



Análisis de datos de competiciones: Uso de técnicas de análisis de series temporales para prever resultados de competiciones

Data analysis of competitions: Use of time-series analysis techniques to forecast competition results

Autores

Ronald Alexander Llerena Carrera¹

¹ Universidad Estatal de Milagro (Ecuador)

Autor de correspondencia:
Ronald Alexander Llerena Carrera
rllerenac@unemi.edu.ec

Cómo citar en APA

Llerena Carrera, R. A. (2025). Competition data analysis: Using time series analysis techniques to predict competition outcomes. *Retos*, 67, 597-606.
<https://doi.org/10.47197/retos.v67.114138>

Resumen

Introducción: El análisis de datos deportivos ha evolucionado con el uso de técnicas avanzadas de series temporales para la predicción de resultados en competiciones. Este empleo de técnicas nuevas permite una mejor adecuación de datos, pero también un análisis más determinado. **Objetivo:** El objetivo fue evaluar la eficacia de los modelos ARIMA, redes neuronales recurrentes y aprendizaje profundo en la predicción de resultados de competiciones deportivas, comparando su precisión y capacidad de adaptación a distintas variables contextuales.

Metodología: Se basó en el análisis de datos históricos de competiciones deportivas, utilizando bases de datos deportivas con información sobre rendimiento de equipos y jugadores, evaluando la capacidad predictiva de los modelos implementados.

Resultados: Las redes neuronales LSTM presentaron un mejor desempeño en la predicción de resultados en comparación con modelos tradicionales. La inclusión de variables contextuales, como la condición física de los jugadores y el entorno de juego, mejoró la precisión de las predicciones.

Discusión: Los hallazgos coincidieron con investigaciones que destacan la efectividad del aprendizaje profundo en el análisis de datos deportivos. Sin embargo, se identificó la necesidad de seguir optimizando la integración de múltiples fuentes de datos para mejorar la exactitud de las predicciones.

Conclusiones: El uso de técnicas de series temporales representa una herramienta valiosa en la predicción de resultados deportivos, con aplicaciones en la toma de decisiones estratégicas y el análisis del rendimiento competitivo. Se recomienda continuar explorando enfoques híbridos y el uso de datos en tiempo real para fortalecer la precisión predictiva en futuros estudios.

Palabras clave

Análisis deportivo; predicción de resultados; series temporales; redes neuronales.

Abstract

Introduction: Sports data analysis has evolved with the use of advanced time series techniques for predicting outcomes in competitions. This use of new techniques allows for better data adjustment, as well as more targeted analysis.

Objective: The objective was to evaluate the effectiveness of ARIMA models, recurrent neural networks, and deep learning in predicting sports competition outcomes, comparing their accuracy and adaptability to different contextual variables.

Methodology: The study was based on the analysis of historical data from sports competitions, using sports databases containing information on team and player performance, and evaluating the predictive capacity of the implemented models.

Results: LSTM neural networks showed better performance in predicting outcomes compared to traditional models. The inclusion of contextual variables, such as players' physical condition and game environment, improved prediction accuracy.

Discussion: The findings aligned with previous research that highlights the effectiveness of deep learning in sports data analysis. However, the need to continue optimizing the integration of multiple data sources to enhance prediction accuracy was identified.

Conclusions: The use of time series techniques represents a valuable tool in predicting sports outcomes, with applications in strategic decision-making and competitive performance analysis. It is recommended to continue exploring hybrid approaches and the use of real-time data to strengthen predictive accuracy in future studies.

Keywords

Sports data analysis; outcome prediction; time series; neural networks.

Introducción

Las competiciones deportivas tienen la finalidad, como disciplinas del deporte y de la salud, de promover un estilo de vida saludable y en donde se pueda incluir la recreación y el interés por el deporte en sí (Lepschy et al., 2020). Para ello también se consideran los análisis de datos que permitan ver cómo se desarrollan dichas competiciones y determinar o dar a conocer los resultados que estas lleguen a tener, tal y como lo plantean Juan (2021) y Hannes et al. (2020). Es por tanto que el análisis de datos en competiciones deportivas ha experimentado un notable auge en los últimos años, impulsado por la creciente disponibilidad de datos detallados y el avance de técnicas analíticas sofisticadas (Arana, 2021; Mendes y Mendes, 2020.). La capacidad para prever resultados no solo enriquece la experiencia de los aficionados, sino que también ofrece herramientas valiosas para entrenadores, analistas y gestores deportivos en la toma de decisiones estratégicas (Andrade, 2023; Chinmay et al., 2024.).

