RECOMMENDATION SYSTEM FOR TOP MOVIES

GROUP 4

HELLEN MUANKI
JESICAH MUTISO
ENDALKACHEW DESSALEGNE
BRIAN WAWERU

DATE: 20TH JUNE 2025

PROJECT GOAL

The goal of this project is to build a personalized recommendation system for the users of a movie streaming company.

Outcome: To improve user engagement and retention through tailored movie suggestions

BUSINESS PROBLEM AND OBJECTIVE

Users are overwhelmed with choices in a competitive streaming market hence the need for a recommendation system to guide them in content discovery.

Objectives:

- Boost engagement by offering tailored movie suggestions.
- Increase user retention through consistent, relevant recommendations.
- Solve the cold start problem for new users with no ratings.

DATA OVERVIEW

Source of dataset: MovieLens website

4 datasets of which 2 were used - movies dataset and rating dataset

Key variables:

- User ID
- Movie ID
- Ratings
- Genres

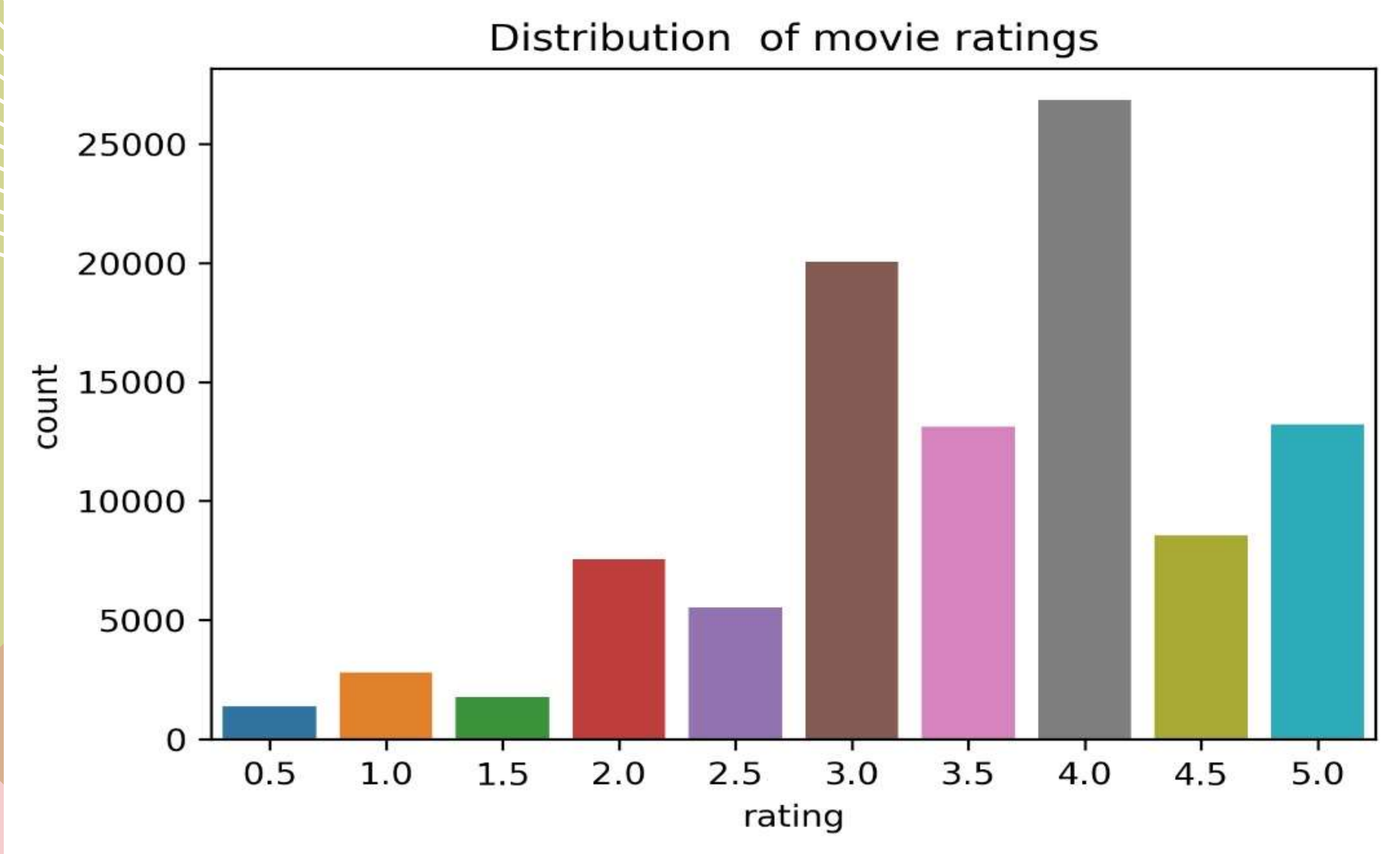
METHODOLOGY

- Exploratory data analysis
- Data cleaning and preprocessing
- Model Development
  - Collaborative Filtering(SVD)
  - Content based Recommendation
- Evaluation Metrics(RMSE, Precision)

