campo de la clasificación y asimismo la precisión de la predicción. Como en la predicción deportiva la cual es un área en expansión en cuanto a la buena precisión de la predicción, dado que el fútbol es una de las áreas de investigación interesante la cual se considera complejo y dinámico en comparación con otros deportes. (Ajgaonkar et al., 2021)

Asimismo, Rabindra Lamsal y Choudhary (2020) que Machine Learning se puede utilizar eficazmente en diversas ocasiones en el deporte, ya sea en el campo como fuera de él. Donde se aplica análisis del estado físico de los jugadores, como también el diseño de tácticas ofensivas o el resultado de un partido. (Lamsal & Choudhary, 2020)

Para, Azeman, et al. (2020) La predicción del fútbol se ha convertido en un problema interesante, donde los investigadores tratan de encontrar soluciones diferentes, donde la idea de modelar y construir un modelo inteligente que analice los datos disponibles y anticipe los resultados de los partidos, la cual se ha hecho cada vez más popular en los últimos años, el aprendizaje automático es una de las ramas de la inteligencia artificial que se ocupa de diseñar algoritmos que permitan a los ordenadores tener la ventaja de aprender sin necesidad de programar las reglas de cada cuestión, la cual se han desarrollado diferentes métodos de predicción para evaluar las características que hacen que un equipo de fútbol pierda un partido o lo gane. (Azeman et al., 2020)

Para Keshtkar y Yamaghani (2019) comenta que los enfoques de minería de datos y el aprendizaje automático tienen como finalidad descubrir información o conocimientos implícitos, previamente desconocidos y potencialmente útiles, dado que las relaciones entre los resultados deportivos y otros elementos de los datos que se ven afectados por varios factores, como el tipo de deporte, el entorno y los objetivos de los jugadores, la cual se han recomendado varios métodos para predecir los resultados de fútbol a partir de los datos disponibles. (Keshtkar Langaroudi & Yamaghani, 2019)

Así mismo Wei Yang (2021) en el campo del deporte, dada la imprevisibilidad de los resultados de partidos de fútbol y la predicción de cada partido jugado ha despertado una gran inquietud entre los medios de comunicación, aficionados, expertos en deportes y otras partes interesadas, asimismo es un tema importante en los círculos académicos, donde la fuerza del equipo está conformada por los jugadores el cual el rendimiento es un factor clave para ganar un partido de fútbol. (Wei, 2021)

Según, Giménez et al. (2020) El fútbol es un deporte de equipo de gran complejidad que se caracteriza por acciones de alta intensidad, en las cuales el rendimiento se basa de diferentes técnicas, tácticas, biomecánicas, psicológicas y fisiológicas. Debido que es una amplia gama de factores las cuales contribuyen al rendimiento en el fútbol, tanto que ha puesto un relieve de importancia de tener en cuenta los indicadores claves del rendimiento durante el entrenamiento y las

competencias con el fin de mejorar los ejercicios de entrenamiento y controlar los requerimientos de las competiciones. (Giménez et al., 2020)

Asimismo, Steffen Lang et al (2022) una característica importante de la estructuración de un partido de fútbol es el estado del juego, si el partido está detenido o en juego, es necesario calcular los indicadores de rendimiento en relación al tiempo de juego eficaz y así encontrar situaciones estándar, acciones con el balón y otras estructuras tácticas en los datos temporales.(Lang et al., 2022)

Para, Daniel Berrar et al (2019) El fútbol es probablemente el deporte de equipo más popular del mundo, en gran parte de la fascinación por el fútbol se debe que casi la mayoría de los partidos jugados terminan en empates o se ganan por sólo dos o menos goles, además de una complejidad de mezcla de factores y sucesos de fortuitos juegan un papel importante en la determinación del resultado final de un partido de fútbol. (Berrar et al., 2019)

Almulla Jassim y Alam Tanvir (2020) Afirman que dado que el fútbol es uno de los deportes más populares en todo el mundo, el cual ganar partidos de fútbol se está volviendo en un aspecto esencial para los clubes, donde analizar el rendimiento colectivo de los jugadores en posiciones de juego se convirtió la clave para comprender su eficacia a la hora de ganar partidos, las cuales se pueden utilizar un modelo de aprendizaje automático para identificar varias características y predecir el ganador de un partido. (Almulla & Alam, 2020)

Capobianco, G, et al. (2019) para determinar el resultado de un partido de futbol muchas veces es muy difícil ya que existen varios factores que de alguna u otra forma puede cambiar el resultado de los partidos y los cuales están relacionados con el comportamiento del jugador. (Capobianco et al., 2019)

## II. METODOLOGÍA

### 2.1. Tipo y diseño de investigación

#### Tipo de investigación

La investigación es de tipo aplicada como indica Hernández et al, (2016) “La investigación aplicada produce conocimiento para que sea aplicable directamente a los problemas de la sociedad” Este tipo de investigación proviene de la investigación básica y gracias a esto se genera un gran valor agregado por la utilización de conocimiento. (Hernández et al., 2016)

#### Diseño de investigación

El diseño de investigación no experimental como indica Hernández et al, (2016) Es el diseño donde no se manipulan las variables de manera intencional y se basa totalmente en la observación de fenómenos para luego analizarlos(Hernández et al., 2016).

La investigación será de tipo longitudinal. Según Roehl Sybing (2019) un estudio longitudinal es una forma de investigación que examina y analiza los cambios de un fenómeno a lo largo del tiempo, los cuales puede arrojar resultados de gran impacto.(Sybing, s. f.2019)

#### Variables y operacionalización

- **Variable:**

Modelo predictivo

#### Definición conceptual

Un modelo predictivo es una de las técnicas inteligentes para demostrar resultados optimistas en el campo de la clasificación y asimismo la precisión de la predicción (Ajgaonkar et al., 2021)

#### Definición operacional

Para la medición del modelo predictivo se evaluará el porcentaje de precisión y se analizaran el número de datos por importancia.

#### Indicadores

Porcentaje de precisión

Datos obtenidos

#### Escala

De Razón

- **Variable:**  
Resultados de futbol

### **Definición conceptual**

El fútbol es probablemente el deporte de equipo más popular del mundo, en gran parte de la fascinación por el fútbol se debe que casi la mayoría de los partidos jugados terminan en empates o se ganan por sólo dos o menos goles, además de una complejidad de mezcla de factores y sucesos de fortuitos juegan un papel importante en la determinación del resultado final de un partido de fútbol. (Berrar et al., 2019)

### **Definición operacional**

Para determinar los resultados de un partido se evalúa mediante los goles marcados, los partidos, jugados, corners, como también las tarjetas amarillas y rojas

### **Indicadores**

Goles marcados

Tiros

Tarjetas

Corners

### **Escala**

De razón

## **2.2. Población, muestra y muestreo**

### **2.2.1. Población**

Como nos dice López-Roldán & Fachelli, (2017) La población se refiere al conjunto de individuos o elementos que componen un grupo de estudio de interés analítico. En general la población es un grupo de unidades donde se extrae la muestra. (López-Roldán & Fachelli, 2017)

La población de este estudio comprendió todos los partidos de fútbol disputados en La Liga EA Sports en las temporadas 2019-2024, que incluyó un total de 27 equipos y 1900 partidos (Hubáček et al., 2019). Se consideraron todos los partidos oficiales de la temporada, excluyendo partidos de copa y de supercopa (Geurkink et al., 2021). La elección de esta población se basó en la relevancia y actualidad de los datos, así como en la importancia de La Liga EA Sports como una de las principales ligas de fútbol a nivel mundial (Berrar et al., 2019)."

### **2.2.2. Muestra**

Dado que se utilizarán todos los datos disponibles de la temporada 2019-2024 de La Liga EA Sports, el tamaño de la muestra incluye todos los partidos oficiales de liga jugados durante estas temporadas. En total, se analizarán 1900 partidos (27 equipos jugando 38 partidos cada uno).

### **2.2.3. Muestreo**

El método de muestreo empleado en este estudio es un muestreo censal, ya que se incluyen todos los elementos de la población relevante para el análisis. En este caso, todos los partidos de liga oficiales de las temporadas 2019-2024 de La Liga EA Sports forman parte de la muestra.

Se justifica que la muestra es totalmente representativa de la población, ya que incluye todos los partidos oficiales de la temporada 2024 de La Liga EA Sports. Esto garantiza que las conclusiones y resultados obtenidos del análisis sean aplicables a todas las temporadas y reflejen adecuadamente las dinámicas y patrones observados en los partidos de fútbol de esta liga.

## **2.3. Técnicas e instrumentos de recolección de datos**

Para la investigación se emplea la técnica del análisis documental, Un análisis documental está conformado por el conjunto de operaciones que difieren el contenido de documentos verídicos ya sea rehaciéndolos o convirtiéndolos en otros documentos con un carácter instrumental, con el fin de dar al investigador la facilidad de obtener información precisa de estos documentos. (Pinto, 1989)

## **2.4. Procedimientos**

Para el desarrollo del proyecto de investigación, se inició por identificar la problemática en estudio, el cual consiste en determinar el porcentaje de la precisión de un modelo predictivo para determinar el resultado de un partido de fútbol, es por ello que se han planteado el uso de dos variables de estudios , una independiente (modelo predictivo) y una dependiente(resultados de fútbol), una vez identificadas las viarias de estudio, se procedió a indagar situaciones similares en diferentes artículos científicos como revistas, repositorios nacionales e internacionales y toda investigación que tenga suficientemente similitud con el proyecto investigado.

