Streamflix Movie Recommender System

Created by:
Evaclaire Munyika
Simon Makumi

Table of Contents

Business Understanding

Business Understanding

Overview

Streamflix aims to enhance user experience with a movie recommendation system that will leverage collaborative and content-based methods for more personalized top 5 movie suggestions.

Business Problem

User retention and engagement issues are impacting company revenue, customer satisfaction and Streamflix's competitive edge in the streaming industry.

Objective

To develop and deploy a hybrid recommendation system to improve prediction accuracy by capturing user-item interactions and movie attributes for a personalized experience.

Data Understanding

Data Understanding

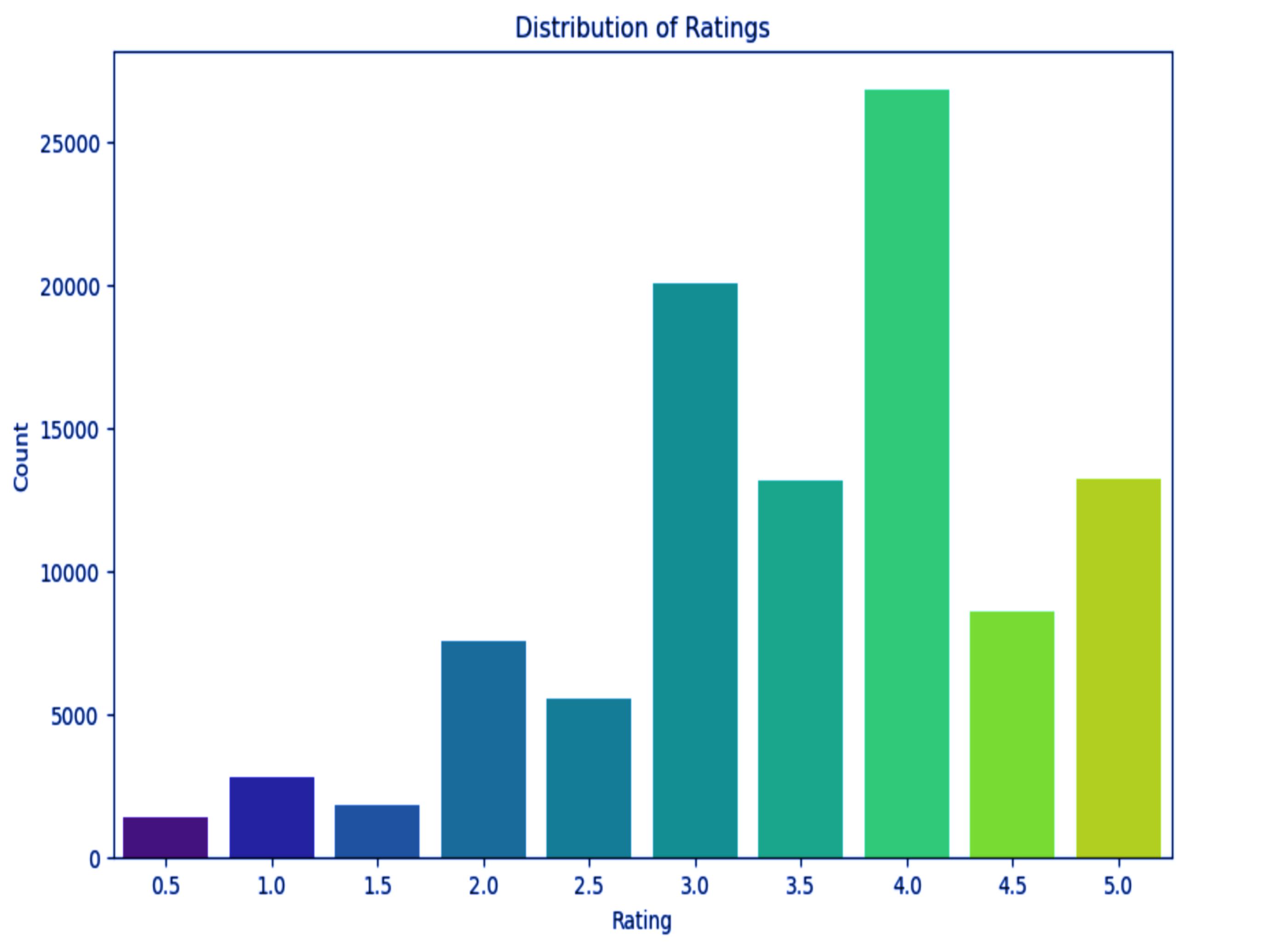
Data Source

Our data is from the MovieLens dataset by the GroupLens research lab at the University of Minnesota.

