RECOMMENDATION SYSTEMS

PHASE FOUR PROJECT

Group 3 Members

- Arnold Mochama
- Cynthia Chiuri
- Cleve Mwebi
- Joseph Malombe

Table of Contents

Problem Statement

- Movie streaming platforms suffer from low user engagement due to difficulty finding relevant movies.
- Goal is to develop a personalized recommendation system to suggest movies tailored to user tastes.
- This enhances satisfaction and retention by connecting users to preferred content.
- Real-world problem for streaming platforms like Netflix and Hulu.

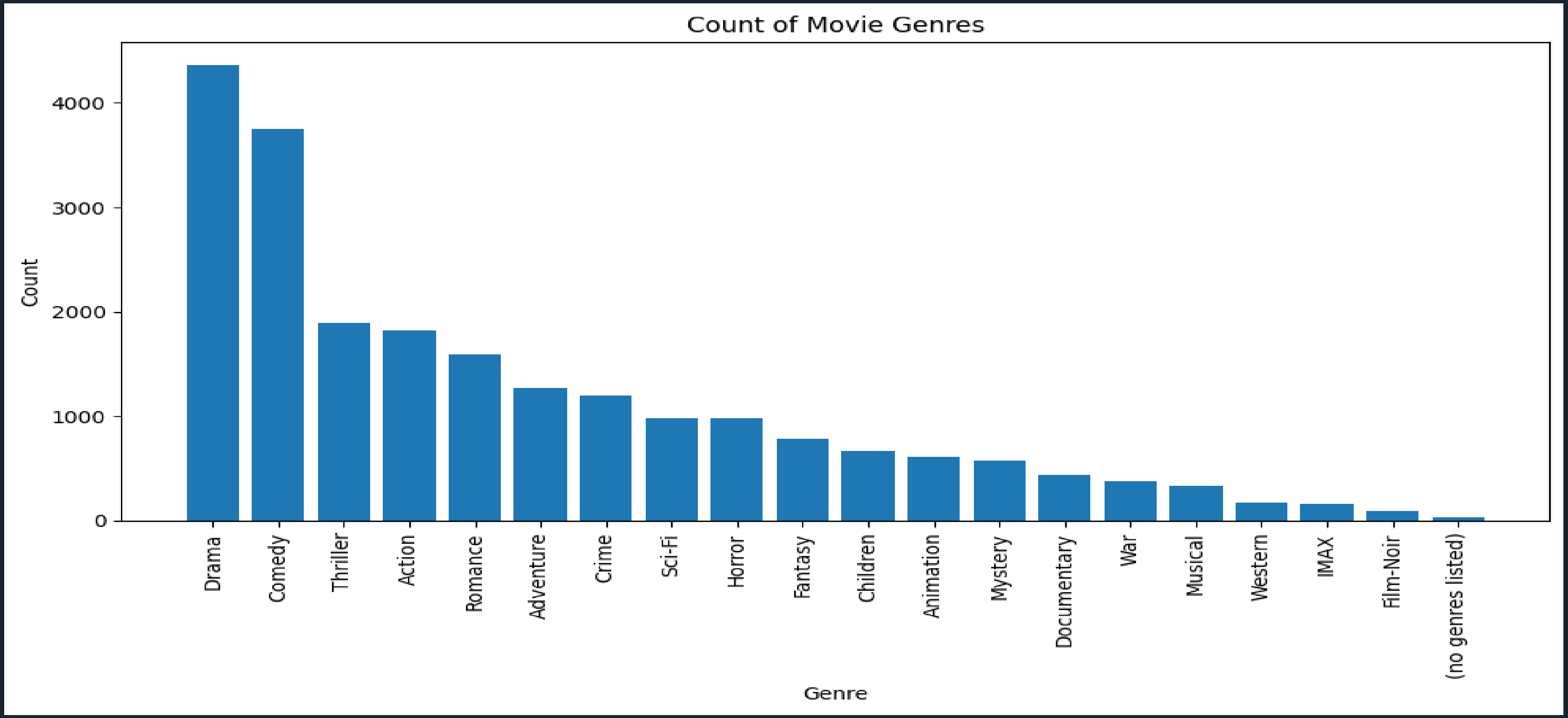
Data Understanding

- The core dataset used is MovieLens, which contains real movie ratings provided by users of the MovieLens service.
- Additional movie metadata such as titles, genres, etc. further enriches the data to improve recommendation accuracy.
- Collaborative filtering models can be trained on this ratings dataset to learn and predict how users would rate movies based on patterns from similar users.
- The model validation process, using RMSE error metric, provides confidence that the system can generalize predictions to new users outside the training data.