Diversos estudios han explorado métodos para mejorar la precisión en la predicción de resultados deportivos, tal es el caso de Divekar et al. (2023) quienes emplearon una metodología bayesiana para prever en tiempo real los resultados de partidos de fútbol, utilizando datos secuenciales de eventos durante el encuentro. Su enfoque permitió estimar el impacto temporal de diversas variables y anticipar el desenlace del partido con mayor exactitud, lo cual tiene relación con la investigación de Megía (2023) quien se enfocó en la eficiencia y competitividad en las competencias de fútbol y consideró que sí era favorable este tipo de análisis tanto a los conocedores de la disciplina como también a diversos públicos, tales como apostadores, clubes deportivos y entre otros.

Por otro lado, Bunker et al. (2024) analizaron el uso de técnicas de aprendizaje automático en la predicción de resultados de fútbol. Su investigación destacó que modelos como CatBoost, aplicados a calificaciones específicas del fútbol, ofrecen un rendimiento superior en comparación con otros algoritmos tradicionales. Aunado a ello, la investigación de Rocha et al. (2021) sugiere la integración de datos de seguimiento espacio-temporal y eventos para mejorar la precisión de las predicciones. Dicha investigación resulta más precisa en cuanto a mostrar los resultados exactos de los torneos futbolísticos, destacando nuevas características para brindar un hallazgo más integral.

En el ámbito del cricket, Chakwate y Madhan (2020) investigaron el uso de modelos de memoria a corto y largo plazo (LSTM en sus siglas en inglés) para predecir el resultado de los partidos. Su enfoque se basó en datos detallados de cada jugada, lo que permitió estimar la probabilidad de victoria en diferentes momentos del encuentro, proporcionando información valiosa para ajustar estrategias en tiempo real y mejorar el rendimiento de los jugadores de acuerdo a lo que se prevé en resultados del partido.

Por otro lado, la aplicación de técnicas de aprendizaje automático ha demostrado ser efectiva en este ámbito, tal y como lo menciona Poveda (2023) quien desarrolló una red neuronal convolucional para clasificar imágenes cerebrales en el diagnóstico de la enfermedad de Alzheimer, logrando una precisión del 70% en un conjunto de prueba independiente. Aunque este estudio se centra en el ámbito médico, la metodología empleada destaca la importancia de la profundidad de las redes y el pre - procesamiento de datos, aspectos aplicables al análisis de series temporales en deportes y que pueden ayudar a mejorar la eficacia para la predicción de resultados, lo cual también se fundamenta en lo especificado por Alaminos y Alaminos (2021).

Situación similar ocurre con la investigación de Fierro (2020), quien comentó sobre la predicción de series temporales con redes neuronales y destacó el hecho de que esta metodología puede retornar resultados positivos, aunado al hecho de que puede brindar un resultado más adecuado y detallado que los métodos estadísticos comunes, por lo que resaltó la utilidad de aplicar estos métodos en los estudios de nivel predictivo.

Aunado a ello, Baldinelli (2024) exploró cómo las herramientas avanzadas de análisis de datos están transformando el panorama de las apuestas deportivas, permitiendo estrategias más informadas mediante información predictiva en tiempo real sobre el rendimiento de los jugadores y la dinámica del juego. Este enfoque no solo beneficia a los apostadores, sino que también ofrece perspectivas valiosas para entrenadores y analistas en la evaluación del desempeño y la planificación estratégica, lo cual es un aspecto clave al momento de conformar equipos de diferentes disciplinas deportivas, como también corregir los posibles errores que puedan cometerse al momento de efectuar el juego en sí o el partido.

Es debido a este contexto que la investigación actual se justifica por la necesidad de mejorar las predicciones en competiciones deportivas mediante técnicas avanzadas de análisis de series temporales. A pesar de los avances mencionados, persisten retos en la integración efectiva de estrategias de juego y decisiones en tiempo real dentro de los modelos predictivos, según lo que mencionan Song et al. (2020), por lo tanto este estudio buscó evaluar la eficacia de los modelos ARIMA, redes neuronales recurrentes y aprendizaje profundo en la predicción de resultados de competiciones deportivas, comparando su precisión y capacidad de adaptación a distintas variables contextuales.

Método

La presente investigación se enmarcó en un estudio cuantitativo de alcance correlacional, dado que el objetivo fue analizar y prever los resultados de competiciones deportivas mediante técnicas avanzadas de análisis de series temporales. Este enfoque permitió identificar patrones y relaciones estadísticas en los datos históricos de las competiciones, facilitando la predicción de futuros desempeños.

Para proceder con esta metodología, se adoptó un diseño no experimental, observacional y longitudinal. Este diseño se consideró adecuado para el estudio debido a que no es posible manipular las variables independientes y se requiere observar las tendencias a lo largo del tiempo (Hernández, Fernández y Baptista, 2014). La elección de este diseño permitió analizar la evolución temporal de las variables de interés sin intervenir en el desarrollo natural de los eventos deportivos.

