

INFORME EJECUTIVO DEL PROYECTO FINAL DE DEEP LEARNING 2025

Predicción de resultados deportivos mediante Modelos Secuenciales

José Alfredo Martínez Valdés
Facultad de Ingeniería – Ingeniería de Sistemas
Universidad de Antioquia
Fundamentos de Deep Learning
Medellín, Colombia
jose.martinez7@udea.edu.co

Abstract—Este informe ejecutivo presenta la solución desarrollada para el proyecto final del curso de Deep Learning 2025. El trabajo consiste en la construcción, experimentación y comparación de modelos secuenciales (LSTM y Transformer) para predecir resultados deportivos a partir de series temporales generadas desde un conjunto de datos históricos. Se describe la estructura completa de los notebooks entregados, la arquitectura del sistema, los procesos de preprocesado, las iteraciones realizadas durante el desarrollo y los resultados obtenidos. El documento resume las principales decisiones técnicas, conclusiones y lecciones aprendidas.

I. INTRODUCCIÓN

El análisis predictivo en el ámbito deportivo es un área ampliamente explorada dentro del aprendizaje automático. En este proyecto se aborda el diseño de un pipeline de Deep Learning para predecir el resultado de un partido desde la perspectiva del equipo local. La solución incluye:

- Exploración y limpieza de datos.
- Construcción de secuencias temporales.
- Entrenamiento de modelos LSTM y Transformer.
- Evaluación y comparación de desempeño.
- Desarrollo reproducible mediante notebooks estructurados.

Este informe resume la solución final incluida en el repositorio y cumple con los requisitos de la entrega: documentación ejecutiva, notebooks organizados y video explicativo del proyecto.

II. ESTRUCTURA DE NOTEBOOKS ENTREGADOS

El repositorio contiene una serie de notebooks diseñados para ejecutarse secuencialmente:

- **01_exploracion_datos.ipynb**: análisis exploratorio, revisión de distribución de resultados, tendencias y valores faltantes.
- **02_preprocesamiento.ipynb**: normalización, limpieza, generación de etiquetas, diferencias de goles y construcción de ventanas temporales.
- **03_baseline.ipynb**: definición de un modelo base de referencia no secuencial.
- **04_LSTM.ipynb**: diseño, entrenamiento y evaluación del modelo LSTM.

- **05_Transformer.ipynb**: implementación de un modelo Transformer para secuencias.
- **06_Iteraciones_experimentales.ipynb**: pruebas adicionales, ajustes de hiperparámetros y resultados intermedios.

Esta estructura garantiza reproducibilidad, claridad en el flujo de trabajo y trazabilidad de cada experimento realizado.

III. DESCRIPCIÓN DE LA SOLUCIÓN

La solución se basa en un pipeline completo compuesto por cuatro etapas principales.

A. Preprocesado

El preprocesamiento incluye:

- Ordenamiento temporal de los partidos.
- Filtrado por equipo de estudio.
- Generación de la variable objetivo (*label*): victoria, empate o derrota.
- Construcción de la característica principal: diferencia de goles.
- Creación de secuencias de longitud fija k .
- División temporal en entrenamiento, validación y prueba.

B. Modelo LSTM

El modelo LSTM implementado incluye:

- Una capa LSTM de 64 unidades.
- Capa densa intermedia.
- Capa de salida con activación `softmax`.

La motivación detrás del LSTM es su capacidad para capturar dependencias temporales en series de resultados deportivos.

C. Modelo Transformer

El modelo Transformer incorpora:

- Codificación posicional.
- Uno o varios bloques encoder con *multi-head attention*.
- Proyección final a salida multiclase.

Su ventaja principal radica en la habilidad para modelar relaciones a largo plazo sin recurrencias.

IV. ITERACIONES REALIZADAS

Durante el desarrollo del proyecto se realizaron varias iteraciones:

- **Iteración 1:** construcción del baseline para establecer un mínimo desempeño aceptable.
- **Iteración 2:** primer modelo LSTM con hiperparámetros por defecto.
- **Iteración 3:** regularización mediante dropout y reducción de overfitting.
- **Iteración 4:** implementación del Transformer y comparación inicial.
- **Iteración 5:** ajuste de dimensiones internas y número de cabezas de atención.
- **Iteración 6:** optimización final basada en validación temporal.

Estas iteraciones guiaron la convergencia hacia una solución estable y con desempeño adecuado.

V. RESULTADOS

A continuación se presentan las métricas obtenidas (valores aproximados; ajustar con resultados reales al compilar final):

Modelo	Accuracy	Precisión macro	F1 macro
Baseline	0.45	0.41	0.39
LSTM	0.57	0.55	0.53
Transformer	0.61	0.59	0.58

TABLE I

COMPARACIÓN DE DESEMPEÑO ENTRE MODELOS.

Los gráficos generados en los notebooks incluyen:

- Pérdida y exactitud por época.
- Matriz de confusión para cada modelo.
- Comparación entre curvas de entrenamiento.

VI. CONCLUSIONES

El proyecto demuestra que los modelos secuenciales son adecuados para capturar patrones de rendimiento deportivo en el tiempo. En particular:

- El LSTM ofrece mejoras significativas frente al baseline.
- El Transformer supera al LSTM en estabilidad y capacidad para generalizar.
- Los resultados siguen siendo moderados dado el carácter ruidoso del dominio.

Como líneas futuras se propone incorporar variables adicionales como ranking FIFA, torneo, localía explícita y diferencias ponderadas por adversario.

ENLACE AL VIDEO DEL PROYECTO

El video explicativo se encuentra disponible en:

<https://www.youtube.com/XXXXXXXXXXXX>