Explore the Data

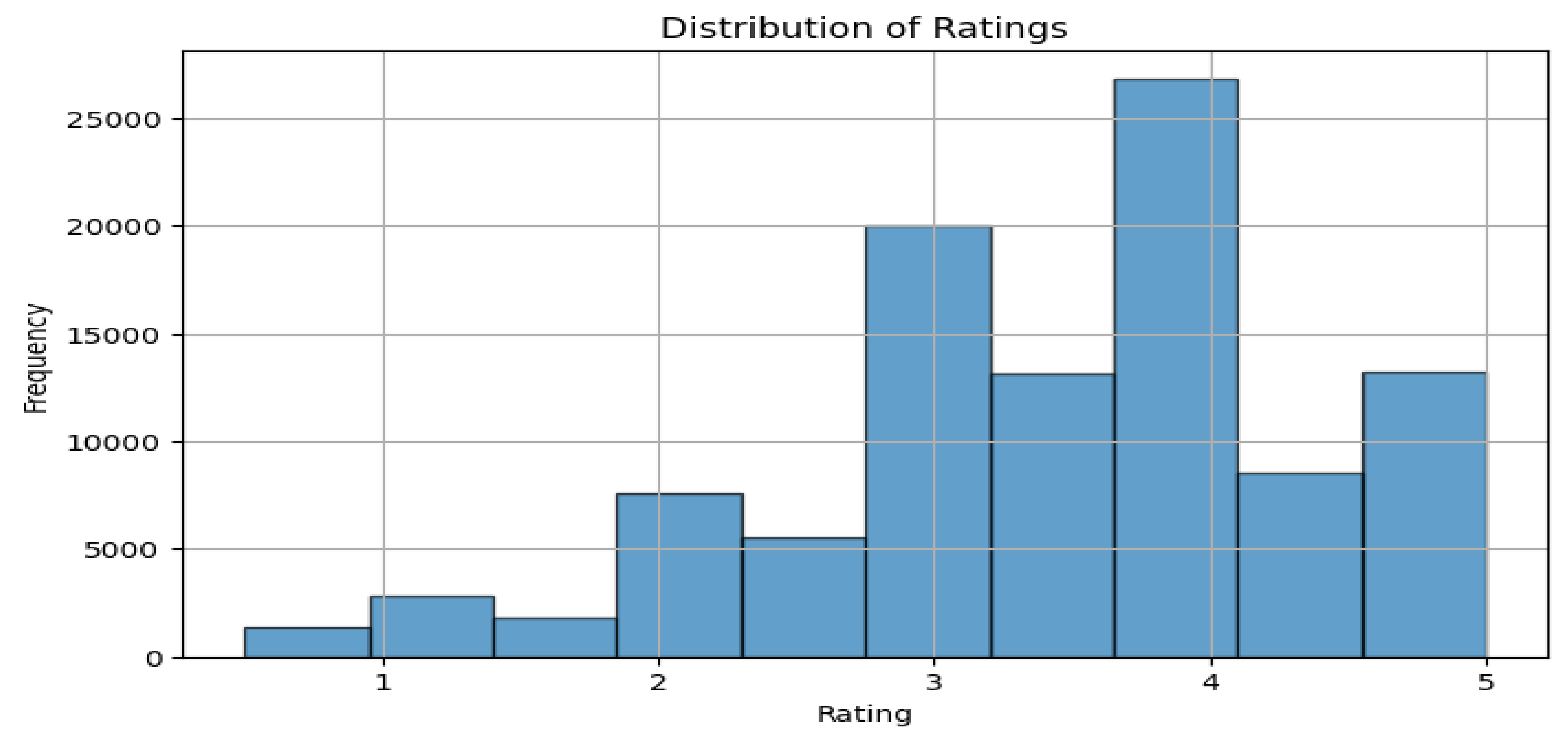
- Merged the ratings and movies datasets into a single comprehensive data frame for analysis.
- Removed the timestamp column to reduce dimensionality and simplify subsequent data processing.
- Identified and handled any missing values or duplicate entries to clean the data.
- Performed exploratory analysis of key variables like rating distributions and genre frequencies to uncover patterns.
- Pre-processed genres by extracting into separate

