Informe Final - Proyecto 3

Sistema de Recomendación Híbrido con Deep Learning

# Integrantes

* Alejandra Montenegro
* Maria Sol Vásquez
* Maria Fernanda Aguirre
* Ricardo Garcia
* Yadira Paguay
* Sebastián Díaz

# 1. Introducción

Este proyecto tiene como objetivo desarrollar un sistema de recomendación híbrido con deep learning utilizando el dataset MovieLens. A lo largo del proyecto, se han implementado múltiples componentes avanzados, incluyendo filtrado colaborativo profundo (NeuralMF), análisis basado en contenido (TF-IDF + posters), factores contextuales (interacción, rating, tiempo de visualización), análisis secuencial (LSTM), aprendizaje multitarea, API REST con FastAPI, métricas online y A/B testing. El objetivo es maximizar la precisión de las recomendaciones y la relevancia contextual para usuarios en entornos reales.

# 2. Arquitectura del Sistema

El sistema de recomendación se compone de los siguientes módulos:  
- NeuralMF (Filtrado Colaborativo Profundo)  
- TF-IDF + CNN (Contenido textual y visual)  
- Análisis temporal con LSTM  
- Factores contextuales (ratings, tiempo de visualización, interacción)  
- Aprendizaje multitarea (recomendación + predicción de métricas online)  
- Exposición mediante API REST con FastAPI y dashboard con Streamlit

# 3. Entrenamiento del Modelo

La red se entrenó utilizando MSE como función de pérdida para la tarea principal de predicción de rating. En el enfoque multitarea, se añadieron cabezales adicionales para predecir el tiempo de visualización y nivel de interacción. El modelo fue entrenado por 20 épocas con early stopping.

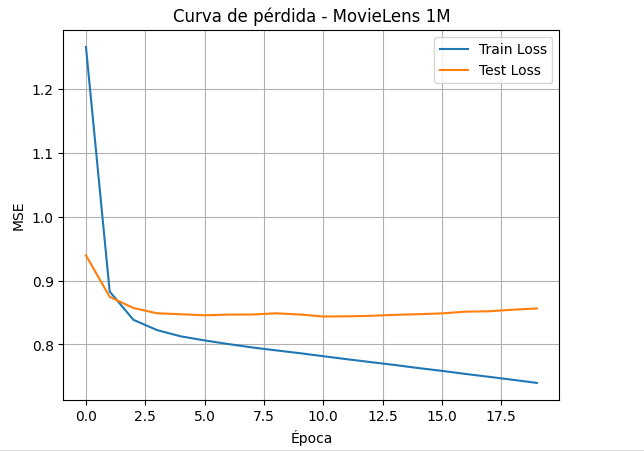


Figura 1. Curva de pérdida para entrenamiento y validación

# 4. Resultados del Sistema de Recomendación

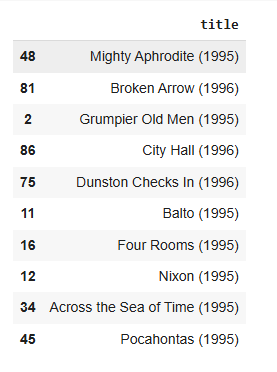


Figura 2. Películas con mayor nivel de recomendación para un usuario.



Figura 3. Recomendaciones visuales usando similitud de vectores de CNN sobre posters.

# 5. Resultados Online y Visualización de Métricas

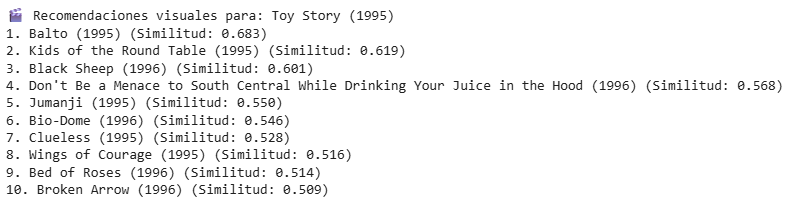


Figura 4. Comparación A/B de tiempo de visualización.

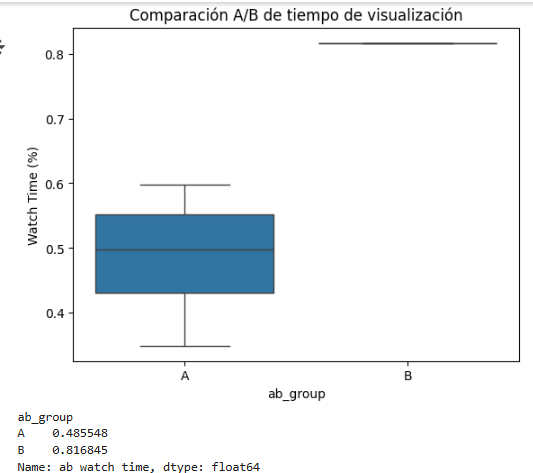


Figura 5. Distribución de Watch Time y Ratings en usuarios.

# 6. API REST para Producción

Se construyó una API REST con FastAPI y fue desplegada en Google Colab utilizando Ngrok para generar un endpoint público. La API cuenta con los siguientes endpoints:  
- `/recommend/{user\_id}`: Recomendaciones personalizadas  
- `/popular`: Películas populares  
- `/poster/{movie\_id}`: Poster de película  
- `/metrics`: Métricas promedio online  
- `/ab\_test`: Resultados de prueba A/B  
  
Endpoint activo de prueba: https://6eb8d6cf9e17.ngrok-free.app/

# 7. Conclusiones

El sistema de recomendación desarrollado combina múltiples fuentes de información y técnicas de deep learning para generar recomendaciones personalizadas, visualmente enriquecidas y contextualmente relevantes. Gracias al aprendizaje multitarea y a la integración de métricas online, se optimiza no solo la precisión de las predicciones, sino también el engagement real del usuario. Esta arquitectura permite escalar el sistema a producción a través de una API REST y monitorizar el rendimiento mediante dashboards interactivos.