

The background of the slide is a dense field of 3D-rendered numbers (0-9) in various shades of blue and white, creating a sense of depth and data. A solid black rectangular box is positioned on the right side of the slide, containing the title and project information in white text.

Recommendation Systems.

Group 8 – Phase 4
Project

Group 8 – Phase 4 Project.

Recommendation Systems.



Team Member Names

- 1. Stanley Maina
- 1.2. Brenda Kurgat
- 3. Susan Mworio
- 4. Chepkwony Joy
- 5. Ezekiel Ngungu
- 6. Emmanuel Korir

Introduction

- This project aims to develop a movie recommendation system.
- The system will utilize machine learning techniques to provide personalized movie suggestions based on user ratings and tags.
- The dataset used for this project is the MovieLens dataset, which contains ratings, tags, movie metadata, and links.

Project Objectives

- Personalized movie recommendations based on the user's history
- Investigate how the system performs for new users and new movies (analysis of the cold start problem)
- Build a deep learning-based model that learns from user ratings and movie features to improve recommendation accuracy
- Optimize recommendation algorithms
- A dashboard to explore recommendations and engagement insights.

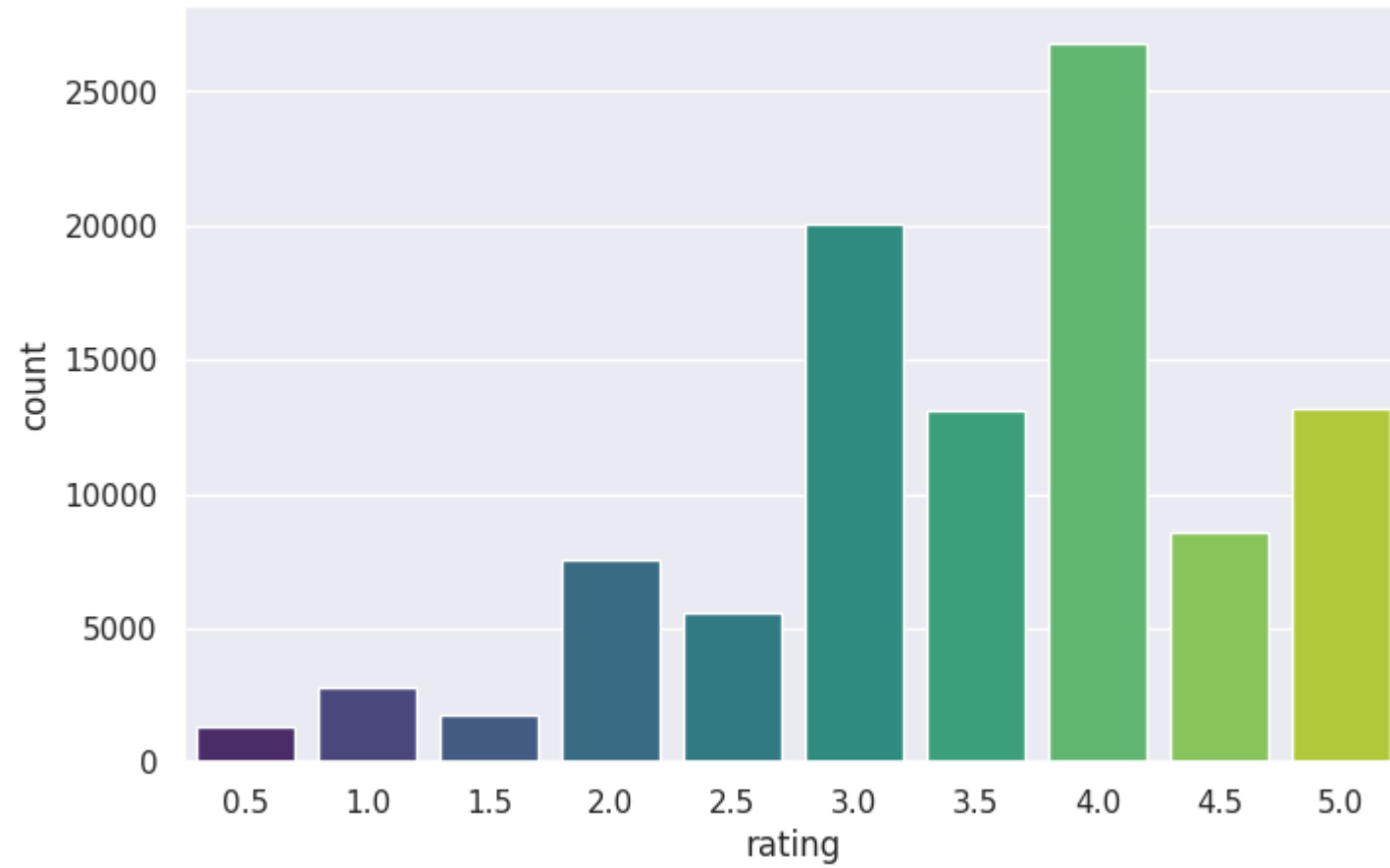
Dataset Overview

The dataset used for this project is the MovieLens dataset, which contains ratings, tags, movie metadata, and links.

