

ANÁLISIS DE DATOS MASIVOS

MINI PROYECTO 4

13 de noviembre de 2024

El objetivo de este miniproyecto es desarrollar un sistema de recomendación usando la base de datos que te fue asignada. De los métodos que se analizaron dentro de la unidad, selecciona el método de tu preferencia es importante que justifiques el método de tu selección.

Por ejemplo, en algunas base de datos es más apropiado usar filtrado colaborativo , sin embargo en algunas bases de datos este método no es el ideal. Para la implementación del proyecto, debes usar pyspark.

- Abigail: Last.FM
- Isaac: 50,000+ Indian Movies Data set
- Bruno: MovieLens 1M Dataset

El proyecto se evaluará con los siguientes puntos:

- Análisis exploratorio/ pre-procesamiento de los datos: 2 puntos
- Justificación e implementación del algoritmo: 3 puntos
- Visualización de los resultados: 2 puntos
- Análisis y redacción de los resultados: 2 puntos
- Documentación del código: 1 punto

- Fecha límite: **viernes 29 de noviembre**
- Forma de entrega: el miniproyecto se presentará durante la clase (puedes usar Databricks o Google Colab).
- Trabajo individual
- Recuerda enviar la libreta por correo antes de la fecha límite.