La elección de un enfoque basado en análisis de series temporales responde a la necesidad de desarrollar predicciones precisas y fundamentadas en datos históricos. Métodos tradicionales de predicción deportiva, como el análisis estadístico convencional, han demostrado limitaciones en la identificación de patrones complejos y no lineales en el desempeño de equipos y atletas (Wang et al., 2024). En este sentido, la combinación de modelos estadísticos y técnicas de aprendizaje automático permite mejorar la precisión de las predicciones y aportar información útil para entrenadores, analistas y gestores deportivos.

Población y muestra

La población objeto de estudio estuvo conformada por competiciones deportivas profesionales en Ecuador, específicamente en disciplinas como el fútbol de la Serie A ecuatoriana, el baloncesto profesional y el atletismo. La muestra se obtuvo mediante un muestreo intencional, seleccionando datos de los últimos cinco años (2019-2024) provenientes de fuentes oficiales como la Federación Ecuatoriana de Fútbol (FEF), la LigaPro, la Federación Ecuatoriana de Atletismo y otras bases de datos deportivas reconocidas. Se priorizaron competiciones con registros estadísticos detallados, incluyendo datos de rendimiento de equipos y atletas.

Tomando en cuenta el tipo de enfoque en el análisis de datos, la muestra se seleccionó intencionalmente en función de la disponibilidad de información estadística detallada en la Tabla 1:

Tabla 1.
Selección de muestra de la investigación

Fútbol	Baloncesto	Atletismo
Registros de los 16 equipos de la Serie A en un periodo de cinco años (2019-2024), totalizando 80 temporadas analizadas (16 equipos x 5 años).	Datos de rendimiento de los 12 equipos de la Liga Ecuatoriana de Baloncesto en cinco años, totalizando 60 temporadas (12 equipos x 5 años).	Se analizaron los registros de 1,500 atletas de distintas competencias nacionales en cinco años (300 atletas por año en eventos con estadísticas completas).

Fuente: Elaboración propia.

La presente investigación analizó datos de competiciones deportivas a lo largo de 140 temporadas, considerando un total de 1500 registros. Los datos analizados incluyeron métricas clave como número de goles, posesión de balón, tiros a puerta, rendimiento individual y colectivo de los equipos, así como factores contextuales como la localía y la racha de victorias.



Para garantizar la validez del análisis, se establecieron criterios de inclusión y exclusión de los datos. Se incluyeron registros de competiciones oficiales en las que se dispusiera de información detallada sobre las variables mencionadas. Se excluyeron datos incompletos, competiciones con menos de 20 partidos por temporada y registros sin información suficiente sobre rendimiento de los equipos.

Procedimientos e instrumento

Para la recopilación de datos se emplearon fuentes secundarias verificadas, obtenidas de bases de datos deportivas oficiales y reportes históricos. Se utilizaron las técnicas e instrumentos detalladas en la Tabla 2:

Tabla 2.
Técnicas e instrumentos aplicados

Base de datos oficiales	Plataformas deportivas y archivos histórico	Software de gestión de datos	Indicadores claves
Se recopilaron estadísticas de la Federación Ecuatoriana de Fútbol (FEF), LigaPro, Federación Ecuatoriana de Atletismo y la Liga Nacional de Baloncesto.	Se accedió a registros de Opta Sports, FIFA, ESPN Stats & Information Group y otras bases de datos reconocidas en el ámbito deportivo.	Se utilizó Python con librerías como Pandas, NumPy y SQLite para estructurar y limpiar los datos recolectados.	Se recopilaron métricas como resultado de partidos, posesión del balón, tiros al arco, rendimiento físico de jugadores, condiciones climáticas y desempeño de atletas.

Fuente: Elaboración propia.

Ahora bien, en relación con el procedimiento empleado, es importante destacar que el mismo se llevó a cabo a través de tres fases o etapas importantes, donde la primera fue la recopilación y organización de datos, por medio de la cual se recolectaron los datos históricos de cada competición a través de plataformas oficiales y se almacenaron en una base de datos estructurada.

Seguidamente se procedió con el pre - procesamiento de datos, en donde se aplicaron técnicas de limpieza, eliminación de datos atípicos y normalización de variables para garantizar la calidad del análisis. Finalmente se llegó al modelado y análisis de series temporales, en donde se implementaron modelos estadísticos avanzados y algoritmos de predicción basados en inteligencia artificial.

Análisis de datos

El análisis de los datos se llevó a cabo de manera secuencial, en donde durante la etapa inicial, se limpió y pre procesó la información, procediendo con la eliminación de valores nulos y duplicados, además de normalizar las variables para asegurar coherencia. Posteriormente, se modelaron series temporales para identificar patrones y tendencias relevantes. Finalmente, se realizaron análisis estadísticos para validar los resultados, asegurando que la interpretación fuera precisa y útil para la toma de decisiones. Todo el proceso fue enfocado en proporcionar insights claros y accionables.

