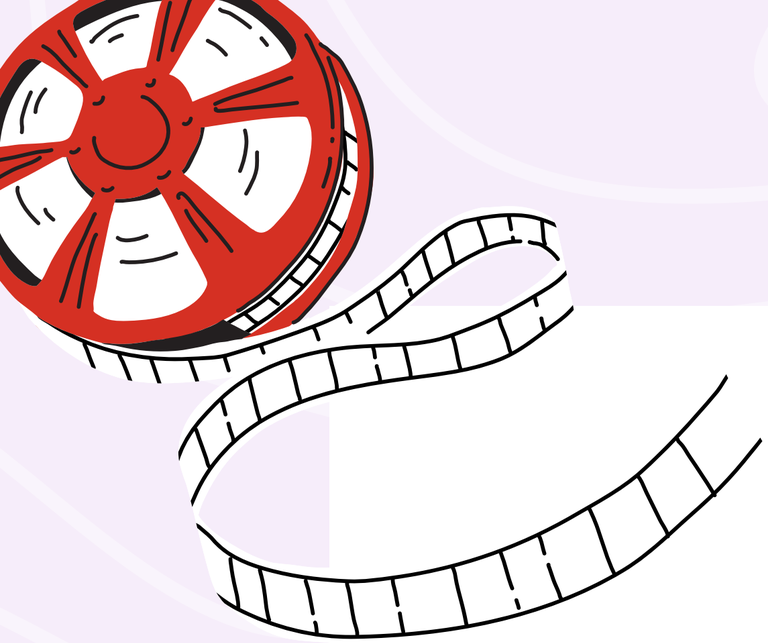


Novelle

movie Recommendation system

Project Overview

- Unlike established streaming services, small businesses in Kenya that sell movies have high customer churn and dissatisfaction rates because they cannot recommend movies to their customers.
- The Novelle Movies recommendation system project will develop a personalized movie recommendation system that will enable small business owners to improve sales and retention of their customers through improved customer experience.

Business Understanding

- Small businesses in Kenya that sell movies have high customer churn and dissatisfaction rates.
- They spend a lot of time watching all the movies to give recommendations.
- They miss opportunities to sell movies that vendors have never watched.
- They have limited resources to develop a personalized movie recommendation system.

Project Objectives

- Increase customer engagement by recommending movies based on user preference
- Increase sales by supporting customers to find movies of their taste
- Be able to make recommendations to new customers

Data Understanding

- The project will use the MovieLens small dataset from the GroupLens research lab at the University of Minnesota.
- It contains 100,863 ratings ,3683 tag applications and 9742 movies.

It contains the following files

- movies.csv contain movies details (movieId, title, genres)
- ratings.CSV which contains user ratings (userId, movieId, rating, timestamp)
- tags.CSV contains user-generated movie tags (userId, movieId, tag, timestamp)

Data preprocessing

1. The small Movie Lens Dataset was used due to computational capability.
2. Data sets were merged and all the Nan features were converted to empty strings.
3. The column timestamp which was not necessary was dropped.
4. The Merged dataset was explored for patterns