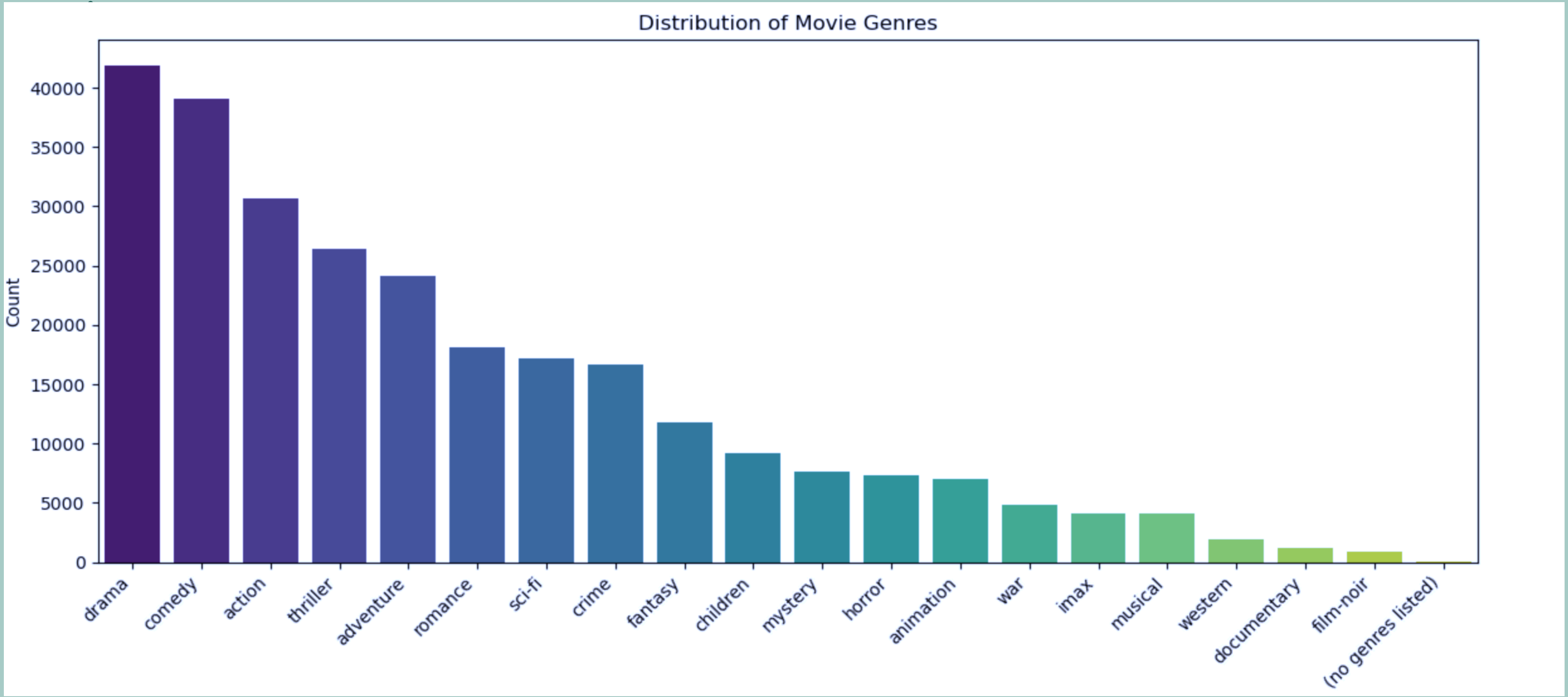
Features of Importance

- Movie_id - unique movie identifier.
- Title - movie title.
- Genres - movie genres.
- User_id - unique identifier for users .
- Rating - ratings given by the users.