Así mismo se hará uso de recolección datos de los partidos jugados de La Liga EA Sports en las temporadas 2019-2024, el cual va a permitir obtener información sobre los goles anotados, victorias, tarjetas, tiros, corners, para el cual se utilizarán algoritmos que permitirá comparar resultados mediante la implementación de un software, la cual mostrara los resultados mediante graficas conteniendo información fundamental de cómo se obtuvo el resultado.

Se realizó análisis de los diferentes antecedentes de cada artículos y tesis relacionadas con la presente investigación acerca del modelo predictivo y resultados de fútbol recolectando información de los métodos, algoritmos y situaciones la cual implementaron para determinar el ganador de un partido de fútbol.

## **2.5. Método de análisis de datos**

Para analizar los datos obtenidos en la investigación, se aplicó regresión logística, que es un tipo de análisis utilizado para predecir el resultado de una variable categórica en función de las variables independientes, la cual podemos medir el nivel de asociación entre estos factores y el resultado de los partidos. Con el algoritmo regresión logística es útil para modelar la probabilidad de un evento ocurriendo en función de otros factores, como partidos jugados, goles anotados, tarjetas amarillas, corners, tiros al arco.

Para realizar el análisis es necesario estudiar detalladamente la relación que existen entre las variables de investigación, la cual se añadieron hipótesis general, específicas. Donde la dimensión Machine Learning cuenta como indicadores, el porcentaje de precisión y datos obtenidos por importancia y una escala de

medición de razón, la dimensión resultados de fútbol tiene como indicadores como, goles marcados, tiros, corners, tarjetas amarillas como rojas con una escala de razón.

## **2.6. Aspectos éticos**

El presente estudio utilizo la base en fundamentos éticos como la privacidad de los datos obtenidos respaldada por la ley de protección de datos personales, Ley N°29733, la cual tiene como objetivo garantizar el derecho de la protección de su privacidad.

La investigación respecto a la autoría intelectual de las fuentes citando y mencionando a ellas, De acuerdo a las leyes establecidas por la Universidad Cesar Vallejo. Según el artículo 15 del código de ética en investigación. La universidad procura la originalidad de la investigación estableciendo el programa de evaluación de la investigación, Turnitin, a disposición de la comunidad universita

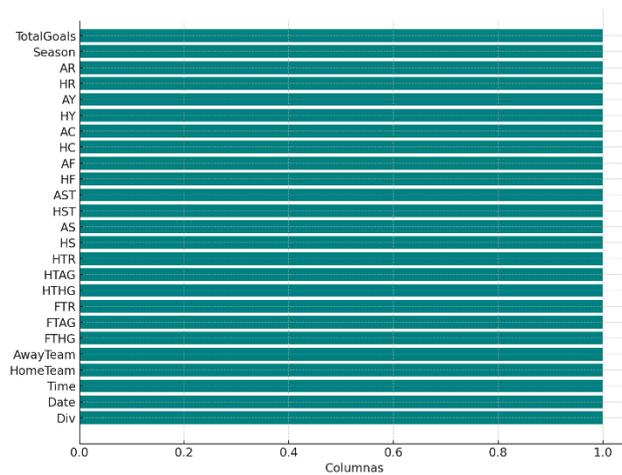
### III. RESULTADOS

Este capítulo presenta los resultados obtenidos del estudio sobre Modelo Predictivo Para Determinar Resultados De Partidos De Fútbol En La Liga EA Sports 2024, utilizando data histórica, indicadores como, tarjetas amarillas, tarjetas rojas, goles anotados, tiros a puerta, corners.

#### Estadística descriptiva

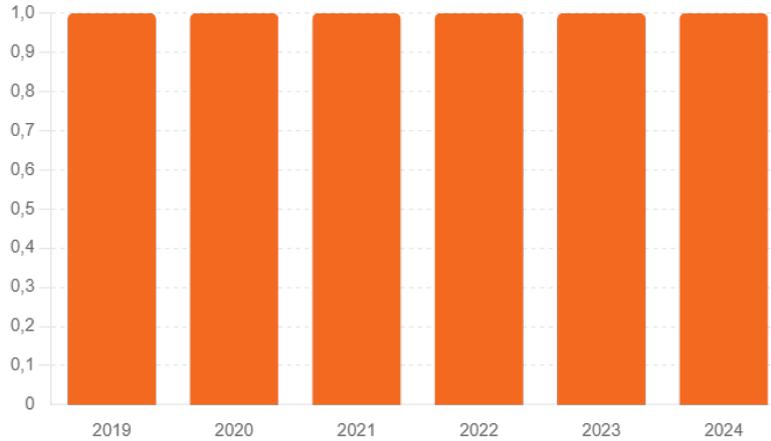
**Objetivo general:** Desarrollar e implementar un modelo predictivo para determinar el resultado de un en La Liga EA Sports 2024

Figura 1:Columnas del conjunto de datos



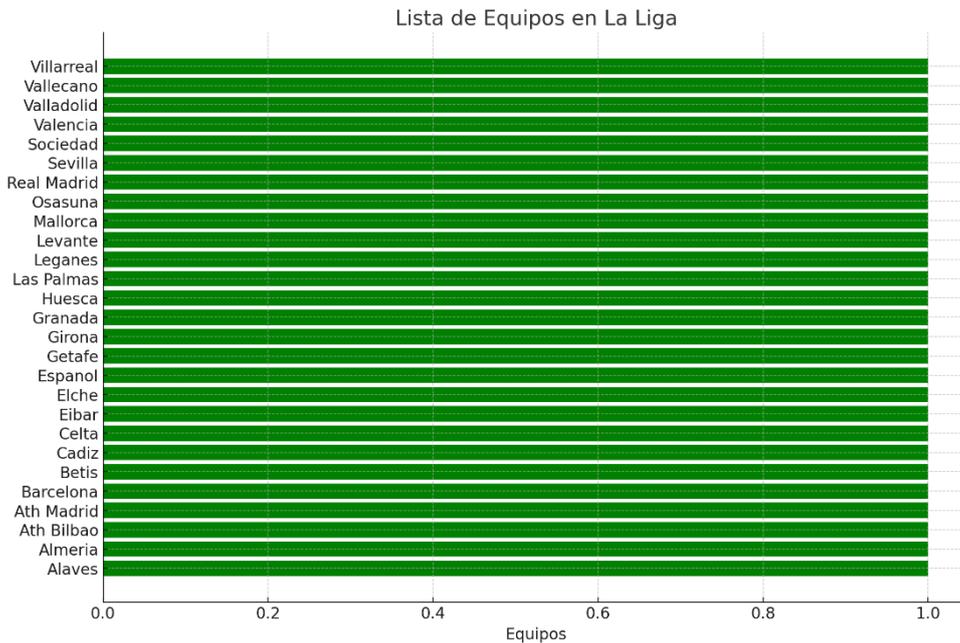
En la figura N°1 se procedió a enumerar todas las columnas disponibles en el conjunto de datos. Cada columna representaba un atributo diferente relacionado con los partidos de la liga, como los goles totales, tarjetas amarillas y rojas, tiros a puerta, equipos locales y visitantes, entre otros. Este gráfico mostraba que los datos eran muy detallados y abarcaban diversos aspectos del juego.

Figura 2:Numero de temporadas de la liga EASports



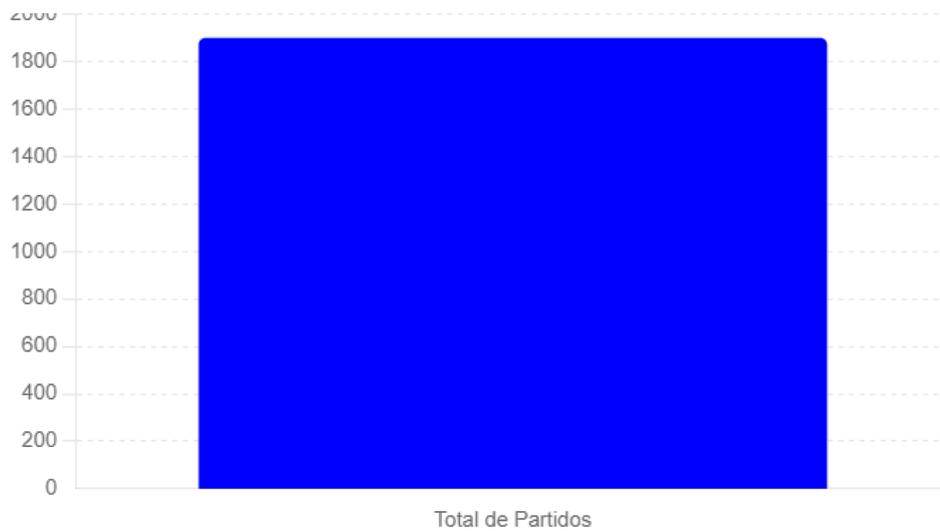
En la figura N°2 se procedió a mostrar las temporadas de La Liga desde 2019 hasta 2024. Cada barra representaba una temporada, y todas las barras alcanzaban el mismo nivel, indicando que había datos disponibles para todas esas temporadas de manera uniforme. Esto quería que el conjunto de datos estaba bien equilibrado temporalmente.

