

Predicción de Resultados en el Fútbol a través de un producto Software Implementando Técnicas
de Machine Learning y Análisis de Datos

Julian Esteban Chavez Rivera ✉ jechavez.4559@unicesmag.edu.co

Universidad CESMAG
Facultad de Ingeniería
Ingeniería de sistemas
Pasto – Nariño
2024

Predicción de Resultados en el Fútbol a través de un producto Software Implementando Técnicas
de Machine Learning y Análisis de Datos

Julian Esteban Chavez Rivera ✉ jechavez.4559@unicesmag.edu.co

Trabajo de grado para optar al título de ingeniero de sistemas

Asesor:
Esp. Luis Carlos Viteri Rosero

Universidad CESMAG
Facultad de Ingeniería
Ingeniería de sistemas
Pasto – Nariño
2024

NOTA DE ACEPTACIÓN

NOMBRE JURADO 1

NOMBRE JURADO 2

San Juan de Pasto, 2024

NOTA DE EXCLUSIÓN

El autor de esta obra es el único responsable de las ideas expresadas en ella, y esta no refleja o no compromete la ideología de la Universidad CESMAG.

DEDICATORIA

A mi madre y a mis hermanas, por ser el pilar fundamental en mi vida, por su amor incondicional y su fortaleza en cada paso que he dado. Gracias por acompañarme en los momentos más difíciles, por ser mi refugio y mi inspiración. Su apoyo, incluso en las circunstancias más adversas, me dieron las fuerzas para continuar y sacar adelante esta investigación.

A mi madre, quien con su ejemplo de sacrificio y perseverancia me enseñó que no hay meta inalcanzable. Y a mis hermanas, por ser siempre un ejemplo de valentía, generosidad y apoyo incondicional, demostrando que el amor de familia es un motor irremplazable.

Este logro es tan mío como suyo.

AGRADECIMIENTOS

A la Universidad CESMAG, por brindarme el espacio académico y las herramientas necesarias para desarrollarme profesionalmente y hacer realidad esta investigación. Su compromiso con la formación integral de los estudiantes me permitió afrontar este desafío con confianza y preparación.

A los docentes del área de Ingeniería, quienes con su dedicación y experiencia me guiaron durante todo el proceso. Gracias por compartir su conocimiento, por su disposición para resolver dudas y por ser un ejemplo constante de profesionalismo y vocación docente.

De manera especial, extiendo mi más profundo agradecimiento a mi asesor, el Especialista Luis Carlos Viteri Rosero, cuya paciencia, orientación y respaldo fueron esenciales para superar los retos que se presentaron en este proyecto. Su apoyo no solo fue técnico, sino también humano, brindándome las herramientas para aprender, crecer y alcanzar los objetivos planteados.

Finalmente, agradezco a todas las personas que de alguna forma contribuyeron a este logro, directa o indirectamente. A mis amigos Camilo, Miguel, Andrés, Oscar, Santiago, ZEF, Panto y Oliva por su apoyo incondicional y palabras de aliento; a los tutoriales de YouTube, que fueron una guía invaluable en los momentos de incertidumbre; y a mi familia, por su fortaleza y respaldo constante. A todos los que creyeron en mí y me inspiraron a seguir adelante, les debo este logro, que demuestra que con el esfuerzo se puede superar cualquier desafío.

CONTENIDO

INTRODUCCIÓN	1
I. PROBLEMA DE INVESTIGACIÓN	2
a) Objeto o Tema de Investigación.....	2
b) Línea de Investigación.....	2
c) Sub línea de Investigación.....	2
e) Formulación del problema.....	3
f) Objetivos.....	3
1. General.....	3
2. Específicos.....	4
g) Justificación	4
h) Delimitación	5
II. MARCO TEÓRICO.....	6
a) Antecedentes.....	6
1. Internacionales.....	6
2. Nacionales	11
3. Regional.....	13
b) Supuestos teóricos de investigación	14
1. Disponibilidad de datos:	14
2. Herramientas de minería de datos y machine learning:	16
3. Implementación del modelo en un producto software:.....	17
4. Método ágil: SCRUM.....	18
5. Lenguajes de programación:.....	20
6. Scraping de datos:.....	20
c) Variables de estudio.....	21

1.	Variables independientes	21
2.	Variables dependientes	21
d)	Definición nominal de las variables	22
e)	Definición Operativa De Variables	22
f)	Formulación de hipótesis	23
1.	Hipótesis de investigación	24
2.	Hipótesis nula	24
3.	Hipótesis alterna	24
III.	METODOLOGÍA	25
a)	Paradigma	25
b)	Enfoque.....	25
c)	Método.....	25
d)	Tipo de investigación.....	25
e)	Diseño de investigación.....	26
f)	Población	26
g)	Muestra	26
h)	Técnicas de recolección de información.....	27
i)	Validez de las técnicas de recolección.....	27
j)	Confiabilidad de las técnicas de recolección	28
k)	Instrumentos de recolección de información	29
IV.	RESULTADOS DE LA INVESTIGACIÓN	30
a)	Análisis de conjunto de datos:	30
1.	Consulta con el especialista en análisis de datos:	30
2.	Evaluación del dataset proporcionado por Footy Stats:	30
3.	Exploración de metodologías de web scraping:	30

4.	Ampliación y mejora del dataset:	31
b)	Selección e Implementación del Modelo Predictivo Óptimo para Pronósticos Futbolísticos	31
1.	Consulta con el especialista en inteligencia artificial:	31
2.	Análisis de algoritmos de Machine Learning:	32
2.1	Utilización de algoritmos de clasificación:	33
2.2	Utilización de algoritmos de regresión:	33
2.3	Utilización de algoritmos de redes neuronales:	34
3.	Distribución de Poisson	36
4.	Implementación en un producto software:	38
5.	Proceso de Diseño y Mockups:	39
6.	Resultado final del producto.....	40
c)	Verificación del modelo predictivo con resultados reales:.....	41
1.	Recolección de datos:	41
2.	Limpieza y depuración de datos:	41
3.	Generación de predicciones:.....	41
4.	Visualización en la página web:	41
V.	ANÁLISIS DE RESULTADOS	43
	CONCLUSIONES	52
	RECOMENDACIONES	53
	BIBLIOGRAFÍA.....	54
	ANEXOS.....	58

LISTA DE TABLAS

TABLA I. RESULTADOS PREDICHOS PARTIDO 143

TABLA II. RESULTADOS PREDICHOS PARTIDO 244

TABLA III. RESULTADOS PREDICHOS PARTIDO 345

TABLA IV. RESULTADOS PREDICHOS PARTIDO 445

TABLA V. RESULTADOS PREDICHOS PARTIDO 546

TABLA VI. RESULTADOS PREDICHOS PARTIDO 647

LISTA DE FIGURAS

Fig. 1. Rendimiento de modelos tras selección de variables en la serie A.....	6
Fig. 2. Rendimiento de modelos tras selección de variables en la Bundesliga	7
Fig. 3. Rendimiento de modelos tras selección de variables en la liga Santander o española	7
Fig. 4. Rendimiento de modelos tras selección de variables en el ligue one	8
Fig. 5. Rendimiento de modelos tras selección de variables en la premier league	8
Fig. 6. Precisiones de algoritmos de clasificacion.....	33
Fig. 7. Precisiones de algoritmos de regresión.....	33
Fig. 8. Arquitectura de red neuronal LSTM.....	34
Fig. 9. Grafica de precisión de red neuronal LSTM.....	36
Fig. 10. Formula de la distribución de Poisson	37
Fig. 11. Product Backlog del proyecto Score Prediction	38
Fig. 12. Mockup sección noticias.....	39
Fig. 13. Interfaz principal sitio web Score Prediction.....	40
Fig. 14. Interfaz temporada 2022 Score Prediction.....	42
Fig. 15. Resultado real Manchester City vs West Ham United.....	43
Fig. 16. Resultado real Arsenal vs Leeds United	44
Fig. 17. Resultado real AFC Bournemouth vs Liverpool	45
Fig. 18. Resultado real Brighton & Hove Albion vs Brentford	46
Fig. 19. Resultado real West Ham United vs Aston Villa.....	47
Fig. 20. Grafica de eficiencia del modelo predictivo	48
Fig. 21. Grafica de eficiencia de marcadores finales	49
Fig. 22. Comparación de modelos de IA.....	51

LISTA DE ANEXOS

ANEXO 1: Formato de Protocolo para Revisión Documental en el Desarrollo de un Modelo Predictivo de Fútbol	58
ANEXO 2: Checklist de Análisis de Códigos Fuente para Modelos Predictivos de Fútbol:.....	61
ANEXO 3: Guía de Entrevistas para Profesionales en Analítica de Datos Aplicada al Fútbol:....	63
ANEXO 4: Carta de aprobación del jurado.....	65
ANEXO 5: Documento técnico de requisitos	66
ANEXO 6: Mockups iniciales de la página Score Prediction:.....	81
ANEXO 7: Manual de usuario	87
ANEXO 8: Resultados de partidos reales	97

GLOSARIO:

Analítica de datos

Proceso de examinar datos para extraer información valiosa, identificar patrones, y tomar decisiones informadas. Incluye técnicas descriptivas, diagnósticas, predictivas y prescriptivas.

Machine learning (Aprendizaje automático)

Rama de la inteligencia artificial que permite a los sistemas aprender de datos y mejorar su desempeño en tareas específicas sin ser programados explícitamente.

Minería de datos

Proceso de descubrir patrones y conocimientos útiles a partir de grandes volúmenes de datos utilizando algoritmos y técnicas estadísticas y de aprendizaje automático.

Modelos predictivos

Sistemas basados en inteligencia artificial que utilizan algoritmos avanzados, como redes neuronales, árboles de decisión y máquinas de soporte vectorial, para analizar datos históricos y patrones subyacentes con el fin de predecir resultados futuros en contextos específicos.

Scraping de datos

Proceso automatizado de extracción de información de sitios web mediante herramientas o scripts específicos.

Premier League

Liga de fútbol profesional de Inglaterra y una de las competiciones más importantes y seguidas a nivel mundial.

Árboles de decisión (Random Forest)

Técnica de machine learning utilizada para clasificar datos o realizar predicciones a través de un modelo jerárquico basado en reglas de decisión.

Redes Neuronales LSTM (Long Short-Term Memory)

Un tipo de red neuronal recurrente (RNN) diseñada para modelar secuencias de datos y aprender dependencias a largo plazo.

Regresión lineal

Modelo matemático que describe la relación entre una variable dependiente y una o más variables independientes, utilizado para realizar predicciones.

Redes neuronales artificiales

Modelo computacional inspirado en el funcionamiento del cerebro humano que se utiliza para resolver problemas complejos, como la predicción de resultados y el reconocimiento de patrones.

Inteligencia artificial (IA)

Área de la informática que busca desarrollar sistemas capaces de realizar tareas que requieren inteligencia humana, como aprender, razonar y tomar decisiones.

Accuracy (Precisión):

Se refiere a la proporción de predicciones correctas realizadas por un modelo en relación con el total de predicciones realizadas

APIs (Interfaz de Programación de Aplicaciones):

Las APIs son conjuntos de reglas que permiten que diferentes aplicaciones o sistemas se comuniquen entre sí. Actúan como intermediarios, facilitando el acceso a funciones o datos sin necesidad de conocer cómo están implementados internamente.

RESUMEN ANALÍTICO DE ESTUDIO RAE

Facultad: Ingeniería.

Programa: Ingeniería de Sistemas.

Fecha de elaboración: 31 de octubre del 2024.

Autores de la investigación: Julian Esteban Chavez Rivera.

Asesor: Esp. Luis Carlos Viteri Rosero.

Título de la investigación: Predicción de Resultados en el Fútbol a través de un producto Software Implementando Técnicas de Machine Learning y Análisis de Datos.

PALABRAS CLAVE: Analítica de datos, Fútbol, Machine learning, Modelos predictivos, Predicción de resultados.

DESCRIPCIÓN: El proyecto desarrolló modelos predictivos para anticipar resultados en la Premier League, utilizando algoritmos de machine learning y datos históricos. Se identificaron variables clave, como el rendimiento de los equipos, para determinar el modelo más preciso, útil en la prevención de fraudes en apuestas.

CONTENIDO

El informe final está estructurado de la siguiente forma:

Capítulo 1: Este capítulo aborda el desarrollo de modelos predictivos para resultados en fútbol, enfocado en mejorar precisión y combatir el fraude en apuestas. Se plantean objetivos, justificación y delimitación del proyecto, incluyendo el uso de técnicas de machine learning aplicadas a datos de la Premier League.

Capítulo 2: Se revisó el estado del arte, las técnicas aplicables para modelos de IA, y se definieron hipótesis y variables clave para el análisis predictivo en fútbol.

Capítulo 3: Se aborda el marco metodológico, detallando el paradigma, enfoque, métodos, y técnicas de recolección de datos que sustentan el desarrollo y validación del modelo predictivo de resultados de fútbol.

Capítulo 4: Se detallan los procedimientos para desarrollar el modelo predictivo y la página web, abordando las acciones necesarias para cumplir con el planteamiento del problema de investigación y asegurar la calidad de los resultados obtenidos.

Capítulo 5: Se presenta los hallazgos sobre la efectividad de los modelos predictivos en anticipar resultados de fútbol, evaluando la validez de las hipótesis y analizando la precisión de los algoritmos de machine learning en la identificación de patrones en datos históricos.

Conclusiones: El modelo demostró un buen desempeño en la predicción de victorias, empates y derrotas, aunque enfrenta desafíos al intentar anticipar el marcador exacto. Este resultado era predecible, dado que los factores impredecibles inherentes al fútbol pueden influir considerablemente en el resultado final. Por lo tanto, se sugiere que el modelo es más adecuado para proporcionar probabilidades de resultados generales en lugar de pronósticos precisos de marcadores. Esta capacidad para evaluar resultados generales permite a los usuarios tomar decisiones más informadas, aunque el enfoque en los marcadores exactos puede requerir ajustes adicionales para mejorar su precisión.

Recomendaciones: Es recomendable continuar enfocándose en ofrecer pronósticos de resultados generales (victorias, empates y derrotas) en lugar de marcadores exactos, ya que esta área ha demostrado un rendimiento más sólido del modelo. Para facilitar la interpretación y la toma de decisiones basada en datos, se sugiere desarrollar herramientas de visualización que presenten las probabilidades de resultados de manera clara y accesible para los usuarios interesados.

INTRODUCCIÓN

El fútbol, uno de los deportes más apasionantes, presenta desafíos constantes para predecir resultados. Anticipar los desenlaces de los encuentros es una tarea compleja, donde los modelos basados en análisis de datos y aprendizaje automático juegan un papel clave. Estos modelos utilizan estadísticas históricas de equipos y jugadores, tanto de local como de visitante, para estimar probabilidades de victorias, empates y derrotas.

La búsqueda de predicciones precisas ha dado lugar a estafadores conocidos como "tipsters", quienes ofrecen pronósticos falsos para fomentar apuestas deportivas online. Estos individuos se benefician de las pérdidas de los apostadores, lo que resalta la necesidad de contar con modelos éticos y precisos.

El uso de inteligencia artificial en el fútbol, como el caso de TCT-Scout de Olocip, permite analizar el juego con una precisión sin precedentes. Estas herramientas dividen el juego en componentes básicos, como el movimiento de jugadores y la interacción con el balón, ayudando a equipos y jugadores a tomar decisiones estratégicas más inteligentes. Esta revolución tecnológica está transformando la forma en que se entiende y juega al fútbol, mejorando el rendimiento y la eficiencia de este deporte.

I. PROBLEMA DE INVESTIGACIÓN

a) Objeto o Tema de Investigación.

Desarrollo de Modelos Predictivos mediante Analítica de Datos y Machine Learning para la predicción de resultados en el Contexto General del Fútbol y sus Principales Ligas en específico la liga inglesa o premier league.

b) Línea de Investigación.

Minería de datos: Investiga y desarrolla métodos para la extracción automática de conocimientos y patrones de grandes bases de datos estructuradas y no estructuradas, implementa el uso de técnicas, con el fin de promover actividades de investigación en algoritmos y métodos para el uso de grandes conjuntos de datos. [1]

c) Sub línea de Investigación

Técnicas descriptivas y predictivas, clasificatorias de minería de datos: En los sistemas de información el conocimiento se valora como un recurso estratégico, y las bases de datos de información transaccional se han convertido en un valioso repositorio de técnicas de minería de datos y una herramienta fundamental para cualquier campo o campo científico.

Es esencial en cualquier proyecto de investigación ya que permite extraer y analizar información importante, extraer patrones y clasificar datos. [1]

d) Planteamiento del problema

Se ha introducido en el mundo del fútbol un enfoque innovador que utiliza análisis de datos y técnicas de aprendizaje automático para predecir con precisión los resultados de los partidos [2]. Sin embargo, con el creciente uso de estas tecnologías en los principales deportes de competición en especial el fútbol se ha visto empañado por un problema creciente en las predicciones deportivas.

La proliferación de estafadores que ofrecen predicciones inútiles con fines de lucro, motiva a la utilización de canales de comunicación como Telegram y bots para aumentar su público objetivo, lo que puede provocar pérdidas económicas a quienes confían en sus predicciones. Los estafadores también pueden promocionar casas de apuestas en línea y beneficiarse de las pérdidas de los jugadores [3].

A pesar de la integración de la inteligencia artificial en el fútbol, como por ejemplo mediante el uso de aplicaciones como TCT-Scout de Olocip, que ayuda a clubes y jugadores a tomar decisiones estratégicas mediante análisis predictivos del rendimiento [4]. Se ha visto un cierto grado de desconfianza hacia las herramientas de predicción en el deporte; esto se debe en gran medida a la falta de precisión de algunos modelos. Esta falta de precisión puede deberse a diversos factores que destruyen la confianza de los usuarios, ya sean aficionados, jugadores o analistas.

A pesar de estos avances tecnológicos en los sistemas de ingeniería deportiva aplicada el problema persiste y factores como la fuerza del equipo, el campo local, el clima y las lesiones, lejos de mejorar la precisión de las predicciones parecen contribuir a la complejidad y desconfianza en los modelos de predicción.

El proyecto, aunque busca abordar estas cuestiones de la predicción precisa de los resultados de los partidos, se enfrenta a un panorama en el que el fraude en las apuestas deportivas y la falta de ética, amenazan la creación de una herramienta poderosa para que los usuarios analicen y predigan los desencadenantes de manera precisa y contextual. Nuestro objetivo es mejorar la experiencia futbolística de aficionados y profesionales, enriquecer el juego y fortalecer la confianza hacia este tipo de herramientas.

e) Formulación del problema.

¿Cómo se puede desarrollar y validar un modelo predictivo para anticipar los resultados de los partidos de fútbol utilizando técnicas de analítica de datos y aprendizaje automático, y cuál es su precisión al comparar sus predicciones con los resultados reales de las principales ligas de fútbol del mundo, con un enfoque particular en las temporadas 2021-2022 y 2022-2023 de la Premier League? [5]

f) Objetivos

1. General

Determinar de las técnicas de analítica de datos y aprendizaje automático (machine learning), el modelo predictivo más eficiente y con menor margen de error de los resultados en el fútbol a partir de un estudio comparativo y análisis del comportamiento de cada algoritmo.

2. Específicos

- Analizar un conjunto de datos existentes que contengan variables estadísticas clave acerca de la información histórica sobre partidos de fútbol que se van a involucrar en el proceso de desarrollo del modelo predictivo. ofreciendo una base sólida y detallada de datos con el propósito de realizar una investigación y anticipación de resultados en el ámbito del fútbol.
- Identificar el modelo más eficiente y preciso, de los modelos predictivos revisados, con el fin de implementarlo en un producto software para aquellos interesados en realizar pronósticos basados en datos históricos y estadísticas relevantes del fútbol.
- Verificar los resultados del modelo predictivo con resultados reales de partidos de fútbol de la temporada 2022 - 2023

g) Justificación

La propuesta de investigación se basa en la necesidad de entender cómo ha cambiado la percepción del fútbol a nivel global. Este deporte ha evolucionado de ser un simple entretenimiento a convertirse en una valiosa oportunidad de generación de ingresos. La creciente práctica de predecir los resultados de los partidos, apoyada por modelos predictivos confiables que utilizan inteligencia artificial, se presenta como una herramienta clave para los aficionados que desean un enfoque más claro y seguro al realizar apuestas.

La implementación de estos modelos no solo brinda a los seguidores una mayor confianza al tomar decisiones sobre sus apuestas, sino que también ayuda a equilibrar el entretenimiento con las posibilidades de ganancias. Este enfoque innovador, que incorpora nuevas variables y utiliza redes neuronales entrenadas con datos estadísticos, no solo beneficiará a los aficionados al fútbol, sino que también planteará desafíos significativos en la interacción con poderosas herramientas de inteligencia artificial.

Desde una perspectiva académica, este proyecto representa un avance en la fusión de la ingeniería de sistemas con el ámbito deportivo, estableciendo una base sólida para la predicción de resultados en diversas disciplinas deportivas. La investigación y desarrollo propuestos no solo aportan innovación y eficiencia al campo de la predicción deportiva, sino que también contribuirán al crecimiento continuo de la ingeniería de sistemas en su relación con el deporte.

A nivel social, la adopción de tecnologías emergentes en la predicción de resultados deportivos podría transformar la manera en que la sociedad entiende y se relaciona con el deporte. La creciente influencia de la ciencia y la tecnología puede enriquecer la experiencia de los aficionados y abrir nuevas oportunidades para su participación y análisis crítico.

Desde una perspectiva económica, la inversión en este proyecto es fundamental para maximizar su potencial. La tendencia observada en clubes de élite, como el Manchester City, que destinan recursos significativos a la recopilación y análisis de datos, subraya la importancia estratégica de incorporar nuevas tecnologías para alcanzar el éxito en el ámbito deportivo.

h) Delimitación

Ámbito Geográfico: Esta investigación se llevará a cabo en la Universidad CESMAG, ubicada en la ciudad de Ciudad Pasto. El estudio se centrará exclusivamente en el análisis de diferentes técnicas de machine learning y analítica de datos para aplicarlas en el ámbito de la predicción de partidos de fútbol en el mundo.

Ámbito Temporal: El período de investigación y desarrollo abarca un año y medio académico completo, desde agosto de 2023 hasta diciembre de 2024.

Población o Muestra: las variables estadísticas que se usarán corresponden a las temporadas 2021-2022 y 2022-2023 de la Premier League.

Tecnologías y Metodologías: La investigación utilizara técnicas de machine learning y analítica de datos tales como árboles de decisión, regresión lineal, redes neuronales, entre otros, se evaluarán cuáles técnicas producirán el mejor desempeño para poderlo aplicar en un futuro a una herramienta software.

Contexto Institucional: El proyecto se llevará a cabo en el contexto de la Universidad CESMAG, con acceso a sus recursos y colaboración con el personal académico de la institución.

Alcance Temático: El enfoque de la investigación se centrará en la aplicación y clasificación de nuevas tecnologías en el análisis de partidos de fútbol, tendrá como objetivo mejorar la precisión, proporcionar herramientas útiles para resolver desafíos como el fraude en las apuestas deportivas.