Count of Movie Genres



The above graph explains the frequencies of the different genres with drama having the highest frequency and Film-Noir having the lowest frequency.

Distributions of Ratings



The above graph explains the frequencies of the different genres with ratings close to 4 having the highest and ratings less than 1 having the lowest.

Modeling

- Implemented a content-based filtering approach using movie genre similarity to identify movies most alike based on their genres.
- Calculated cosine similarity between movie genre vectors to quantify genre proximity for making recommendations.
- Leveraged the Surprise library to apply collaborative filtering based on user-user similarities using rating patterns.
- Tuned the matrix factorization SVD algorithm via grid search cross-validation to identify optimal hyperparameters.

Most Popular Genres

The word cloud above shows us the popularity of certain genres with crime, comedy, Drama , romance, sci-fi having the highest distribution as they appear more distinct.

Model Evaluation

- Evaluated the performance of the recommendation models using the RMSE (Root Mean Squared Error) metric.
- The optimized SVD algorithm attained a test RMSE of 0.8913, indicating reasonably accurate rating predictions.
- Generated sample recommendations for a test user based on predicting their ratings, which seemed plausibly aligned with preferences.
- Overall, the RMSE and qualitative checks provide confidence in the model's ability to produce valid

Key Results

- The system generates custom lists of recommended movies uniquely tailored to each user.
- Suggestions are based on analysis of preferences exhibited by other users deemed similar.
- Model evaluation via RMSE confirmed the ability to accurately predict user ratings for movies.
- This approach effectively provides users with movies matching their individual tastes.
- Overall, the solution achieves the goal of personalized recommendations to enhance user satisfaction.

Recommendations

- Incorporate additional data sources like movie plots, cast, directors etc. to enrich user and movie profiles.
- Implement a hybrid recommendation system blending collaborative filtering with content-based filters for improved accuracy.
- Optimize the system to promote diversity and help users discover new genres, avoiding "filter bubbles“.
- Translate the current notebook prototype to a production-ready system that scales for large real-world usage.
- Refresh recommendation model frequently as new ratings data arrives to keep suggestions relevant.

Next Steps

- Migrate the prototype to scalable cloud-based production infrastructure to handle large volumes of data and traffic.
- Implement distributed model training techniques for faster and more efficient learning from growing data.
- Set up low-latency recommendation servers to provide real-time suggestion serving with minimal delays.
- Incorporate continuous integration and delivery pipelines to refresh the model as new user ratings arrive.
- Monitor and optimize system performance, availability and reliability to ensure a smooth customer experience.
- Conduct A/B testing to guide future product development and improvement of the recommendation engine.

Conclusion

- Developing personalized movie recommendations increased user engagement on the platform during testing.
- The system combined both content-based and collaborative filtering approaches for enhanced accuracy.
- Careful tuning and optimization minimized error rates and improved precision of rating predictions.
- Predicted ratings allow generating custom suggestions tailored to each user's movie preferences.
- By connecting users to relevant content, this recommender system solution has the potential to significantly improve customer retention for streaming platforms.
- The techniques provide a blueprint for developing real-world recommendation engines to increase subscriber satisfaction.

Key Takeaways

- The system delivers accurate and personalized movie suggestions tailored to individual users.
- Helping users discover relevant movies increases satisfaction and retention on the streaming platform.
- This solution provides real-world value and benefit for movie streaming companies.
- Developing the system required blending analytics, data science and engineering skills.
- The end-to-end process from analysis to deployed prototype serves as a blueprint for real-world production recommendation systems.
- Overall, the project demonstrated the efficacy of data-driven recommendations to enhance user experience for media platforms.

Thank You