Figura 3:Lista de equipos en la liga EASports



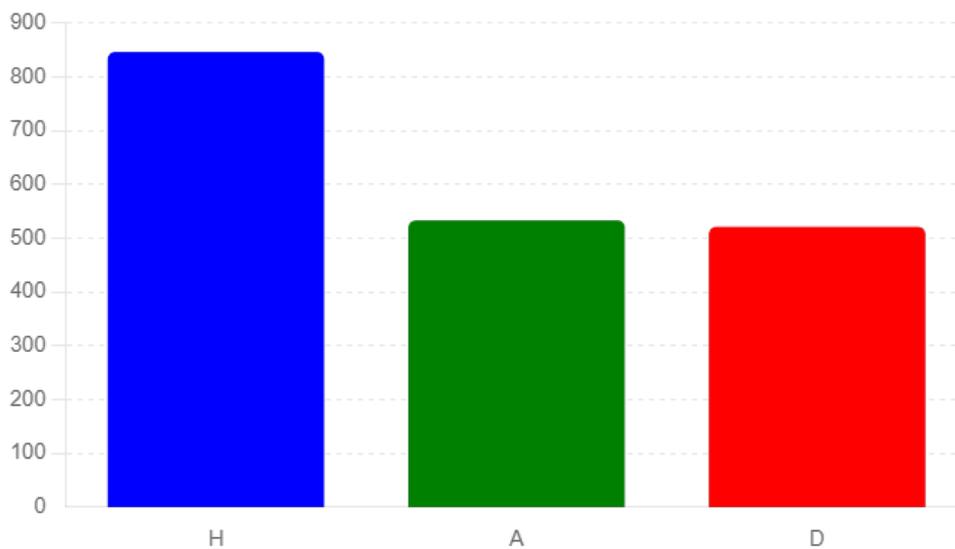
En la figura N°3 se procedió a mostrar los equipos que participaron en La Liga. Desde la temporada 2019 hasta la temporada 2024.Cada barra representaba un equipo diferente, indicando que la liga contaba con una diversidad significativa de equipos. La lista incluía equipos reconocidos como Real Madrid, Barcelona, Villarreal, Valencia, entre otros.

Figura 4: Total de partidos en la liga EASports



En la figura N°4 se procedió a mostrar el total de partidos jugados en la Liga EA Sports. El número total de partidos representaba un valor elevado, alcanzando casi los 2000 partidos. El gráfico indicaba que la liga tenía un gran volumen de partidos, reflejando una temporada completa con todos los equipos participantes.

Figura 5: Frecuencia de partidos



En la figura N° 5 el gráfico mostró que las victorias del equipo local fueron el resultado más común en la Liga EA Sports, seguidas por las victorias del equipo visitante y los empates.

Tabla 1:Frecuencias y Porcentajes de Resultados de Partidos

Frecuencias y Porcentajes de Resultados de Partidos		
Resultado del Partido	Frecuencia	Porcentaje
H	846.0	44.53
A	533.0	28.05
D	521.0	27.42

En la tabla N°1 La tabla se mostró las frecuencias y porcentajes de los diferentes resultados de los partidos en la Liga EA Sports. Los resultados se dividieron en tres categorías: victorias del equipo local (H), victorias del equipo visitante (A) y empates (D). Las victorias del equipo local (H) fueron las más comunes, con una frecuencia de 846 partidos, lo que representó el 44.53% de todos los partidos jugados. Esto significó que casi la mitad de los partidos fueron ganados por el equipo que jugaba en casa. Las victorias del equipo visitante (A) ocurrieron en 533 partidos, lo que correspondió al 28.05% del total. Esto indicó que más de una cuarta parte de los partidos fueron ganados por el equipo que jugaba fuera de casa. Los empates (D) ocurrieron en 521 partidos, representando el 27.42% del total de los partidos. Esto mostró que aproximadamente uno de cada cuatro partidos no tuvo un ganador definido.

**Objetivo específico:** Determinar qué factores influyen en el resultado de un partido

Tabla 2: Goles

Estadísticas Descriptivas de Goles (Equipo Local y Visitante)							
Variable	N	Perdidos	Media	Mediana	DE	Mínimo	Máximo
Goles Local (FTHG)	1900	0	1.433	1.000	1.235	0.000	7.000
Goles Visitante (FTAG)	1900	0	1.096	1.000	1.067	0.000	7.000

En la tabla N°2 el análisis de los goles anotados por los equipos locales, la media de goles fue de 1.433, con una mediana de 1.000. La desviación estándar (DE) de 1.235 indicó una considerable variabilidad en los goles anotados por el equipo local. El número mínimo de goles registrados fue 0, mientras que el máximo fue 7.

En cuanto a los equipos visitantes, la media de goles fue de 1.096, con una mediana de 1.000. La desviación estándar (DE) fue de 1.067, lo que también reflejó variabilidad en los goles anotados por el equipo visitante. El número mínimo de goles fue 0 y el máximo fue 7.

Tabla 3: Tiros

Estadísticas Descriptivas							
Variable	N	Perdidos	Media	Mediana	DE	Mínimo	Máximo
Tiros a Puerta Local (HST)	1900	0	4.491	4.000	2.467	0.000	17.000
Tiros a Puerta Visitante (AST)	1900	0	3.553	3.000	2.097	0.000	15.000

En la tabla N°3 el análisis de los tiros a puerta realizados por los equipos locales, la media de tiros a puerta fue de 4.491, con una mediana de 4.000. La desviación estándar (DE) de 2.467 indicó una considerable variabilidad en los tiros a puerta realizados por el equipo local. El número mínimo de tiros a puerta registrados fue 0, mientras que el máximo fue 17.

En cuanto a los equipos visitantes, la media de tiros a puerta fue de 3.553, con una mediana de 3.000. La desviación estándar (DE) fue de 2.097, lo que también reflejó variabilidad en los tiros a puerta realizados por el equipo visitante. El número mínimo de tiros a puerta fue 0 y el máximo fue 15.

Tabla 4: Tarjetas

Estadísticas Descriptivas							
Variable	N	Perdidos	Media	Mediana	DE	Mínimo	Máximo
Tarjetas Amarillas Local (HY)	1900	0	2.384	2.000	1.533	0.000	9.000
Tarjetas Amarillas Visitante (AY)	1900	0	2.520	2.000	1.543	0.000	8.000
Tarjetas Rojas Local (HR)	1900	0	0.121	0.000	0.348	0.000	2.000
Tarjetas Rojas Visitante (AR)	1900	0	0.128	0.000	0.355	0.000	2.000

En la tabla N° 4 el análisis de las tarjetas amarillas recibidas por los equipos locales, la media fue de 2.384, con una mediana de 2.000. La desviación estándar (DE) de 1.533 indicó una considerable variabilidad en las tarjetas amarillas recibidas por el equipo local. El número mínimo de tarjetas amarillas registradas fue 0 y el máximo fue 9. En cuanto a los equipos visitantes, la media de tarjetas amarillas fue de 2.520, con una mediana de 2.000. La desviación estándar (DE) fue de 1.543, lo que también reflejó variabilidad en las tarjetas amarillas recibidas por el equipo visitante. El número mínimo de tarjetas amarillas fue 0 y el máximo fue 8.

Para las tarjetas rojas recibidas por los equipos locales, la media fue de 0.121, con una mediana de 0.000. La desviación estándar (DE) de 0.348 indicó una variabilidad en las tarjetas rojas recibidas por el equipo local. El número mínimo de tarjetas rojas registradas fue 0 y el máximo fue 2. En cuanto a los equipos visitantes, la media de

tarjetas rojas fue de 0.128, con una mediana de 0.000. La desviación estándar (DE) fue de 0.355, reflejando variabilidad en las tarjetas rojas recibidas por el equipo visitante. El número mínimo de tarjetas rojas fue 0 y el máximo fue 2.

Tabla 5: Corners

Estadísticas Descriptivas							
Variable	N	Perdidos	Media	Mediana	DE	Mínimo	Máximo
Córners Local (HC)	1900	0	5.112	5.000	2.826	0.000	19.000
Córners Visitante (AC)	1900	0	4.147	4.000	2.484	0.000	15.000

En la tabla N° 5 del análisis de los córners realizados por los equipos locales, la media de córners fue de 5.538, con una mediana de 5.000. La desviación estándar (DE) de 2.877 indicó una considerable variabilidad en los córners realizados por el equipo local. El número mínimo de córners registrados fue 0, mientras que el máximo fue 17. En cuanto a los equipos visitantes, la media de córners fue de 4.367, con una mediana de 4.000. La desviación estándar (DE) fue de 2.467, lo que también reflejó variabilidad en los córners realizados por el equipo visitante. El número mínimo de córners fue 0 y el máximo fue 15.