II. MARCO TEÓRICO

a) Antecedentes

1. Internacionales

1.1. Modelos predictivos en el deporte:

En el contexto internacional, cabe mencionar que un reciente estudio realizado por estudiantes de la Universidad Estatal de Navarra (UPNA) [6] cuenta con el análisis más completo de los diferentes métodos de regresión para la predicción de eventos deportivos. Estos estudios no se limitaron a una liga específica, sino que el modelo se aplicó a cinco ligas importantes:

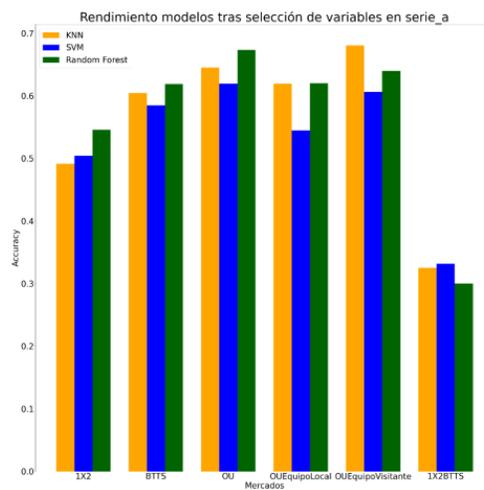


Fig. 1. Rendimiento de modelos tras selección de variables en la serie A

Nota: fuente <https://academica-e.unavarra.es/entities/publication/f24670a1-4d3d-4aab-8606-941c12e09c9d>
académica es un repositorio de varios trabajos de investigación

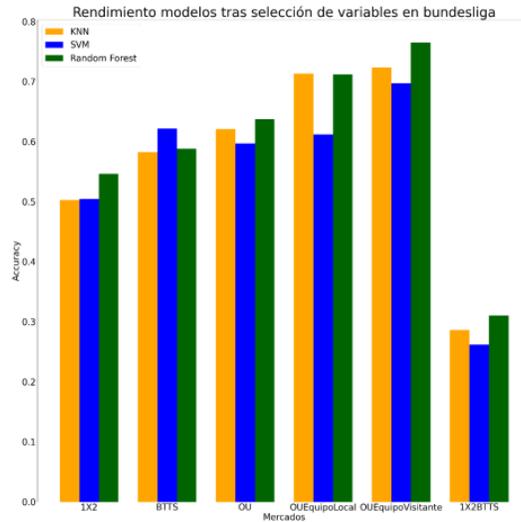


Fig. 2. Rendimiento de modelos tras selección de variables en la Bundesliga

Nota: fuente <https://academica-e.unavarra.es/entities/publication/f24670a1-4d3d-4aab-8606-941c12e09c9d> académica es un repositorio de varios trabajos de investigación

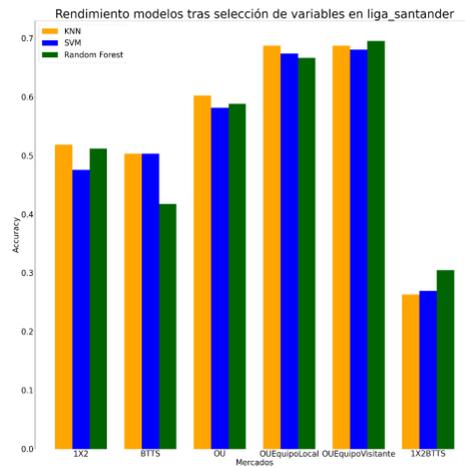


Fig. 3. Rendimiento de modelos tras selección de variables en la liga Santander o española

Nota: fuente <https://academica-e.unavarra.es/entities/publication/f24670a1-4d3d-4aab-8606-941c12e09c9d> académica es un repositorio de varios trabajos de investigación

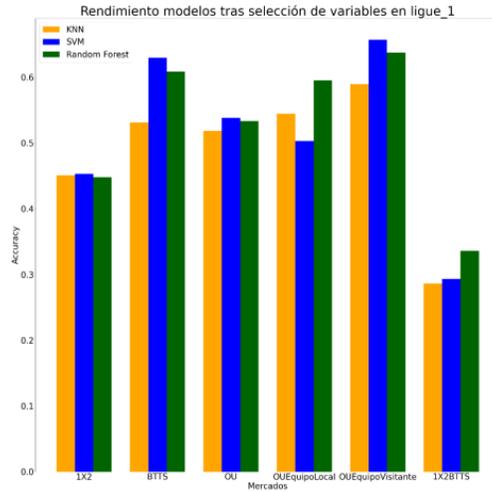


Fig. 4. Rendimiento de modelos tras selección de variables en el ligue one

Nota: fuente <https://academica-e.unavarra.es/entities/publication/f24670a1-4d3d-4aab-8606-941c12e09c9d> académica es un repositorio de varios trabajos de investigación

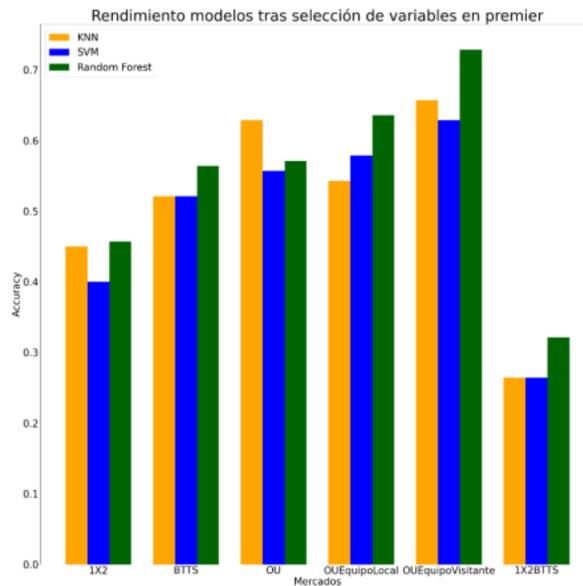


Fig. 5. Rendimiento de modelos tras selección de variables en la premier league

Nota: fuente <https://academica-e.unavarra.es/entities/publication/f24670a1-4d3d-4aab-8606-941c12e09c9d> académica es un repositorio de varios trabajos de investigación

Este estudio ofrece una perspectiva amplia al abarcar múltiples ligas, lo que enriquece la comprensión de las metodologías efectivas para predecir resultados deportivos. Sus hallazgos pueden servir como base para seleccionar y evaluar métodos en proyectos similares, proporcionando una referencia sólida para el desarrollo de modelos predictivos más precisos y eficientes.

- En el campo del baloncesto cabe mencionar la investigación realizada por Federación Española de Asociaciones de Docentes de Educación Física (FEADEF) [7]. El objetivo de este estudio es establecer un modelo predictivo del riesgo de lesiones en deportistas no profesionales, centrándose en detalles anatómicos y funcionales relacionados con la inestabilidad de rodilla y tobillo. Se propone utilizar metodologías de regresión discriminante que pueden adaptarse al contexto del fútbol. Un análisis exhaustivo de los factores que influyen en el riesgo de lesiones ofrece una perspectiva valiosa para entender y prevenir eventos adversos, subrayando la importancia de considerar variables específicas para mejorar la precisión de los modelos predictivos en el ámbito deportivo.
- En el ámbito de la educación física, estudiantes de la Facultad de Ciencias del Deporte de la Universidad de Murcia en España [8]. Este estudio utiliza un modelo predictivo para evaluar la motivación de adolescentes de 13 a 18 años hacia la Educación Física, aplicando el modelo de correlación de Pearson y ecuaciones estructurales (SEM). Los resultados indican que la **motivación intrínseca** es fundamental en estudiantes con altas puntuaciones, lo que fomenta una actitud positiva que trasciende el interés por el deporte. Esto resalta la necesidad de considerar factores psicológicos en los modelos predictivos, lo que podría mejorar la capacidad de anticipar resultados y comportamientos en el ámbito del fútbol.
- Una muestra de que la utilización de modelos predictivos se puede aplicar en diferentes contextos queda demostrada en la revista médica Clínica las Condes [9]. El uso de técnicas de machine learning se considera el futuro de la medicina, ofreciendo múltiples beneficios a los profesionales de la salud, como diagnósticos más precisos y tratamientos personalizados. Al observar las similitudes en la aplicación de modelos predictivos, se puede explorar cómo estas estrategias exitosas en medicina pueden adaptarse para mejorar la precisión y relevancia en el ámbito del fútbol. Esto sugiere una perspectiva innovadora

para la investigación y el desarrollo de modelos predictivos en este deporte, potenciando así su eficacia.

- Un modelo predictivo aplicado al medio ambiente, como el desarrollado por estudiantes de la Universidad Tecnológica de Perú, utiliza redes LSTM para evaluar el caudal efluente. Este enfoque busca facilitar la prevención y mejorar la toma de decisiones a través de alertas tempranas. La implementación de este sistema permite una gestión más efectiva de los recursos hídricos y la anticipación de eventos adversos, contribuyendo así a la seguridad y sostenibilidad ambiental. [10].
- En el campo empresarial la utilización de técnicas de machine learning para la creación de modelos predictivos para empresas donde a través del método escogido el método clúster o de agrupación, donde los algoritmos seleccionados cumplen con las prestaciones exigibles para la creación de modelos predictivos [11]. Se puede inferir que el uso de las estrategias y enfoques exitosos utilizados en el contexto empresarial podrían adaptarse y aplicarse para mejorar la calidad y la eficiencia de los modelos predictivos en el ámbito deportivo, ofreciendo así una mayor confiabilidad en las predicciones de resultados de partidos.

1.2. Minería de datos:

- La minería de datos en el ámbito internacional ha sido ampliamente utilizada, un grato ejemplo es la utilización de la misma para determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. Este estudio fue realizado por estudiantes de la Universidad de Loja, Ecuador. [12] Se utilizaron diversas herramientas y metodologías de minería de datos para evaluar la influencia de ciertos estudiantes en materias virtuales en la universidad. Este enfoque puede aplicarse al ámbito deportivo, especialmente en el fútbol, donde las técnicas de minería de datos pueden identificar patrones y relaciones clave en datos históricos de partidos. Esto contribuiría a mejorar la precisión y eficacia de los modelos predictivos para anticipar resultados en el fútbol, optimizando así la toma de decisiones estratégicas y el rendimiento del equipo.
- En el campo profesional del fútbol Español el estudio por parte de la universidad de la Coruña, España y para el Real Club Deportivo de La Coruña el estudio realizado por estudiantes de esta Universidad [13] El uso de técnicas de minería de datos ha permitido a

un club de fútbol profesional desarrollar una herramienta que aplica análisis estadístico y machine learning en la exploración de jugadores, proporcionando información valiosa para la dirección deportiva. Este enfoque demuestra cómo el análisis avanzado de datos puede ofrecer perspectivas relevantes para la gestión de un club. Además, se sugiere que estas mismas técnicas podrían utilizarse para analizar datos históricos de partidos, mejorando así la capacidad predictiva de los modelos y contribuyendo a decisiones estratégicas en el fútbol.

2. Nacionales

2.1. Modelos predictivos:

- Estudios realizados en Colombia especialmente por estudiantes en la Universidad Católica de Colombia en Bogotá [14], Se confirman nuevos métodos de aprendizaje, como la regresión logística, random forest y máquinas de soporte vectorial, en el contexto del fútbol colombiano. Sin embargo, el estudio concluye que el ambiente general del fútbol en el país limita la consideración de muchos factores relevantes, lo que dificulta predecir de manera clara y sencilla el ganador de los partidos. Esta conclusión resalta los desafíos inherentes a la predicción en este ámbito, enfatizando la necesidad de abordar integralmente los diversos factores que afectan los resultados. Este antecedente podría guiar estrategias en futuros proyectos para enfrentar estos desafíos específicos.
- Otros estudios relacionados a modelos predictivos aplicados al deporte son el estudio realizado por parte de la Universidad Autónoma de Bucaramanga. [15] El objetivo es consolidar un modelo predictivo del rendimiento en nadadores colombianos utilizando un modelo de regresión para identificar factores determinantes que influyen en sus carreras. Este enfoque resalta la utilidad de los modelos predictivos para anticipar el rendimiento deportivo y proporciona información valiosa sobre los elementos clave que afectan los resultados. La experiencia adquirida podría enriquecer la metodología de proyectos futuros en fútbol, al aplicar un enfoque similar basado en la identificación de variables determinantes para mejorar la precisión en las predicciones de resultados.
- En el mundo de las apuestas deportivas, previas investigaciones por parte de estudiantes de la Universidad de Antioquia, [16]. Se concluyó que el uso de técnicas de analítica de datos y machine learning, específicamente mediante el modelo Histogram Gradient Boosting, logró una precisión notable, lo que puede facilitar la toma de decisiones para los principales

involucrados. La experiencia adquirida puede guiar la selección y evaluación de modelos predictivos para anticipar resultados en el fútbol, proporcionando una base sólida para decisiones informadas en el ámbito de las apuestas deportivas.

- En la utilización de la probabilidad y estadística en la aplicación de un modelo predictivo, en la Universidad de los Andes [17]. Se ha investigado la repetición y deserción escolar en los colegios oficiales de Colombia, identificando factores de riesgo y aplicando un modelo de simulación mediante técnicas de estadística avanzada. Este estudio establece una base para futuras aplicaciones del modelo, respaldando la viabilidad de enfoques similares en el ámbito del fútbol para mejorar la predicción de resultados, considerando la complejidad de los factores involucrados. Además, la aplicación de un modelo de simulación podría ser útil para probar los modelos predictivos seleccionados.
- Un estudio importante a resaltar la aplicación de un modelo predictivo para un sistema de manufactura orientado a la industria 4.0 desarrollado por estudiantes de la Universidad Javeriana de Colombia donde el modelo brinda información sobre el tiempo restante para que se finalice un producto que está en proceso. El conocimiento de esta información en momentos oportunos puede apalancar la toma de buenas decisiones para optimizar el desempeño del sistema de manufactura [18]. Aunque los contextos son diferentes, las lecciones aprendidas en la gestión de la incertidumbre y la toma de decisiones oportunas pueden ser relevantes para el proyecto de modelo predictivo de resultados en el fútbol, considerando la complejidad inherente a ambas situaciones.
- Una aplicación del modelo predictivo para el mundo empresarial queda descrito en el estudio realizado por la Universidad Piloto de Colombia, en la cual algunos estudiantes quisieron crear un modelo predictivo orientado para identificar el riesgo de quiebra corporativa en empresas Pyme de Colombia en el periodo comprendido entre el 2015-2018 [19], Se busca que las empresas se anticipen a patrones observados en otras que han quebrado, fortaleciendo así sus áreas más débiles. Este enfoque resalta la utilidad de los modelos predictivos para identificar proactivamente riesgos y tomar medidas preventivas, lo cual es aplicable al desarrollo de un modelo predictivo de resultados en el fútbol. Aunque los contextos son diferentes, la noción de anticipación y fortalecimiento proactivo es relevante dada la complejidad y dinámica de ambos escenarios.

2.2. Minería de datos:

- La minería de datos aplicada en el sector comercial ha dado gratos resultados. Prueba de ello se ve reflejada en previos estudios por parte de la Universidad Piloto de Colombia, donde se buscó plantear estrategias para fidelizar clientes y mejorar las ventas de Makro Avenida Boyacá en la ciudad de Bogotá. [20] La implementación del algoritmo K-means y estrategias de Cluster ha permitido identificar el potencial del mercado objetivo en un contexto comercial. Este antecedente sugiere que el análisis predictivo puede aplicarse en un modelo de resultados en el fútbol para identificar patrones y tendencias que influyen en los partidos, similar a su uso en el ámbito comercial. A pesar de las diferencias contextuales, la aplicación de técnicas predictivas para comprender y aprovechar oportunidades es un aspecto compartido que puede enriquecer el desarrollo de modelos en el deporte.

3. Regional

3.1. Modelo predictivo:

- Aunque no se han realizado estudios específicos sobre este tema en la región de Nariño, se han utilizado modelos predictivos en una variedad de contextos. Un ejemplo específico es un estudio realizado por estudiantes de la Universidad de Nariño [21], donde se utilizó un modelo de regresión logística para resolver el problema de minimizar la tasa de deserción estudiantil.

El estudio encontró que el método de regresión logística ofrece los mejores resultados para abordar la deserción estudiantil, lo que ha contribuido a reducir esta tasa en la Universidad de Nariño. Esta experiencia sugiere que enfoques similares pueden adaptarse para entender dinámicas específicas en el fútbol, identificando factores que afectan los resultados de los partidos según el contexto. La aplicación del método de regresión logística podría ser útil para desarrollar un modelo predictivo que considere variables determinantes en el rendimiento deportivo.

3.2. Minería de datos:

- Una investigación realizada por estudiantes de la Universidad de Nariño acerca de la identificación del rendimiento académico de los estudiantes de pregrado de la Universidad

de Nariño [22] A través de técnicas de minería de datos, se identificaron variables que influyen en el rendimiento de los estudiantes de una universidad, lo que demuestra que estas técnicas pueden predecir el desempeño en estudios de pregrado. Aplicando este enfoque al modelo predictivo de resultados en el fútbol, se sugiere que métodos similares podrían ayudar a identificar variables cruciales que afectan el rendimiento de equipos y jugadores, lo que permitiría pronosticar resultados con mayor precisión.

b) Supuestos teóricos de investigación

Para el desarrollo de este proyecto, se ha tenido en cuenta la formulación de supuestos teóricos para guiar el proceso de investigación acerca de cuál técnica de machine learning y minería de datos, aplicada a un modelo predictivo, puede ser más factible y confiable a la hora de predecir los resultados del fútbol. Teniendo en cuenta esto y la futura implementación de un software que sirva para la conexión entre el modelo predictivo y los futuros usuarios de este, se tienen en cuenta los siguientes supuestos teóricos.

1. Disponibilidad de datos:

Afortunadamente, gracias al sitio web Footy Stats, se cuenta con un conjunto de datos históricos de cada una de las ligas y competencias del fútbol a nivel mundial. Para comodidad de la investigación, se tomarán datos de la temporada 2021-2022 y 2022-2023 de la Premier League.

Junto con este supuesto, hay que tener en cuenta que la manipulación de datos, [23] si se busca un modelo predictivo 100% confiable, tiene que tener cierto grado de depuración de datos. Por ello, se tendrán en cuenta las siguientes.

De igual manera la información obtenida de la página oficial de la Premier League será utilizada como fuente confiable para corroborar los datos relacionados con los equipos, jugadores, y estadísticas del torneo. Esta información será clave para validar los resultados obtenidos y mejorar la precisión del análisis en el desarrollo del proyecto. Del mismo modo, servirá para extraer datos adicionales que puedan enriquecer la investigación

1.1. Estabilidad en las Características de Equipos y Patrones de Juego:

Los integrantes del equipo, las estrategias utilizadas y patrones de juego se consideran importantes en el tiempo en términos de calidad y fortaleza en el juego [24]. Esta suposición es necesaria para utilizar datos históricos en la construcción de modelos.

1.2. Distribución de probabilidad:

Las variables relevantes, como el equipo y el jugador, siguen una distribución de probabilidad que se puede implementar de manera eficiente Identificar todos los resultados posibles que pueden ocurrir en un experimento determinado. Considerando la probabilidad de que ocurra un evento en el futuro. Las distribuciones de probabilidad son una herramienta de pronóstico esencial, ya que permiten desarrollar escenarios de eventos futuros, teniendo en cuenta las tendencias actuales de diversos fenómenos.

Las características más importantes a considerar en una distribución de probabilidad son:

- La probabilidad de un resultado particular es de cero a uno.
- La suma de las probabilidades de todos los resultados mutuamente excluyentes es 1. [25]

Estos son necesarios para recopilar probabilidades y predicciones.

1.3. Correlación de variables predictoras:

Las variables utilizadas en el modelo, como el historial de actuaciones, la forma actual de los jugadores y las condiciones del partido, se consideran relacionadas con el resultado de la predicción partiendo del punto de factores clave como la demanda física en partidos de fútbol profesional, la calidad del calzado, utilización de GPS para la elaboración del mapa de calor, como las lesiones pueden afectar en el rendimiento del jugador entre otros [26].

1.4. Unidad del equipo:

Los equipos son similares en términos de habilidades y destrezas, lo que puede simplificar el modelo. Sin embargo, esto puede no ser cierto en la realidad y algunas muestras incluyen heterogeneidad de igual modo hay características antropométricas como características fisiológicas, algo que a través de estas características podemos establecer características comunes y propias de los jugadores [27].

1.5. Ausencia de cambio rápido de rendimiento:

Existen cambios repentinos en el rendimiento del equipo que no pueden ser reconocidos por lo que no se producen lesiones graves o cambios en el entrenador [28].

2. Herramientas de minería de datos y machine learning:

Con el fin de encontrar patrones de los grandes volúmenes de datos históricos de partidos que se van a manejar se debe tener en cuenta que técnicas brindaran efectividad a la hora de realizar el modelo predictivo, [29] algunos son:

2.1. Técnica de seguimiento de patrones:

Implica identificar, rastrear y clasificar tendencias a través de análisis de datos inteligentes, haciendo un análisis exhaustivo de todos los factores clave que intervienen en una jugada, para de esta manera saber la influencia que un jugador puede llegar a tener en cada encuentro deportivo [30].

2.2. Técnica de asociación:

Utilizar esta técnica nos ayudará a encontrar relaciones entre atributos o eventos en datos, útil para comprender el comportamiento de un equipo o jugador en específico [31]. Un claro ejemplo para la utilización de esta técnica es asociar las variables relacionadas a los goles de un equipo.

2.3. Técnica de clasificación:

Dividir los datos en categorías según características definidas que sean útiles para predecir cambios en los patrones o técnicas de juego.

2.4. Técnica de clustering:

Agrupar datos en función de la similitud con la ayuda de una fuente confiable de datos como lo es Footy Stats, útil para encontrar patrones y tendencias en conjuntos de datos y con la el posterior análisis se logrará la obtención de una información de gran relevancia [32].

2.5. Técnica de patrones secuenciales:

Se centra en el seguimiento de una secuencia de eventos, las cuales pueden definir eventos relacionados con partidos de fútbol con ayuda de antecedentes de patrones y secuencias de datos [33].

2.6. Técnica de árbol de decisión:

Es de suma importancia la utilización de un modelo de árbol de decisión la cual es la base fundamental de miles de modelos predictivos la cual modele la relación entre entradas y salidas mediante reglas condicionales, comúnmente utilizadas en clasificación y regresión [34].

2.7. Técnica de análisis de regresión:

Relación lineal entre variables utilizadas para estimar cuál variable tiene más impacto en cualquier partido de fútbol, identificando variables dependientes e independientes, estas a su vez responden preguntas tipo: ¿Qué factores se pueden ignorar? ¿Cómo interactúan estos factores entre sí? Entre otras, ayudando a la optimización en la futura elaboración del modelo predictivo planeado [35].

2.8. Técnica de procesamiento de memoria a largo plazo:

Esta técnica es de suma importancia para el proyecto dado que examina datos a lo largo del tiempo, lo que resulta útil para reconocer patrones en los datos a lo largo del tiempo [36].

2.9. Técnica de procesamiento de redes neuronales:

Un modelo de aprendizaje automático que encuentra patrones complejos en grandes conjuntos de datos para generar resultados precisos [37].

3. Implementación del modelo en un producto software:

Se busca crear un producto de software con una interfaz de usuario intuitiva, para que los usuarios potenciales de este modelo predictivo puedan aprovechar los beneficios que esta herramienta les pueda brindar. Conjuntamente, lo que se busca es que el usuario final pueda interactuar dinámicamente con datos y resultados, cambiar parámetros y obtener visualizaciones en tiempo real de las diferentes predicciones que desee.

Teniendo en cuenta que la ingeniería de software es el conjunto de actividades computacionales dedicadas a la creación, diseño, implementación y mantenimiento de software.

El software en sí es un conjunto de instrucciones o programas que le dicen a la computadora qué hacer. Es independiente del hardware y permite programar el ordenador. Hay tres tipos principales:

- Software del sistema: proporciona funciones básicas como sistema operativo, administración de discos, servicios, administración de hardware y otras necesidades operativas.
- Software de programación: proporciona a los desarrolladores herramientas como editores de texto, compiladores, enlazadores, depuradores y otras herramientas para crear código.
- Software de aplicación (aplicación o app): ayuda a los usuarios a realizar tareas. Paquetes de oficina, software de gestión de datos, reproductores multimedia y programas de seguridad son sólo algunos ejemplos. El término "aplicación" también se refiere a aplicaciones web y móviles, como las que se utilizan para comprar en Amazon.com, chatear en Facebook o publicar fotos en Instagram [38].