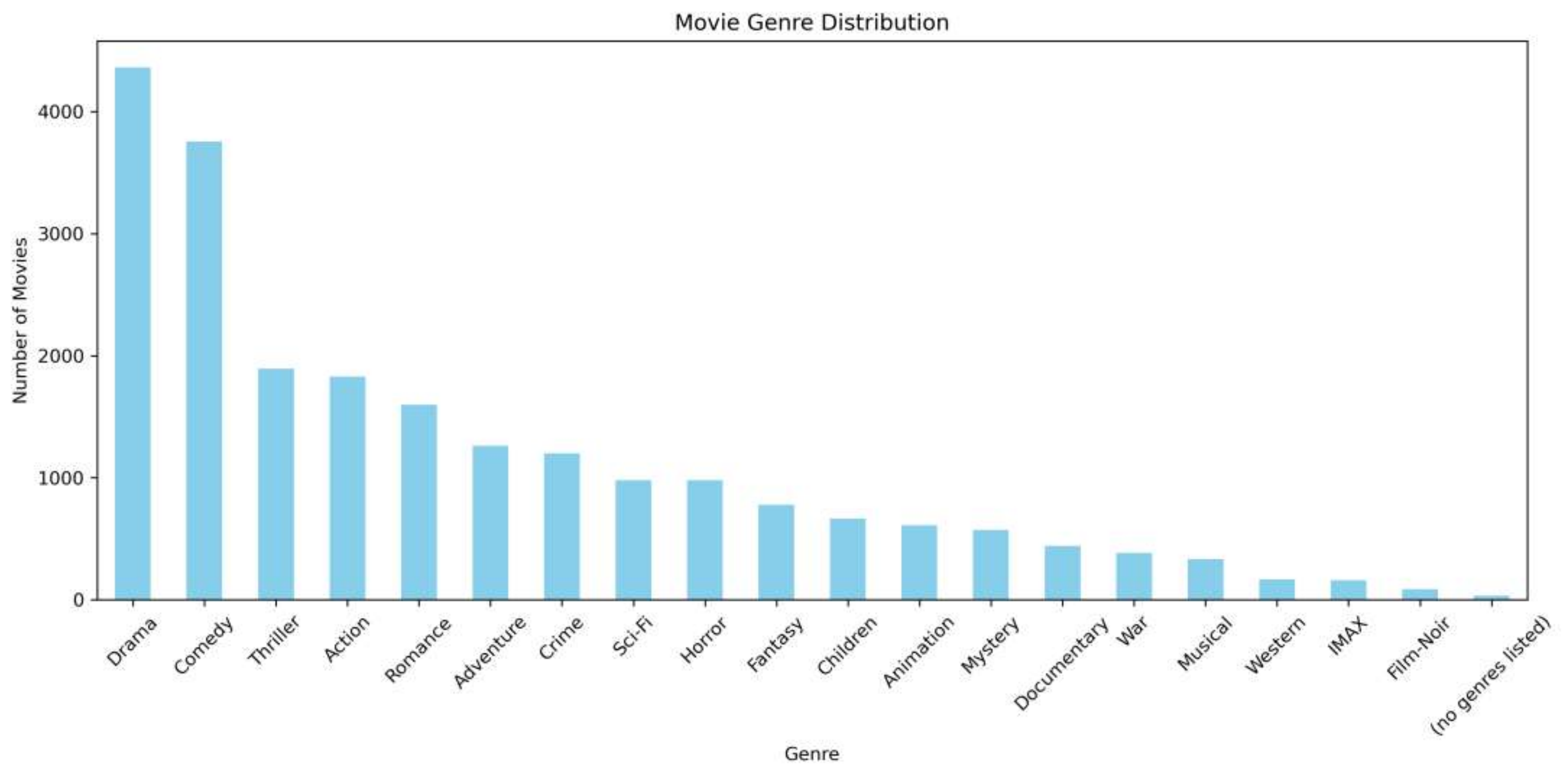
EXPLORATORY INSIGHTS

- Entries in the dataset- 100836
- Users in the dataset- 610
- Number of movies- 9724
- Average number of ratings per user: 165.3
- Average number of ratings per movie: 10.37
- Average rating- 3.5

DISTRIBUTION OF MOVIE RATINGS



MOVIE GENRE DISTRIBUTION



RECOMMENDATION MODELS

- **Collaborative Filtering(SVD)**
- Learns user preferences based on similar users
- Personalized top 5 movie list per user

- **Content-based recommendation**
- Recommends top-rated movies overall
- Recommends similar movies for a new user –based on genres
- This model addresses cold start (when user has not rated movies yet)
- Used TF-IDF vectorization

RECOMMENDATION MODELS

- **Hybrid CF-CBF Recommendation Model**
- Normalizes CF and CBF scores for a common scale
- Blend the two models using a weighting factor alpha(e.g., 0.5)
- Sorts and returns top 5 movie recommendations

MODEL PERFOMANCE

- We implemented Collaborative Filtering(CF) using SVD from the Surprise library. It performs better for active users.
- Evaluation matrix RMSE = 0.87 for Collaborative Filtering is achieved. This is acceptable for rating prediction.
- Collaborative Filtering is ideal for personalized recommendations once the user has rated many items/movies.
- We used TF-IDF vectorization of genre for Content-Based Filtering(CBF) to recommend top N unseen movies for all users, addresses cold-start users

BUSINESS IMPACT

- Improved user experience
- Higher Retention and engagement
- Competitive edge in the market
- Foundation for advanced features e.g. hybrid systems

RECOMMENDATIONS

- Hyper-parameter tuning for CF to improve RMSE
- To include more features such as movie description for TF-IDF to improve the model.
- Optimize the SVD model by using different similarity measures
- Optimize alpha to balance the right weight of CF and CBF

THANK YOU!