### **Análisis inferencial**

Se presentan el siguiente análisis inferencial para el Objetivo específico: Determinar qué factores influyen en el resultado de un partido

H1: El número de goles influye significativamente en el resultado de un partido de fútbol.

H0: El número de goles no influye significativamente en el resultado de un partido de fútbol.

Tabla 6: Prueba normalidad goles

Prueba de Normalidad (Shapiro Wilk)			
		W	p
Goles Local		0.895	< .001
(FTHG)			
Empate (D)		0.832	< .001
Goles Visitante		0.880	< .001
(FTAG)			

En la tabla N° 6 para los goles anotados por el equipo local, la estadística W fue 0.895 con un p-valor <0.001, lo que sugirió que la distribución de los goles no fue normal. En los partidos que terminaron en empate, la estadística W fue 0.832 con un p-valor <0.001, indicando que la distribución de goles en estos partidos tampoco fue normal. Similarmente, para los goles anotados por el equipo visitante, la estadística W fue 0.880 con un p-valor <0.001, lo que indicó una distribución no normal.

Tabla 7: ANOVA goles

ANOVA de Un Factor (Welch)				
	F	gl1	gl2	p
ANOVA	62.931	2	1897	< .001

En la tabla N°7 el valor de F es 62.931 con grados de libertad (gl1) de 2 y gl2 de 1231. El p-valor es <0.001, lo que indica que hay una diferencia significativa en el número de goles según el resultado del partido. Esto apoya la hipótesis alternativa H1: El número de goles influye significativamente en el resultado de un partido de fútbol.

H1: El número de tiros a puerta influye significativamente en el resultado de un partido de fútbol.

H0: El número de tiros a puerta no influye significativamente en el resultado de un partido de fútbol.

Tabla 8: Prueba normalidad tiros

Prueba de Normalidad (Shapiro Wilk)		
	W	p
Tiros a Puerta Local (HST)	0.968	< .001
Empate (D)	0.977	< .001
Tiros a Puerta Visitante (AST)	0.971	< .001

En la tabla N°8 para los tiros a puerta realizados por el equipo local, la estadística W fue 0.968 con un p-valor < 0.001, lo que sugirió que la distribución de los tiros a puerta no fue normal. En los partidos que terminaron en empate, la estadística W fue 0.977 con un p-valor < 0.001, indicando que la distribución de tiros a puerta en estos partidos tampoco fue normal. Similarmente, para los tiros a puerta realizados por el equipo visitante, la estadística W fue 0.971 con un p-valor menor a 0.001, lo que indicó una distribución no normal.

Tabla 9: ANOVA tiros

Prueba de Hipótesis (ANOVA)				
	F	gl1	gl2	p
ANOVA	26.953	2	1897	< 0.001

En la tabla N°9 el valor de F es 26.953 con grados de libertad (gl1) de 2 y gl2 de 1897. El p-valor es < 0.001 lo que indica que hay una diferencia significativa en el número de tiros a puerta según el resultado del partido. Esto apoya la hipótesis alternativa H1: El número de tiros a puerta influye significativamente en el resultado de un partido de fútbol.

H1: El número de tarjetas (amarillas y rojas) influye significativamente en el resultado de un partido de fútbol.

H0: El número de tarjetas (amarillas y rojas) no influye significativamente en el resultado de un partido de fútbol.

Tabla 10: Prueba normalidad tarjetas

Prueba de Normalidad (Shapiro Wilk)		
	W	p
Tarjetas	0.965	< .001
Local (HY + HR)		< .001
Empate (D)	0.973	< .001
Tarjetas	0.971	
Visitante (AY + AR)		

En la tabla N°10 para las tarjetas recibidas por el equipo local, la estadística W fue 0.965 con un p-valor < 0.001, lo que sugirió que la distribución de las tarjetas no fue normal. En los partidos que terminaron en empate, la estadística W fue 0.973 con un p-valor < 0.001, indicando que la distribución de tarjetas en estos partidos tampoco fue normal. Similarmente, para las tarjetas recibidas por el equipo visitante, la estadística W fue 0.971 con un p-valor < 0.001, lo que indicó una distribución no normal.

Tabla 11: ANOVA tarjetas

Prueba de Hipótesis (ANOVA)				
	F	gl1	gl2	p
ANOVA	9.314	2	1897	< .001

En la tabla N°11 el valor de F es 9.314 con grados de libertad (gl1) de 2 y gl2 de 1897. El p-valor es < 0.001, lo que sugiere que hay diferencias significativas en el

número de tarjetas según el resultado del partido. Esto apoya la hipótesis alternativa  
H1: El número de tarjetas influye significativamente en el resultado de un partido de fútbol.

H1: El número de córners influye significativamente en el resultado de un partido de fútbol.

H0: El número de córners no influye significativamente en el resultado de un partido de fútbol.

Tabla 12: Prueba normalidad corners

Prueba de Normalidad (Shapiro Wilk)		
	W	p
Córners Local (HC)	0.979	< .001
Empate (D)	0.984	< .001
Córners Visitante (AC)	0.986	< .001

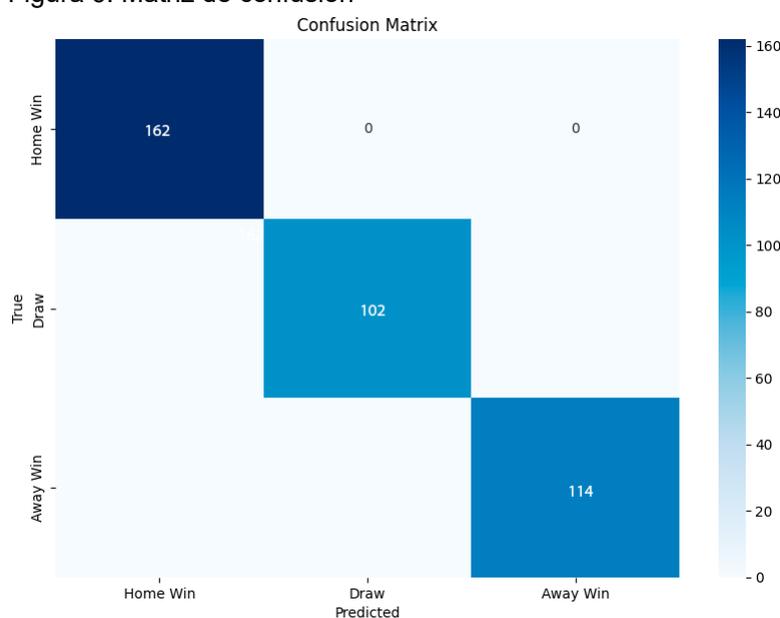
En la tabla N°12 para los córners realizados por el equipo local, la estadística W fue 0.979 con un p-valor < 0.001, lo que sugirió que la distribución de los córners no fue normal. En los partidos que terminaron en empate, la estadística W fue 0.984 con un p-valor < 0.001, indicando que la distribución de córners en estos partidos tampoco fue normal. Similarmente, para los córners realizados por el equipo visitante, la estadística W fue 0.986 con un p-valor < 0.001, lo que indicó una distribución no normal.

Tabla 13:ANOVA corners

Prueba de Hipótesis (ANOVA)				
	F	gl1	gl2	p
ANOVA	0.607	2	1897	0.545

En la tabla N° 13 el valor de F es 0.607 con grados de libertad (gl1) de 2 y gl2 de 1897. El p-valor es 0.545, lo que sugiere que no hay diferencias significativas en el número de córners según el resultado del partido. Esto no apoya la hipótesis alternativa H1: El número de córners influye significativamente en el resultado de un partido de fútbol, y respalda la hipótesis nula H0: El número de córners no influye significativamente en el resultado de un partido de fútbol.

Figura 6: Matriz de confusión



En la figura 6 la matriz de confusión muestra que el modelo de predicción de resultados deportivos tiene un desempeño excelente para predecir victorias locales y visitantes, pero no puede identificar empates. Específicamente, predice correctamente 162 victorias locales y 114 victorias visitantes sin cometer errores en estas categorías. Sin embargo, no logra predecir ningún empate, ya que todos los 102 partidos que fueron empates no fueron clasificados correctamente. Esto sugiere que el modelo necesita ajustes significativos para mejorar su capacidad de reconocer empates. Se puede deducir que el modelo tiene un porcentaje de precisión del 73.02%.

#### IV. DISCUSIÓN

A través de los resultados obtenidos, la presente investigación proporciona un análisis detallado de las comparaciones basadas en las dimensiones propuestas; con esto en mente, se formuló una propuesta a través de dos modelos tecnológicos (Excel y Jamovi). Se llevaron a cabo pruebas de normalidad utilizando la prueba de (Shapiro-Wilk) para examinar la distribución de los resultados obtenidos y luego, se aplicaron análisis estadísticos inferenciales, como la prueba de ANOVA y la prueba A continuación, se procede a discutir los resultados a fin de encontrar la coherencia científica de los mismos.

