

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

José Alfredo Martínez Valdés\*, Dorian Alexander Jaramillo Rivas<sup>†</sup>, Wilmer Mario Leiva Esteban<sup>‡</sup>

Facultad de Ingeniería, Ingeniería de Sistemas

Universidad de Antioquia

Fundamentos de Deep Learning

Medellín, Colombia

\*jose.martinez7@udea.edu.co

<sup>†</sup>dorian.jaramillo@udea.edu.co

<sup>‡</sup>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].

## 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  $\sim 3$  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 [2] 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 [3], [4].

Estos hallazgos sugieren que enfoques basados en deep learning superan a los métodos tradicionales en la predicción de resultados deportivos.

## REFERENCES

- [1] Martj42, "International football results from 1872 to 2017," Kaggle, 2018, <https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017>.
- [2] N. Tax and Y. Joutsa, "Predicting football results using machine learning techniques," *Machine Learning and Data Mining for Sports Analytics*, 2017.

- [3] R. Bunker and T. Susnjak, "A survey of sports prediction using machine learning," *Journal of Artificial Intelligence Research*, vol. 66, pp. 341–376, 2019.
- [4] A. Constantinou and N. Fenton, "Football match outcome prediction: a systematic review," *Statistics in Sports*, 2022.