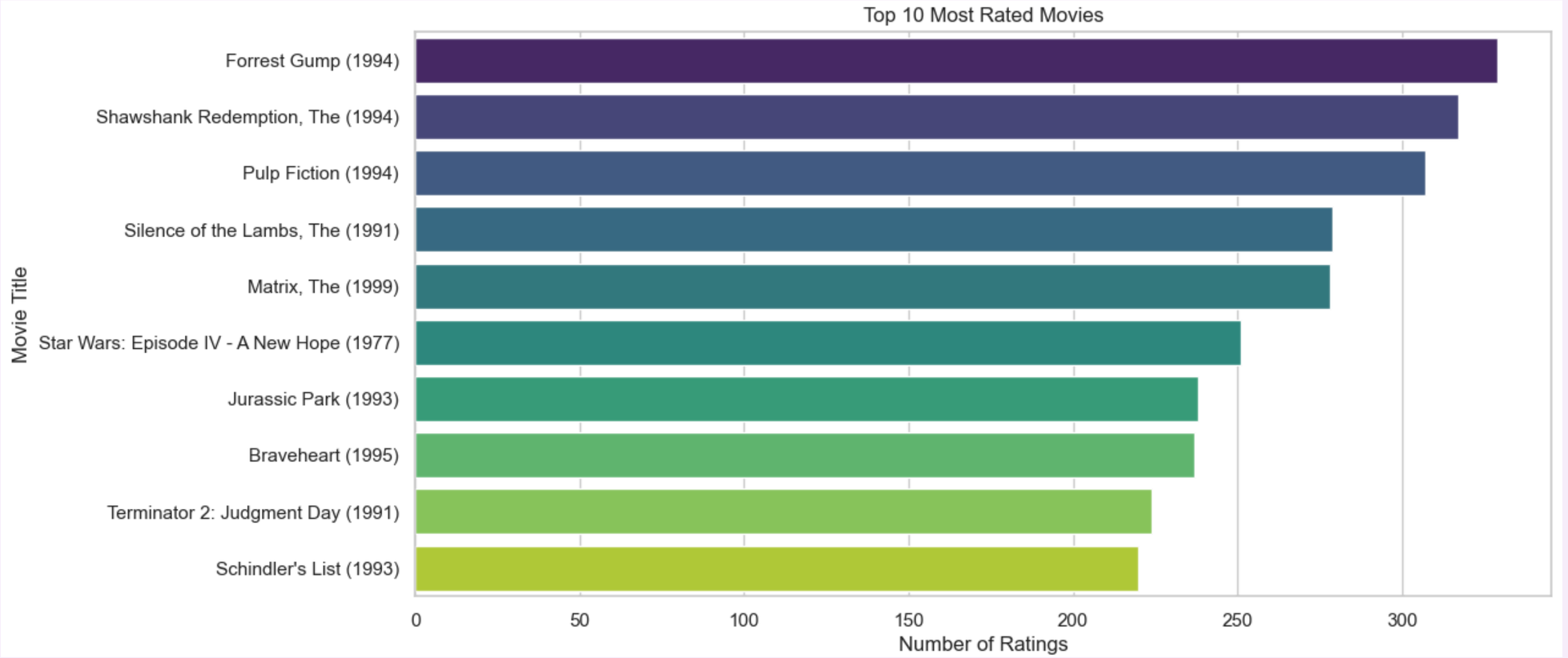
Modelling Approach

- Collaborative filtering (SVD) modeling using surprise
- Content-based filtering: The TD_IF and cosine similarity to recommend movies
- Hybrid recommendation that combines collaborative filtering (SVD) and content-based filtering
- Model improvement by hyperparameter tuning using GridSearchCv and SVDpp

Model Performance Evaluation

- Data was split into training and testing with test_size of 0.2 and a random _state of 42.
- The model was evaluated using root mean square error (RMSE) and mean absolute error (MAE)

