

Entrega 1

Deep Learning

I. Contexto

El éxito comercial y crítico de una película es un fenómeno complejo influenciado por múltiples factores: el argumento, el género, el director, el reparto, el presupuesto y las tendencias culturales de cada época. Sin embargo, ciertos patrones narrativos y estilísticos pueden correlacionarse con un mayor atractivo para el público. En este proyecto, nos enfocamos en analizar sinopsis, géneros y directores como variables principales para entrenar un modelo de *deep learning* capaz de predecir el potencial de éxito de una película. Este enfoque no solo puede ser útil para estudios de cine y productores al evaluar proyectos en fase de desarrollo, sino también para plataformas de streaming que deseen anticipar la recepción de sus contenidos.

II. Objetivo

Queremos desarrollar un modelo de aprendizaje profundo supervisado que, dado el resumen (texto), el género y el director de una película, prediga su nivel de éxito representado como una puntuación del 1 al 10. El objetivo final es:

- Establecer vínculos entre el argumento y el éxito potencial.
- Detectar patrones narrativos o temáticas recurrentes asociadas a una buena recepción.
- Integrar variables adicionales (año de estreno, duración, premios, etc.) para aumentar la precisión.

III. Dataset

El conjunto de datos principal proviene de Kaggle:
[IMDb Dataset of Top 1000 Movies and TV Shows](#)

- **Tamaño:** 1000 películas y series.
- **Características incluidas:**
 - Nombre de la obra.
 - Año de estreno.
 - Duración.
 - Género(s).
 - Sinopsis (texto en lenguaje natural).

- Director.
- Reparto principal.
- Banda sonora / puntuación IMDb (éxito percibido).
- **Formato:** CSV (aprox. 250 KB).
- **Limitaciones:** El tamaño del dataset (1000 entradas) puede resultar insuficiente para entrenar un modelo de *deep learning* complejo. En consecuencia, se considerará ampliar la base de datos con otros datasets públicos (ej. IMDb datasets oficiales, TMDb, u otros conjuntos de Kaggle con decenas de miles de películas).

IV. Métricas

Métricas de machine learning:

- RMSE (Root Mean Squared Error): para evaluar la precisión de las predicciones numéricas en la escala 1–10.
- MAE (Mean Absolute Error): para medir el error medio y facilitar la interpretación.
- R^2 (coeficiente de determinación): para medir la proporción de varianza explicada.

Métricas de negocio:

- Capacidad de identificar películas con alto potencial de éxito antes de su estreno.
- Detección de géneros o patrones narrativos emergentes.
- Utilidad práctica para productores, distribuidores y plataformas de streaming.

V. Referencias

Kaggle Dataset: IMDb Dataset of Top 1000 Movies and TV Shows (Harshit Shankhdhar, 2022).

Trabajos relacionados:

- Experimentos previos de **sentiment analysis en reseñas de IMDb** (e.g., *IMDb Movie Reviews Dataset*).
- Modelos de predicción de taquilla basados en metadatos de películas.
- Aplicaciones de *Natural Language Processing (NLP)* para analizar tramas y sinopsis.