Phase-4-Project-Movie-Recommendation-System

This project delves into the classic MovieLens dataset to build a robust movie recommendation system. Leveraging various machine learni and User-Based Collaborative Filtering as well as following the CRISP-DM methodology, the aim is to predict user preferences and deliver the end-to-end process from data understanding to model evaluation, emphasizing model training best practices with pipelines and hype

1. Business Understanding

The core objective is to develop a system that provides users with tailored movie recommendations based on their past ratings. Imagine tl engagement by surfacing films users are most likely to enjoy.

2. Data Understanding

The data source is https://grouplens.org/datasets/movielens/latest/ Initial exploration of the MovieLens "small" dataset (movies.csv, rating

- Movies: Over 9,700 titles with genres.
- Ratings: Over 100,000 ratings from 610 users, showcasing a clear positive bias towards higher scores (3.0, 4.0, 5.0 are most common).
- Popularity Skew: A few blockbusters receive a disproportionately high number of ratings, while many films have very few. Visual analy reviewed movies can be misleadingly extreme, necessitating data filtering.

Data Challenges

- Filtering is needed to exclude movies with few ratings(<20) to mitigate noice in recommendation mode
- Potential bias with popular movies dominating ratings --> niche titles maybe underrepresented.