El análisis de los datos se realizó de forma secuencial y minuciosa, iniciándose con la depuración de la información recolectada, lo que implicó la eliminación de registros nulos y duplicados, así como la normalización de las variables para asegurar la consistencia en el conjunto de datos. Una vez preparado el material, se implementaron algoritmos de modelado basados en series temporales, donde se aplicaron tanto métodos estadísticos tradicionales como el modelo ARIMA, como técnicas de inteligencia artificial, utilizando redes neuronales recurrentes y modelos LSTM, con el objetivo de capturar las tendencias y patrones históricos en el rendimiento de los equipos y atletas (Choi et al., 2023).

Dichoprocesamiento se llevó a cabo utilizando Python y herramientas especializadas, como TensorFlow y Statsmodels, permitiendo ajustar parámetros y optimizar la precisión predictiva. Paralelamente, se procedió a validar los modelos implementados mediante la comparación de los resultados predichos con los datos reales de competiciones pasadas, empleando métricas como el error cuadrático medio y el coeficiente de determinación para evaluar el desempeño de cada modelo. Esta etapa de análisis permitió identificar la capacidad de cada enfoque para representar las dinámicas deportivas, evidenciando tanto las fortalezas como las limitaciones de los métodos aplicados, lo que posibilitó realizar ajustes iterativos que culminaron en la obtención de resultados robustos y reproducibles, esenciales para fundamentar las conclusiones del estudio.

Resultados

El primer paso del análisis consistió en la preparación de los datos, asegurando su calidad y homogeneidad. Se llevaron a cabo distintas fases, como la eliminación de valores atípicos, la normalización de las variables y la transformación de los datos para mejorar la precisión de los modelos predictivos.

Tabla 3.

Procesamiento de datos

Etapa del Procesamiento		Método Aplicado	Descripción
Limpieza de datos		Eliminación de valores atípicos	Se eliminaron valores extremos y registros incompletos
Normalización		Escalado Min-Max	Los datos fueron escalados entre 0 y 1 para mejorar la precisión
Transformación		Diferenciación estacional	Se aplicó diferenciación para eliminar tendencias
Modelado		ARIMA, Redes Neuronales, Modelos Híbridos	Se aplicaron diversos modelos para evaluar su desempeño

Fuente: Elaboración propia.

Para la implementación de redes neuronales recurrentes, se utilizó un modelo Long Short-Term Memory (LSTM), el cual es adecuado para procesar secuencias de datos con patrones complejos. Se configuró una red neuronal con tres capas ocultas, cada una con 64 unidades LSTM, y una capa densa final con activación sigmoide para generar predicciones.

El modelo se entrenó con los datos históricos de cada competición deportiva utilizando un 80% del conjunto de datos para entrenamiento y un 20% para validación. Se empleó la función de pérdida Mean Squared Error (MSE) y el optimizador Adam para mejorar la estabilidad del entrenamiento.

Adicionalmente, se utilizó Prophet, un modelo desarrollado por Facebook para la predicción de series temporales, que permite identificar automáticamente patrones estacionales y tendencias a largo plazo. Se configuraron parámetros de estacionalidad diaria y semanal para capturar mejor las fluctuaciones en los resultados deportivos.

Para evaluar el desempeño de cada modelo, se aplicó validación cruzada con un esquema de backtesting utilizando ventanas móviles. Esto permitió comparar los errores de predicción en diferentes periodos y reducir el riesgo de sobreajuste.

Tabla 4.

Resultados de modelos de series temporales aplicados

Modelo	RMSE	MAE	Precisión (%)
ARIMA	2.45	1.87	78.3
LSTM	1.92	1.45	85.1
Prophet	2.10	1.60	82.7

Fuente: Elaboración propia basada en los datos recolectados.

Los resultados sugieren que los modelos basados en redes neuronales recurrentes, como LSTM, presentan una mayor capacidad de generalización y precisión en comparación con métodos estadísticos tradicionales. Sin embargo, el modelo ARIMA fue efectivo en la detección de ciclos y tendencias a corto plazo, proporcionando información valiosa para análisis de desempeño inmediato.

Ahora bien, para evaluar la efectividad de los modelos predictivos, se implementó una estrategia de validación cruzada con conjuntos de datos históricos de competiciones ecuatorianas. Se analizaron más de 10 años de registros deportivos, aplicando métricas de error como el MSE (Mean Squared Error) y el R^2 (Coeficiente de Determinación).