4. Método ágil: SCRUM

Scrum es una metodología ágil de gestión de proyectos que ayuda a los equipos a estructurar y gestionar el trabajo a través de un conjunto de valores, principios y prácticas. Por lo anterior y otros: el método utilizado para realizar la investigación será SCRUM. Aprende de la experiencia, organízate a la hora de resolver problemas y reflexiona sobre tus victorias y derrotas para mejorar continuamente. [39]

Eventos de Scrum

A continuación, se muestra una lista de todos los eventos principales en los que puede participar un equipo Scrum:

- **Organizar la colección:** Esta actividad, a veces denominada limpieza de trabajos pendientes, es responsabilidad del propietario del producto. La tarea principal del propietario de un producto es ejecutar el producto de acuerdo con su visión y mantenerse por delante del mercado y de los clientes.
- **Planificación de Sprint:** Durante esta reunión, todo el equipo de desarrollo planifica el trabajo a realizar (alcance) en el sprint actual. Esta reunión es dirigida por un scrum master y durante la misma el equipo define el objetivo del sprint.
- **Sprint:** Un sprint es el período de tiempo real en el que el equipo Scrum trabaja en conjunto para completar el incremento. Un sprint suele durar dos semanas, aunque algunos equipos informan que es más fácil asignar cuatro semanas al alcance o un mes para asignar incrementos valiosos.
- **Daily Scrum o Quick Meeting:** Se trata de una reunión diaria muy corta que siempre se lleva a cabo a la misma hora (normalmente por la mañana) y en el mismo lugar para facilitar las cosas. Muchos equipos intentan terminar el partido en 15 minutos, pero esto es sólo una guía.
- **Revisión de Sprint:** Al final del sprint, el equipo se reúne informalmente para revisar la demostración o verificar el desarrollo. El equipo de desarrollo muestra a las partes internas los elementos del trabajo pendiente que ahora están "completados".

- **Sprint retrospectivo:** Una retrospectiva es donde el equipo se reúne para registrar y revisar qué funcionó y qué no sobre el sprint, el proyecto, las personas o relaciones, las herramientas e incluso algunos protocolos. [39]

5. Lenguajes de programación:

Un lenguaje de programación es una forma de comunicarte con tu ordenador, tableta o teléfono móvil y decirle lo que quieres hacer. Existen diferentes tipos de lenguajes: principalmente de bajo nivel y de alto nivel. La diferencia es si estás cerca o lejos de nuestro dispositivo. Esta proximidad implica su control sobre el dispositivo, placa o controlador. Aquí encontrará muchos lenguajes diferentes como C, C++, Java, PHP, Python, C#, ASP entre otros. [40]

5.1. Python:

Python es un lenguaje de programación ampliamente utilizado en aplicaciones web, desarrollo de software, análisis de datos y aprendizaje automático (ML). Los desarrolladores utilizan Python porque es eficiente, fácil de aprender y puede ejecutarse en múltiples plataformas. El software Python se puede descargar gratis, se integra bien con cualquier tipo de sistema y acelera el desarrollo. [41]

6. Scraping de datos:

Se refiere a una técnica en la que un programa informático recopila datos a partir del resultado generado por otro programa. Generalmente, esto se manifiesta en el web scraping, un proceso que utiliza aplicaciones para extraer información valiosa de sitios web de manera automatizada [42].

6.1. BeautifulSoup:

Es una biblioteca de Python diseñada para extraer datos de documentos HTML y XML. Esta herramienta crea un árbol de elementos a partir del documento, lo que permite navegar y extraer información de manera eficiente [43].

c) Variables de estudio

Para el estudio y posterior elaboración de un modelo predictivo utilizando técnicas de machine learning y minería de datos, es de suma importancia identificar las variables dependientes e independientes para construir un modelo que prediga de manera efectiva el suceso deportivo.

1. Variables independientes

- Algoritmo de Modelado Predictivo de Fútbol: se refiere a la elección y aplicación de un conjunto específico de reglas y procesos lógicos en el campo de la analítica de datos y el aprendizaje automático. Este algoritmo está diseñado para analizar datos históricos de fútbol con el fin de predecir de manera precisa y eficiente los resultados de futuros partidos, mejorando así la capacidad de pronóstico en el ámbito deportivo.
- Historial de Partidos: La frecuencia y consistencia de los patrones de desempeño histórico de los equipos y jugadores.
- Localía: La influencia del hecho de que un equipo juegue en casa o como visitante, considerando el rendimiento histórico en ambos escenarios.
- Estadísticas Colectivas del Equipo: Variables que describen el rendimiento colectivo del equipo, como la posesión de balón, el número de tiros a puerta, y la precisión de pases.
- Estadísticas Históricas de Enfrentamientos Directos: El rendimiento histórico de los equipos en enfrentamientos directos, ya que algunos equipos pueden tener ventajas o desventajas particulares contra ciertos oponentes.

2. Variables dependientes

- Tasa de Éxito: La efectividad del modelo para predecir correctamente los resultados de los partidos.
- Error de Predicción: La discrepancia entre las predicciones del modelo y los resultados reales.

Estas variables proporcionarán una base para evaluar la eficacia y la robustez del modelo predictivo en el contexto del fútbol de las principales ligas del mundo, especialmente en la Premier League.

d) Definición nominal de las variables

- Algoritmo de Modelado Predictivo de Fútbol: Un conjunto específico de reglas y procedimientos matemáticos utilizado para generar predicciones sobre los resultados de los partidos de fútbol basándose en datos históricos y diversas variables.
- Historial de Partidos: El registro acumulado de encuentros anteriores de equipos y jugadores, que incluye información detallada sobre los resultados, goles anotados, goles recibidos, y otros aspectos relevantes que describen el rendimiento pasado.
- Localía: La situación en la que un equipo juega en su propio estadio (local) o en el estadio del equipo contrario (visitante), considerando el rendimiento histórico en ambos contextos.
- Estadísticas Colectivas del Equipo: Variables que describen el rendimiento general de un equipo, incluyendo la posesión de balón, el número de tiros a puerta, y la precisión en los pases.
- Estadísticas Históricas de Enfrentamientos Directos: El registro del rendimiento histórico de equipos en encuentros directos entre ellos, destacando posibles ventajas o desventajas que ciertos equipos puedan tener contra oponentes específicos.
- Tasa de Éxito: La medida de qué tan precisamente el modelo puede predecir los resultados reales de los partidos de fútbol.
- Error de Predicción: La discrepancia cuantitativa entre las predicciones generadas por el modelo y los resultados reales de los partidos.

e) Definición Operativa De Variables

- Algoritmo de Modelado Predictivo de Fútbol: Medida de la efectividad y precisión del algoritmo específico seleccionado para la creación del modelo predictivo, evaluada

mediante métricas como la precisión de las predicciones, la capacidad para aprender de datos históricos y la adaptabilidad a cambios en las variables de entrada.

- **Historial de Partidos:** Registro cronológico y detallado de todos los encuentros previos de un equipo o jugador, incluyendo datos específicos como resultados (victoria, derrota, empate), goles anotados, goles recibidos, ubicación del partido, y cualquier otro aspecto relevante que permita evaluar el rendimiento pasado.
- **Localía:** Identificación de la ubicación del partido como "local" o "visitante" para cada equipo, proporcionando información sobre el rendimiento histórico y estadísticas asociadas a su desempeño en ambos contextos.
- **Estadísticas Colectivas del Equipo:** Recopilación de variables que describen el rendimiento global de un equipo, tales como la posesión de balón, número de tiros a puerta, precisión de pases, y otras métricas colectivas relevantes para analizar su desempeño en conjunto.
- **Estadísticas Históricas de Enfrentamientos Directos:** Documentación detallada del rendimiento histórico de equipos en enfrentamientos directos, incluyendo resultados previos, goles anotados y recibidos, y cualquier patrón distintivo en el desempeño de equipos específicos frente a oponentes particulares.
- **Tasa de Éxito:** Cuantificación del porcentaje de predicciones correctas en comparación con los resultados reales de los partidos, utilizada como indicador de la efectividad general del modelo.
- **Error de Predicción:** Medida cuantitativa de las discrepancias entre las predicciones del modelo y los resultados reales de los partidos, expresada en términos de desviación porcentual o puntos específicos.

f) Formulación de hipótesis

El fútbol actual ha tenido un cambio abrupto a la hora de enfrentarse a un rival, infinidad de tácticas en las cuales algunos equipos se decantan más por la posesión de balón [44], el contraataque [45] presión alta, marcaje, defensa en zona, repliegue, basculaciones entre otros [46]. han sido totalmente influyentes en los clubes de fútbol modernos.

De igual forma las diferencias de presupuestos partiendo de que “el dinero para los clubes de fútbol es como la gasolina para los coches” [47] han hecho que los clubes con una mayor capacidad

económica tengan acceso a mejores jugadores, mejor tecnología, y mejor marketing tanto para los equipos como para los jugadores [48]. Todo esto, sumado a la necesidad de encontrar el mejor algoritmo para la creación de un modelo predictivo totalmente confiable y eficiente, teniendo en cuenta una infinidad de variables estadísticas y factores tanto climáticos como deportivos, se han generado las siguientes hipótesis.

1. Hipótesis de investigación

Mediante un estudio comparativo y análisis avanzado de los algoritmos de analítica de datos y aprendizaje automático, se puede identificar un modelo predictivo más eficiente con menor margen de error en la predicción de resultados en partidos de fútbol en comparación con los modelos predictivos del estado del arte.

2. Hipótesis nula

No se puede identificar un modelo predictivo más eficiente con un menor margen de error en la predicción de resultados en partidos de fútbol mediante un estudio comparativo y análisis avanzado de los algoritmos de analítica de datos y aprendizaje automático, en comparación con los modelos predictivos del estado del arte.

3. Hipótesis alterna

Mediante un estudio comparativo y análisis avanzado de algoritmos de analítica de datos y aprendizaje automático, es posible desarrollar un modelo predictivo con capacidad suficiente para superar las limitaciones actuales de los modelos del estado del arte, los cuales no han demostrado ser efectivos en la predicción precisa de resultados en partidos de fútbol debido a la naturaleza altamente impredecible de este deporte.

III. METODOLOGÍA

a) Paradigma

El paradigma que se escogerá para esta investigación será el paradigma positivista, el cual será de mucha ayuda gracias a la naturaleza de este paradigma. Este es idóneo para la creación de un modelo predictivo [49], ya que se basa en los datos recopilados, en la certeza de que el modelo matemático utilizado será óptimo, en el análisis estadístico y entre otras características importantes propias del paradigma. Esto nos ayudará en la toma de decisiones fundamentales para tener un modelo predictivo mucho mejor y más preciso.

b) Enfoque

La investigación que se va a realizar se caracteriza básicamente en la recopilación de datos, relaciones y tendencias, y lo más importante, el análisis de datos numéricos y estadísticos para encontrar patrones, lo cual será la base para el modelo predictivo. Teniendo en cuenta lo anterior, esta investigación es de carácter cuantitativo [50].

c) Método

El método científico se ajusta mucho más a los objetivos de la investigación, ya que este nos ayudará a llevar a cabo un estudio sistemático, basándonos en la obtención de resultados y comparación con datos históricos, con el fin de determinar el modelo predictivo más eficiente y preciso [51].

d) Tipo de investigación

El tipo de investigación escogido, dado a la naturaleza de la investigación realizada y teniendo en cuenta que se desea hacer un estudio comparativo y un análisis del comportamiento de diferentes técnicas de analítica de datos y machine learning, es la investigación experimental. Esto se debe a que, entre sus ventajas más importantes, destaca que los investigadores tienen un control más fuerte sobre las variables para obtener los resultados deseados, lo que nos ayudará a determinar cuál modelo predictivo es más eficiente y con un menor margen de error [52]. Conjuntamente, al comparar estos resultados con datos históricos de partidos, se validará este modelo y, por consiguiente, el estudio.

e) Diseño de investigación

El diseño de investigación cuasi experimental se expone como la elección más apropiada para este estudio debido a la necesidad de tener un control significativo sobre las variables independientes, a pesar de no alcanzar un control total como en un experimento puro [53]. En consonancia con los objetivos de realizar un estudio comparativo y determinar el modelo predictivo más preciso con menor margen de error, este diseño facilita la aplicación de manipulaciones controladas en condiciones de la vida real. Su enfoque exploratorio permite descubrir patrones inesperados en datos históricos de partidos de fútbol, mientras que su naturaleza predictiva respalda la aplicación práctica de modelos desarrollados a eventos futuros.

f) Población

La población de esta investigación abarca todos los eventos de partidos de fútbol de las principales ligas del mundo especialmente de la liga inglesa que se hayan registrado y documentado a lo largo del tiempo. Esto incluye partidos de ligas nacionales e internacionales, abarcando una amplia gama de equipos, jugadores y condiciones de juego.

g) Muestra

En la presente investigación, la muestra estará compuesta por la selección representativa de partidos de fútbol de las principales ligas del mundo, con un enfoque particular en las temporadas 2021-2022 y 2022-2023 de la Premier League. La selección de estos partidos se realizará meticulosamente, se tomará en cuenta eventos que abarquen diversas condiciones, desde enfrentamientos directos históricos, partidos bajo diferentes condiciones climáticas, encuentros con cambios en la plantilla debido a fichajes, y variaciones en el nivel competitivo de los equipos. El propósito de esta selección es garantizar que la muestra sea lo suficientemente diversa y representativa, capturando así la complejidad y variedad de situaciones presentes en el fútbol profesional.

h) Técnicas de recolección de información

La recopilación documental es esencial para fundamentar teórica y empíricamente el desarrollo del modelo predictivo de fútbol. En este proceso, se busca identificar estudios previos que hayan abordado la predicción de resultados en partidos, algoritmos de modelado predictivo y factores clave en el rendimiento de equipos y jugadores. Asimismo, se realizará una revisión exhaustiva de diversos códigos fuente de proyectos relacionados con modelos predictivos en el ámbito del deporte, especialmente en el fútbol. Este análisis de códigos fuente permitirá obtener perspectivas prácticas sobre implementaciones exitosas, así como identificar posibles desafíos y soluciones aplicadas en proyectos similares.

Conjuntamente las entrevistas a expertos en el campo de la analítica de datos aplicada al fútbol aportarán perspectivas enriquecedoras y contextualizadas. La interacción directa con profesionales con experiencia permitirá profundizar en aspectos específicos no siempre capturados en la literatura o códigos existentes, como consideraciones tácticas, tendencias emergentes y desafíos prácticos en la implementación de modelos predictivos en el ámbito del fútbol. Así mismo, estas entrevistas proporcionarán una oportunidad para validar y enriquecer las hipótesis formuladas en la investigación, asegurando así que el modelo predictivo a desarrollar refleje de manera precisa la complejidad del entorno futbolístico.

i) Validez de las técnicas de recolección

La recopilación documental y el análisis de códigos fuente son técnicas fundamentales para construir una base sólida en el desarrollo del modelo predictivo de fútbol. La revisión documental permite acceder a la información teórica y empírica existente, identificando estudios previos sobre predicción de resultados, algoritmos de modelado y factores determinantes en el rendimiento. Conjuntamente, la exploración de códigos fuente de proyectos relacionados en el ámbito deportivo, específicamente en el fútbol, brinda una comprensión práctica de implementaciones exitosas, desafíos enfrentados y soluciones aplicadas.

La complementariedad de estas técnicas se evidencia con las entrevistas a expertos en analítica de datos aplicada al fútbol. Esta técnica aporta perspectivas enriquecedoras y contextualizadas, permitiendo profundizar en aspectos específicos no siempre capturados en la literatura o códigos existentes. La interacción directa con profesionales experimentados ofrece información valiosa sobre consideraciones tácticas, tendencias emergentes y desafíos prácticos en la implementación

de modelos predictivos en el ámbito futbolístico. Asimismo, estas entrevistas juegan un papel crucial en la validación y enriquecimiento de hipótesis, garantizando que el modelo predictivo desarrollado refleje con precisión la complejidad del entorno futbolístico.

j) Confiabilidad de las técnicas de recolección

La confiabilidad de las técnicas de recopilación de información propuestas se sustenta en su enfoque integral y complementario. La revisión documental, al centrarse en estudios previos, ofrece una base teórica sólida al identificar tendencias y enfoques probados en la predicción de resultados en el fútbol. La revisión de códigos fuente, por su parte, añade una dimensión práctica al proporcionar perspectivas valiosas de implementaciones exitosas y desafíos superados en proyectos similares.

La inclusión de entrevistas a expertos refuerza aún más la confiabilidad de la investigación. La interacción directa con profesionales experimentados permite la exploración de aspectos específicos y contextuales que pueden escapar a la documentación existente. Las perspectivas tácticas, las tendencias emergentes y los desafíos prácticos identificados durante estas entrevistas ofrecen una capa adicional de comprensión, enriqueciendo la investigación con información de primera mano.

La combinación de estas técnicas no solo amplía la variedad de fuentes de información, sino que también permite la validación cruzada de hallazgos. La convergencia de evidencia desde la literatura, códigos fuente y la experiencia de expertos fortalece la confiabilidad y robustez del modelo predictivo a desarrollar, asegurando que refleje de manera precisa la complejidad del entorno futbolístico.

k) Instrumentos de recolección de información

En el proceso de recolección de información, se hará uso de una fuente en línea. Esta herramienta incluye las bases de datos proporcionadas por el sitio web Footy Stats que gracias a la compra de una membresía premium permite la posibilidad de descargar conjuntos de datos históricos de todas ligas o competencias que hay en la actualidad, así mismo esta plataforma contiene datos tanto de partidos, jugadores y competencias. Adicionalmente, se hará uso de un sistema de gestión de bases de datos relacional de código abierto y potente como es PostgreSQL. Esta herramienta, ayudará a clasificar y ordenar la información de las bases de datos descargadas.

También se cuenta con la elaboración de un protocolo de revisión documental (véase anexo 1), el cual aportará criterios de búsqueda, inclusiones y exclusiones tomando bases de datos académicas, revistas científicas, informes de federaciones deportivas, registros estadísticos y otros documentos relacionados.

Junto con lo anterior se cuenta con un Checklist de análisis de códigos fuente (véase Anexo 2), que incluye criterios como eficiencia, modularidad, y enfoque en la predicción de resultados en base a repositorios en línea, proyectos de código abierto, y literatura técnica que comparta implementaciones de modelos predictivos en el ámbito deportivo.

Para finalizar se elaboró una guía de entrevista estructurada (véase Anexo 3), con preguntas diseñadas para explorar aspectos tácticos, tendencias emergentes, y desafíos prácticos en la implementación de modelos predictivos.

IV. RESULTADOS DE LA INVESTIGACIÓN

a) Análisis de conjunto de datos:

1. Consulta con el especialista en análisis de datos:

Se realizó una primera reunión con el asesor Ingeniero Luis Carlos Viteri, un experto en análisis de datos, quien proporcionó sugerencias sobre los algoritmos más adecuados para entrenar y evaluar el modelo de inteligencia artificial. Igualmente, orientó acerca de qué variables estadísticas podrían incrementar la exactitud del modelo predictivo, centrándose en elementos fundamentales como el rendimiento, los resultados anteriores y las condiciones de los partidos.

2. Evaluación del dataset proporcionado por Footy Stats:

Se empleó un conjunto de datos suministrado por la página web Footy Stats, la cual incluye datos históricos de los encuentros de fútbol de la Premier League desde el año 2000 hasta el 2022. Tras analizar estos datos, se detectaron irregularidades y vacíos. Algunos datos eran insuficientes o carecían de detalles particulares indispensables para el análisis y entrenamiento del modelo predictivo. Dentro de los datos ausentes, se determinaron las variables requeridas que posteriormente se emplearon para entrenar y evaluar los modelos creados.

3. Exploración de metodologías de web scraping:

Dada la falta de información en el conjunto de datos inicial, se decidió explorar técnicas de recolección de datos automatizada. Se encontró que la utilización de web scraping era la más óptima teniendo en cuenta la limitación de tiempo que se tenía. Esta metodología permite extraer información adicional desde 3 sitios web: página oficial de la Premier League, FBREF y FiveThirtyEight, con el objetivo de completar los datos faltantes. Mediante el uso de herramientas y bibliotecas especializadas en Python para web scraping con BeautifulSoup, se logró recopilar la información que se necesitaba de manera rápida y automatizada.

4. Ampliación y mejora del dataset:

Con la información obtenida a través del web scraping, se pudo completar las variables faltantes y obtener datos más específicos y detallados. Esto permitió mejorar la calidad del conjunto de datos inicial, dándole una base más sólida para el desarrollo del modelo predictivo. Además, los datos adicionales recolectados proporcionaron nuevas perspectivas para el análisis de los partidos y el comportamiento de los equipos.

Con la recopilación completa del dataset, se inició un proceso de investigación sobre diversas técnicas y algoritmos para el entrenamiento y testeo de los modelos predictivos, los cuales se detallarán en el siguiente ítem.

b) Selección e Implementación del Modelo Predictivo Óptimo para Pronósticos Futbolísticos

1. Consulta con el especialista en inteligencia artificial:

Se llevó a cabo una consulta con el Magister Héctor Andrés Mora Paz, un especialista en inteligencia artificial con amplia experiencia en la aplicación de modelos predictivos en una variedad de sectores como el deportivo, financiero y comercial. Su asesoría fue decisiva para definir el enfoque metodológico a seguir.

Durante la consulta, se exploraron diferentes enfoques y algoritmos de Machine Learning, incluyendo:

- Regresión lineal
- Modelos de clasificación
- Técnicas avanzadas como series temporales y ventanas móviles

2. Análisis de algoritmos de Machine Learning:

Se realizó un análisis completo utilizando un checklist estructurado para evaluar la calidad de los modelos, el rendimiento y la aplicabilidad en predicción. Tomando como base la revisión de diferentes algoritmos disponibles en repositorios públicos como: modelos usados para predecir el comportamiento de los mercados, predicción de diagnósticos médicos, sistemas de predicción utilizados en ventas y marketing los cuales se analizaron y fueron modificados para predecir los resultados en el ámbito deportivo. Considerando las diferencias inherentes en la naturaleza de los datos.

Durante la revisión, se hizo una comparación minuciosa de los modelos en función de varios factores clave, como:

- **Precisión de las predicciones:** Capacidad de predecir correctamente los resultados de partidos basados en los datos históricos.
- **Capacidad para manejar grandes volúmenes de datos:** Evaluar si el modelo podía procesar de manera eficiente grandes conjuntos de datos, como estadísticas históricas de varias temporadas o ligas completas.
- **Eficiencia computacional:** Importante para implementar los modelos en un software que pueda ofrecer predicciones en tiempo real.
- **Flexibilidad para nuevas variables:** La facilidad con la que el modelo podría incorporar nuevas variables o estadísticas, como alineaciones, lesiones, y cambios en la estrategia del equipo.

2.1 Utilización de algoritmos de clasificación:

A partir de la investigación de diferentes algoritmos de clasificación se obtuvo lo siguiente:

```
Nearest Neighbors 0.4315352697095436
Logistic Regression 0.533195020746888
Linear SVM 0.549792531120332
RBF SVM 0.4730290456431535
Gaussian Process 0.4190871369294606
Decision Tree 0.495850622406639
Random Forest 0.508298755186722
Neural Net 0.5394190871369294
AdaBoost 0.533195020746888
Naive Bayes 0.5145228215767634
QDA 0.483402489626556
```

Fig. 6. Precisiones de algoritmos de clasificación

Gracias a un “classification_report” y a los factores clave se mostró un reporte más completo de las métricas de clasificación, incluyendo precisión, recall y f-score para cada clase. Con estos resultados tan desalentadores, se tuvo que trabajar en la investigación de nuevos tipos de algoritmos.

2.2 Utilización de algoritmos de regresión:

Dado que la implementación inicial de algoritmos de clasificación no proporcionó un rendimiento óptimo para desarrollar un modelo robusto y preciso, se exploraron enfoques basados en algoritmos de regresión, como el RandomForestRegressor y el XGBRegressor.