Para nuestro objetivo general que se denomina “Desarrollar e implementar un modelo predictivo para determinar el ganador de un partido en La Liga EA Sports 2024” el cual se alinea con la creciente tendencia en la investigación deportiva de emplear modelos predictivos basados en aprendizaje automático para pronosticar resultados de partidos. Nuestra investigación busca desarrollar un modelo predictivo para la Liga EA Sports 2024 basado en Redes Neuronales Artificiales (RNA), utilizando datos históricos y diversos indicadores de rendimiento. Tanto esta investigación como el de (Geurkink et al., 2021) se enfocaron en desarrollar modelos predictivos para determinar el resultado de partidos de fútbol, pero difieren en el alcance y la metodología. Asimismo, (Berrar et al., 2018) estos autores proponen dos métodos para crear características predictivas: la extracción de características de actualidad (basada en el rendimiento reciente) y el aprendizaje de características de clasificación (basado en un modelo de calificación de equipos). Utilizan k-NN y XGBoost para construir modelos predictivos. De igual manera (Azeman et al., 2020) los autores implementaron algoritmos como Decision Forest (DF), Artificial Neural Network (ANN) y Support Vector Machine (SVM). Aunque las técnicas específicas difieren, ambos estudios demuestran la aplicabilidad del aprendizaje automático en la predicción de resultados deportivos. Como también (Anfilets et al. 2020) estos autores utilizaron una red neuronal profunda multicapa basada en la regularización de redes elásticas para predecir el ganador de los partidos de la Premier League inglesa. También se puede comparar nuestra investigación con (Stübinger et al., 2020) ya que estos autores amplían su análisis a las cinco ligas europeas más importantes y sus respectivas segundas divisiones tomando temporadas entre 2006 y 2018. Esta diferencia en el alcance podría explicar las variaciones en los resultados, ya que el modelo

de(Stübinger et al., 2020) tiene una muestra más grande y diversa. También (Badiella et al., 2023) utilizaron un Modelo Lineal Generalizado Mixto (GLMM). Se compara también con (Capobianco et al., 2019) los autores también utilizan el aprendizaje automático para predecir los resultados de los partidos de fútbol, pero se centran en la liga italiana Serie A 2017-2018 ellos utilizan y evalúan varios algoritmos de clasificación para predecir los resultados de los partidos de fútbol, incluyendo J48, SMO, RepTree, RandomForest y MLP (Perceptrón multicapa). El algoritmo RandomForest.

Para nuestro objetivo específico “Determinar qué factores influyen en el resultado de un partido” (Geurkink et al., 2021) utilizó una amplia gama de 100 variables, incluyendo indicadores técnicos, tácticos, físicos y contextuales, para predecir victorias y derrotas en la primera división Belga, mientras que nuestra investigación se centró en un conjunto más reducido de variables, principalmente estadísticas de goles, tiros a puerta, tarjetas y córners, para predecir el ganador de partidos en La Liga EA Sports 2024. A pesar de las diferencias, nuestras investigaciones destacan la importancia de considerar una variedad de factores al predecir el resultado de un partido de fútbol.(Geurkink et al., 2021) encontró que los tiros a puerta desde el área penal ofensiva eran el predictor más fuerte, pero también señaló la importancia de indicadores físicos y contextuales. Nuestra investigación, aunque se centró en un conjunto más limitado de variables, también encontró que los goles, tiros a puerta, tarjetas influyen significativamente en el resultado de un partido. Asimismo, (Azeman et al., 2020) , encontró que los tiros a puerta son un factor influyente en el resultado del partido. Por otro lado, en cuanto a las tarjetas en su investigación no encontró evidencia suficiente para respaldar esta afirmación en su contexto.

Asimismo, la investigación de (Tsokos et al., 2018) el cual se centra en la predicción de resultados de partidos de fútbol, utilizando datos históricos y modelos estadísticos. Nuestra investigación coincide en la importancia de factores como los goles marcados y los tiros a puerta como predictores significativos del resultado de un partido, pero los corners no son tan influyentes en el resultado de un partido. Sin embargo, difieren en la complejidad de los modelos utilizados y en los métodos de validación. Mientras que nuestro estudio se basa en un modelo de clasificación multiclase el de estos autores se basa en la comparación de diversas extensiones de modelos como Bradley-Terry y un modelo jerárquico Poisson log-lineal para predecir resultados de

partidos de fútbol. Su investigación explora una gama más amplia de modelos, incluyendo modelos jerárquicos y enfoques no paramétricos, y utiliza un marco de validación temporal más riguroso para evaluar el rendimiento de los modelos. Por el contrario,(Berrar et al., 2018) los autores destacan la importancia de incorporar el conocimiento del dominio del fútbol en el proceso de modelado, incluyendo factores como el rendimiento ofensivo y defensivo, el rendimiento reciente, la fuerza de la oposición y la ventaja de jugar en casa. Mientras que(Anfilets et al., 2020) estos autores utilizaron un conjunto de datos abierto que incluía tiros, tiros a puerta, tarjetas amarillas y rojas, entre otros. En cuanto a las tarjetas amarillas,(Badiella et al., 2023) estos autores observaron una interacción entre las tarjetas amarillas y la diferencia de goles, aunque estas tienen un efecto menos pronunciado que las rojas.

Para nuestro objetivo específico “Determinar el porcentaje de precisión de los resultados de partidos mediante el modelo predictivo” A comparación con nuestro objetivo específico número dos, ambos estudios (tanto el nuestro como el de Geurkink) empleamos técnicas de aprendizaje automático para construir nuestros modelos predictivos.(Geurkink et al., 2021) utilizó Extreme Gradient Boosting (GB) y logró una precisión del 89.6%, mientras que nuestro algoritmo de Redes Neuronales Artificiales (RNA), reportó una precisión del 73.02%. La diferencia en precisión podríamos atribuirlo a la selección de variables, la calidad de los datos o el algoritmo de aprendizaje automático empleado ya que estos son diferentes. Asimismo, en comparación con la investigación de (Berrar et al., 2019) El mejor modelo (k-NN entrenado con características de clasificación) logró un RPSavg de 0.2054 en el conjunto de predicción. Igualmente (Azeman et al., 2020) en su investigación llegó a tener resultados de los algoritmos que utilizó estos muestran que el algoritmo Decision Forest logra la mayor precisión (89%), seguido por ANN (72%) y SVM (70%). Esto indica que el modelo basado en Decision Forest es el más efectivo para predecir correctamente los resultados de los partidos en el conjunto de datos utilizado. Asimismo, los resultados tienen cierta diferencia con Anfilets et al. (2020) los autores llegaron a tener una precisión del 61,14 % en su modelo, que superó a otros enfoques mencionados en su estudio. Nuestro modelo comparado con el de estos autores tienen la necesidad de mejorar más que todo a la hora de predecir empates. En cuanto con (Badiella et al., 2023) estos no reportaron explícitamente un porcentaje de eficacia general, pero sí cuantificaron el impacto esperado de las expulsiones en

términos de goles, lo que podría ser utilizado para evaluar la eficacia de su modelo en situaciones específicas. En términos de precisión (Capobianco et al., 2019) obtienen como resultado de su investigación los siguientes resultados para la predicción de partidos J48(0.735 para predicción de partidos y 0.749 para predicción de goles),SMO(0.843 para predicción de partidos y 0.860 para predicción de goles),RepTree(0.796 para predicción de partidos y 0.842 para predicción de goles),RandomTree(0.744 para predicción de partidos y 0.862 para predicción de goles),RandomForest(0.843 para predicción de partidos y 0.862 para predicción de goles),MultilayerPerceptron(0.754 para predicción de partidos y 0.842 para predicción de goles) teniendo como resultado que el algoritmo RandomForest obtuvo la mayor precisión tanto en la predicción de partidos como en la predicción de goles.

## **V. CONCLUSIONES**

El presente estudio tuvo como objetivo general desarrollar e implementar un modelo predictivo para determinar el resultado de los partidos en la Liga EA Sports 2024. Este objetivo se logró con éxito, ya que se desarrolló un modelo predictivo utilizando técnicas de machine learning y análisis estadístico, alcanzando una precisión del 73.02%.

En cuanto a los objetivos específicos, el primero fue determinar los factores que influyen en el resultado de un partido. Se identificaron varios factores significativos, tales como goles, tiros a puerta, tarjetas amarillas y rojas. Los análisis estadísticos realizados indicaron que estos factores tienen un impacto considerable en los resultados de los partidos.

El segundo objetivo específico era determinar el porcentaje de precisión del modelo predictivo. El modelo implementado alcanzó una precisión del 73.02%, demostrando ser relativamente eficaz en la predicción de los resultados de los partidos de fútbol en la Liga EA Sports 2024.

Los análisis estadísticos apoyaron la hipótesis de que factores como los goles y tiros a puerta influyen significativamente en los resultados de los partidos, mientras que los córners no mostraron una influencia significativa.

## **VI. RECOMENDACIONES**

Es recomendable mejorar la capacidad del modelo para predecir empates, ya que este es un aspecto crítico que no fue abordado con precisión en la investigación actual. Se recomienda realizar un análisis más profundo de las características de los partidos que terminan en empate y ajustar los algoritmos de machine learning para mejorar su desempeño en este aspecto.

Para aumentar la precisión del modelo predictivo, se sugiere la inclusión de más variables que podrían influir en el resultado de un partido. Estas podrían incluir factores como la forma física y psicológica de los jugadores, las condiciones climáticas, el estado del campo de juego, y la presión de los aficionados.