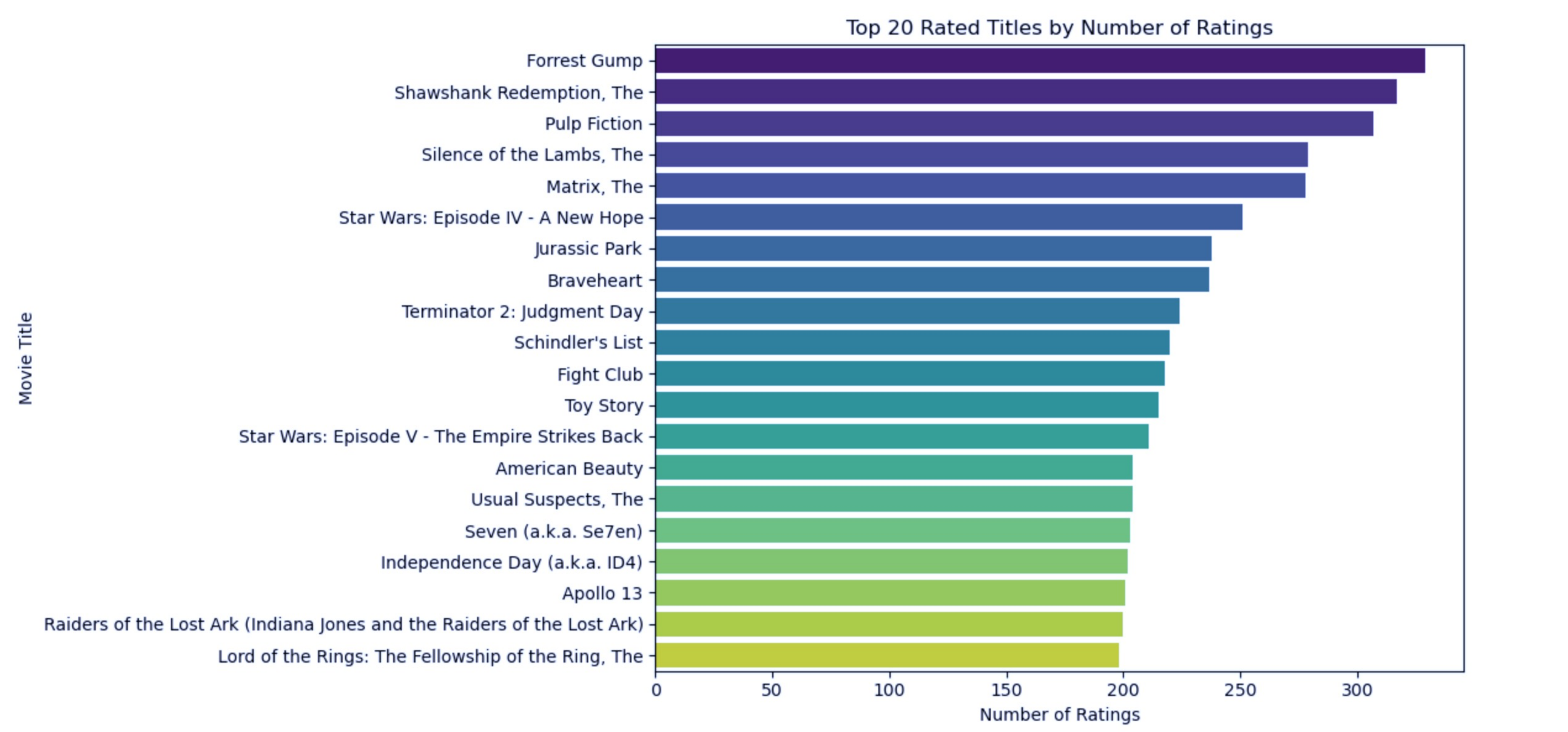
Observations



- The most common rating is 4.0.
- The distribution is positively skewed towards higher ratings.



- Drama is the most preferred genre followed closely by comedy.
- Niche genres like animation, war, IMAX, musical and Western appear to have fewer than 10,000 movies each.



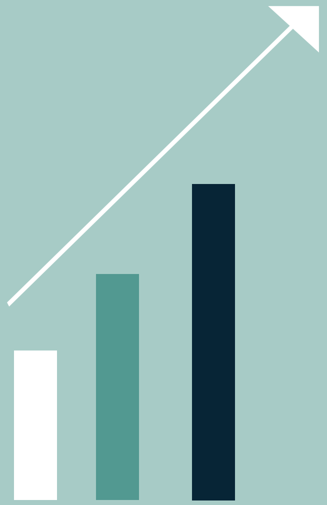
- 'Forrest Gump' has the highest number of ratings followed by 'The Shawshank Redemption' and 'Pulp Fiction’.

Modeling Results

- Baseline dummy model yielded an RMSE of 1.43, SVD model outperformed KNN with an RMSE of 0.862 vs. 0.975.

- Developed a CollabBasedModel for collaborative filtering and a ContentBasedModel for content-based filtering and combined both techniques creating a hybrid model.

- The hybrid model tested various collaborative filtering weights, showing improved RMSE scores with higher weights reaching 1.1221 at a weight of 0.8 indicating better prediction accuracy with higher collaborative filtering emphasis.

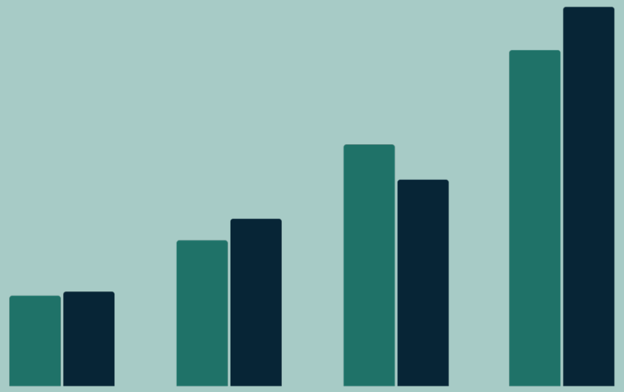


Conclusions

Conclusions

- The collaborative filtering model with an RMSE of 0.86 outperforms the hybrid model which shows higher RMSE of 1.24

- This illustrates that collaborative filtering in a hybrid approach yields better accuracy and recommendation quality.



Recommendations

Recommendations

Experiment with smaller increments around the optimal collaborative filtering weight ($0.6 > k < 0.8$)

Use cross-validation to confirm that improvements in RMSE are consistent and not due to random variations or overfitting.

Integrate deep learning-based models and other advanced methods to enhance model capabilities..

Next Steps

Next steps

1. Set up an A/B test comparing the collaborative filtering and hybrid systems.

2. Set up a robust monitoring system to track performance of the deployed model over time.

3. Implement a mechanism for users to rate the quality of recommendations.

Thank You!

For follow-up questions
Reach out to us on:
evamunyika@gmail.com
simonmakumi5@gmail.com