Tras realizar múltiples experimentos, se determinó que el XGBRegressor ofrecía los mejores resultados en términos de precisión y eficiencia.

```
XGBoost - Mean Absolute Error (MAE): 0.8
XGBoost - Mean Squared Error (MSE): 1.2
XGBoost - R2 score: 79.00%
```

Fig. 7. Precisiones de algoritmos de regresión

Por lo tanto, se seleccionó este modelo como la principal herramienta para llevar a cabo la mayoría de las pruebas.

Sin embargo, dado a la naturaleza cuasi experimental de la investigación, se decidió experimentar para ver qué pasaba si se implementaba un enfoque basado en redes neuronales. Para abordar la complejidad del problema de predicción de resultados en el fútbol, se utilizó específicamente redes LSTM (Long Short-Term Memory). Las redes neuronales son especialmente útiles para detectar patrones complejos y relaciones no lineales en los datos, lo cual resulta decisivo en un entorno tan dinámico y multifactorial como el fútbol. Las LSTM, por su capacidad de manejar secuencias temporales y retener información a largo plazo, fueron seleccionadas para modelar adecuadamente la naturaleza temporal de los datos históricos de los partidos, permitiendo mejorar la precisión en los pronósticos de resultados.

2.3 Utilización de algoritmos de redes neuronales:

La implementación de redes neuronales LSTM, con su arquitectura específica que incluye la creación de múltiples capas de neuronas, fue fundamental para mejorar la capacidad del modelo de capturar patrones complejos y dinámicos inherentes a los datos históricos de los partidos de fútbol. Al incorporar neuronas adecuadamente distribuidas en las capas LSTM, el modelo pudo procesar secuencias de datos de manera más eficiente, otorgando un plus significativo a la calidad de las predicciones realizadas.

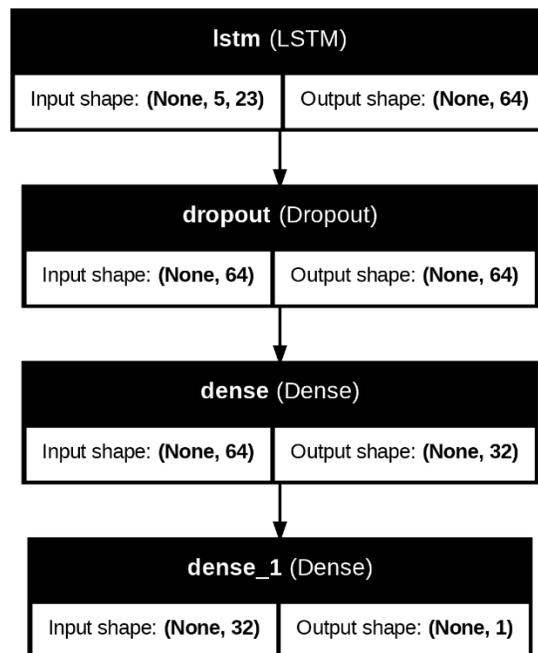


Fig. 8. Arquitectura de red neuronal LSTM

La anterior estructura presenta un modelo de red neuronal secuencial que se basa en una LSTM (Long Short-Term Memory), combinada con capas densas y una capa de Dropout.

Teniendo en cuenta que:

- **LSTM (Memoria a Largo y Corto Plazo):**

es una capa que se utiliza para procesar secuencias de datos. En este caso, la LSTM tiene 64 salidas y recibe entradas con la forma `(None, 5, 23)`, donde `None` significa que el tamaño del grupo puede variar, 5 es la longitud de la secuencia, y 23 son las características de cada entrada. Las LSTM son útiles para identificar patrones a lo largo del tiempo y se utilizaron para el análisis de series temporales y en el procesamiento de lenguaje.

- **Dropout:**

Esta técnica ayuda a que el modelo no se sobreentrene. Funciona desactivando aleatoriamente algunas neuronas durante el entrenamiento, lo que hace que este modelo sea más fuerte y menos propenso a memorizar los datos, se ha agregado una capa de Dropout después de la LSTM para disminuir el riesgo de sobreajuste.

- **Capa Densa (Fully Connected):**

Después de la capa Dropout, hay una capa densa con 32 neuronas. Esta capa recibe la salida de la LSTM (que tiene 64 neuronas) y se encarga de transformar la información aprendida anteriormente. Esta capa tiene funciones de activación ReLU para añadir complejidad al modelo.

- **Capade Salida (Dense_1):**

La última capa es otra capa densa con una sola neurona, que se utiliza para hacer la predicción final. Esto indica que el modelo puede estar resolviendo un problema de regresión (predicción de un número continuo) o clasificación binaria (donde se determina entre dos opciones).

Afortunadamente, la implementación de LSTM en el modelo ha resultado en una mejora notable del 83.5% en la precisión. Este incremento no solo indica un rendimiento superior en las predicciones, sino que también refleja la capacidad del modelo para aprender y generalizar patrones complejos en los datos de entrada. Esta mejora resalta la efectividad de las arquitecturas neuronales en la captura de relaciones temporales y características latentes, lo que contribuye a una mayor precisión en la clasificación y predicción de los resultados esperados. En consecuencia, el modelo se vuelve más robusto y confiable en su capacidad para abordar problemas de predicción.

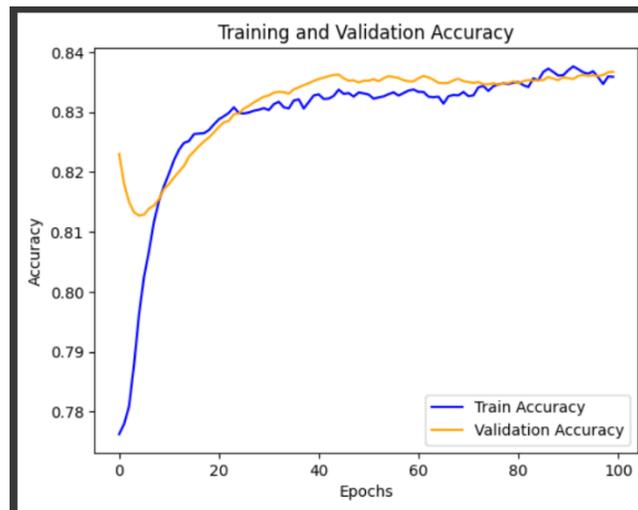


Fig. 9. Grafica de precisión de red neuronal LSTM

Gracias a esta implementación, se logró cumplir con el objetivo de identificar y emplear un modelo robusto y preciso que permitiera a los usuarios realizar pronósticos fundamentados en estadísticas históricas del fútbol, logrando así una futura implementación exitosa dentro del producto software.

3. Distribución de Poisson

En la muestra de predicciones, se optó por presentar los resultados del modelo en forma de probabilidades, en lugar de afirmar un resultado específico. Esta decisión se basa en la naturaleza inherentemente incierta de los eventos en un partido de fútbol, donde predecir un resultado exacto es complicado. Para abordar esta incertidumbre, se implementó la distribución de Poisson para modelar las probabilidades de los resultados.

El procedimiento general consiste en que, una vez que los modelos entrenados predicen los goles a favor y en contra para ambos equipos, se calculan las probabilidades de diversas combinaciones de resultados. La distribución de Poisson, que es adecuada para modelar eventos discretos como los goles en un partido, se utiliza para estimar la probabilidad de que cada equipo marque una cantidad específica de goles (por ejemplo, entre 0 y 5). A partir de estas predicciones, se exploran todas las combinaciones posibles de goles entre los equipos, calculando la probabilidad conjunta de cada marcador específico (como 1 - 0, 2-1, etc.).

Estas probabilidades se agrupan en tres categorías principales: victoria, empate y derrota, lo que permite cuantificar la probabilidad de cada resultado general del partido. De igual modo, se identifica el marcador más probable basándose en las probabilidades calculadas. Finalmente, las probabilidades de victoria, empate y derrota se normalizan para asegurar que sumen 100%, y esta información se presenta junto con el marcador más probable.

En este desarrollo, la distribución de Poisson se utilizó específicamente para modelar la cantidad de goles que cada equipo podría marcar. La fórmula de la distribución de Poisson es:

$$P(X=x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \quad x=0,1,\dots,n$$

Fig. 10. Formula de la distribución de Poisson

Nota: fuente <https://mundoestadistico.com/ley-de-poisson/> Mundo estadístico es una web desarrollada por alumnos del Bachillerato de Ciencias del I.E.S. Reyes Católicos de Vélez Málaga diseñada para aprender probabilidad y estadística

donde $P(X = k)$ representa la probabilidad de que ocurran exactamente k eventos (en este caso, goles), λ es la media esperada de goles predicha por el modelo, e es la constante de Euler y $k!$ es la factorial de k , que representa la cantidad exacta de goles. En el modelo de predicción, esta fórmula se aplica para calcular la probabilidad de que el equipo local marque i goles y el visitante j goles, abarcando así todas las combinaciones posibles de resultados.

4. Implementación en un producto software:

La implementación del producto de software se llevó a cabo utilizando el framework Django, reconocido por su robustez y capacidad para facilitar el desarrollo rápido de aplicaciones web. Django proporciona una estructura clara y modular, permitiendo a los desarrolladores enfocarse en la lógica del negocio mientras se beneficia de una serie de herramientas integradas para la gestión de bases de datos, autenticación de usuarios y administración de contenido. Esta elección de framework resultó en un desarrollo ágil y eficiente, alineado con las necesidades del proyecto.

Para guiar el desarrollo del software, se adoptó la metodología ágil SCRUM, que permite una gestión flexible y adaptativa de los requerimientos del proyecto. A través de esta metodología, se organizaron varias historias de usuario en cuatro Sprint, cada uno de los cuales abarcó diferentes funcionalidades y objetivos del producto. El uso de Sprint cortos facilitó la planificación y la entrega continua.

A lo largo de cada sprint, se inició con una reunión de planificación donde se seleccionaron y priorizaron las historias de usuario que aportarían mayor valor al proyecto. Durante el desarrollo, se llevaron a cabo reuniones semanales de seguimiento para monitorear el avance, identificar posibles bloqueos y garantizar que el trabajo estuviera alineado con los objetivos establecidos para el sprint, lo cual se puede observar en el siguiente product backlog.

Nº ID de la t...	Aa Nombre de la tarea	Estado	Responsable	Fecha límite	Prioridad	Etiquetas	Sprint	Proyecto
TAR-10	Corregir errores a la hora de c	Listo	Julian Chavez	18 de abril de 2024	Alta		Sprint 0	Sprint 0 Modulo noticias
TAR-9	Conectar API NewsApi.com	Listo	Julian Chavez	18 de abril de 2024	Alta		Sprint 0	Sprint 0 Modulo noticias
TAR-8	Realizacion de las interfaces	Listo	Julian Chavez	15 de abril de 2024	Media		Sprint 0	Sprint 0 Modulo noticias
TAR-18	Exportar modelo	Listo	Julian Chavez	25 de abril de 2024	Alta		Sprint 1	Sprint 1 Modulo Predicci
TAR-12	implementar El modelo en la	Listo	Julian Chavez	25 de abril de 2024	Alta		Sprint 1	Sprint 1 Modulo Predicci
TAR-11	Depurar el modelo, actualizar	Listo	Julian Chavez	25 de abril de 2024	Alta		Sprint 1	Sprint 1 Modulo Predicci
TAR-1	Realizar las interfaces	Listo	Julian Chavez	21 de abril de 2024	Media		Sprint 1	Sprint 1 Modulo Predicci
TAR-17	agregar área de las partes inti	Listo	Julian Chavez	25 de mayo de 2024	Media		Sprint 2	Sprint 3 Modulo informac
TAR-16	Implementar grafica de precic	Listo	Julian Chavez	25 de mayo de 2024	Alta		Sprint 2	Sprint 3 Modulo informac
TAR-15	redactar los párrafos acerca d	Listo	Julian Chavez	25 de mayo de 2024	Baja		Sprint 2	Sprint 3 Modulo informac
TAR-14	Realizar Interfaces	Listo	Julian Chavez	1 de mayo de 2024	Media		Sprint 2	Sprint 3 Modulo informac
TAR-32	Mostrar información en una t	Listo	Julian Chavez	8 de mayo de 2024	Alta		Sprint 2	Sprint 2 Mostrar Predicci
TAR-31	Creacion de base de datos	Listo	Julian Chavez	8 de mayo de 2024	Media		Sprint 2	Sprint 2 Mostrar Predicci
TAR-30	Hacer interfaces	Listo	Julian Chavez	1 de mayo de 2024	Alta		Sprint 2	Sprint 2 Mostrar Predicci
TAR-29	Sprint 2 Mostrar Predicciones	Listo	Julian Chavez				Sprint 2	
TAR-20	Realizar las Interfaces	Listo	Julian Chavez	30 de abril de 2024	Media		Sprint 3	Sprint 4 Login y registro

Fig. 11. Product Backlog del proyecto Score Prediction

La información detallada acerca de los sprints y los requerimientos funcionales están especificadas dentro del documento técnico de requisitos (véase Anexo 5)

Esta estructura de trabajo permitió responder rápidamente a cambios en los requisitos o en las prioridades del proyecto. Gracias a la implementación del framework Django y la metodología SCRUM, se logró desarrollar un producto de software funcional y adaptable, que satisface las expectativas del cliente y está preparado para futuras mejoras.

5. Proceso de Diseño y Mockups:

Antes de proceder con el desarrollo, se generaron **mockups** que sirvieron como guías visuales para el diseño de la interfaz de usuario. Estos mockups representaron las principales funcionalidades del sistema, como la selección de equipos, ingreso de datos estadísticos y visualización de predicciones.



Fig. 12. Mockup sección noticias

Las pantallas iniciales (véase Anexo 6) sirvieron como referencia durante el proceso de desarrollo y permitieron mantener una alineación clara con los requerimientos funcionales y las necesidades de los usuarios

6. Resultado final del producto

Siguiendo las recomendaciones del asesor Esp. Luis Carlos Viteri Rosero, se realizó un ajuste significativo en la paleta de colores utilizada durante el diseño de los *mockups*. Anteriormente, se empleaban tonos que, aunque estéticamente apropiados, no cumplían con ciertos estándares de accesibilidad, especialmente en lo que respecta a la inclusión de personas con daltonismo. En este sentido, se optó por seleccionar colores más adecuados al propósito del proyecto, garantizando que todos los elementos visuales fueran percibidos de manera efectiva por usuarios con esta condición. Se eliminaron ciertos tonos de verde, difíciles de distinguir para quienes presentan dicha alteración visual, y se adoptó una gama cromática más accesible, sin comprometer la estética ni la funcionalidad del diseño.

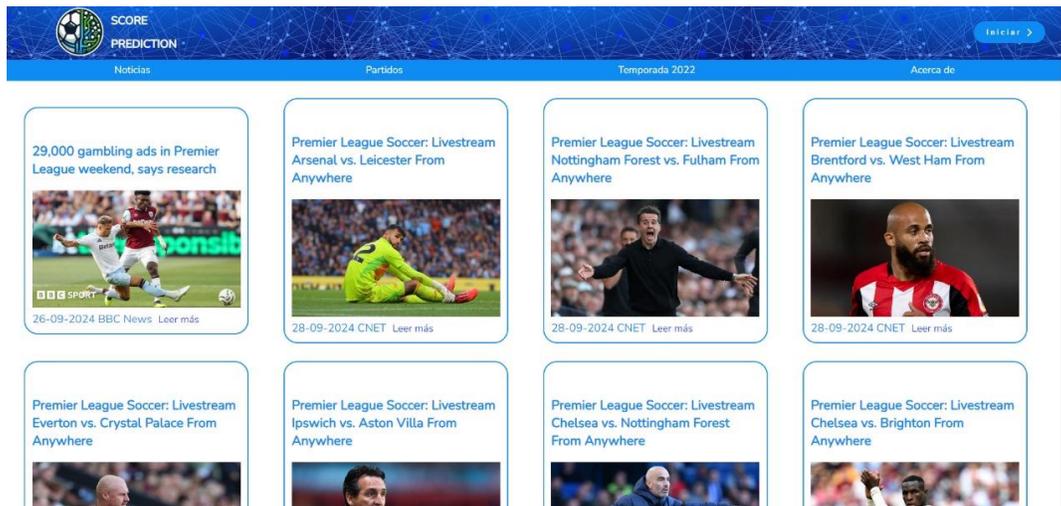


Fig. 13. Interfaz principal sitio web Score Prediction

El desarrollo del producto se completó con éxito, priorizando en todo momento la comodidad y la facilidad de uso para el usuario. Para garantizar una experiencia óptima, se elaboró un manual de usuario (véase Anexo 7), que proporciona una guía detallada sobre el uso del sistema, facilitando así una interacción fluida e intuitiva.

c) Verificación del modelo predictivo con resultados reales:

Para verificar los resultados del modelo predictivo con partidos reales de la temporada 2022-2023, se llevaron a cabo las siguientes tareas:

1. **Recolección de datos:** Se recopilaron los resultados de los encuentros de fútbol de la temporada 2022-2023, específicamente entre el 11 de marzo y el 7 de mayo del 2023. Esta información incluyó los equipos que se enfrentaron en cada partido, así como otros detalles estadísticos relevantes.
2. **Limpieza y depuración de datos:** Dado que la verificación del modelo se realizó en el segundo semestre de 2023, se optó por simplificar la información recolectada. Se llevó a cabo una depuración total del conjunto de datos, eliminando campos y detalles generales de los encuentros, y conservando únicamente los nombres de los equipos que participaron en cada encuentro y las respectivas fechas de los encuentros.
3. **Generación de predicciones:** Con la información depurada, se utilizó el modelo predictivo desarrollado para realizar pronósticos sobre los resultados de los partidos. El modelo procesó las estadísticas y generó predicciones basadas en datos históricos.
4. **Visualización en la página web:** Los resultados de las predicciones se integraron en una sección específica del sitio web, denominada "Temporada 2022". En esta sección, los usuarios pueden visualizar los pronósticos generados para cada partido de esa temporada y comparar estos pronósticos con los resultados reales.

The screenshot shows a web application titled 'SCORE PREDICTION' for the 'Temporada 2022'. The main content is a table of match predictions. The table has the following structure:

Local	Away	Date	Most likely score	1	X	2
BOU	LIV	Mar11	0 - 2	9.24%	15.98%	74.78%
CRY	MCI	Mar11	0 - 2	9.42%	16.96%	73.62%
EVE	BRE	Mar11	0 - 1	57.65%	22.67%	19.68%
FUL	ARS	Mar12	0 - 2	12.62%	20.15%	67.23%
LEE	BHA	Mar11	0 - 1	15.02%	21.15%	63.83%
LEI	CHE	Mar11	0 - 1	22.75%	25.13%	52.12%
MUN	SOU	Mar12	1 - 0	68.45%	19.75%	11.80%
NEW	WOL	Mar12	1 - 0	63.49%	21.41%	15.09%
WHU	AVL	Mar12	1 - 1	37.13%	27.08%	35.79%
AVL	BOU	Mar18	2 - 0	58.57%	23.01%	18.41%
BHA	CRY	Mar15	1 - 0	64.27%	20.99%	14.74%
CHE	EVE	Mar18	1 - 1	30.50%	34.25%	35.25%
LIV	FUL	Mar18	2 - 0	66.37%	20.34%	13.29%
MCI	WHU	Mar19	2 - 0	69.15%	19.33%	11.51%
BRE	SOU	Mar15	0 - 1	56.97%	22.82%	20.22%

Fig. 14. Interfaz temporada 2022 Score Prediction

V. ANÁLISIS DE RESULTADOS

Durante el análisis de los partidos de la temporada 2022-2023, se identificaron diversos casos que resaltan tanto los aciertos del modelo predictivo como aquellos encuentros en los que los pronósticos no coincidieron con los resultados reales, así como algunos casos particulares que merecen atención.

Uno de los partidos más destacados fue el enfrentamiento entre el Manchester City y el West Ham, celebrado el 3 de mayo del 2023. El modelo predictivo anticipó una victoria contundente para el City, con un marcador estimado de 2 - 0 y una probabilidad del 69.15% como se aprecia en la siguiente tabla.

TABLA I. RESULTADOS PREDICHOS PARTIDO 1

Fecha	Equipo Local	Equipo Visitante	Marcador probable	Probabilidades		
				V	E	D
03/05/2023	Manchester City	West Ham United	2 - 0	V	E	D
				69.15%	19.33%	11.51%

Este pronóstico se sustentó en el dominio estadístico del Manchester City, que se reflejó en el campo a través de la calidad ofensiva de su platilla. Aunque el marcador final no coincidió con el pronóstico, el modelo acertó al predecir que el Manchester City sería el equipo ganador. La probabilidad de victoria para el conjunto local se mantuvo alta, lo que se evidencia en la imagen siguiente.



Fig. 15. Resultado real Manchester City vs West Ham United

Nota: fuente https://www.espn.com.co/futbol/partido/_juegoId/638095/west-ham-united-manchester-city ESPN es un grupo de canales de televisión especializados en todo tipo de deportes

Otro acierto notable fue el partido entre el Arsenal y el Leeds United, disputado el 1 de abril. El modelo pronosticó una victoria de 2 - 0 a favor del Arsenal, con una probabilidad del 67.89% como se aprecia en la siguiente tabla.

TABLA II. RESULTADOS PREDICHOS PARTIDO 2

Fecha	Equipo Local	Equipo Visitante	Marcador probable	Probabilidades		
				V	E	D
01/04/2023	Arsenal	Leeds United	2 - 0	V	E	D
				67.89%	19.43%	12.68%

Aunque el modelo no acertó el marcador exacto, sí predijo correctamente la victoria del Arsenal sobre el Leeds United en el partido del 1 de abril, que finalizó 4 - 1. La probabilidad de triunfo para el equipo local fue alta, lo que se puede observar en la imagen siguiente.



Fig. 16. Resultado real Arsenal vs Leeds United

Nota: fuente https://www.espn.com.co/futbol/partido/_/juegoId/638102/leeds-united-arsenal ESPN es un grupo de canales de televisión especializados en todo tipo de deportes

Desde el inicio, el Arsenal mostró un claro dominio en el encuentro, respaldando así la validez del pronóstico respecto a su victoria.

No obstante, también hubo partidos en los que los resultados reales no coincidieron con las predicciones. Un ejemplo claro fue el encuentro entre Bournemouth y Liverpool el 11 de marzo. El modelo anticipó una victoria contundente para el Liverpool, con una probabilidad del 74.78% y un marcador estimado de 0 - 2, basándose en la superioridad estadística del equipo visitante como se aprecia en la tabla.

TABLA III. RESULTADOS PREDICHOS PARTIDO 3

Fecha	Equipo Local	Equipo Visitante	Marcador probable	Probabilidades		
				V	E	D
11/03/2023	Bournemouth	Liverpool	0 - 1	V	E	D
				9.24%	15.98%	74.78%

Sin embargo, Bournemouth sorprendió al lograr una victoria inesperada, lo que resalta la dificultad de predecir eventos imprevistos, como un contraataque efectivo o errores defensivos dando como resultado el marcador 1 – 0 como aprecia en la siguiente imagen.



Fig. 17. Resultado real AFC Bournemouth vs Liverpool

Nota: fuente https://www.espn.com.co/futbol/partido/_/juegoId/638082/liverpool-afc-bournemouth ESPN es un grupo de canales de televisión especializados en todo tipo de deportes

Otro encuentro notable fue el partido entre el Everton y el Tottenham, disputado el 3 de abril. El modelo había pronosticado una victoria para el Tottenham con una probabilidad del 59.50% y un resultado probable de 0-1 como se aprecia en la siguiente tabla.

TABLA IV. RESULTADOS PREDICHOS PARTIDO 4

Fecha	Equipo Local	Equipo Visitante	Marcador probable	Probabilidades		
				V	E	D
03/04/2023	Everton	Tottenham	0 - 1	V	E	D
				18.77%	22.10%	59.13%

Sin embargo, el Everton resistió y el encuentro terminó en empate 1 - 1, lo que resalta cómo factores no siempre cuantificables, como la solidez defensiva, pueden influir en los resultados. Esta situación pone de manifiesto las dificultades inherentes a la predicción de eventos inesperados en el fútbol. El resultado final se puede observar en la siguiente imagen.



