

ENTREGA 1 - Predicción de Resultados en Fútbol Internacional usando Deep Learning

José Alfredo Martínez Valdés*, Dorian Alexander Jaramillo Rivas†, Wilmer Mario Leiva Esteban‡

Facultad de Ingeniería, Ingeniería de Sistemas

Universidad de Antioquia

Fundamentos de Deep Learning

Medellín, Colombia

*jose.martinez7@udea.edu.co

†dorian.jaramillo@udea.edu.co

‡wilmer.leiva@udea.edu.co

I. CONTEXTO DE APLICACIÓN

El fútbol es uno de los deportes más populares a nivel mundial, con millones de partidos jugados a lo largo de la historia. La predicción de resultados en competiciones internacionales tiene relevancia tanto académica como práctica, dado que impacta en áreas como el análisis deportivo, las casas de apuestas y la gestión de estrategias de juego.

El dataset *International football results from 1872 to 2017*, disponible en Kaggle, contiene más de 40,000 registros de partidos internacionales de selecciones nacionales, incluyendo información sobre equipos, fechas, sedes y resultados [1].

Además, se han utilizado varios notebooks de Google Colab como referencia técnica para implementar experimentos, procesamiento de datos y visualizaciones de modelos en este proyecto [2], [3], [4], [5], [6], [7], [8].

II. OBJETIVO DE MACHINE LEARNING

El objetivo de este proyecto es construir un modelo de aprendizaje profundo capaz de predecir el resultado de un partido internacional (victoria local, empate o victoria visitante), dadas características estáticas (equipos, sede) y secuenciales (rendimiento histórico de los equipos en partidos previos).

Formalmente, se busca aproximar una función:

$$f : (X_{static}, X_{seq}) \rightarrow y$$

donde X_{static} representa información como país local, visitante y estadio, X_{seq} resume el rendimiento histórico reciente (goles anotados, recibidos, resultados pasados), y $y \in \{0, 1, 2\}$ indica el resultado del partido.

III. DATASET

El dataset proviene de Kaggle e incluye partidos desde 1872 hasta 2017 [1].

- **Tipo de datos:** Tabular, con variables categóricas (equipos, torneos, países) y numéricas (goles anotados, año).
- **Tamaño:** Aproximadamente 43,000 partidos, archivo CSV de ~ 3 MB.
- **Variables principales:** fecha, equipo local, equipo visitante, goles local, goles visitante, sede, torneo.

- **Distribución de clases:** Históricamente, la proporción aproximada es: 46% victorias locales, 25% empates y 29% victorias visitantes, lo cual refleja una distribución desbalanceada.

IV. MÉTRICAS DE DESEMPEÑO

Se emplearán dos tipos de métricas:

A. Métricas de machine learning

- **Exactitud (Accuracy):** Medirá el porcentaje total de predicciones correctas.
- **F1-score macro:** Para mitigar el desbalance de clases y evaluar el rendimiento equilibrado entre victoria local, empate y visitante.
- **Matriz de confusión:** Para analizar errores específicos (ej. confusión entre empate y victoria visitante).

B. Métricas de negocio

- **Precisión en victorias locales:** relevante para análisis de equipos anfitriones.
- **Predicción de empates:** especialmente útil en apuestas deportivas, donde empatar es un evento de baja probabilidad pero alto valor.

V. REFERENCIAS Y RESULTADOS PREVIOS

Existen múltiples estudios en la literatura académica que exploran la predicción de resultados en fútbol mediante técnicas estadísticas y de aprendizaje automático.

En [9] se evaluaron diferentes modelos de machine learning aplicados al fútbol holandés, concluyendo que los modelos de clasificación logística y árboles de decisión ofrecen un rendimiento aceptable pero limitado en tareas predictivas.

Investigaciones recientes muestran que las redes neuronales recurrentes (RNN, LSTM y GRU) y arquitecturas híbridas (CNN-LSTM, Transformers) permiten modelar secuencias de partidos con mayor precisión [10], [11].

Además, los notebooks de Colab usados en este proyecto documentan paso a paso los procesos de análisis exploratorio, preprocesamiento, implementación de modelos y evaluación experimental [2], [3], [4], [5], [6], [7], [8].

REFERENCES

- [1] *International football results from 1872 to 2017*, Dataset, Kaggle, Accedido el 2 de octubre de 2025, 2018. [Online]. Available: <https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017>.
- [2] “Notebook colab – experimento 1,” Accessed: Oct. 2, 2025. [Online]. Available: <https://colab.research.google.com/drive/15vpPGMI8PTqjWUYNNOM9zCkZSmR2h2Cq?usp=sharing>.
- [3] “Notebook colab – experimento 2,” Accessed: Oct. 2, 2025. [Online]. Available: https://colab.research.google.com/drive/1Cahvog-u-L3QFR6g4nKqyt3405tKN_1H?usp=sharing.
- [4] “Notebook colab – experimento 3,” Accessed: Oct. 2, 2025. [Online]. Available: <https://colab.research.google.com/drive/1Iyi4g4w1k65FRxyGJyKsHnvm3fDWNA4f?usp=sharing>.
- [5] “Notebook colab – experimento 4,” Accessed: Oct. 2, 2025. [Online]. Available: <https://colab.research.google.com/drive/1TBpNKa-Tht9IB27EIMM2ITR9THaAHend?usp=sharing>.
- [6] “Notebook colab – experimento 5,” Accessed: Oct. 2, 2025. [Online]. Available: <https://colab.research.google.com/drive/1Q9qnKpH2yFsp0QNr6kkrJni1LKqTbmye?usp=sharing>.
- [7] “Notebook colab – experimento 6,” Accessed: Oct. 2, 2025. [Online]. Available: <https://colab.research.google.com/drive/187BhRFabGzCiWccRMbZaw3cA2Ppp4XvK?usp=sharing>.
- [8] “Notebook colab – experimento 7,” Accessed: Oct. 2, 2025. [Online]. Available: <https://colab.research.google.com/drive/1Gx9Rq11XJdMlIOIkTkN6oR7taaQc7G8M?usp=sharing>.
- [9] N. Tax and Y. Joustra, *Predicting football match results with machine learning*, University of Amsterdam, 2016. [Online]. Available: <https://www.researchgate.net/publication/307864150>.
- [10] R. P. Bunker and F. Thabtah, *A machine learning framework for sport result prediction*, Applied Computing and Informatics, 2019. DOI: 10.1016/j.aci.2018.12.002.
- [11] D. Rudrapal, S. Boro, J. Srivastava, and S. Singh, “A deep learning approach to predict football match result,” in *Advances in Data and Information Sciences*, C. H. V. Reddy, S. Bhatia, M. Tiwari, and V. Singh, Eds. Springer, Singapore, Jan. 2020, pp. 93–99, ISBN: 978-981-13-8675-6. DOI: 10.1007/978-981-13-8675-6_3_9.