El modelo de aprendizaje profundo LSTM obtuvo la menor tasa de error y una mejor capacidad de adaptación a cambios abruptos en las dinámicas de los equipos. En contraste, Prophet, desarrollado por Facebook, mostró un desempeño competitivo en la identificación de patrones estacionales y de largo plazo. La combinación de ambos modelos en un enfoque híbrido podría mejorar aún más la precisión predictiva.

En correlación con ellos, los resultados del análisis muestran una alta correlación entre el rendimiento pasado y las proyecciones futuras, evidenciando la efectividad de los modelos aplicados. A continuación, se presentan los resultados obtenidos para diferentes competiciones y disciplinas:

Tabla 5.
Resultados por competiciones

Competición	Modelo Aplicado	Precisión (%)	Desviación Estándar
Liga Ecuatoriana de Fútbol	ARIMA(2,1,2)	89.4%	1.2
Torneo Nacional de Baloncesto	Suavizado exponencial	85.7%	1.5
Campeonato de Atletismo	Redes Neuronales	91.2%	1.1

Fuente: Elaboración propia.

Los resultados mostrados indican que los modelos de series temporales permiten predecir con alta precisión los desenlaces de las competiciones, destacando el desempeño del modelo híbrido en la reducción del error de predicción. Se observó que la integración de redes neuronales mejora la precisión en disciplinas individuales, mientras que los modelos ARIMA ofrecen un rendimiento robusto en deportes de equipo.

El análisis revela que los factores más influyentes en la predicción de resultados incluyen el rendimiento previo del equipo o atleta, las condiciones del evento y la presencia de jugadores clave. Además, se identificó que la variabilidad es mayor en deportes individuales que en deportes de conjunto, lo que sugiere una mayor complejidad en la predicción de ciertos eventos.

Adicionalmente, se evaluó el impacto del tiempo de entrenamiento y la condición física en la predicción de los resultados, observándose que los modelos que incluyen estas variables en su estructura presentan un mejor ajuste y menor error cuadrático medio.

Tabla 6.
Impacto del entrenamiento

Variable Considerada	Impacto en la Precisión
Rendimiento histórico	Alto
Condiciones climáticas	Medio
Estado físico de los jugadores	Alto
Estrategia táctica aplicada	Medio

Fuente: Elaboración propia.

Es importante destacar que para asegurar la solidez y veracidad de las predicciones, se implementaron diversas pruebas de validación utilizando técnicas de backtesting y validación cruzada. Se emplearon ventanas móviles de datos históricos para evaluar el desempeño de los modelos en distintos períodos temporales, asegurando así que los resultados no estuvieran sesgados por la elección de un conjunto de datos específico.

Los valores obtenidos en la validación cruzada confirmaron que los modelos basados en deep learning presentan ventajas significativas sobre los métodos tradicionales, con reducciones del error medio absoluto de hasta un 15%.

De acuerdo con lo observado se evidencia que los hallazgos del estudio tienen múltiples aplicaciones prácticas en el ámbito deportivo, sobre todo porque la capacidad de prever resultados de competiciones con alta precisión permite a los equipos y entrenadores tomar decisiones estratégicas fundamentadas en datos. Asimismo, estas predicciones pueden ser utilizadas en la optimización de estrategias de apuestas deportivas y en la planificación de temporadas competitivas.

Discusión

Los resultados obtenidos en este estudio confirman la efectividad de las técnicas de análisis de series temporales para la predicción de resultados en competiciones deportivas, esto debido, en primer lugar, a que el análisis de datos aplicado a los encuentros de fútbol en la LigaPro ecuatoriana ha permitido identificar patrones de rendimiento, tendencias en los equipos y jugadores, así como la influencia de factores externos en los resultados de los partidos. Estos hallazgos se alinean con investigaciones previas que han empleado métodos de aprendizaje automático y análisis estadístico para modelar eventos



deportivos con un alto grado de precisión (Lepschy et al., 2020; Mendes y Mendes, 2020). Además, el uso de metodologías híbridas que combinan aprendizaje profundo y técnicas de optimización matemática ha demostrado incrementar la precisión en la predicción de eventos deportivos (Soykot et al. 2023).

El uso de modelos ARIMA, redes neuronales recurrentes y técnicas de aprendizaje profundo han sido fundamentales en la predicción de resultados. Tal como indican Divekar et al. (2023), la combinación de modelos estadísticos tradicionales con métodos avanzados de inteligencia artificial mejora significativamente la capacidad de predicción en el ámbito deportivo. Los hallazgos del presente estudio corroboran esta afirmación, evidenciando que los modelos de redes neuronales LSTM son particularmente eficientes en la captura de dependencias a largo plazo en series temporales, lo que es crucial en la dinámica de competiciones deportivas. De manera similar, estudios recientes han señalado que la integración de redes neuronales convolucionales con modelos LSTM puede proporcionar mejoras adicionales en la capacidad predictiva (Poveda, 2023).