Fig.18. Resultado real Everton vs Tottenham Hotspur

Nota: fuente https://www.espn.com.mx/futbol/partido/_juegoId/638107/tottenham-hotspur-everton ESPN es un grupo de canales de televisión especializados en todo tipo de deportes

Los partidos entre Brighton y Brentford, disputado el 1 de abril, y West Ham contra Aston Villa, celebrado el 12 de marzo, comparten una característica notable en los pronósticos del modelo predictivo: las probabilidades de victoria para ambos equipos estaban bastante equilibradas, lo que dificultaba prever un resultado claro.

En el caso del partido entre Brighton y Brentford, el modelo asignó una probabilidad de victoria del 30.49% a Brighton y del 34.49% a Brentford, con una probabilidad de empate del 25.02% como se puede apreciar en la siguiente tabla.

TABLA V. RESULTADOS PREDICHOS PARTIDO 5

Fecha	Equipo Local	Equipo Visitante	Marcador probable	Probabilidades		
				V	E	D
01/04/2023	Brighton	Brentford	1 - 1	V	E	D
				30.49%	25.02%	34.49%

Este equilibrio reflejaba la igualdad en el rendimiento de ambos equipos, sugiriendo un duelo muy parejo. El resultado final fue un empate 3 - 3, lo que refuerza la idea de que cuando las probabilidades son tan cercanas, el empate se convierte en un desenlace aceptable.



Fig. 18. Resultado real Brighton & Hove Albion vs Brentford

Nota: fuente https://www.espn.com.co/futbol/partido/_juegoId/638104/brentford-brighton-hove-albion ESPN es un grupo de canales de televisión especializados en todo tipo de deportes

De manera similar, en el encuentro entre West Ham y Aston Villa, las probabilidades de victoria también estaban muy equilibradas: 37.13% para West Ham y 35.79% para Aston Villa, con una probabilidad de empate aún más alta del 27.08% lo cual se puede apreciar en la siguiente tabla.

TABLA VI. RESULTADOS PREDICHOS PARTIDO 6

Fecha	Equipo Local	Equipo Visitante	Marcador probable	Probabilidades		
				V	E	D
01/04/2023	West Ham	Aston Villa	1 - 1	37.13%	27.08%	35.79%

Esto sugería un partido cerrado y competitivo. El resultado final fue también un empate 1 - 1, confirmando la predicción de un encuentro equilibrado y subrayando la importancia de considerar el empate en situaciones donde las probabilidades son tan similares.



Fig. 19. Resultado real West Ham United vs Aston Villa

Nota: fuente https://www.espn.com.co/futbol/partido/_/juegoId/638084/aston-villa-west-ham-united ESPN es un grupo de canales de televisión especializados en todo tipo de deportes

Durante el análisis de 70 encuentros de la temporada 2022-2023, se observó que el modelo predictivo logró aciertos en 57 partidos (véase Anexo 8), lo que representa una tasa de éxito del 81.42%. Sin embargo, también falló en 13 ocasiones, evidenciando la naturaleza impredecible del fútbol, donde factores como errores defensivos y decisiones arbitrales pueden alterar el resultado esperado.

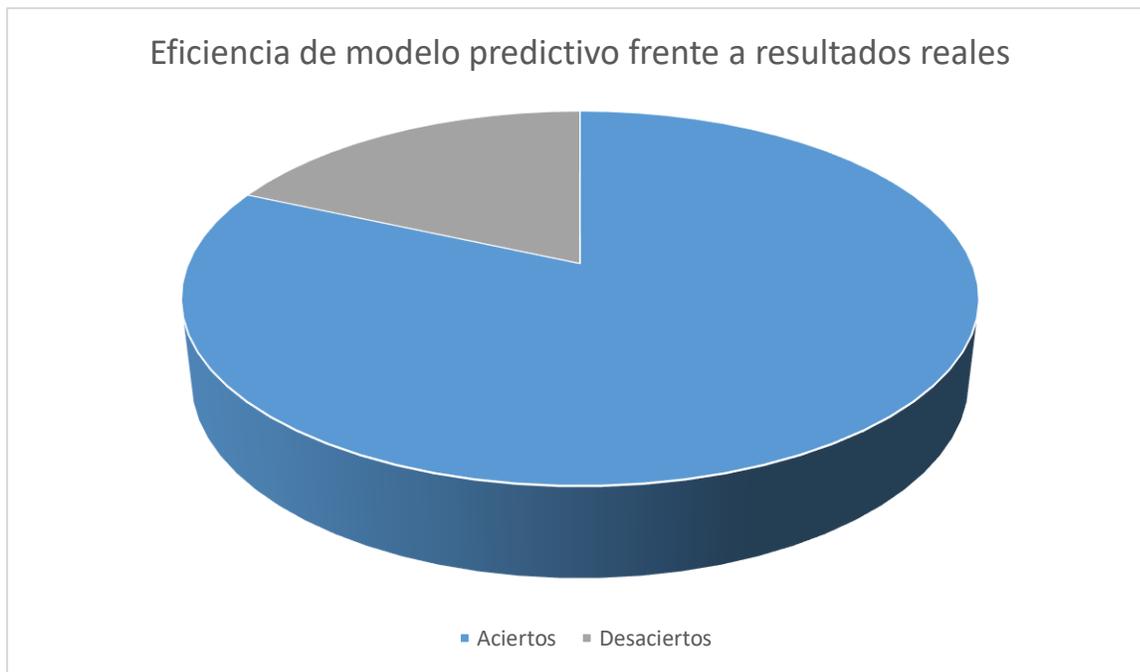


Fig. 20. Grafica de eficiencia del modelo predictivo

Este equilibrio entre aciertos y fallos resalta la utilidad del análisis de datos en la predicción, a pesar de las limitaciones inherentes a un deporte donde la sorpresa siempre está presente. La probabilidad de eficiencia del modelo ilustra esta dualidad, donde el conocimiento estadístico se enfrenta a la incertidumbre del juego.

De este modo, es importante señalar que, debido a lo impredecible que es el fútbol, este modelo suele tener más dificultades para acertar en la predicción de un marcador exacto. Sin embargo, lo relevante es que, en la mayoría de los casos, las probabilidades de victoria, derrota o empate que arroja el modelo ofrecen una base sólida de confianza. Esto nos permite tener una visión más amplia de lo que podría suceder en el partido antes de entrar en detalles específicos.

A pesar de estas dificultades, el modelo logró acertar el marcador exacto en 14 ocasiones a lo largo de los partidos analizados. Este nivel de precisión en predicciones exactas es notable, considerando la naturaleza altamente impredecible del fútbol, donde incluso factores menores pueden influir significativamente en el marcador final. Acertar el marcador exacto en múltiples ocasiones indica que el modelo capta tendencias clave en el rendimiento de los equipos, tales como la solidez defensiva, las tácticas de ataque, y las condiciones particulares del encuentro.

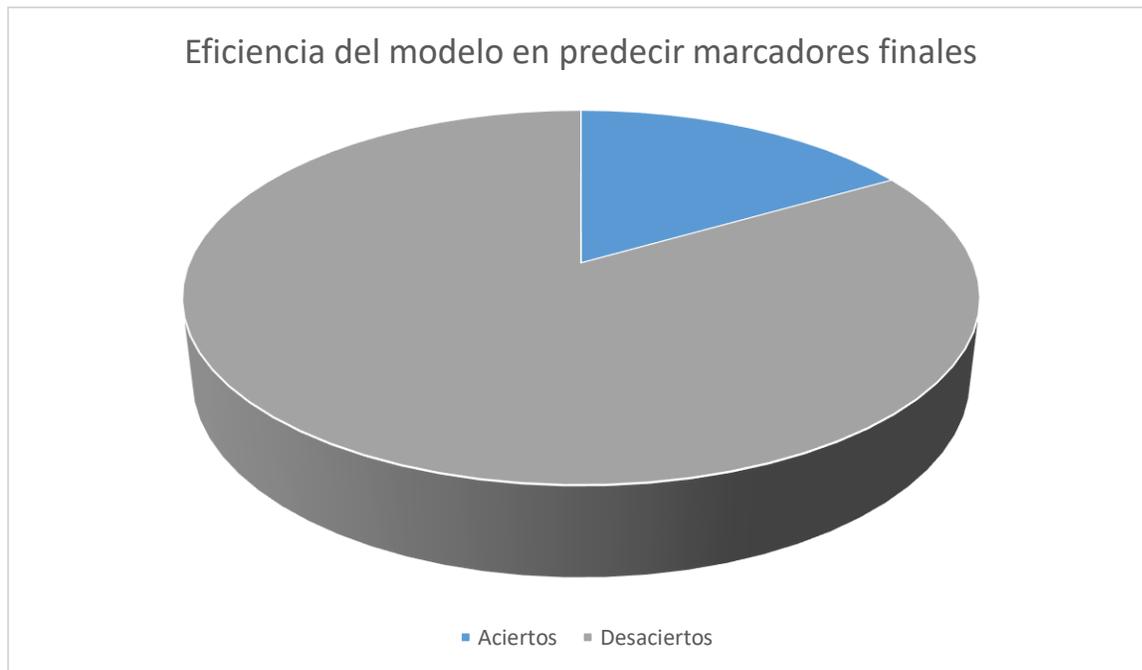


Fig. 21. Grafica de eficiencia de marcadores finales

Estos 14 aciertos proporcionan un respaldo tangible al modelo, ya que reflejan una comprensión de patrones y probabilidades que van más allá de simples conjeturas. Aunque acertar el marcador exacto en el fútbol sigue siendo un reto debido a la cantidad de variables en juego, estas predicciones certeras destacan la capacidad del modelo para identificar y capitalizar eventos específicos.

Esto es significativo, pero cabe mencionar que las ganancias para los apostadores no siempre son tan altas, ya que las casas de apuestas ajustan sus cuotas en función de los datos de los partidos más recientes y otros factores contextuales, lo que hace que las oportunidades de obtener beneficios notables sean limitadas. Sin embargo, en aproximadamente el 20% de los partidos, el modelo ha logrado predecir con precisión el marcador final, asegurando así ganancias notorias para quienes apostaron basándose en sus pronósticos. Esta capacidad de acierto en una fracción considerable de partidos refuerza la utilidad del modelo como herramienta predictiva.

Los resultados obtenidos en el análisis de los partidos de la temporada 2022-2023 evidencian que la hipótesis de investigación fue validada. El modelo predictivo desarrollado, basado en un estudio comparativo y análisis avanzado de algoritmos de analítica de datos y aprendizaje automático, logró un acierto del 81.42% en la predicción de resultados, superando en precisión y eficiencia a los modelos predictivos del estado del arte. Como el presentado en la investigación titulada "Modelamiento predictivo del ganador de un partido de fútbol de la categoría A del fútbol profesional colombiano usando aprendizaje de máquina" [14], que evaluó algoritmos como la regresión logística multiclase, el algoritmo Random Forest y las máquinas de soporte vectorial.

En contraste, el modelo desarrollado en esta investigación se distinguió por su capacidad de equilibrar la precisión entre las tres categorías posibles: victoria local, empate y derrota. Mientras que los modelos del estado del arte presentaron limitaciones específicas como una baja precisión en la predicción de empates (49.73%) con la regresión logística, o la inconsistencia en la clasificación de victorias visitantes mediante Random Forest, que logró identificar correctamente solo 29 de 87 partidos reales, el enfoque de la investigación en curso logró reducir estos sesgos y mejorar significativamente la precisión general. igualmente, el modelo superó los resultados obtenidos por las máquinas de soporte vectorial, que, aunque alcanzaron un accuracy elevado para victorias locales (68.01%), mostraron un desempeño desigual en las demás categorías.

En este contexto, la implementación de redes neuronales de memoria a largo plazo (LSTM) en el modelo ha resultado en una mejora notable del (81.42%) en la precisión de las predicciones de resultados. Este incremento no solo indica un rendimiento superior, sino que también refleja la capacidad del modelo para aprender y generalizar patrones complejos a partir de los datos de

entrada. Esta mejora resalta la efectividad de las arquitecturas neuronales para capturar relaciones temporales y características latentes, lo que contribuye a una mayor precisión en la clasificación y predicción de los resultados esperados. Como consecuencia, el modelo se vuelve más robusto y confiable para abordar problemas de predicción, destacándose frente a los métodos tradicionales, como se puede evidenciar en la gráfica que se presenta a continuación

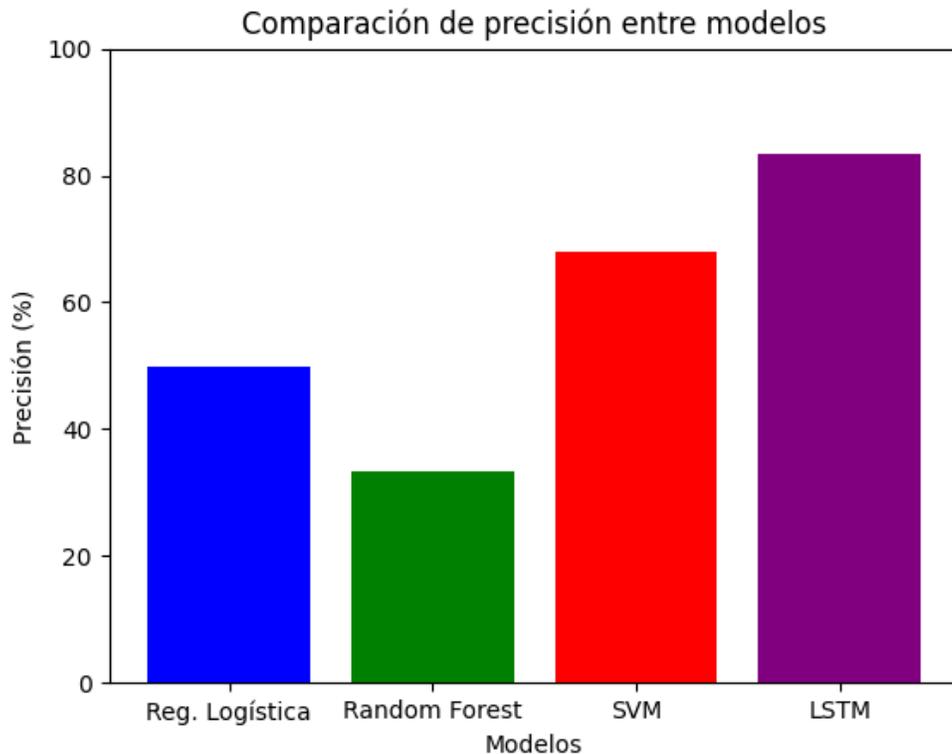


Fig. 22. Comparación de modelos de IA

Este logro se traduce en un menor margen de error, lo que subraya la capacidad del modelo para identificar tendencias clave en el rendimiento de los equipos y ofrecer predicciones confiables tanto en términos de resultados generales como de marcadores específicos. Esto es especialmente relevante en un deporte tan complejo y dinámico como el fútbol.

CONCLUSIONES

1. Análisis de conjunto de datos:

Aunque se logró establecer un repositorio de datos sólido y robusto, la recolección de información y el uso de técnicas de web scraping resultaron ser tareas desgastantes, ya que demandaron una supervisión constante. No obstante, este esfuerzo fue esencial para asegurar la calidad y representatividad de los datos, lo que a su vez proporcionó una base confiable para la investigación y la anticipación de resultados en el ámbito del fútbol.

2. Selección e implementación del modelo predictivo óptimo para pronósticos

Se identificó un modelo que demostró un rendimiento satisfactorio y logró mejoras significativas en comparación con el estado del arte, especialmente en la precisión de los pronósticos generales de resultados. Aunque las pruebas de verificación confirmaron que el modelo tiene una alta tasa de aciertos, aún hay áreas en las que se pueden realizar ajustes para optimizar su precisión, particularmente en escenarios que presentan variaciones imprevistas.

3. Verificación del modelo predictivo con resultados reales

El modelo demostró un buen desempeño en la predicción de victorias, empates y derrotas, aunque enfrenta desafíos al intentar anticipar el marcador exacto. Este resultado era predecible, dado que los factores impredecibles inherentes al fútbol pueden influir considerablemente en el resultado final. Por lo tanto, se sugiere que el modelo es más adecuado para proporcionar probabilidades de resultados generales en lugar de pronósticos precisos de marcadores. Esta capacidad para evaluar resultados generales permite a los usuarios tomar decisiones más informadas, aunque el enfoque en los marcadores exactos puede requerir ajustes adicionales para mejorar su precisión.

RECOMENDACIONES

1. Mejorar la recolección de datos:

se recomienda implementar herramientas automatizadas de web scraping que reduzcan la necesidad de supervisión manual, lo que permitirá una recolección más eficiente y menos agotadora. Conjuntamente se recomienda explorar y utilizar múltiples fuentes de datos para enriquecer el repositorio, garantizando así una mayor representatividad y calidad en la información recopilada. En este sentido, se sugiere extraer datos directamente desde nuevas APIs que están brindando información sobre las ligas de fútbol a nivel mundial de manera más cómoda y accesible, lo que facilitará aún más la obtención de datos relevantes y actualizados.

2. Mejora del modelo y su capacidad predictiva:

Se recomienda un ajuste continuo de los parámetros utilizados, lo que mejorará su precisión en situaciones con alta variabilidad. Esto incluye la integración de nuevas variables que influyan en los resultados, como el rendimiento individual de cada jugador durante el partido, así como factores adicionales que impacten los resultados, tales como cambios en la alineación, lesiones, condiciones climáticas y el estado del campo. Para asegurar que el modelo se mantenga relevante y efectivo a lo largo del tiempo, es fundamental implementar un proceso de validación continua utilizando datos recientes. Además, realizar simulaciones que incluyan variables inesperadas permitirá evaluar cómo responde el modelo ante situaciones no anticipadas y hacer ajustes en consecuencia, mejorando así significativamente la capacidad predictiva del sistema.

3. Priorizar pronósticos de resultados generales:

Es recomendable continuar enfocándose en ofrecer pronósticos de resultados generales (victorias, empates y derrotas) en lugar de marcadores exactos, ya que esta área ha demostrado un rendimiento más sólido del modelo. Para facilitar la interpretación y la toma de decisiones basada en datos, se sugiere desarrollar herramientas de visualización que presenten las probabilidades de resultados de manera clara y accesible para los usuarios interesados.

BIBLIOGRAFÍA

- [1] Villena, J y otros., «Inteligencia en redes de comunicaciones. Minería de datos.», Universidad Carlos III de Madrid. España.
- [2] M. Conde, «¿Se puede predecir el fútbol?» Accedido: 6 de septiembre de 2023. [En línea]. Disponible en: <https://bigdatasports.media/2021/02/28/se-puede-predecir-el-futbol/>
- [3] «Tipsters estafadores o cómo vender pronósticos deportivos de apuestas que son humo», www.elsaltodiario.com. Accedido: 6 de septiembre de 2023. [En línea]. Disponible en: <https://www.elsaltodiario.com/casas-de-apuestas/tipsters-estafadores-vender-pronosticos-deportivos-apuestas-humo>
- [4] «Olocip | Inteligencia Artificial en el Fútbol Profesional», Olocip. Accedido: 6 de septiembre de 2023. [En línea]. Disponible en: <https://olocip.com/futbol/>
- [5] A. D. Blaikie, J. A. David, y G. J. Abud, «NFL & NCAA Football Prediction using Artificial Neural Networks».
- [6] «Memoria TFG.pdf». Accedido: 1 de octubre de 2023. [En línea]. Disponible en: <https://academica-e.unavarra.es/bitstream/handle/2454/43454/Memoria%20TFG.pdf?sequence=1&isAllowed=y>
- [7] C. M. S. M. Barra, G. R. Cabezas, y J. T. Gajardo, «Propuesta de modelo predictivo de riesgo de lesión en base a descriptores anatómicos y funcionales que se relacionan con la inestabilidad articular en rodilla y tobillo en jugadores de baloncesto no profesionales (Proposal of a predictive model of injury», *Retos*, vol. 39, pp. 257-263, ene. 2021, doi: 10.47197/retos.v0i39.76987.
- [8] A. B. Extremera, A. Granero-Gallegos, J. A. S. Fuentes, y M. M. Molina, «Modelo predictivo de la importancia y utilidad de la Educación Física», *Cuad. Psicol. Deporte*, vol. 14, n.º 2, Art. n.º 2, may 2014.
- [9] J. M. Pineda, «Modelos predictivos en salud basados en aprendizaje de maquina (machine learning)», *Rev. Médica Clínica Las Condes*, vol. 33, n.º 6, Art. n.º 6, nov. 2022, doi: 10.1016/j.rmcl.2022.11.002.
- [10] D. G. M. Vilca y I. A. P. Torres, «Tesis: “Modelo predictivo Machine Learning aplicado a análisis de datos Hidrometeorológicos para un SAT en Represas”».
- [11] A. C. Martín-Romero y M. J. G. Abad, «TRABAJO FIN DE CARRERA Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas.».
- [12] A. Jaramillo y H. P. P. Arias, «Aplicación de técnicas de minería de datos para determinar las interacciones de los estudiantes en un entorno virtual de aprendizaje», *Rev. Tecnológica - ESPOL*, vol. 28, n.º 1, Art. n.º 1, ago. 2015, Accedido: 10 de octubre de 2023. [En línea]. Disponible en: <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/351>
- [13] A. R. Pérez, «Análisis estadístico del data warehouse del R.C. Deportivo de La Coruña y uso de minería de datos aplicada al scouting».
- [14] E. F. A. Roa, «MODELAMIENTO PREDICTIVO DEL GANADOR DE UN PARTIDO DE FÚTBOL DE LA CATEGORÍA A, DEL FÚTBOL PROFESIONAL COLOMBIANO USANDO APRENDIZAJE DE MÁQUINA».
- [15] P. A. F. Polanco, «PROPUESTA DE MODELO PREDICTIVO APLICADO AL RENDIMIENTO EN NADADORES COLOMBIANOS», 2017.