Recomendamos explorar y aplicar técnicas más avanzadas de machine learning, como redes neuronales profundas, aprendizaje por refuerzo y técnicas de ensemble. Estas técnicas tienen el potencial de capturar relaciones más complejas entre las variables y mejorar la precisión del modelo predictivo.

También se realizar una validación continua del modelo desarrollado y compararlo con otros modelos predictivos existentes. Esto permitirá identificar fortalezas y debilidades relativas y ajustar el modelo para mejorar su desempeño. Se sugiere la participación en competiciones de predicción deportivas para evaluar el modelo en un entorno competitivo.

## REFERENCIAS

- Ajgaonkar, Y., Bhoyar, K., Patil, A., & Shah, J. (2021). Prediction of Winning Team using Machine Learning. *International Journal of Engineering Research*, 9(3), Article 3.
- Almulla, J., & Alam, T. (2020). Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League. *IEEE Access*, 8, 213695-213705. <https://doi.org/10.1109/ACCESS.2020.3038601>
- Anfilets, S., Bezobrazov, S., Golovko, V., Sachenko, A., Komar, M., Dolny, R., Kasyanik, V., Bykovyy, P., Mikhno, E., & Osolinskyi, O. (2020). DEEP MULTILAYER NEURAL NETWORK FOR PREDICTING THE WINNER OF FOOTBALL MATCHES. *International Journal of Computing*, 70-77. <https://doi.org/10.47839/ijc.19.1.1695>
- Arntzen, H., & Hvattum, L. M. (2021). Predicting match outcomes in association football using team ratings and player ratings. *Statistical Modelling*, 21(5), Article 5. <https://doi.org/10.1177/1471082X20929881>
- Azeman, A., Mustapha, A., Mostafa, S. A., AbuSalim, S. W. G., Jubair, M. A., & Hassan, M. H. (2020). Football match outcome prediction by applying three machine learning algorithms. *International Journal of Emerging Trends in Engineering Research*, 8(1 Special Issue 1), Article 1 Special Issue 1. Scopus. <https://doi.org/10.30534/ijeter/2020/1181.12020>
- Badiella, L., Puig, P., Lago-Peñas, C., & Casals, M. (2023). Influence of Red and Yellow cards on team performance in elite soccer. *Annals of Operations Research*, 325(1), 149-165. <https://doi.org/10.1007/s10479-022-04733-0>
- Bai, L., Gedik, R., & Egilmez, G. (2022). What does it take to win or lose a soccer game? A machine learning approach to understand the impact of game and team statistics. *Journal of the Operational Research Society*, 0(0), Article 0. <https://doi.org/10.1080/01605682.2022.2110001>

- Berrar, D., Lopes, P., & Dubitzky, W. (2019). Incorporating domain knowledge in machine learning for soccer outcome prediction. *Machine Learning*, 108(1), Article 1. <https://doi.org/10.1007/s10994-018-5747-8>
- Caballero, P., García-Rubio, J., & José Ibáñez, S. (2017). Influence of situational variables on the U'18 soccer performance analysis: Análisis de la influencia de las variables situacionales en el rendimiento en fútbol U'18. *Retos: Nuevas Perspectivas de Educación Física, Deporte y Recreación*, 32, 224-227. <https://doi.org/10.47197/retos.v0i32.56071>
- Capobianco, G., Di Giacomo, U., Mercaldo, F., Nardone, V., & Santone, A. (2019). Can Machine Learning Predict Soccer Match Results?: *Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, 458-465. <https://doi.org/10.5220/0007307504580465>
- Cho, Y., Yoon, J., & Lee, S. (2018). Using social network analysis and gradient boosting to develop a soccer win–lose prediction model. *Engineering Applications of Artificial Intelligence*, 72, 228-240. <https://doi.org/10.1016/j.engappai.2018.04.010>
- Connor, M., Mernagh, D., & Beato, M. (2022). Quantifying and modelling the game speed outputs of English Championship soccer players. *Research in Sports Medicine*, 30(2), Article 2. <https://doi.org/10.1080/15438627.2021.1888108>
- Dijkhuis, T. B., Kempe, M., & Lemmink, K. A. P. M. (2021). Early Prediction of Physical Performance in Elite Soccer Matches—A Machine Learning Approach to Support Substitutions. *Entropy*, 23(8), Article 8. <https://doi.org/10.3390/e23080952>
- Diniz, M. A., Izbicki, R., Lopes, D., & Salazar, L. E. (2019). Comparing probabilistic predictive models applied to football. *Journal of the Operational Research Society*, 70(5), Article 5. <https://doi.org/10.1080/01605682.2018.1457485>
- Geurkink, Y., Boone, J., Verstockt, S., & Bourgois, J. G. (2021). Machine Learning-Based Identification of the Strongest Predictive Variables of Winning and

- Losing in Belgian Professional Soccer. *Applied Sciences*, 11(5), Article 5. <https://doi.org/10.3390/app11052378>
- Giménez, J. V., Jiménez-Linares, L., Leicht, A. S., & Gómez, M. A. (2020). Predictive modelling of the physical demands during training and competition in professional soccer players. *Journal of Science and Medicine in Sport*, 23(6), Article 6. <https://doi.org/10.1016/j.jsams.2019.12.008>
- Hassan, A., Akl, A.-R., Hassan, I., & Sunderland, C. (2020). Predicting wins, losses and attributes' sensitivities in the soccer world cup 2018 using neural network analysis. *Sensors*, 20(11), Article 11.
- Hernández, R., Fernández, C., & Baptista, P. (2016). Metodología de la investigación. 6ta Edición Sampieri. Soriano, RR (1991). *Guía para realizar investigaciones sociales*. Plaza y Valdés.
- Hubaček, O., Šourek, G., & Zálezny, F. (2019). Score-based soccer match outcome modeling – an experimental review. *Conference Proceedings*, 9.
- Keshtkar Langaroudi, M., & Yamaghani, M. (2019). Sports Result Prediction Based on Machine Learning and Computational Intelligence Approaches: A Survey. *Journal of Advances in Computer Engineering and Technology*, 5(1), Article 1.
- Konefał, M., Chmura, P., Tessitore, A., Melcer, T., Kowalczyk, E., Chmura, J., & Andrzejewski, M. (2020). The Impact of Match Location and Players' Physical and Technical Activities on Winning in the German Bundesliga. *Frontiers in Psychology*, 11, 1748. <https://doi.org/10.3389/fpsyg.2020.01748>
- Lamsal, R., & Choudhary, A. (2020). *Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning* (arXiv:1809.09813; Número arXiv:1809.09813). arXiv. <https://doi.org/10.48550/arXiv.1809.09813>
- Lang, S., Wild, R., Isenko, A., & Link, D. (2022). Predicting the in-game status in soccer with machine learning using spatiotemporal player tracking data. *Scientific Reports*, 12(1), Article 1. <https://doi.org/10.1038/s41598-022-19948-1>

- Lee, J., Kim, J., Kim, H., & Lee, J.-S. (2022). A Bayesian Approach to Predict Football Matches with Changed Home Advantage in Spectator-Free Matches after the COVID-19 Break. *Entropy*, 24(3), Article 3. <https://doi.org/10.3390/e24030366>
- López-Roldán, P., & Fachelli, S. (2017). El diseño de la muestra. *Metodología de la investigación social cuantitativa*. <https://ddd.uab.cat/record/185163>
- Malamatinos, M.-C., Vrochidou, E., & Papakostas, G. A. (2022). On Predicting Soccer Outcomes in the Greek League Using Machine Learning. *Computers*, 11(9), Article 9. <https://doi.org/10.3390/computers11090133>
- Manack, H., & Van Zyl, T. L. (2020). Deep Similarity Learning for Soccer Team Ranking. *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 1-7. <https://doi.org/10.23919/FUSION45008.2020.9190564>
- Manderson, A. A., Murray, K., & Turlach, B. A. (2018). Dynamic Bayesian forecasting of AFL match results using the Skellam distribution. *Australian & New Zealand Journal of Statistics*, 60(2), Article 2. <https://doi.org/10.1111/anzs.12225>
- Medina, P., Carrasco, S., Rogan, J., Montes, F., Meisel, J. D., Lemoine, P., Lago Peñas, C., & Valdivia, J. A. (2021). Is a social network approach relevant to football results? *Chaos, Solitons & Fractals*, 142, 110369. <https://doi.org/10.1016/j.chaos.2020.110369>
- Pinto, M. (1989). Introducción al análisis documental y sus niveles: El análisis de contenido. *Boletín de la ANABAD, ISSN 0210-4164, Tomo 39, Nº 2, 1989, pags. 323-342.*
- Razali, N., Mustapha, A., Utama, S., & Din, R. (2018). A Review on Football Match Outcome Prediction using Bayesian Networks. *Journal of Physics: Conference Series*, 1020, 012004. <https://doi.org/10.1088/1742-6596/1020/1/012004>
- Stübinger, J., Mangold, B., & Knoll, J. (2020). Machine Learning in Football Betting: Prediction of Match Results Based on Player Characteristics. *Applied Sciences*, 10(1), Article 1. <https://doi.org/10.3390/app10010046>