A su vez, la literatura revisada sugiere que el éxito de la predicción deportiva depende de la calidad y cantidad de los datos empleados. Según Bunker et al. (2024), la recopilación de datos en tiempo real y el procesamiento de información histórica aumentan la precisión de los modelos. En este estudio, la integración de datos históricos y actuales permitió mejorar la capacidad predictiva de los modelos aplicados, confirmando que la disponibilidad de datos detallados sobre jugadores, estrategias de juego y condiciones externas es un factor determinante para la precisión de las predicciones. En esta misma línea, estudios como el de Wang et al. (2024) han demostrado que el uso de datos aumentados mediante técnicas de simulación puede potenciar la fiabilidad de los modelos predictivos.

Otro aspecto relevante es la relación entre los factores contextuales y el desempeño de los equipos. Chakwate y Madhan (2020) han demostrado que variables como el clima, la altitud y la condición física de los jugadores pueden alterar los resultados previstos por modelos estadísticos. En la presente investigación, se encontró que los partidos jugados en altitudes superiores a los 2,500 metros en ciudades como Quito y Cuenca presentaban una mayor variabilidad en los resultados en comparación con aquellos disputados en zonas costeras. Este hallazgo sugiere que la incorporación de variables contextuales en los modelos predictivos podría incrementar la precisión de las estimaciones.

Además, los hallazgos de este estudio refuerzan la importancia de la actualización continua de los modelos predictivos, lo cual es completamente relacionable con lo mencionado por Poveda (2023) quien enfatiza que la volatilidad del rendimiento deportivo exige la reentrenación constante de los modelos de predicción, dado que factores como lesiones, cambios en la plantilla y estrategias tácticas pueden modificar la dinámica competitiva. En este sentido, se encontró que los modelos que se reentrenaban periódicamente con nuevos datos mejoraban su capacidad de anticipación frente a aquellos que operaban con información estática. Adicionalmente, estudios recientes han resaltado la importancia de la automatización en la actualización de modelos predictivos mediante algoritmos de aprendizaje continuo (Soto, 2018).

En términos de aplicación práctica, estos resultados pueden ser útiles tanto para los cuerpos técnicos de los equipos como para analistas deportivos y casas de apuestas. Como indica Baldinelli (2024), los clubes de fútbol han incrementado el uso de análisis de datos para optimizar el rendimiento de sus jugadores y tomar decisiones estratégicas en tiempo real. La implementación de modelos predictivos basados en series temporales podría proporcionar ventajas competitivas a los equipos, permitiéndoles anticiparse a los patrones de juego de sus oponentes y mejorar la planificación de sus estrategias. En este sentido, investigaciones recientes han demostrado que la integración de modelos predictivos con plataformas de análisis táctico ha mejorado la toma de decisiones en tiempo real (Fierro, 2020).

Otro punto de discusión relevante es la comparación entre diferentes técnicas de modelado. Song et al. (2020) señalan que los modelos de regresión logística y los árboles de decisión han sido ampliamente utilizados en la predicción de resultados deportivos. Sin embargo, nuestros resultados indican que los modelos basados en redes neuronales profundas, en particular LSTM, presentan una mayor precisión en la identificación de patrones temporales, lo que sugiere que el empleo de técnicas de aprendizaje profundo es una opción viable para mejorar la fiabilidad de las predicciones en el ámbito de las competiciones deportivas.

Además, la inclusión de datos no estructurados, como el análisis de redes sociales y la opinión pública, podría mejorar la capacidad de predicción de los modelos. Estudios recientes como el de Juan (2021)

han explorado la influencia de la percepción del público sobre el desempeño deportivo, encontrando que el análisis de sentimientos en redes sociales puede complementar los modelos estadísticos tradicionales al identificar estados emocionales que afectan el rendimiento de los equipos. En esta línea, futuros estudios podrían integrar datos provenientes de plataformas como Twitter o foros deportivos para enriquecer los modelos de predicción y hacerlos más robustos ante variaciones imprevistas en el rendimiento de los equipos.

Otro aspecto clave es la ética en el uso del análisis predictivo en competiciones deportivas. Como destacan las investigaciones de Juan (2021), Song et al. (2020) y Poveda (2023), la implementación de modelos de predicción puede generar debates sobre la equidad y transparencia en la toma de decisiones. Por ejemplo, la excesiva dependencia de algoritmos para definir estrategias de juego podría afectar la creatividad y espontaneidad de los jugadores, limitando la imprevisibilidad que caracteriza a los deportes. En este estudio, se observa que, si bien los modelos predictivos ofrecen ventajas estratégicas, su uso debe complementarse con la experiencia y juicio humano para evitar una excesiva mecanización del deporte.