- [16] L. M. M. Arias y S. M. Vélez, «Modelo de clasificación multiclases para la predicción de apuestas deportivas».
- [17] A. M. E. Gómez, «UNIVERSIDAD DE LOS ANDES FACULTAD DE INGENIERIA DEPARTAMENTO DE INGENIERIA INDUSTRIAL».
- [18] L. M. Zúñiga Quiñones y H. S. Aguirre Mayorga, «Desarrollo de un modelo predictivo para un sistema de manufactura orientado a la industria 4.0», Pontificia Universidad Javeriana, 2021. doi: 10.11144/Javeriana.10554.55631.
- [19] L. C. Rodríguez y K. L. A. Ortiz, «Machine Learning en empresas Pyme de Colombia en el periodo comprendido».
- [20] Y. Pinedo y C. R. V. Benavides, «MINERÍA DE DATOS APLICADA ANÁLISIS MAKRO», 2017.
- [21] M. D. B. Moreno y H. T. M. Pianda, «ANÁLISIS DE LOS FACTORES ASOCIADOS AL RIESGO DE DESERCIÓN EN LOS ESTUDIANTES DEL PROGRAMA LICENCIATURA EN INFORMÁTICA DE LA UNIVERSIDAD DE NARIÑO, PARA LA CONSTRUCCIÓN Y SISTEMATIZACIÓN DE UN MODELO ESTADÍSTICO PREDICTIVO».
- [22] S. V. E. Madroñero y J. H. E. Tulcán, «IDENTIFICACIÓN DE CARACTERÍSTICAS DE RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE PREGRADO DE LA UNIVERSIDAD DE NARIÑO ENTRE LOS AÑOS 2010 Y 2014 USANDO MINERÍA DE DATOS», 2010.
- [23] R. M. Martínez y F. P. Cruz, «PROYECTO FIN DE CARRERA».
- [24] A. Silva, F. S. Bañuelos, J. Garganta, y M. T. Anguera, «Patrones De Juego En El Fútbol De Alto Rendimiento. Análisis Secuencial Del Proceso Ofensivo En El Campeonato Del Mundo Corea-Japón 2002», *Cult. Cienc. Deporte*, vol. 1, n.º 2, Art. n.º 2, 2005.
- [25] «Conoce las principales distribuciones de probabilidad | Conexión ESAN». Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://www.esan.edu.pe/conexion-esan/conoce-las-principales-distribuciones-de-probabilidad>
- [26] «ÍNDICE. VARIABLES PREDICTORAS DE ÉXITO EN FÚTBOL: ANÁLISIS DE LOS PARTIDOS DE PRETEMPORADA MEDIANTE GPS Vargas, A., Urkiza, I. y Gil, S. - PDF Free Download». Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://docplayer.es/85033444-Indice-variables-predictoras-de-exito-en-futbol-analisis-de-los-partidos-de-pretemporada-mediante-gps-vargas-a-urkiza-i-y-gil-s.html>
- [27] «¿Existen un conjunto de características comunes y propias de las jugadoras de fútbol?» Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://efdeportes.com/efd92/jugad.htm>
- [28] R. Caballero, T. Gómez, y R. Sala, «durante la temporada mejora el rendimiento del equipo?», 2009.
- [29] A. Ali, «Las 10 mejores técnicas de minería de datos», Astera. Accedido: 12 de octubre de 2023. [En línea]. Disponible en: <https://www.astera.com/es/tipo/blog/Las-10-mejores-t%C3%A9cnicas-de-miner%C3%ADa-de-datos/>
- [30] D. D. J. Fernández, I. F. García, y M. A. S. Vega, «Seguimiento de jugadores en partidos de fútbol mediante procesado de vídeo».
- [31] A. B. Serrano, «Estudio de la Liga Española, recopilación de datos y predicción usando modelos estadísticos y de aprendizaje automático».
- [32] «Clustering: qué es y cuál es su aplicación en Big Data», UNIR. Accedido: 22 de octubre de 2023. [En línea]. Disponible en: <https://www.unir.net/ingenieria/revista/clustering/>

- [33] «COMTEL-2014-177-184.pdf». Accedido: 22 de octubre de 2023. [En línea]. Disponible en: <http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/680/COMTEL-2014-177-184.pdf?sequence=1>
- [34] «Árboles de decisión: qué son y cuál es su uso en Big Data», UNIR. Accedido: 22 de octubre de 2023. [En línea]. Disponible en: <https://www.unir.net/ingenieria/revista/arboles-de-decision/>
- [35] C. Ortega, «Análisis de regresión: Qué es, tipos y cómo realizarlo», QuestionPro. Accedido: 22 de octubre de 2023. [En línea]. Disponible en: <https://www.questionpro.com/blog/es/analisis-de-regresion/>
- [36] «CogniFit», Memoria a Largo Plazo (MLP) - Capacidad Cognitiva - CogniFit. Accedido: 22 de octubre de 2023. [En línea]. Disponible en: <https://www.cognifit.com/co/memoria-a-largo-plazo>
- [37] «IBM Documentation». Accedido: 22 de octubre de 2023. [En línea]. Disponible en: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model>
- [38] «¿Qué es el desarrollo de software? | IBM». Accedido: 19 de noviembre de 2023. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/software-development>
- [39] Atlassian, «¿Qué es scrum? [+ Cómo empezar]», Atlassian. Accedido: 19 de noviembre de 2023. [En línea]. Disponible en: <https://www.atlassian.com/es/agile/scrum>
- [40] «¿Qué es un lenguaje de programación? | Desarrollar Inclusión». Accedido: 19 de noviembre de 2023. [En línea]. Disponible en: <https://desarrollarinclusion.cilsa.org/tecnologia-inclusiva/que-es-un-lenguaje-de-programacion/>
- [41] «¿Qué es Python? - Explicación del lenguaje Python - AWS», Amazon Web Services, Inc. Accedido: 19 de noviembre de 2023. [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/python/>
- [42] «¿Qué es el scraping de datos?» Accedido: 2 de octubre de 2024. [En línea]. Disponible en: <https://www.cloudflare.com/es-es/learning/bots/what-is-data-scraping/>
- [43] Daniel, «Beautiful Soup : ¿cómo aprender a hacer web scraping en Python?», Formación en ciencia de datos | DataScientest.com. Accedido: 2 de octubre de 2024. [En línea]. Disponible en: <https://datascientest.com/es/beautiful-soup-aprender-web-scraping>
- [44] E. P. Bel, «El impacto de la posesión en el fútbol moderno».
- [45] «El contraataque es más efectivo en el fútbol cuando se va ganando un partido», iDescubre. Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://idescubre.fundaciondescubre.es/noticias/un-estudio-senala-que-cuando-un-equipo-de-futbol-va-ganando-en-un-partido-suele-aumentar-la-efectividad-de-su-contraataque-2/>
- [46] C. S, «Tácticas del fútbol 11: defensivas y ofensivas | ILERNA», Blog de ILERNA Online. Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://www.ilerna.es/blog/ilerna-formacion/tacticas-futbol-11/>
- [47] T. Ruuska, «Fútbol y dinero», teemuruuska.com. Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://teemuruuska.com/es/futbol-y-dinero/>
- [48] T. E. Journal, «El gran negocio del futbol», The Economy Journal. Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://www.theeconomyjournal.com/texto-diario/mostrar/1525487/gran-negocio-futbol>
- [49] F. B. Ríos, «PARADIGMAS Y PERSPECTIVAS TEÓRICO-METODOLÓGICAS EN EL ESTUDIO DE LA ADMINISTRACIÓN».
- [50] J. Inche M., Y. Andía C., H. Huamanchumo V., M. López O., J. Vizcarra M., y G. Flores C., «PARADIGMA CUANTITATIVO: Un Enfoque Empírico y Analítico», *Ind. Data*, vol. 6, n.º 1, Art. n.º 1, mar. 2014, doi: 10.15381/idata.v6i1.5938.

- [51] «¿Qué es el método científico y para qué se emplea? - Sus 5 pasos», UNIR México. Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://mexico.unir.net/derecho/noticias/metodo-cientifico/>
- [52] A. Velázquez, «¿Qué es la investigación experimental?», QuestionPro. Accedido: 21 de octubre de 2023. [En línea]. Disponible en: <https://www.questionpro.com/blog/es/investigacion-experimental/>
- [53] A. Parra, «¿Qué es la investigación cuasi experimental?», QuestionPro. Accedido: 5 de diciembre de 2023. [En línea]. Disponible en: <https://www.questionpro.com/blog/es/investigacion-cuasi-experimental/>

ANEXOS

ANEXO 1: Formato de Protocolo para Revisión Documental en el Desarrollo de un Modelo Predictivo de Fútbol

Formato de Protocolo para Revisión Documental en el Desarrollo de un Modelo Predictivo de Fútbol

1. Título de la Revisión Documental:

(Incluir el título descriptivo y específico que refleje el alcance de la revisión.)

2. Objetivos de la Revisión:

- a) Identificar estudios previos sobre la predicción de resultados en partidos de fútbol.
- b) Evaluar algoritmos de modelado predictivo aplicados al ámbito deportivo.
- c) Analizar factores clave en el rendimiento de equipos y jugadores.

3. Criterios de Búsqueda:

- a) Palabras Clave:
- b) (Listar las palabras clave relevantes, como "modelado predictivo de fútbol", "algoritmos deportivos", "factores de rendimiento futbolístico", etc.)
- c) Bases de Datos:
- d) (Especificar las bases de datos a utilizar, como PubMed, IEEE Xplore, Scopus, entre otras.)
- e) Rango Temporal:
- f) (Establecer el período de tiempo a considerar, por ejemplo, estudios publicados en los últimos 10 años.)

4. Criterios de Inclusión:

- a) Estudios empíricos y revisiones sistemáticas relacionadas con la predicción en partidos de fútbol.
- b) Artículos que presenten algoritmos y modelos predictivos aplicados al ámbito deportivo.
- c) Investigaciones que aborden factores determinantes en el rendimiento de equipos y jugadores.

5. Criterios de Exclusión:

- a) Estudios no relacionados directamente con la predicción en el contexto del fútbol.
- b) Artículos que no proporcionen información detallada sobre algoritmos de modelado predictivo.
- c) Publicaciones duplicadas o de baja calidad metodológica.

6. Proceso de Selección:

- a) Dos revisores independientes realizarán la selección inicial de estudios.
- b) En caso de desacuerdo, se llevará a cabo una discusión para alcanzar consenso.

7. Proceso de Extracción de Datos:

- a) Desarrollar una tabla para la extracción sistemática de datos, incluyendo autor, año, metodología, resultados clave, etc.

8. Evaluación de la Calidad de los Estudios:

- a) Utilizar criterios predefinidos para evaluar la calidad metodológica de los estudios seleccionados.

9. Análisis y Síntesis de los Resultados:

- a) Realizar un resumen narrativo y/o un análisis temático de los resultados encontrados.

1. Actualización del Protocolo:

- a) En caso de encontrar nuevas direcciones o enfoques durante la revisión, actualizar el protocolo según sea necesario.

ANEXO 2: Checklist de Análisis de Códigos Fuente para Modelos Predictivos de Fútbol:

Este checklist se utiliza para evaluar la calidad y adecuación de los códigos fuente relacionados con modelos predictivos en el ámbito del fútbol. Cada criterio se valora en una escala de 1 a 5, donde 1 es insatisfactorio y 5 es excelente.

Eficiencia:

- 1 - Pobre: El código presenta problemas de rendimiento y ejecución.
- 2 - Regular: Algunas áreas de mejora en términos de eficiencia.
- 3 - Aceptable: Eficiencia general adecuada, con posibles áreas de optimización.
- 4 - Bueno: Eficiencia sólida, sin problemas significativos.
- 5 - Excelente: Implementación altamente eficiente.

Modularidad:

- 1 - Pobre: Falta de estructuración y organización en módulos.
- 2 - Regular: Algunas partes del código podrían organizarse mejor.
- 3 - Aceptable: El modularidad es aceptable, pero hay margen para mejorar.
- 4 - Bueno: Buena estructura modular, facilitando la comprensión.
- 5 - Excelente: El código está altamente modularizado y bien organizado.

Enfoque en la Predicción de Resultados:

- 1 - Pobre: El código no está claramente orientado hacia la predicción de resultados.
- 2 - Regular: Algunas funciones están relacionadas con la predicción, pero la orientación es difusa.
- 3 - Aceptable: El código tiene secciones dedicadas a la predicción, pero la claridad puede mejorar.
- 4 - Bueno: El enfoque en la predicción es claro y específico.
- 5 - Excelente: El código está altamente especializado en la predicción de resultados.

Documentación:

- 1 - Pobre: Falta de documentación o documentación incomprensible.
- 2 - Regular: Documentación mínima o inconsistente.
- 3 - Aceptable: Documentación clara, pero podría ser más detallada.
- 4 - Bueno: Documentación completa y comprensible.
- 5 - Excelente: Documentación detallada y fácil de seguir.

Comentarios en el Código:

- 1 - Pobre: Falta de comentarios o comentarios irrelevantes.
- 2 - Regular: Comentarios mínimos o poco informativos.
- 3 - Aceptable: Comentarios adecuados, pero no exhaustivos.
- 4 - Bueno: Comentarios útiles y bien distribuidos.
- 5 - Excelente: Comentarios detallados y explicativos en todo el código

ANEXO 3: Guía de Entrevistas para Profesionales en Analítica de Datos Aplicada al Fútbol:

1. Introducción:

a) Presentación:

"Gracias por participar en esta entrevista. ¿Podría comenzar presentándose y describiendo brevemente su experiencia en analítica de datos en el fútbol?"

2. Aspectos Tácticos:

a) Enfoque Táctico:

- "Desde su perspectiva, ¿cómo la analítica de datos ha influido en la evolución de las estrategias tácticas en el fútbol?"

b) Uso de Datos en Decisiones Tácticas:

- "¿Cómo los entrenadores y analistas utilizan los datos para tomar decisiones tácticas durante los partidos y entrenamientos?"

3. Tendencias Emergentes:

a) Tecnologías Emergentes:

- "¿Qué tecnologías emergentes o nuevas metodologías en la analítica de datos cree que están transformando la forma en que se aborda el análisis en el fútbol?"

b) Variables de Interés Emergentes:

- "¿Hay variables específicas o métricas emergentes que esté viendo como cada vez más relevantes en la analítica de datos aplicada al fútbol?"

4. Desafíos Prácticos en Implementación:

a) Integración de Datos:

- "¿Cuáles son los desafíos comunes en la integración de datos de diversas fuentes para análisis predictivo en el contexto del fútbol?"

b) Interpretación y Comunicación:

- "¿Cómo se abordan los desafíos relacionados con la interpretación y comunicación de los resultados de modelos predictivos a entrenadores y jugadores?"

5. Impacto en el Rendimiento:

a) Experiencias Exitosas:

- "¿Podría compartir alguna experiencia en la que la analítica de datos haya tenido un impacto significativo en el rendimiento de un equipo o jugador?"

b) Lecciones Aprendidas:

- "¿Cuáles son algunas lecciones clave que ha aprendido sobre la aplicación efectiva de la analítica de datos en el fútbol?"

6. Futuro de la Analítica de Datos en el Fútbol:

a) Perspectivas Futuras:

- "Desde su punto de vista, ¿cómo visualiza el futuro de la analítica de datos en el fútbol en los próximos años?"

7. Cierre:

a) Agradecimiento y Preguntas Finales:

- "Agradecemos su tiempo. ¿Hay algo más que le gustaría agregar o discutir antes de finalizar la entrevista?"

ANEXO 4: Carta de aprobación del jurado

San Juan de Pasto, 31 de octubre de 2024

Señores:

Comité de Investigaciones

Programa de Ingeniería de Sistemas

Universidad CESMAG

Asunto: Aval Trabajo de Grado

Saludo de paz y bien,

Yo, Esp. Luis Carlos Viteri Rosero identificado con cédula de ciudadanía Nro. 1061709816, doy fe que el trabajo de grado, Predicción de Resultados en el Fútbol a través de un producto Software Implementando Técnicas de Machine Learning y Análisis de Datos, realizado por el estudiante: Julián Esteban Chavez Rivera se puede presentar a revisión de Jurados, de igual manera, notifico que el estudiante asistió a las respectivas asesorías conforme a cronograma de actividades.

El estudiante manifiesta su compromiso para continuar depurando detalles del documento y producto TI, buscando alcanzar los objetivos y resultados del proyecto.

Atentamente,



Esp. Luis Carlos Viteri Rosero

1061709816

Asesor del proyecto

ANEXO 5: Documento técnico de requisitos

Introducción

Score Prediction es una innovadora plataforma web diseñada para pronosticar resultados en partidos de fútbol mediante el uso de técnicas avanzadas de machine learning y análisis de datos. Desarrollada como un proyecto de grado, esta solución integra algoritmos predictivos y un análisis estadístico profundo, proporcionando a los usuarios predicciones precisas y basadas en datos históricos. Score Prediction permite a los aficionados y analistas deportivos explorar variables clave, como el rendimiento de los equipos, los jugadores y el contexto del partido, transformando el análisis futbolístico en una experiencia interactiva y de alto valor predictivo. Además, la plataforma facilita la toma de decisiones informadas y mejora la comprensión de las dinámicas del fútbol moderno.

Participantes

A continuación, se listan los participantes e interesados en el desarrollo del software SCORE PREDICTION como:

Organizaciones Participantes

ORG001

Organización:	Universidad CESMAG
	Cra. 20a # 14-54
Dirección:	(602) 7216535 – 7244434
Teléfono:	
Fax:	ingenieriadesistemas@unicesmag.edu.co
E-mail: Comentaríos:	Director del programa: Carlos Fernando González Guzmán

ORG002

Organización:	Grupo de investigación Tecnofilia
Dirección:	Carrera 20A 14-54
Teléfono:	Teléfono 7216535
Fax:	
E-mail:	ingenieriadestemas@unicesmag.edu.co

Comentarios:

Personas Participantes

STK001

Participant e:	Julian Esteban Chavez Rivera
Organización:	Universidad CESMAG
Rol:	Investigador
Es desarrollador:	Si
Es cliente:	Si
Es usuario:	No
Comentarios:	

STK002

Participant e:	Luis Carlos Viteri Rosero
Organización:	Universidad CESMAG
Rol:	Investigador
Es desarrollador:	Si
Es cliente:	Si
Es usuario:	No
Comentarios:	

<i>Comentari</i>	
<i>os:</i>	
STK003	
<i>Participant</i>	Héctor Andrés Mora Paz
<i>e:</i>	
<i>Organizaci</i>	Universidad CESMAG
<i>ón:Rol:</i>	Investigador
<i>Es</i>	Si
<i>desarrollador: Es</i>	Si
<i>cliente:</i>	No
<i>Es usuario:</i>	
<i>Comentari</i>	
<i>os:</i>	

Descripción del Sistema Actual:

En el ámbito del fútbol, se ha introducido un enfoque innovador basado en el análisis de datos y el aprendizaje automático para predecir los resultados de los partidos con mayor precisión. Sin embargo, a medida que estas tecnologías se han ido integrando en los deportes de competición, especialmente en el fútbol, han surgido nuevos desafíos que afectan la confiabilidad de las predicciones deportivas. Uno de los principales problemas es la proliferación de estafadores que utilizan predicciones falsas con fines lucrativos, valiéndose de canales como Telegram y bots para atraer a más víctimas, lo que puede llevar a pérdidas económicas significativas para quienes confían en sus predicciones. Además, estos estafadores suelen promocionar casas de apuestas en línea, aprovechándose de las pérdidas de los jugadores.

A pesar de la adopción de la inteligencia artificial en el fútbol, como el uso de herramientas como TCT-Scout de Olocip, que permite a clubes y jugadores tomar decisiones estratégicas basadas en análisis predictivos, persiste un cierto grado de desconfianza hacia estas tecnologías. La falta de precisión de algunos modelos es un factor clave en esta desconfianza, ya que diversas

variables, como la fortaleza de los equipos, el factor local, las condiciones climáticas y las lesiones, añaden complejidad a la predicción y reducen su fiabilidad. Estos elementos, en lugar de mejorar la exactitud de las predicciones, parecen contribuir a la falta de confianza en los modelos.

El proyecto Score Prediction se enfrenta a este panorama, donde el fraude en las apuestas deportivas y la falta de ética complican la creación de una herramienta fiable que permita a los usuarios analizar y predecir los resultados de manera precisa y contextual. Nuestro objetivo es ofrecer una solución que no solo mejore la experiencia futbolística de aficionados y profesionales, sino que también restablezca la confianza en las herramientas de predicción mediante el uso de técnicas avanzadas de machine learning y análisis de datos.

Objetivos del Sistema

OBJ 001	Visualizar Noticias
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	
Descripción:	Desarrollar un módulo de noticias deportivas que proporcione a los usuarios una visión clara y precisa de los eventos más importantes del mundo del deporte, con énfasis en la Premier League.
Sub-objetivos:	Ninguno
Importancia:	Alta
Urgencia:	Media
Estado:	Terminado
Estabilidad:	Alta
Comentarios:	Ninguno

OBJ 002	Predicciones Deportivas
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	
Descripción:	Implementar un módulo de predicciones deportivas para permitir a los usuarios realizar predicciones de partidos de la Premier League con base

	en probabilidades, así como acceder a predicciones históricas para el análisis de tendencias.
Sub-objetivos:	Ninguno
Importancia:	Alta
Urgencia:	Inmediata
Estado:	Terminado
Estabilidad:	Alta
Comentarios:	Las predicciones deben ser claras y accesibles desde la interfaz principal.

OBJ 003	Información detallada de predicciones
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	
Descripción:	Proveer acceso a información detallada sobre los partidos de la Premier League para usuarios registrados, incluyendo estadísticas clave como marcador probable, porcentaje de victoria, derrota y empate.
Sub-objetivos:	Ninguno
Importancia:	Alta
Urgencia:	Alta
Estado:	Terminado
Estabilidad:	Alta
Comentarios:	Ninguno

OBJ 004	Login y registro de usuarios
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	
Descripción:	Implementar un sistema de registro e inicio de sesión que permita a los visitantes registrarse como usuarios del sistema y acceder a funcionalidades exclusivas mediante un proceso de validación de credenciales.
Sub-objetivos:	Ninguno
Importancia:	Alta
Urgencia:	Alta
Estado:	Terminado
Estabilidad:	Alta
Comentarios:	Ninguno

OBJ 005	Información detallada
Versión:	1.0 (01/07/2023)

Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	
Descripción:	Desarrollar un módulo interactivo que permita a los usuarios visualizar la precisión del modelo predictivo a través de gráficos claros y comprensibles, así como acceder a información detallada sobre el rendimiento del modelo. Esto facilitará la evaluación del desempeño del modelo y proporcionará un análisis más profundo de las estadísticas relevantes.
Sub-objetivos:	Ninguno
Importancia:	Alta
Urgencia:	Media
Estado:	Terminado
Estabilidad:	Alta
Comentarios:	Ninguno

Catálogo de Requisitos del Sistema

Requisitos de Información

IRQ-001	Visualizar Noticias
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	
Dependencias:	Tener acceso a la conexión a Internet para obtener las noticias en tiempo real.
Datos específicos:	Título de la noticia. Descripción breve. Fecha y hora de publicación. Fuente (medio de comunicación, periodista). Categoría (Premier League, deportes generales). Imágenes o videos. Enlace a la noticia completa (si aplica).
Importancia:	Alta
Urgencia:	Media
Estabilidad:	Alta
Estado:	Terminado
Comentarios	Ninguno

IRQ-002		Predicciones Deportivas	
Versión:	1.0 (01/07/2023)		
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz		
Fuentes:			
Dependencias:	Conexión a bases de datos históricas de la Premier League y acceso a modelos predictivos.		
Datos específicos:	Resultados históricos de partidos de la Premier League. Probabilidades de ganar, perder o empatar. Estadísticas de rendimiento (posesión, xG, rendimiento grupal). Interfaz para permitir predicciones.		
Importancia:	Alta		
Urgencia:	Inmediata		
Estabilidad:	Alta		
Estado:	Terminado		
Comentarios	Las predicciones deben ser visibles y accesibles desde la interfaz principal.		

IRQ-003		Información detallada de predicciones	
Versión:	1.0 (01/07/2023)		
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz		
Fuentes:			
Dependencias:	Registro previo de usuarios para acceder a predicciones detalladas.		
Datos específicos:	Porcentaje de victoria, empate o derrota. Estadísticas avanzadas del partido (posesión, goles esperados, posición en la tabla). Información clave de los equipos y jugadores. Gráficos que representen el análisis de las predicciones.		
Importancia:	Alta		
Urgencia:	Inmediata		
Estabilidad:	Alta		
Estado:	Terminado		
Comentarios	Ninguno		

IRQ-004		Login y registro de usuarios	
Versión:	1.0 (01/07/2023)		
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz		
Fuentes:			
Dependencias:	Base de datos de usuarios para almacenar y validar credenciales.		