- Sybing, R. (s. f.). *Estudios longitudinales: Métodos, ventajas y retos*. ATLAS.ti. Recuperado 12 de junio de 2024, de <https://atlasti.com/es/research-hub/estudio-longitudinal>
- Tsokos, A., Narayanan, S., Kosmidis, I., Baio, G., Cucuringu, M., Whitaker, G., & Király, F. (2019). Modeling outcomes of soccer matches. *Machine Learning*, 108(1), Article 1. <https://doi.org/10.1007/s10994-018-5741-1>
- Wei, Y. (2021). Predict soccer match outcome based on player performance. *Frontiers in Sport Research*, 3(3), Article 3. <https://doi.org/10.25236/FSR.2021.030314>
- Zhou, C., Calvo, A. L., Robertson, S., & Gómez, M.-Á. (2021). Long-term influence of technical, physical performance indicators and situational variables on match outcome in male professional Chinese soccer. *Journal of Sports Sciences*, 39(6), Article 6. <https://doi.org/10.1080/02640414.2020.1836793>

## ANEXOS

### Anexo 1. Matriz de consistencia

Título: Modelo predictivo para determinar resultados de partidos de fútbol en La Liga EA Sports 2024				
PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS PRINCIPAL	VARIABLES	DISEÑO METODOLOGÍA
¿Puede un modelo predictivo determinar los resultados de los partidos de fútbol?	Desarrollar e implementar un modelo predictivo para determinar el resultado de los partidos en la Liga EA Sports 2024.	Un modelo predictivo basado en aprendizaje automático puede pronosticar los resultados de futuros partidos de fútbol con precisión.	Modelo predictivo	<b>Tipo de investigación:</b> aplicada  <b>Nivel de investigación:</b> Longitudinal descriptiva-correlacional  <b>Diseño de investigación:</b> no experimental  <b>Población:</b> Datos históricos de partidos de la Liga EA Sports de las temporadas 2019-2024.  <b>Técnica:</b> técnica del análisis documental
PROBLEMAS ESPECÍFICOS	OBJETIVOS ESPECIFICOS	HIPOTESIS ESPECIFICAS		
¿Determinar qué factores influyen en el resultado de un partido?	Determinar qué factores influyen en el resultado de un partido	Existen varios factores que influyen en el resultado de un partido de fútbol, como las estadísticas de juego (goles, tiros, tarjetas, córners).	Resultados de fútbol	
¿Cuál es el porcentaje de precisión de un modelo predictivo para determinar un resultado en un partido de fútbol?	Determinar el porcentaje de precisión de los resultados de partidos mediante el modelo predictivo.	Un modelo predictivo basado en datos históricos de equipos y estadísticas de juego tendrá una precisión superior al 60% en la predicción de los resultados de partidos de fútbol.		

Anexo 2. Matriz de operacionalización de variables

VARIABLES DE ESTUDIO	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ESCALA DE MEDICIÓN	COMO MEDIRLO
Modelo predictivo	Un modelo predictivo es una de las técnicas inteligentes para demostrar resultados optimistas en el campo de la clasificación y asimismo la precisión de la predicción (Yash et al 2021)	Para la medición del modelo predictivo se evaluará el porcentaje de precisión y se analizaran los datos por importancia	Machine Learning	Porcentaje de precisión	Razón	Análisis documental
				Datos obtenidos por importancia		
Resultados de futbol	El fútbol es probablemente el deporte de equipo más popular del mundo, en gran parte de la fascinación por el fútbol se debe que casi la mayoría de los partidos jugados terminan en empates o se ganan por sólo dos o menos goles, además de una complejidad de mezcla de factores y sucesos de fortuitos juegan un papel importante en la determinación del resultado final de un partido de fútbol. (Berrar et al., 2019)	Para determinar el resultado de un partido se evalúa mediante los goles marcados, los partidos, jugadores, corners, como también las tarjetas amarillas y rojas.	Resultado	Goles	Razón	Análisis documental
				Tarjetas		
				Tiros		
				Corners		

Fuente: Elaboración propia de los autores.

Anexo 3. Indicadores de variables

OBJETIVO ESPECÍFICO	INDICADOR	DESCRIPCIÓN	TÉCNICA / INSTRUMENTO	TIEMPO EMPLEADO	MODO DE CÁLCULO
Determinar el porcentaje de precisión de los resultados de partidos mediante el modelo predictivo.	<p>Porcentaje de precisión</p> <p>Datos obtenidos</p>	<p>Verificar la precisión del modelo predictivo</p> <p>Datos importantes recolectados</p>	<p>Análisis documental/Ficha de registro</p>		$\text{Precisión (Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN}$ <p>Precisión del modelo = (Total de predicciones correctas / Total de casos) * 100</p>
Determinar los factores que influyen un resultado en los partidos futbolísticos	<p>Goles marcados</p> <p>Tiros</p> <p>Tarjetas</p> <p>Corners</p>	<p>Cantidad de goles marcados de los equipos</p> <p>Cantidad de tiros marcados por equipos</p> <p>Cantidad de tarjetas tanto amarillas como rojas</p> <p>Cantidad de córners por los equipos</p>	<p>Análisis documental/ficha de registro</p>		<p>GEL=GOLES*EL GEV=GOLESEV</p> <p>TPL= TIROS*EL TPV= TIROS*EV</p> <p>TAL=TARJETASY*EL TRL=TARJETASR*EL TAV= TARJETASY*EV TRV= TARJETASR*EV CL= CORNERS*EL CV= CORNERS*EV</p>

Fuente: Elaboración propia de los autores.



Ficha de registro de Machine learning					
<b>Investigadores</b>		Siancas Huamán Daniel Alberto Valencia Yovera Franklin Paul		<b>Método de análisis:</b>	Descriptiva e inferencial
<b>Institución</b>		Universidad Privada César Vallejo			
<b>Dimensión de estudio</b>		Machine Learning		<b>Objetivo: "Determinar el porcentaje de precisión del modelo predictivo"</b>	
<b>Fecha de Inicio</b>		20/05/2024	<b>Fecha final</b>		25/05/2024
<b>Variable</b>		<b>Indicador</b>	<b>Medida</b>		<b>Fórmula</b>
Modelo predictivo		Porcentaje de precisión	Análisis documental		Eficacia del modelo = (Total de predicciones correctas / Total de casos) * 100
<b>#</b>	<b>Fecha de revisión</b>	<b>Predicciones correctas</b>			<b>Resultado</b>
		<b>Vitorias Locales</b>	<b>Empates</b>	<b>Victorias Visitantes</b>	
1					
		<b>Total, de casos (PARTIDOS)</b>			
		<b>Vitorias Locales</b>	<b>Empates</b>	<b>Victorias Visitantes</b>	
1					
TOTAL					

## Anexo 5. Validación de los instrumentos



# UNIVERSIDAD CÉSAR VALLEJO

### CONSTANCIA DE VALIDACIÓN

Yo Iván Michell Castillo Jiménez, con DNI 02883813, de profesión Ingeniero Informático desempeñándome actualmente como Docente a tiempo completo en la Universidad César Vallejo

Por medio de la presente hago constar que he revisado con fines de Validación de la ficha de Registro Ficha de registro para la dimensión Resultado.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Ficha de registro para la dimensión Resultado	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad					X
2. Objetividad			X		
3. Actualidad				X	
4. Organización			X		
5. Suficiencia			X		
6. Intencionalidad					X
7. Consistencia				X	
8. Coherencia				X	
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 20 días del mes de Febrero del Dos mil veinticuatro.

---

Ing. : Iván Michell Castillo Jiménez  
DNI : 02883813  
Especialidad : Ing. Informático  
E-mail : imcj@hotmail.com



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo Iván Michell Castillo Jiménez, con DNI 02883813, de profesión Ingeniero Informático desempeñándome actualmente como Docente a tiempo completo en la Universidad César Vallejo.

Por medio de la presente hago constar que he revisado con fines de Validación de la ficha de Registro Ficha de registro para la dimensión machine learning.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Ficha de registro para la dimensión machine learning	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad					X
2. Objetividad				X	
3. Actualidad			X		
4. Organización			X		
5. Suficiencia			X		
6. Intencionalidad					X
7. Consistencia				X	
8. Coherencia				X	
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 20 días del mes de Febrero del Dos mil veinticuatro.

---

Ing. : Iván Michell Castillo Jiménez  
DNI : 02883813  
Especialidad : Ing. Informático  
E-mail : imcj@hotmail.com



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo Andrew Joaquin Ipanaque Torre , con DNI 48043017 , de profesión Ingeniero de sistemas desempeñándome actualmente como Desarrollador de software en Whaticket LLC

Por medio de la presente hago constar que he revisado con fines de Validación de la ficha de Registro **Ficha de registro para la dimensión Resultado.**

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

<b>Ficha de registro para la dimensión Resultado</b>	<b>DEFICIENTE</b>	<b>ACEPTABLE</b>	<b>BUENO</b>	<b>MUY BUENO</b>	<b>EXCELENTE</b>
1. Claridad				X	
2. Objetividad				X	
3. Actualidad					X
4. Organización				X	
5. Suficiencia			X		
6. Intencionalidad			X		
7. Consistencia				X	
8. Coherencia				X	
9. Metodología				X	

En señal de conformidad firmo la presente en la ciudad de Piura a los 20 días del mes de Febrero del Dos mil veinticuatro.