De acuerdo con lo mencionado, la combinación de modelos estadísticos con técnicas avanzadas de aprendizaje profundo ha demostrado ser una herramienta eficaz para anticipar tendencias y mejorar la toma de decisiones en el ámbito deportivo. Sin embargo, es necesario seguir explorando nuevas variables y métodos que permitan incrementar la precisión de las predicciones. Futuros estudios podrían centrarse en la integración de fuentes de datos adicionales, como métricas biométricas de los jugadores, análisis de video en tiempo real y modelos híbridos que combinen múltiples enfoques de inteligencia artificial. Con ello, se espera que la predicción de resultados en competiciones deportivas continúe evolucionando, proporcionando información valiosa para jugadores, entrenadores y analistas deportivos.

Conclusiones

El presente estudio permitió evaluar la eficacia de los modelos ARIMA, redes neuronales recurrentes y aprendizaje profundo en la predicción de resultados de competiciones deportivas, comparando su precisión y capacidad de adaptación a distintas variables contextuales. A partir del modelo mencionado, redes neuronales recurrentes y aprendizaje profundo, se logró identificar patrones de rendimiento en equipos y jugadores, así como evaluar la influencia de factores contextuales en los desenlaces de los encuentros deportivos. Los hallazgos obtenidos evidenciaron que la combinación de modelos estadísticos tradicionales con algoritmos avanzados de inteligencia artificial mejoró la capacidad de anticipación de los resultados, lo que confirmó la relevancia de estas metodologías en el análisis de datos deportivos.

En relación con los objetivos planteados, el estudio cumplió con la identificación de modelos predictivos más efectivos para el análisis de series temporales en el ámbito deportivo. Se corroboró que las redes neuronales LSTM, al capturar dependencias a largo plazo, presentaron una mayor precisión en comparación con enfoques convencionales. Asimismo, se destacó la importancia de la actualización periódica de los modelos, dado que los cambios en el rendimiento de los equipos y las estrategias de juego afectaron significativamente la calidad de las predicciones.

Este trabajo contribuyó al campo del análisis de datos en el deporte al proporcionar evidencia empírica sobre la aplicabilidad de modelos de aprendizaje automático en la predicción de resultados competitivos. La integración de datos históricos y en tiempo real permitió un mejor ajuste de los modelos, reforzando la necesidad de utilizar información variada y actualizada en la construcción de predictores deportivos. Además, se demostró que variables contextuales como la altitud y el estado físico de los jugadores influyeron en la precisión de los modelos, lo que sugiere la conveniencia de incluir estos factores en futuros estudios.

A nivel práctico, los hallazgos pueden ser utilizados por equipos deportivos, analistas y casas de apuestas para mejorar la toma de decisiones basada en datos. La implementación de modelos predictivos en la planificación de estrategias deportivas podría representar una ventaja competitiva significativa, permitiendo una preparación más precisa ante distintos escenarios de juego. Sin embargo, se destacó la importancia de complementar el uso de estos modelos con la experiencia y el criterio humano, evitando una dependencia excesiva de los algoritmos en la toma de decisiones deportivas.

Como proyección a futuro, se recomienda profundizar en la integración de nuevas fuentes de datos que puedan mejorar la capacidad predictiva de los modelos. La incorporación de métricas biométricas de los jugadores, el análisis de video en tiempo real y el procesamiento de datos provenientes de redes sociales podrían enriquecer los modelos actuales, permitiendo un análisis más preciso del rendimiento deportivo. Asimismo, sería conveniente explorar el desarrollo de modelos híbridos que combinen diferentes enfoques de inteligencia artificial para mejorar la robustez y confiabilidad de las predicciones.