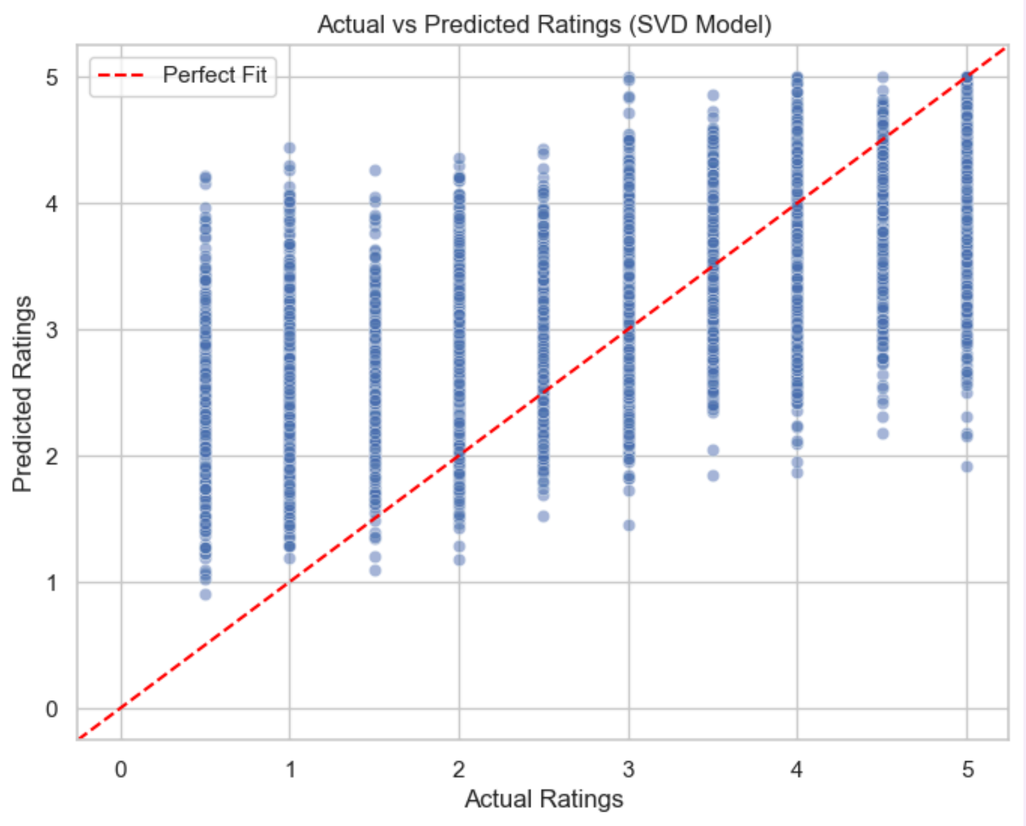
Results and Insights



Results and Insights

Collaborative (SVD) model Evaluations

- The mean absolute error of 0.5781 indicates that on overage the predicted ratings deviate from the actual rating by 0.58 which is significant
- The root mean squared error of 0.75 indicates that some predictions deviate significantly
- lower value so both RMSE and MAE indicate better model performance
- The graph shows that the model has high predicted ratings for low-rated movies and high-rated movies have low predicted ratings.



Results and insights

Consider a Hybrid recommendation system that combines collaborative filtering (SVD) and content-based filtering.

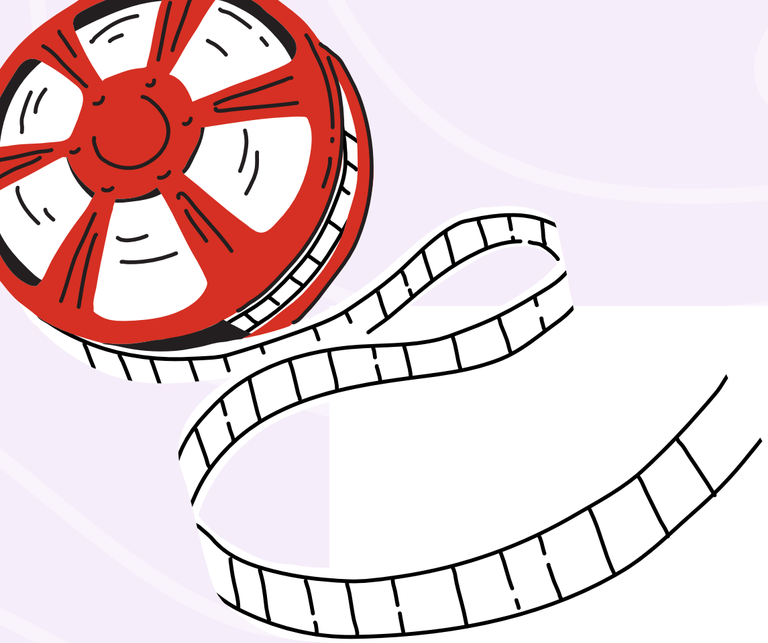
- If the user has rated at least 5 movies, use collaborative filtering (SVD).
- If the user has rated less than 5 movies, use content-based filtering and recommend the top-rated movie.
- If the user has not rated any movies, recommend the top-rated movies to address cold start problem

SVDpp had slightly better performance and needs to be explored further used to improve performance

- RMSE: 0.7194
- MAE: 0.5506

Next steps

- Use the large MovieLens dataset for modeling
- Deploy the updated model for ease of use
- Connect to customer accounts



Thank you