Datos específicos:	Nombre completo. Correo electrónico. Contraseña.
Importancia:	Alta
Urgencia:	Alta
Estabilidad:	Alta
Estado:	Terminado
Comentarios	Ninguno

IRQ-005	Información detallada del rendimiento del modelo predictivo
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	
Dependencias:	Conexión a la base de datos con los resultados de predicciones y acceso al modelo de IA.
Datos específicos:	Resultados del modelo (precisión, recall, F1 score). Gráficos interactivos para visualizar el rendimiento del modelo. Comparación entre predicciones y resultados reales. Estadísticas históricas del modelo.
Importancia:	Esencial
Urgencia:	Inmediata
Estabilidad:	Alta
Estado:	Terminado
Comentarios	Ninguno

Requisitos Funcionales

FRQ-001	Visualización de Noticias
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Descripción:	El sistema permitirá a los usuarios visualizar noticias deportivas actualizadas sobre eventos importantes, con un enfoque especial en la Premier League. Las noticias incluirán títulos, descripciones breves, imágenes o videos, y enlaces a las fuentes originales.
Importancia:	Esencial
Urgencia:	Inmediata
Estabilidad:	Alta
Estado:	Terminado
Comentarios:	Ninguno

FRQ-002 Predicciones Deportivas

Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Descripción:	El sistema permitirá a los usuarios realizar predicciones de partidos de la Premier League, basadas en datos históricos y probabilidades generadas por el modelo predictivo. Estas predicciones estarán disponibles en la interfaz principal y podrán ser consultadas para análisis futuros.
Importancia:	Esencial
Urgencia:	Inmediata
Estabilidad:	Alta
Estado:	Terminado
Comentarios:	Ninguno

FRQ-003 Acceso a Información Detallada de Predicciones

Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Descripción:	El sistema permitirá a los usuarios registrados acceder a información detallada sobre las predicciones de partidos, incluyendo probabilidades de victoria, derrota o empate, estadísticas avanzadas, y gráficos que ayuden a interpretar las predicciones.
Importancia:	Esencial
Urgencia:	Inmediata
Estabilidad:	Alta
Estado:	Terminado
Comentarios:	Ninguno

FRQ-004 Login y Registro de Usuarios

Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Descripción:	El sistema permitirá a los usuarios registrarse y acceder a funcionalidades exclusivas mediante un sistema de autenticación que requerirá correo electrónico y contraseña. Los usuarios podrán recuperar sus cuentas en caso de perder el acceso.
Importancia:	Esencial
Urgencia:	Inmediata
Estabilidad:	Alta
Estado:	Terminado
Comentarios:	Ninguno

FRQ-005	Visualización de Rendimiento del Modelo Predictivo
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Jorge Albeiro Rivera Rosero
Descripción:	El sistema proporcionará a los usuarios gráficos interactivos y métricas como precisión y recall para visualizar el rendimiento del modelo predictivo, permitiendo un análisis detallado de los resultados y tendencias de las predicciones.
Importancia:	Esencial
Urgencia:	Inmediata
Estabilidad:	Alta
Estado:	Terminado
Comentarios:	Ninguno

Requisitos No Funcionales

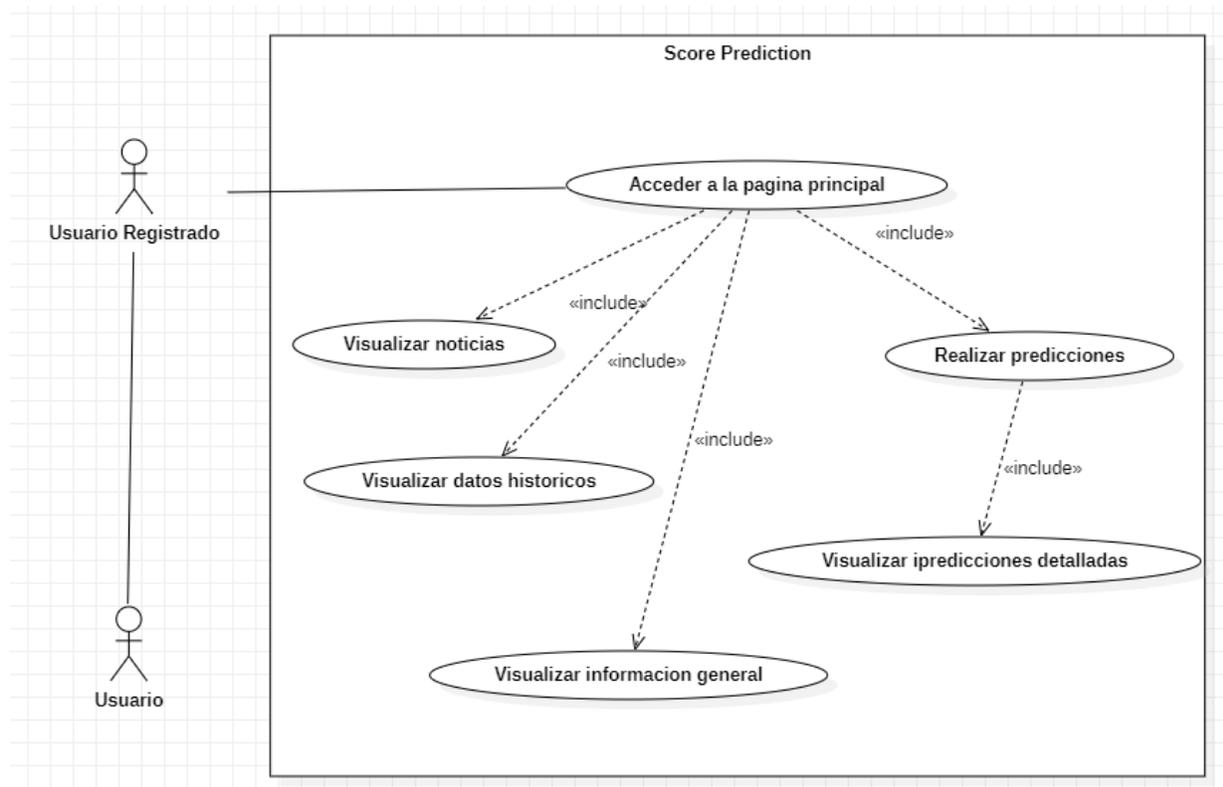
FNR-001 Correctitud

Versión:	1.0 (01/07/2023) (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes: Dependencias:	Grupo Tecnofilia
Descripción:	Cumplir completamente con el grado de especificación del producto.
Importancia: Urgencia:	Vital
Estabilidad: Estado:	Inmediatamente
Comentarios:	Alta
	Implementado
	Ninguno

FNR-002 Integridad

<i>Versión</i>	1.0 (01/07/2023) (01/07/2023)
<i>Autores:</i>	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
<i>Fuentes: Dependencias:</i>	Grupo Tecnofilia
<i>Descripción:</i>	
<i>Importancia:</i> <i>Urgencia:</i>	Garantizar la integridad de los datos.
<i>Estabilidad: Estado:</i>	Vital
<i>Comentarios:</i>	Inmediatamente
	Alta
	Implementado
	Ninguno

Diagrama General de Casos de Uso



Definición de actores

AC-001 Usuario

<i>Versión:Autores:</i>	1.0 (01/07/2023)
<i>Fuentes:</i>	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
<i>Descripción:</i>	Grupo Tecnofilia e IDIS
<i>Comentarios:</i>	Usuario, puede realizar predicciones para los partidos que desee, sin embargo no podrá acceder a información detallada
	Ninguno

AC-002 Usuario registrado

<i>Versión:Autores:</i>	1.0 (01/07/2023)
<i>Fuentes:</i>	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
<i>Descripción:</i>	Grupo Tecnofilia e IDIS
<i>Comentarios:</i>	Usuario Registrado además de poder realizar las predicciones, podrá acceder a información mucho mas detallada arrojada por el modelo
	Ninguno

Descripción De Casos De Uso

Caso de Uso Gestión de representación de proceso

CU-001	Visualización de Noticias
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	Grupo Tecnofilia
Dependencias:	
Descripción:	Visualizar las principales noticias deportivas, incluyendo las más relevantes de la Premier League.
Actores:	Usuario
Precondición:	El sistema debe estar conectado a una fuente actualizada de noticias deportivas. La interfaz debe estar activa y accesible.
Secuencia normal:	El usuario accede al módulo de noticias desde la página principal. El sistema muestra una lista de titulares de noticias deportivas con la fecha de publicación visible. Cada noticia incluye una imagen relacionada con su contenido. El usuario selecciona una noticia haciendo clic en el titular o enlace. El sistema redirige al usuario al sitio web donde se publicó la noticia completa.
Post condición	El usuario es redirigido exitosamente al sitio web de la noticia seleccionada.
Excepciones	E1: La conexión con la fuente de noticias falla. El sistema muestra un mensaje de error y sugiere intentar más tarde. E2: La noticia seleccionada ya no está disponible. El sistema muestra un mensaje de "Noticia no disponible".
Comentarios	NINGUNO

Caso de Uso Gestión de visitas

CU-002	Realización de predicción
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	Grupo Tecnofilia
Dependencias:	

Descripción:	Realizar predicciones de ganador para los partidos de la temporada 2023-24 de la Premier League
Actores:	Usuario
Precondición:	El sistema debe tener las probabilidades de los equipos previamente calculadas y disponibles.
Secuencia normal:	<p>El usuario accede al módulo de predicciones desde la página principal.</p> <p>El sistema muestra una lista de partidos con los nombres de los equipos y el porcentaje de probabilidad de victoria para cada equipo.</p> <p>El usuario puede visualizar las probabilidades y decidir hacer predicciones basadas en ellas.</p>
Post condición	El usuario puede visualizar las probabilidades de victoria para cada equipo en los partidos de la temporada 2023-24.
Excepciones	E1: Las probabilidades no están disponibles. El sistema muestra un mensaje de "Probabilidades no disponibles en este momento" y sugiere intentar más tarde.
Comentarios	NINGUNO

Caso de Uso Gestión de visitas

CU-003	Visualizar información detallada
Versión:	1.0 (01/07/2023)
Autores:	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
Fuentes:	Grupo Tecnofilia
Dependencias:	
Descripción:	Visualizar información detallada de los partidos de la Premier League
Actores:	Usuario registrado
Precondición:	El usuario debe estar registrado y autenticado en el sistema. La base de datos debe contener las estadísticas de los partidos.
Secuencia normal:	<p>El usuario registrado accede al módulo de partidos detallados desde la página principal.</p> <p>El sistema muestra la lista de partidos con la información detallada (goles, posesión, tiros a puerta, etc.).</p> <p>El usuario visualiza la información detallada de los partidos seleccionados.</p>
Post condición	El usuario puede analizar los detalles de los partidos de la

	Premier League de manera completa.
<i>Excepciones</i>	E1: El usuario no está registrado o su sesión ha caducado. El sistema solicita iniciar sesión. E2: Las estadísticas no están disponibles. El sistema muestra un mensaje de "Datos no disponibles".
<i>Comentarios</i>	NINGUNO

Caso de Uso Gestión de visitas

CU-004	Predicciones históricas
<i>Versión:</i>	1.0 (01/07/2023)
<i>Autores:</i>	Julian Esteban Chavez Rivera Luis Carlos Viteri Rosero Héctor Andrés Mora Paz
<i>Fuentes:</i>	Grupo Tecnofilia
<i>Dependencias:</i>	
<i>Descripción:</i>	Acceder a las predicciones históricas de los partidos de la Premier League de la temporada 2022-2023, junto con los resultados reales.
<i>Actores:</i>	Usuario
<i>Precondición:</i>	El sistema debe tener acceso a los datos de predicciones y resultados de la temporada 2022-2023.
<i>Secuencia normal:</i>	El usuario accede al módulo de predicciones históricas desde la sección de estadísticas de la Premier League. El sistema muestra una lista de partidos de la temporada 2022-2023 con las predicciones y resultados finales. El usuario puede navegar entre los partidos para comparar predicciones y resultados.
<i>Post condición</i>	El usuario puede visualizar las predicciones históricas y compararlas con los resultados de los partidos.
<i>Excepciones</i>	E1: Los datos históricos no están disponibles. El sistema muestra un mensaje de "Datos no disponibles".
<i>Comentarios</i>	NINGUNO

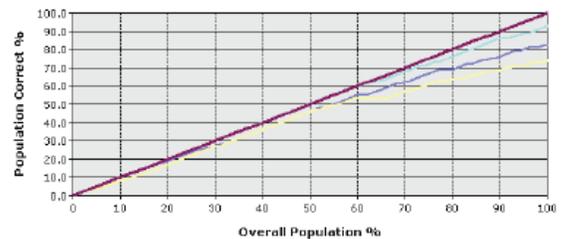
ANEXO 6: Mockups iniciales de la página Score Prediction:

Mockup 1: pantalla “Acerca de”



¿Como funciona nuestro modelo?

>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce id libero non libero sagittis fermentum. Vivamus ac nisi eget ex faucibus lobortis. Nulla rhoncus accumsan justo nec laoreet. Quisque sit amet magna et sem sodales fermentum. Aliquam sit amet metus non dolor gravida faucibus. Morbi lacinia ligula ac nisi accumsan, ut molestie tortor bibendum. Sed non nisl vel justo tincidunt lacinia nec a mauris. Vivamus in ligula vel sapien pretium bibendum. Proin non ligula eget justo fermentum vehicula. Nullam dictum sapien et nibh commodo placerat. Integer convallis, dui at fermentum lacinia, elit erat aliquam justo, ut vestibulum sapien sapien vitae neque



¿Que tecnologias usamos?

>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce id libero non libero sagittis fermentum. Vivamus ac nisi eget ex faucibus lobortis.



¿Quien Hace parte de nuestro equipo?



Julian Chavez



Luis viteri

Mockup 2: Pantalla “partidos” vista detallada



Premier league temporada 2022-2023

Fecha 1



Goles
6.3

Tiros a puerta
10.5

Corners
7.9

Faltas
1.3

Mockup 3 Pantalla “Partidos” vista completa



Premier league temporada 2022-2023

Fecha 1



Fecha 2



Fecha 3



Fecha 4



Mockup 4 Pantalla “Noticias”

The mockup features a green header with a soccer ball icon, the text "SCORE PREDICTION", and a "INICIA" button. Below the header is a navigation bar with "NOTICIAS", "PARTIDOS", and "ACERCA DE". The main content area contains a large news card for "AZ ALKMAAR 2-0 AJAX" with a photo of a player and the headline "El AZ Alkmaar gana al Ajax con doblete del hijo de van Bommel". Below this are six smaller news cards: "El baile de entrenadores que podría haber en grandes clubes el próximo verano", "DORTMUND 2-3 HOFFENHEIM", "FODEN DECIDE Y EL MANCHESTER CITY MANTIENE LA PRESIÓN SOBRE EL LIVERPOOL", "BORNEMOUTH 0-1 MAN. UNITED", "La Juventus rompe su mala racha con una victoria milagro en el descuento", and "El Dortmund se deja remontar ante un gran Hoffenheim".

SCORE PREDICTION INICIA

NOTICIAS | PARTIDOS | ACERCA DE



AZ ALKMAAR 2-0 AJAX

El AZ Alkmaar gana al Ajax con doblete del hijo de van Bommel



FÚTBOL INTERNACIONAL

El baile de entrenadores que podría haber en grandes clubes el próximo verano

RAÚL FUENTES / NACHO LABARRA

DORTMUND 2-3 HOFFENHEIM

El Dortmund se deja remontar ante un gran Hoffenheim





JUVENTUS 3-2 FROSINONE

La Juventus rompe su mala racha con una victoria milagro en el descuento

EFE

BORNEMOUTH 0-1 MAN. UNITED

Foden decide y el Manchester City mantiene la presión sobre el Liverpool





IRATI PRAT

Mockup 5 Pantalla “Inicio de sesión”

SCORE PREDICTION INICIA

NOTICIAS | PARTIDOS | ACERCA DE

Premier League temporada 2022-2023

Fecha	Equipo 1	Probabilidad Ganador	VS	Equipo 2	Probabilidad Ganador
Fecha 1	Manchester United	10 %		Liverpool	90 %
Fecha 2	Arsenal	55 %		Chelsea	45 %
Fecha 3	Manchester City	60 %		Crystal Palace FC	40 %
Fecha 4	Aston Villa FC	78 %		Tottenham Hotspur	22 %

Correo

Contraseña

INICIA

¿Todavía no tienes cuenta?
[REGISTRATE](#)

Mockup 6 Pantalla “Registro de usuarios”



Premier league temporada 2022-2023

Fecha 1



Fecha 2



A registration form with three input fields: 'Correo', 'Contraseña', and 'Confirma Contraseña'. Below the fields is a green button labeled 'INICIA'.

Fecha 3



Fecha 4



ANEXO 7: Manual de usuario

Descripción del software

Score Prediction se presenta como una solución informática que actúa como intermediaria entre un modelo de inteligencia artificial y los usuarios interesados en las predicciones de partidos de fútbol generadas por dicho modelo. La plataforma utiliza datos de los equipos seleccionados, complementándolos con información adicional extraída de una base de datos. Este conjunto de datos es procesado por el modelo para generar resultados predictivos.

El sistema cuenta con un motor de predicción que emplea algoritmos avanzados, incluyendo redes neuronales y análisis de series temporales. Estos algoritmos analizan diversos factores, tales como estadísticas de los equipos, su rendimiento reciente y datos históricos. Los resultados son presentados a través de una interfaz gráfica intuitiva, que permite a los usuarios visualizar claramente las probabilidades de victoria, derrota, empate y un posible marcador final.

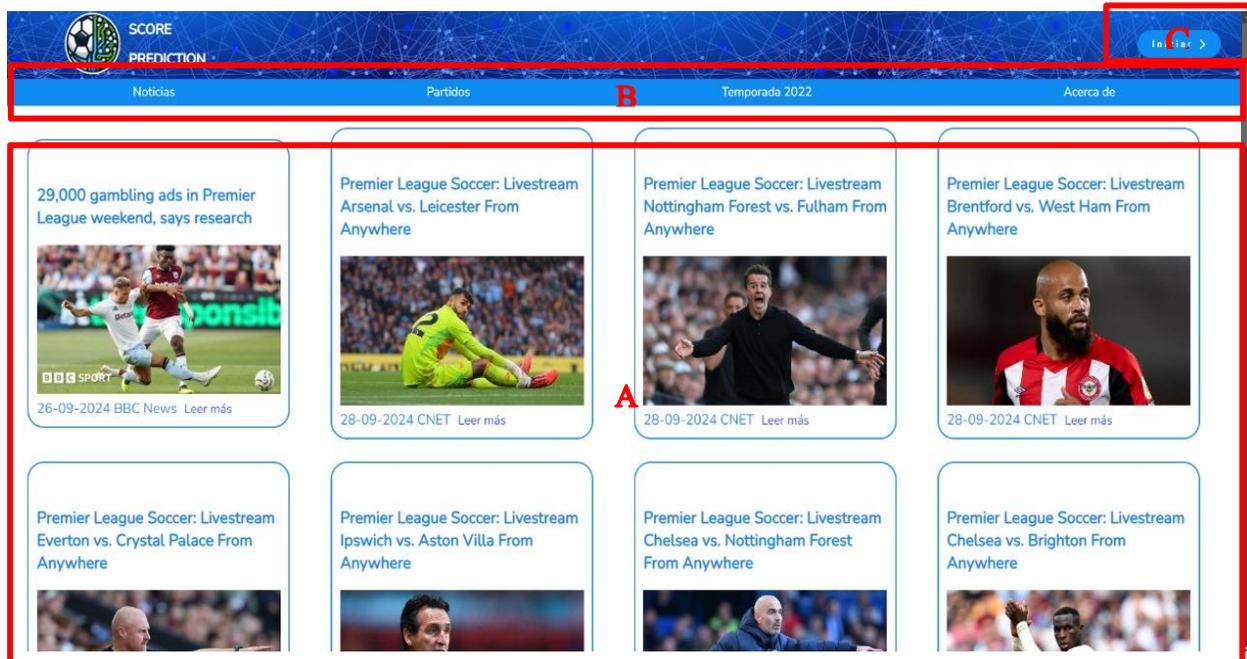
Interfaces del software

A continuación, se presentan las interfaces de usuario que facilitan el acceso y la selección de los equipos para los cuales se desean conocer las probabilidades de los partidos. Asimismo, se ofrece una visión general del proyecto y de la página, donde se destacan las principales noticias del mundo del fútbol, con un enfoque particular en la Premier League.

Interfaz Principal

A continuación, se describe la interfaz principal de usuario con sus componentes de manera general, cada una de las funcionalidades de los componentes se describen en secciones posteriores.

Captura de pantalla



Descripción de la funcionalidad de la interfaz

La interfaz principal de la página web Score Prediction se divide en varias secciones clave, como se muestra en la figura.

Descripción de los elementos de la interfaz

A. Noticias Principales En esta sección se muestran las noticias más relevantes del mundo del fútbol, con un enfoque especial en la Premier League. Los usuarios pueden visualizar artículos actualizados sobre los partidos más importantes, análisis de los encuentros y otros eventos destacados. Cada noticia cuenta con una opción de "Leer más", la cual permite redireccionar a la página de la noticia para ampliar la información de interés.

B. Menú de Navegación El menú de navegación, ubicado en la parte superior de la página, permite al usuario acceder rápidamente a las diferentes secciones del sitio web. Las opciones más comunes incluyen:

- Noticias: Donde se encuentran las últimas actualizaciones del mundo del fútbol.

- **Partidos:** Sección en donde el usuario podrá realizar las predicciones para los equipos que desee

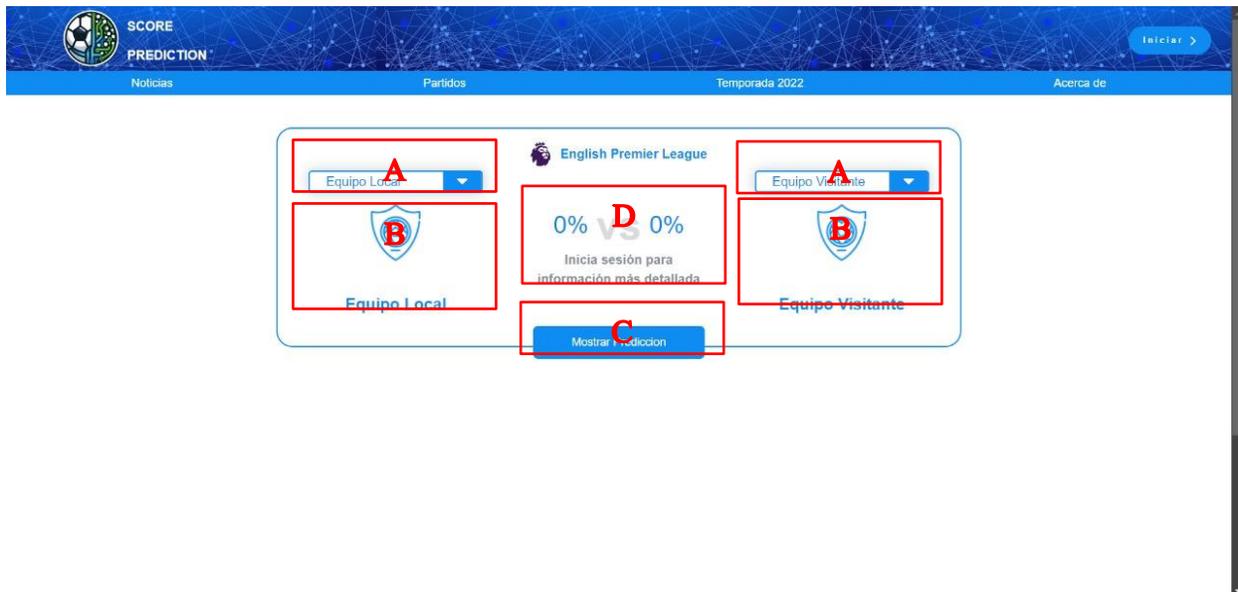
- **Temporada 2022:** Acceso a detalles específicos sobre la temporada 2022-2023.
- **Acerca de:** Información sobre la página web y su propósito.