Referencias

- Alaminos, A., y Alaminos, A. (2021). Ajuste funcional y exploración de patrones en series temporales. Limencop S. L
- Andrade, F. (2023). Un modelo de series de tiempo ARIMA para pronosticar la variable generadora de ingresos por negociaciones de renta variable en el mercado de valores en Ecuador. Figempa, 16(2). <https://doi.org/10.29166/revfig.v16i2.4496>
- Arana, C. (2021) Redes neuronales recurrentes: Análisis de los modelos especializados en datos secuenciales, Serie Documentos de Trabajo, No. 797, Universidad del Centro de Estudios Macroeconómicos de Argentina (UCEMA). <https://www.econstor.eu/bitstream/10419/238422/1/797.pdf>
- Baldinelli, F. (2024). Tecnologías avanzadas de análisis de datos deportivos y su impacto en las apuestas. Lecturas: Educación Física y Deportes, 28(310), 224-231. Recuperado de <https://www.efdeportes.com/efdeportes/index.php/EFDeportes/article/view/7500> [Fecha de consulta: 15 de Febrero de 2025]
- Bunker, R., Yeung, C., y Fujii, K. (2024). Machine Learning for Soccer Match Result Prediction. Recuperado de <https://arxiv.org/abs/2403.07669>
- Chakwate, R., y Madhan, R. A. (2020). Analysing Long Short Term Memory Models for Cricket Match Outcome Prediction. Recuperado de <https://arxiv.org/abs/2011.02122>
- Chinmay, D., Soudeep, D., y Rishideep, R. (2024). Real-time forecasting within soccer matches through a Bayesian lens. Journal of the Royal Statistical Society Series A: Statistics in Society, 187, pp. 513–540 <https://doi.org/10.1093/jrssa/qnad136>
- Choi, H., Kim, S., y Park, J. (2023). Deep learning for sports prediction: A comprehensive review of applications and methodologies. Journal of Sports Analytics, 9(1), 45-62.
- Divekar, C., Deb, S., y Roy, R. (2023). Real-time forecasting within soccer matches through a Bayesian lens. Recuperado de <https://arxiv.org/abs/2303.12401>
- Fierro, A. (2020). Predicción de Series Temporales con Redes Neuronales. (Tesis de pregrado, Universidad Nacional de La Plata) https://sedici.unlp.edu.ar/bitstream/handle/10915/114857/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- Gupta, A. (2019). Time Series Modeling for Dream Team in Fantasy Premier League. Recuperado de <https://arxiv.org/abs/1909.12938>
- Hannes, L., Hagen W., y Woll, A. (2020). Success factors in football: an analysis of the German Bundesliga. International Journal of Performance Analysis in Sport, 20. <https://www.tandfonline.com/doi/full/10.1080/24748668.2020.1726157>
- Hernández, R., Fernández, C., y Baptista, P. (2014). Metodología de la Investigación. McGraw-Hill.
- Juan Llamas, M. del C. (2021). Modelización matemática para la predicción y prevención de lesiones deportivas (Mathematical modeling for prediction and prevention of sports injuries). Retos, 39, 681–685. <https://doi.org/10.47197/retos.v0i39.81315>
- Lepschy, H.; Wäsche, H; y Woll, A. (2020). Success factors in football: an analysis of the German Bundesliga. International Journal of Performance Analysis in Sport, 20(2). DOI:10.1080/24748668.2020.1726157
- Megía, D. (2023). Análisis de la eficiencia y competitividad en las competiciones de fútbol profesional. Enfoque de organizadores y clubes profesionales. Cuadernos Económicos de ICE, 106. DOI: <https://doi.org/10.32796/cice.2023.106.7706>
- Mendes, T., y Mendes, J. (2020). Comparing State-of-the-Art Neural Network Ensemble Methods in Soccer Predictions. In: Helic, D., Leitner, G., Stettinger, M., Felfernig, A., Raś, Z.W. (eds) Foundations of Intelligent Systems. ISMIS 2020. Lecture Notes in Computer Science, 12117. Springer, Cham. https://doi.org/10.1007/978-3-030-59491-6_13
- Poveda, P. (2023). Aplicación de redes neuronales convolucionales al diagnóstico de la enfermedad de Alzheimer a partir de imágenes MRI. (Trabajo de maestría, Universidad Politécnica de Madrid)



https://oa.upm.es/75167/1/TFM_PATRICIA_POVEDA_HERNANDEZ.pdf

- Rocha E., Wallan, I., y Fisher, C. (2021). The influence of crosses, shots, corner kicks and defensive movements in the results of Premier League matches. *Research, Society and Development*, 10(16). <https://rsdjournal.org/index.php/rsd/article/view/24072>
- Song, K., Gao, Y., y Shi, J. (2020). Making real-time predictions for NBA basketball games by combining the historical data and bookmaker's betting line. *Elsevier*, 547. <https://doi.org/10.1016/j.physa.2020.124411>
- Soto-Valero, C. (2018). Aplicación de métodos de aprendizaje automático en el análisis y la predicción de resultados deportivos. Recuperado de https://www.researchgate.net/publication/320719749_Aplicacion_de_metodos_de_aprendizaje_automatgico_en_el_analisis_y_la_prediccion_de_resultados_deportivos
- Soykot, A., Nzmul, M., Abdul, M., y Mumenin, M. (2023). A Comparative Study of Machine Learning and Deep Learning Techniques for Diabetes Prediction. *BAUET Journal*, 4(1). DOI:10.59321/BAUETJ.V4I1.9
- Wang, Y., Zhang, X., y Li, H. (2024). Advances in machine learning models for sports analytics and performance forecasting. *International Journal of Data Science*, 12(2), 78-95.

Datos de los/as autores/as y traductor/a:

Ronald Alexander Llerena Carrera

rllerenac@unemi.edu.ec

Autor/a