Ing. : Andrew Ipanaque Torre  
DNI : 48043017  
Especialidad : Desarrollo de software  
E-mail : itan.root@gmail.com



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo Andrew Joaquin Ipanaque Torre , con DNI 48043017 , de profesión Ingeniero de sistemas desempeñándome actualmente como Desarrollador de software en Whaticket LLC

Por medio de la presente hago constar que he revisado con fines de Validación de la ficha de Registro **Ficha de registro para la dimensión machine learning**.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Ficha de registro para la dimensión machine learning	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1.Claridad				x	
2.Objetividad					x
3.Actualidad					x
4.Organización				x	
5.Suficiencia			x		
6.Intencionalidad				x	
7.Consistencia			x		
8.Coherencia			x		
9.Metodología		x			

En señal de conformidad firmo la presente en la ciudad de Piura a los 20 días del mes de Febrero del Dos mil veinticuatro.

Ing. : Andrew Joaquin Ipanaque Torre  
DNI : 48043017  
Especialidad : Desarrollo de software  
E-mail : itan.root@gmail.com



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, Josefina Gabriela Alvarez Zapata, con DNI 47933710  
de profesión Ing. Computación y Sist. desempeñándome actualmente  
como docente en Universidad  
César Vallejo

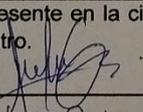
Por medio de la presente hago constar que he revisado con fines de Validación de la  
ficha de Registro Ficha de registro para la dimensión machine learning.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes  
apreciaciones.

Ficha de registro para la dimensión machine learning	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad					X
2. Objetividad				X	
3. Actualidad				X	
4. Organización					X
5. Suficiencia			X		
6. Intencionalidad				X	
7. Consistencia				X	
8. Coherencia				X	
9. Metodología					X

En señal de conformidad firmo la presente en la ciudad de Piura a los 20 días del  
mes de Febrero del Dos mil veinticuatro.

Ing.  
DNI  
Especialidad  
E-mail

  
: Josefina Gabriela Alvarez Zapata  
: 47933710  
: Ing. Computación y Sistemas  
: jjalvarez@ucvvirtual.edu.pe



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, Josefina Gabriela Alvarez Zapata, con DNI 47773710  
de profesión Ing. Computación Sist. desempeñándome actualmente  
como docente en la Univ. César Vallejo

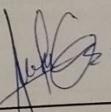
Por medio de la presente hago constar que he revisado con fines de Validación de la ficha de Registro **Ficha de registro para la dimensión Resultado**.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Ficha de registro para la dimensión Resultado	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad				X	
2. Objetividad					X
3. Actualidad					X
4. Organización				X	
5. Suficiencia			X		
6. Intencionalidad				X	
7. Consistencia				X	
8. Coherencia				X	
9. Metodología					X

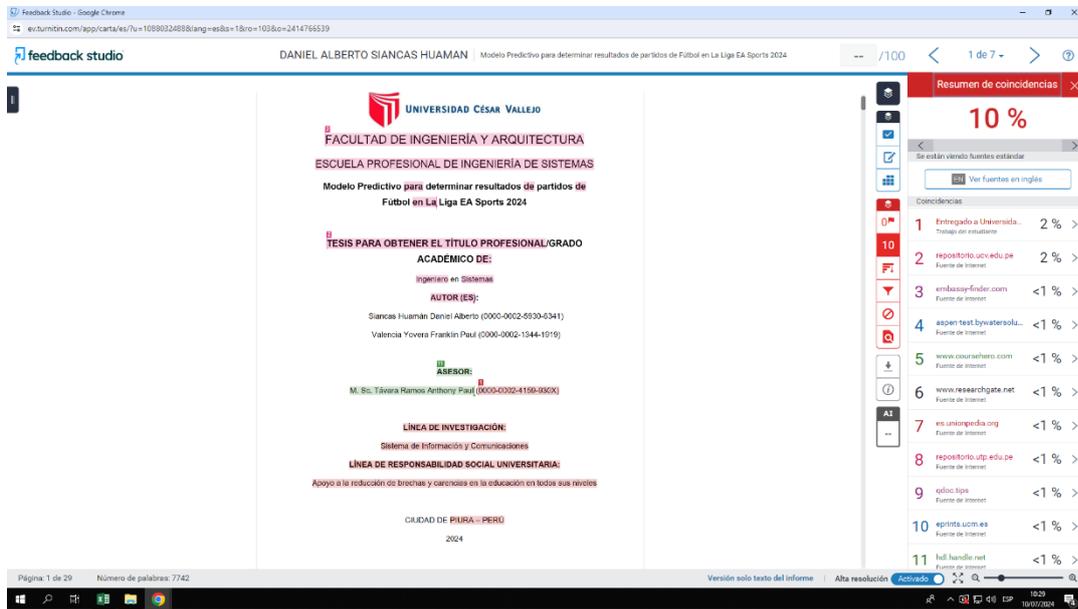
En señal de conformidad firmo la presente en la ciudad de Piura a los 20 días del mes de Febrero del Dos mil veinticuatro.

Ing.  
DNI  
Especialidad  
E-mail

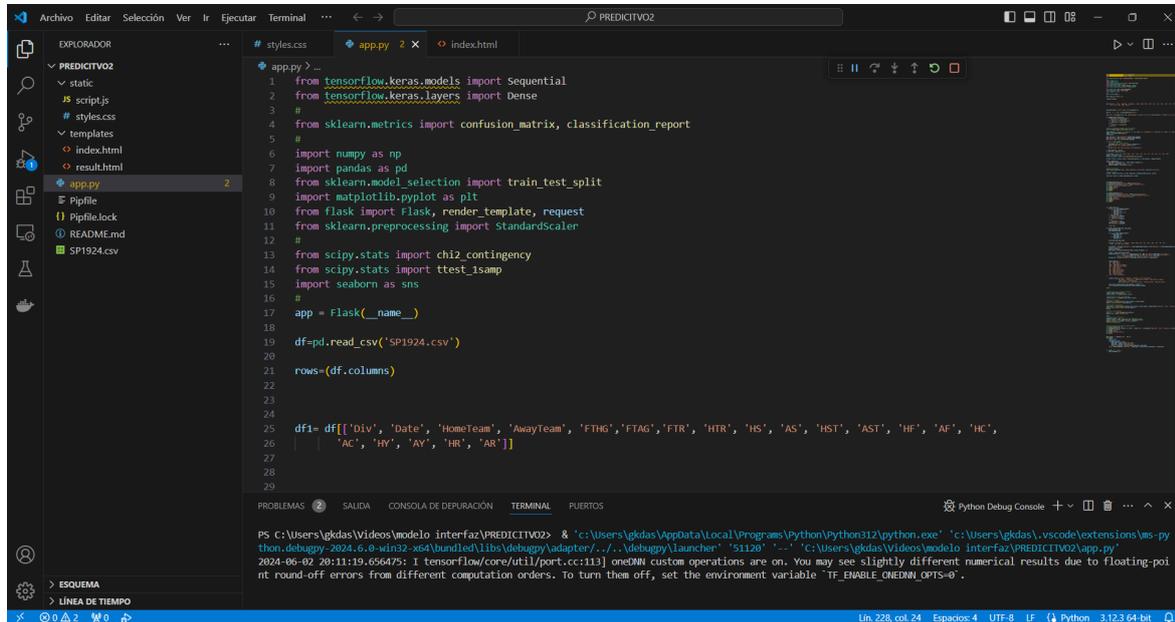
  
: Josefina Gabriela Alvarez Zapata  
: 47773710  
: Ing. Computación y Sistemas  
: jgalaraz@ucvrrval.edu.pe

## Anexo 5. Fotos y documentos

### Reporte turnitin de biblioteca



### Fotos del código del modelo





# Resultado de la predicción de un partido

Predicción de un partido de fútbol

Equipo local:  
Alaves

Equipo visitante:  
Alaves

Predecir

**Resultado de la Predicción:**  
There is a 72.3% chance Real Madrid will win  
There is a 9.28% chance Barcelona will win  
There is a 27.42% chance it will be a draw

**Estadísticas de partidos anteriores**

Date	Home Team	Away Team	Full Time Result	Home Shots	Away Shots	Home Shots on Target	Away Shots on Target	Home Corners	Away Corners	Home Yellow Cards	Away Yellow Cards	Home Red Cards	Away Red Cards
01/03/2020	Real Madrid	Barcelona	H	13	9	5	4	8	2	2	2	0	0
10/04/2021	Real Madrid	Barcelona	H	14	18	3	4	5	6	3	4	1	0
20/03/2022	Real Madrid	Barcelona	A	14	18	4	10	4	3	4	4	0	0
14/10/2022	Real Madrid	Barcelona	H	8	18	4	5	2	3	2	2	0	0
21/04/2024	Real Madrid	Barcelona	H	14	14	8	6	2	8	3	2	0	0