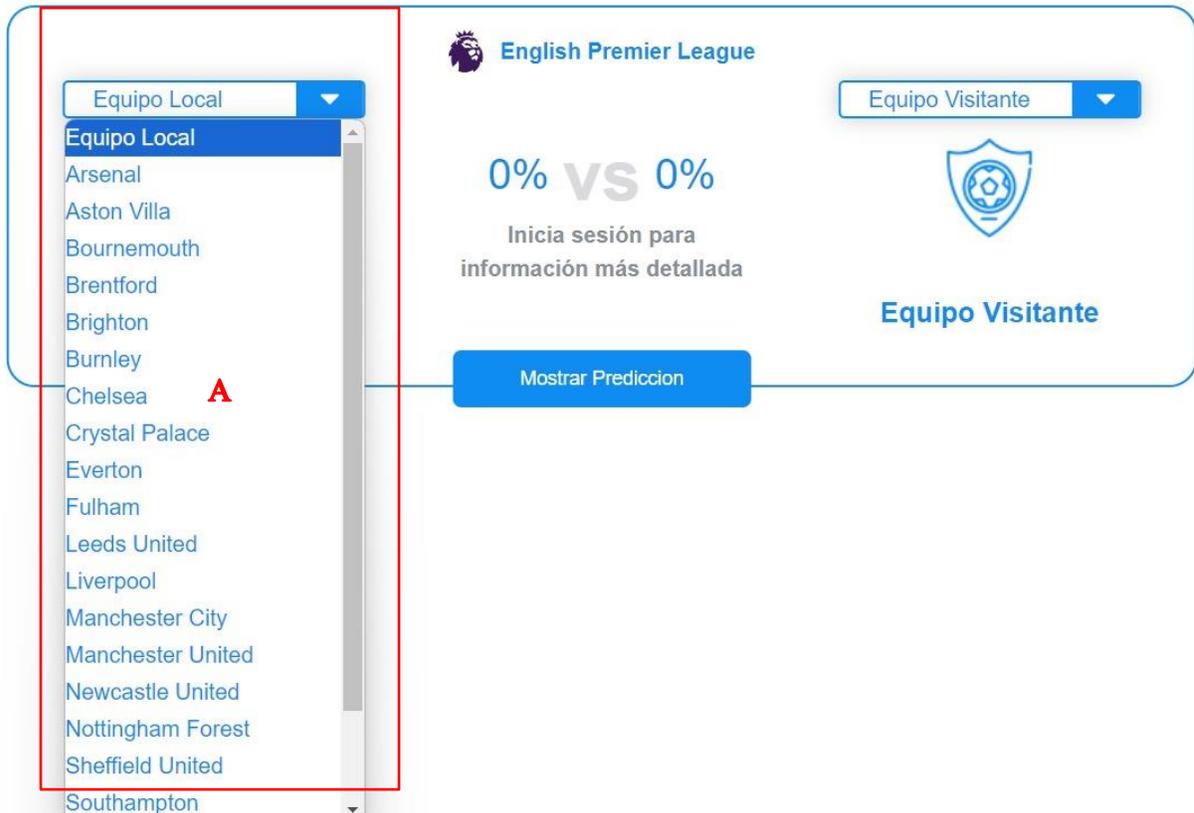
C. **Inicio de Sesión** En la esquina superior derecha se encuentra la opción de Inicio de Sesión. Los usuarios registrados que acceden con sus credenciales pueden disfrutar de funciones adicionales centradas al acceso a predicciones detalladas de los partidos.

Interfaz visualización de partidos

A continuación, se detalla la interfaz de la sección de partidos, donde el usuario podrá realizar predicciones para los equipos que desee.

Captura de pantalla





Descripción de la funcionalidad de la interfaz

La interfaz de predicción permite a los usuarios seleccionar equipos y visualizar predicciones basadas en un modelo de inteligencia artificial. La página está dividida en varias secciones importantes, como se describe a continuación:

Descripción de los elementos de la interfaz

Selección de Equipos: En esta sección, los usuarios encontrarán dos listas desplegables:

Equipo Local: Permite seleccionar el equipo que jugará como local.

Equipo Visitante: Permite seleccionar el equipo que jugará como visitante.

Al desplegar las opciones, el usuario puede elegir cualquier equipo disponible para obtener una predicción personalizada.

Visualización de Escudos: Después de seleccionar los equipos en la Sección A, los escudos de los equipos elegidos se mostrarán automáticamente en esta área. Esto permite al usuario verificar visualmente sus selecciones tanto para el equipo local como para el equipo visitante.

Botón para Generar Predicción: Una vez que los equipos han sido seleccionados, el usuario deberá presionar el botón "Mostrar Predicción" ubicado en el centro de la interfaz. Este botón disparará el proceso de predicción, y la página mostrará el resultado en la siguiente sección.

Detalles de la Predicción: Aquí es donde se presentan los resultados de la predicción. Los usuarios registrados podrán acceder a información adicional, como probabilidad de empate y un marcador posible

Interfaz Sección temporada 2022

A continuación, se describe la interfaz la sección temporada 2022, a la cual se puede acceder desde el menú superior desde cualquier sección detallada en la barra superior.

Captura de pantalla

Local	Away	Date	Most likely score	1	X	2
BOU	LIV	Mar11	0 - 2	9.24%	15.98%	74.78%
CRY	MCI	Mar11	0 - 2	9.42%	16.96%	73.62%
EVE	BRE	Mar11	0 - 1	19.68%	22.67%	57.65%
FUL	ARS	Mar12	0 - 1	12.62%	20.15%	67.23%
LEE	BHA	Mar11	1 - 1	15.02%	21.15%	63.83%
LEI	CHE	Mar11	0 - 1	22.75%	25.13%	52.12%
MUN	SOU	Mar12	2 - 0	68.45%	19.75%	11.80%
NEW	WOL	Mar12	1 - 0	63.49%	21.41%	15.09%
WHU	AVL	Mar12	1 - 1	37.13%	27.08%	35.79%
AVL	BOU	Mar18	1 - 0	58.57%	23.01%	18.41%
BHA	CRY	Mar15	1 - 0	64.27%	20.99%	14.74%

Descripción de los elementos de la interfaz

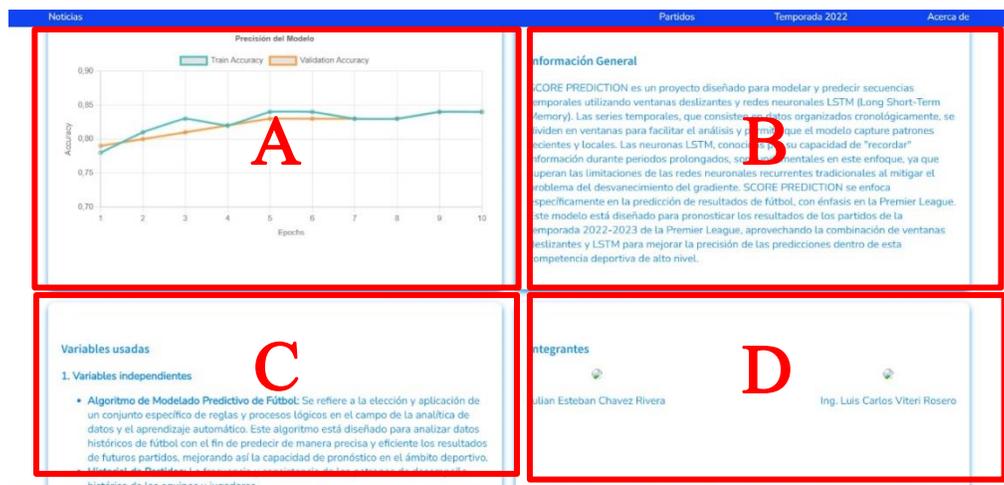
Tabla de Predicciones: los usuarios pueden visualizar una tabla que contiene las predicciones generadas por el modelo para diferentes encuentros de fútbol. La tabla incluye la siguiente información:

- Local: El equipo que juega en casa.
- Away: El equipo visitante.
- Date: La fecha en que se disputará el partido.
- Most likely score: El marcador más probable para el encuentro, basado en los análisis del modelo.
- 1, X, 2: Columnas que muestran las probabilidades correspondientes a los resultados posibles. 1 correspondiente a la probabilidad que el equipo local gane el partido, X a la probabilidad que el partido termine en empate, 2 a la probabilidad de que el equipo visitante gane el partido

Interfaz Sección acerca de

A continuación, se describe la interfaz la función de configuración de salida, a la cual se puede acceder desde la interfaz principal de usuario, este elemento se identifica con el uso de diferentes combinaciones del color naranja.

Captura de pantalla



Descripción de la funcionalidad de la interfaz

En esta sección de la página web se presentan los resultados del modelo de predicción y detalles sobre el proyecto SCORE PREDICTION.

Descripción de los elementos de la interfaz

Gráfica de Precisión del Modelo: presenta una gráfica que compara la precisión del modelo de predicción a lo largo de varias épocas. Las líneas diferenciadas muestran Train Accuracy La cual es la precisión alcanzada durante el entrenamiento del modelo y Validation Accuracy La cual muestra la precisión medida durante la validación del modelo.

Información General del Proyecto: contiene una descripción general del proyecto SCORE PREDICTION, detallando los objetivos principales, que incluyen el desarrollo de un modelo basado en redes neuronales LSTM para la predicción de resultados en la Premier League.

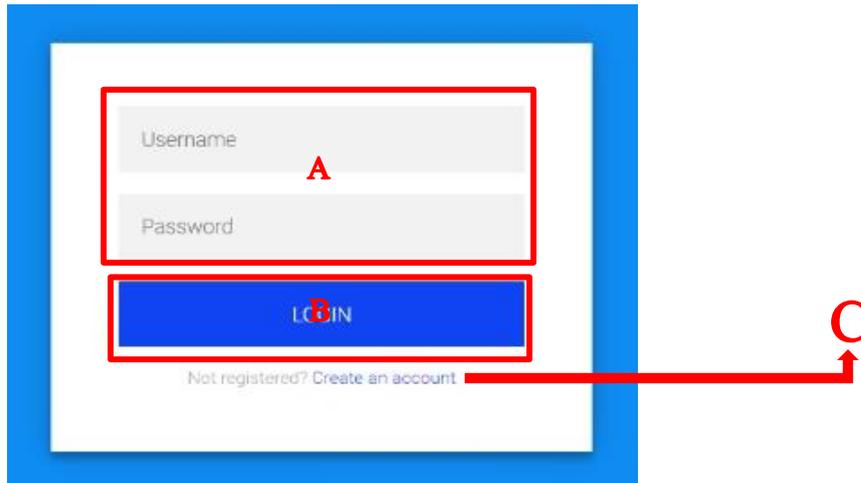
Variables Utilizadas: se destacan las variables que fueron usadas en el desarrollo del modelo predictivo. Esto incluye una descripción técnica de las Variables Independientes y la explicación del Algoritmo de Modelado Predictivo utilizado para el análisis de datos históricos y las predicciones.

Integrantes del Proyecto: muestra los nombres y cargos de los integrantes clave del proyecto, tanto el estudiante investigador como el asesor.

Interfaz inicio de sesión

A continuación, se describe la interfaz la función de inicio de sesión, para poder acceder a información más detallada que se arroja a través del modelo.

Captura de pantalla



Descripción de la funcionalidad de la interfaz

Esta interfaz permite al usuario ingresar sus credenciales para acceder a la plataforma.

Descripción de los elementos de la interfaz

A. Campos de Entrada de Datos: el usuario encontrará los campos de texto para introducir sus credenciales de acceso:

- Username: Espacio donde el usuario debe ingresar su nombre de usuario.
- Password: Espacio donde el usuario debe ingresar su contraseña.

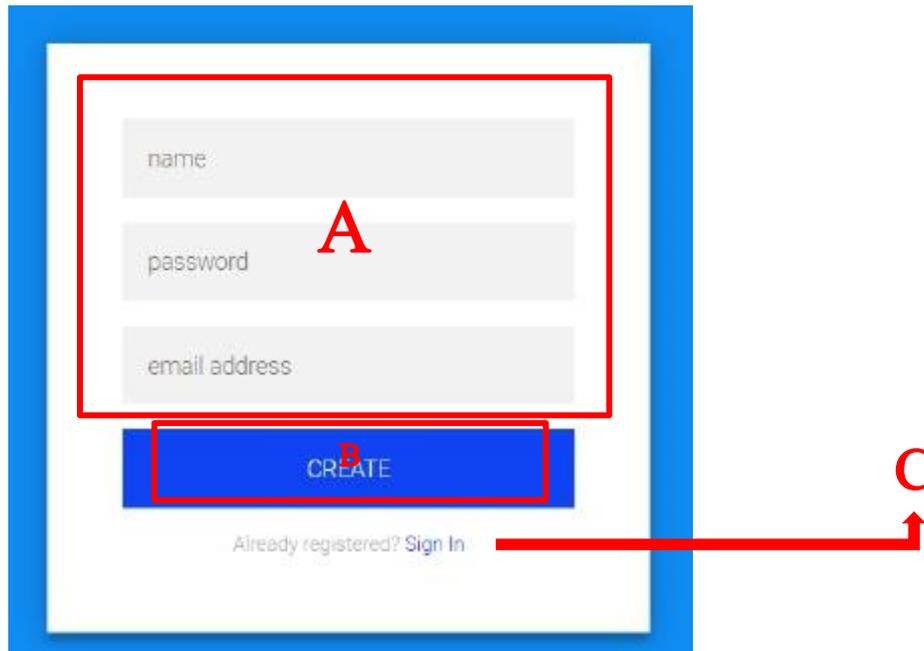
B. Botón de Login: contiene el botón de LOGIN, el cual, al ser presionado, procesará las credenciales y permitirá el acceso a la página principal en caso de que los datos sean correctos.

C. Enlace para Crear una Cuenta: los usuarios que aún no se han registrado pueden hacer clic en el enlace Create an account, el cual redirige a la página de registro, donde podrán crear una nueva cuenta.

Interfaz de Registro de usuarios

A continuación, se describe la interfaz la función de registro de usuarios, para que los usuarios que deseen puedan registrarse y disfrutar de los beneficios de tener una cuenta registrada.

Captura de pantalla



Descripción de la funcionalidad de la interfaz

Esta interfaz permite a los nuevos usuarios registrarse en la plataforma.

Descripción de los elementos de la interfaz

A. Campos de Entrada de Datos: el usuario encontrará los campos de texto para introducir la información requerida para el registro:

- Name: Espacio donde el usuario debe ingresar su nombre.
- Password: Espacio donde el usuario debe ingresar una contraseña segura.
- Email Address: Espacio donde el usuario debe ingresar su dirección de correo electrónico.

B. Botón de Registro: contiene el botón de CREATE, que permitirá al usuario completar el proceso de registro. Al hacer clic, la información proporcionada será enviada, y si es válida, se creará la cuenta y el usuario podrá acceder a la plataforma.

C. Enlace para Iniciar Sesión: los usuarios que ya tienen una cuenta pueden hacer clic en el enlace Sign In, el cual los redirige a la página de inicio de sesión, para que puedan acceder con sus credenciales ya registradas.

ANEXO 8: Resultados de partidos reales

Fecha	Equipo Local	Equipo Visitante	Marcador probable	Probabilidades			¿Acertó?
				V	E	D	
03/11/2023	Bournemouth	Liverpool	0 - 2	9.24%	15.98%	74.78%	No
03/11/2023	Crystal Palace	Manchester City	0 - 2	9.42%	16.96%	73.62%	Sí
03/11/2023	Everton	Brentford	1 - 0	57.65%	22.67%	19.68%	Sí
03/12/2023	Fulham	Arsenal	0 - 3	12.62%	20.15%	67.23%	Sí
03/11/2023	Leeds United	Brighton	1 - 1	30.50%	34.25%	35.25%	Sí
03/11/2023	Leicester City	Chelsea	0 - 1	22.75%	25.13%	52.12%	Sí
03/12/2023	Manchester United	Southampton	1 - 0	68.45%	19.75%	11.80%	Sí
03/12/2023	Newcastle United	Wolverhampton	1 - 0	63.49%	21.41%	15.09%	Sí
03/12/2023	West Ham United	Aston Villa	1 - 1	37.13%	27.08%	35.79%	Sí
03/18/2023	Aston Villa	Bournemouth	2 - 0	58.57%	23.01%	18.41%	Sí
03/15/2023	Brighton	Crystal Palace	1 - 0	64.27%	20.99%	14.74%	Sí
03/18/2023	Chelsea	Everton	1 - 1	30.50%	34.25%	35.25%	Sí
03/18/2023	Liverpool	Fulham	2 - 0	66.37%	20.34%	13.29%	Sí
03/19/2023	Manchester City	West Ham United	2 - 0	69.15%	19.33%	11.51%	Sí
03/15/2023	Brentford	Southampton	0 - 1	56.97%	22.82%	20.22%	Sí
03/18/2023	Wolverhampton	Leeds United	1 - 2	30.55%	27.81%	41.64%	Sí
04/01/2023	Arsenal	Leeds United	2 - 0	67.89%	19.43%	12.68%	Sí
04/01/2023	Brighton	Brentford	1 - 1	30.49%	34.49%	25.02%	Sí
04/01/2023	Bournemouth	Fulham	1 - 0	51.65%	24.99%	23.36%	Sí
04/01/2023	Chelsea	Aston Villa	0 - 2	27.43%	26.27%	46.31%	Sí
15/03/2023	Brighton	Crystal Palace	1 - 0	64.27%	20.99%	14.74%	Sí
18/03/2023	Chelsea	Everton	1 - 1	30.50%	34.25%	35.25%	Sí
18/03/2023	Liverpool	Fulham	2 - 0	66.37%	20.34%	13.29%	Sí
19/03/2023	Manchester City	West Ham United	2 - 0	69.15%	19.33%	11.51%	Sí
15/03/2023	Brentford	Southampton	0 - 1	56.97%	22.82%	20.22%	Sí
18/03/2023	Wolverhampton	Leeds United	1 - 2	30.55%	27.81%	41.64%	Sí
01/04/2023	Arsenal	Leeds United	2 - 0	67.89%	19.43%	12.68%	Sí
01/04/2023	Brighton	Brentford	1 - 1	30.49%	34.49%	25.02%	Sí
01/04/2023	Bournemouth	Fulham	1 - 0	51.65%	24.99%	23.36%	Sí
01/04/2023	Chelsea	Aston Villa	1 - 2	27.43%	26.27%	46.31%	Sí
01/04/2023	Crystal Palace	Leicester City	1 - 0	44.98%	26.81%	28.22%	Sí
03/04/2023	Everton	Tottenham	0 - 1	18.45%	22.05%	59.50%	No

01/04/2023	Manchester City	Liverpool	1 – 0	52.65%	24.61%	22.74%	Sí
02/04/2023	Newcastle United	Manchester United	1 – 1	36.28%	26.88%	36.83%	Sí
02/04/2023	West Ham United	Southampton	1 – 1	51.34%	24.90%	23.76%	Sí
04/04/2023	Bournemouth	Brighton	0 – 2	11.81%	19.53%	68.66%	Sí
04/04/2023	Chelsea	Liverpool	1 – 1	32.17%	25.83%	42.01%	Sí
03/04/2023	Everton	Tottenham	0 – 1	18.77%	22.10%	59.13%	No
04/04/2023	Leicester City	Aston Villa	1 – 1	34.42%	26.71%	38.87%	Sí
05/04/2023	Manchester United	Brentford	1 – 1	39.51%	26.97%	33.53%	Sí
05/04/2023	West Ham United	Newcastle United	1 – 2	27.45%	25.09%	47.46%	Sí
08/04/2023	Aston Villa	Nottingham Forest	2 – 0	64.52%	20.99%	14.49%	Sí
08/04/2023	Brentford	Newcastle United	1 – 2	30.52%	27.23%	42.25%	Sí
08/04/2023	Fulham	West Ham United	1 – 0	51.50%	24.40%	24.10%	No
08/04/2023	Leicester City	Bournemouth	1 – 1	37.28%	25.83%	36.89%	No
08/04/2023	Southampton	Manchester City	0 – 2	13.29%	18.80%	67.91%	Sí
08/04/2023	Tottenham	Brighton	1 – 0	37.53%	24.60%	37.87%	No
08/04/2023	Wolverhampton	Chelsea	0 – 1	24.71%	28.82%	46.47%	No
09/04/2023	Leeds United	Crystal Palace	1 – 1	36.84%	29.88%	33.28%	No
09/04/2023	Liverpool	Arsenal	2 – 2	30.62%	30.03%	39.35%	Si
15/04/2023	Aston Villa	Newcastle United	1 – 0	52.24%	28.88%	18.88%	Si
15/04/2023	Chelsea	Brighton	0 – 1	25.20%	29.44%	45.36%	Si
15/04/2023	Crystal Palace	Southampton	1 – 0	46.28%	27.72%	26.00%	Sí
15/04/2023	Everton	Fulham	1 – 1	36.05%	28.56%	35.39%	No
15/04/2023	Manchester City	Leicester City	3 – 1	67.98%	19.66%	12.36%	Sí
15/04/2023	Tottenham	Bournemouth	2 – 1	57.84%	21.04%	21.12%	No
16/04/2023	Leeds United	Liverpool	0 – 2	11.87%	21.32%	66.81%	Sí
16/04/2023	West Ham United	Arsenal	1 – 2	18.64%	24.71%	56.65%	No
22/04/2023	Arsenal	Southampton	2 – 0	63.87%	21.03%	15.10%	No
22/04/2023	Brentford	Aston Villa	1 – 1	32.47%	29.87%	37.66%	Sí
22/04/2023	Brighton	Chelsea	2 – 1	55.41%	24.14%	20.45%	Sí
22/04/2023	Crystal Palace	Everton	1 – 0	49.38%	25.82%	24.80%	Sí
22/04/2023	Leicester City	Wolverhampton	1 – 1	36.74%	27.82%	35.44%	Sí
23/04/2023	Bournemouth	West Ham United	1 – 2	17.11%	28.67%	54.22%	Si
23/04/2023	Liverpool	Nottingham Forest	2 – 0	64.57%	20.78%	14.65%	Sí

23/04/2023	Newcastle United	Tottenham	2 – 1	48.35%	26.87%	24.78%	Sí
25/04/2023	Leeds United	Leicester City	1 – 1	34.49%	27.38%	38.13%	Sí
26/04/2023	Nottingham Forest	Brighton	0 – 2	14.57%	21.78%	63.65%	No
27/04/2023	West Ham United	Liverpool	1 – 2	20.87%	23.60%	55.53%	Sí
27/04/2023	Manchester City	Arsenal	2 – 1	49.35%	25.84%	24.81%	Sí

 <p>UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)	CÓDIGO: AAC-BL-FR-032
		VERSIÓN: 1
		FECHA: 09/JUN/2022

San Juan de Pasto, 18/02/2025

Biblioteca
REMIGIO FIORE FORTEZZA OFM. CAP.
Universidad CESMAG
Pasto

Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominado **Predicción de Resultados en el Fútbol a través de un producto Software Implementando Técnicas de Machine Learning y Análisis de Datos**, presentado por el (los) autor(es) **Julian Esteban Chavez Rivera**, del Programa Académico Ingeniería de sistemas al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor(a), que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,



Luis Carlos Viteri Rosero
1061709816
Ingeniería de Sistemas
3046358259
Lcviteri@unicesmag.edu.co

 UNIVERSIDAD CESMAG <small>NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</small>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

INFORMACIÓN DEL (LOS) AUTOR(ES)	
Nombres y apellidos del autor: Julian Esteban Chavez Rivera	Documento de identidad: 1004234559
Correo electrónico: Estebanchavez058@gmail.com	Número de contacto: 3113182393
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del asesor: Luis Carlos Viteri Rosero	Documento de identidad: 1061709816
Correo electrónico: Lcviteri@unicesmag.edu.co	Número de contacto: 3046358259
Título del trabajo de grado: Predicción de Resultados en el Fútbol a través de un producto Software Implementando Técnicas de Machine Learning y Análisis de Datos	
Facultad y Programa Académico: Facultad de Ingeniería, programa de Ingeniería de Sistemas	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

- a) La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.

 <p>UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

- b) Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.
- c) Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación, distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.
- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
- e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

NOTA: En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permiso(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 18 días del mes de 02 del año 2025

<i>Julian Chz</i>	Firma del autor
Nombre del autor: Julian Esteban Chavez Rivera	Nombre del autor:
Firma del autor	Firma del autor
Nombre del autor:	Nombre del autor:
 <hr/> Nombre del asesor: Luis Carlos Viteri Rosero	



UNIVERSIDAD
CESMAG
NIT: 800.109.387-7
VIGILADA MINEDUCACIÓN

**AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE
GRADO O TRABAJOS DE APLICACIÓN EN
REPOSITORIO INSTITUCIONAL**

CÓDIGO: AAC-BL-FR-031

VERSIÓN: 1

FECHA: 09/JUN/2022