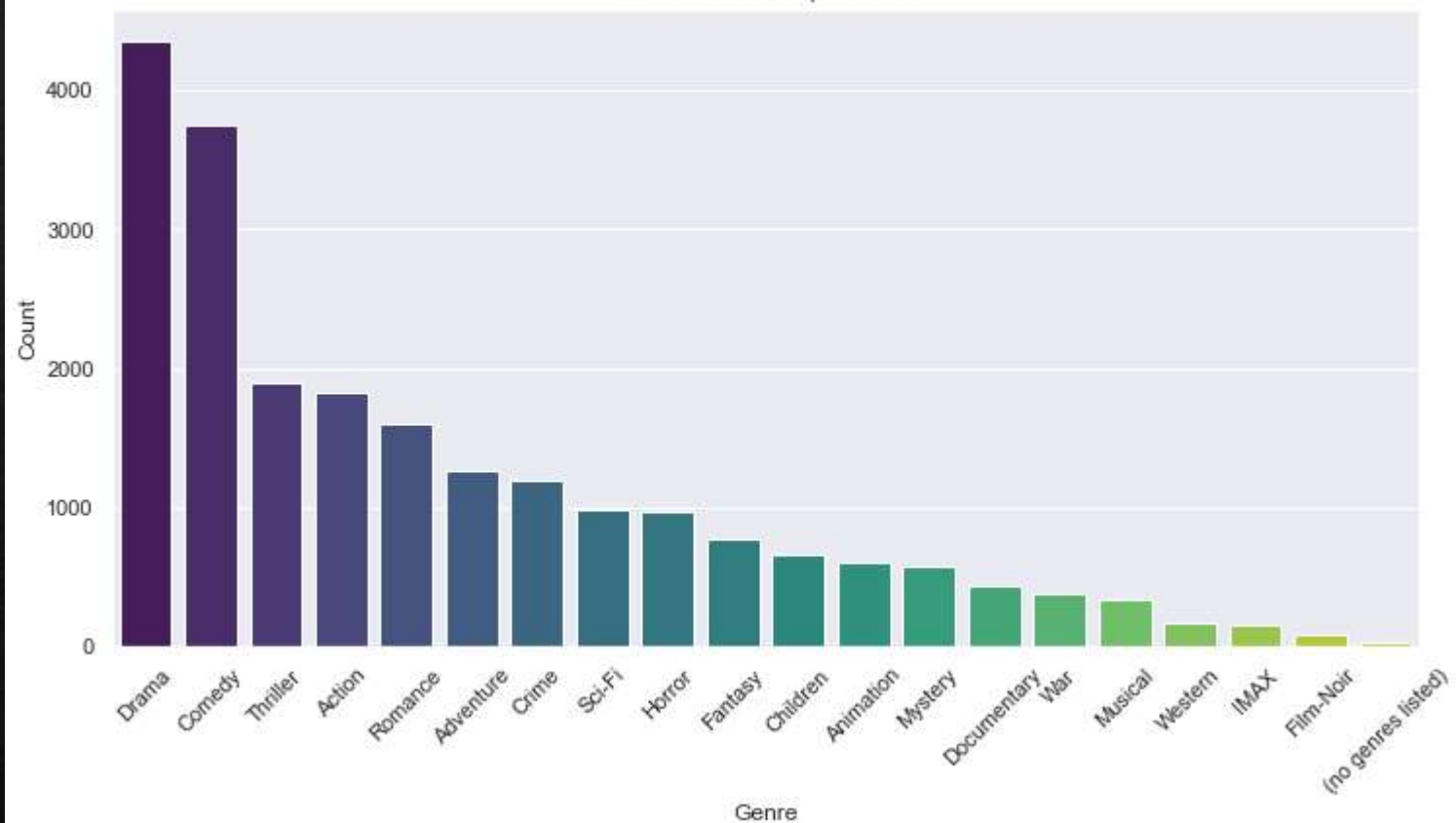
Data Preprocessing & EDA

- Handled missing values and duplicates.
- Merged the datasets into one common file.
- Encoded categorical variables.
- Normalized the rating column.
- Visualized rating distributions and user/item behavior patterns.

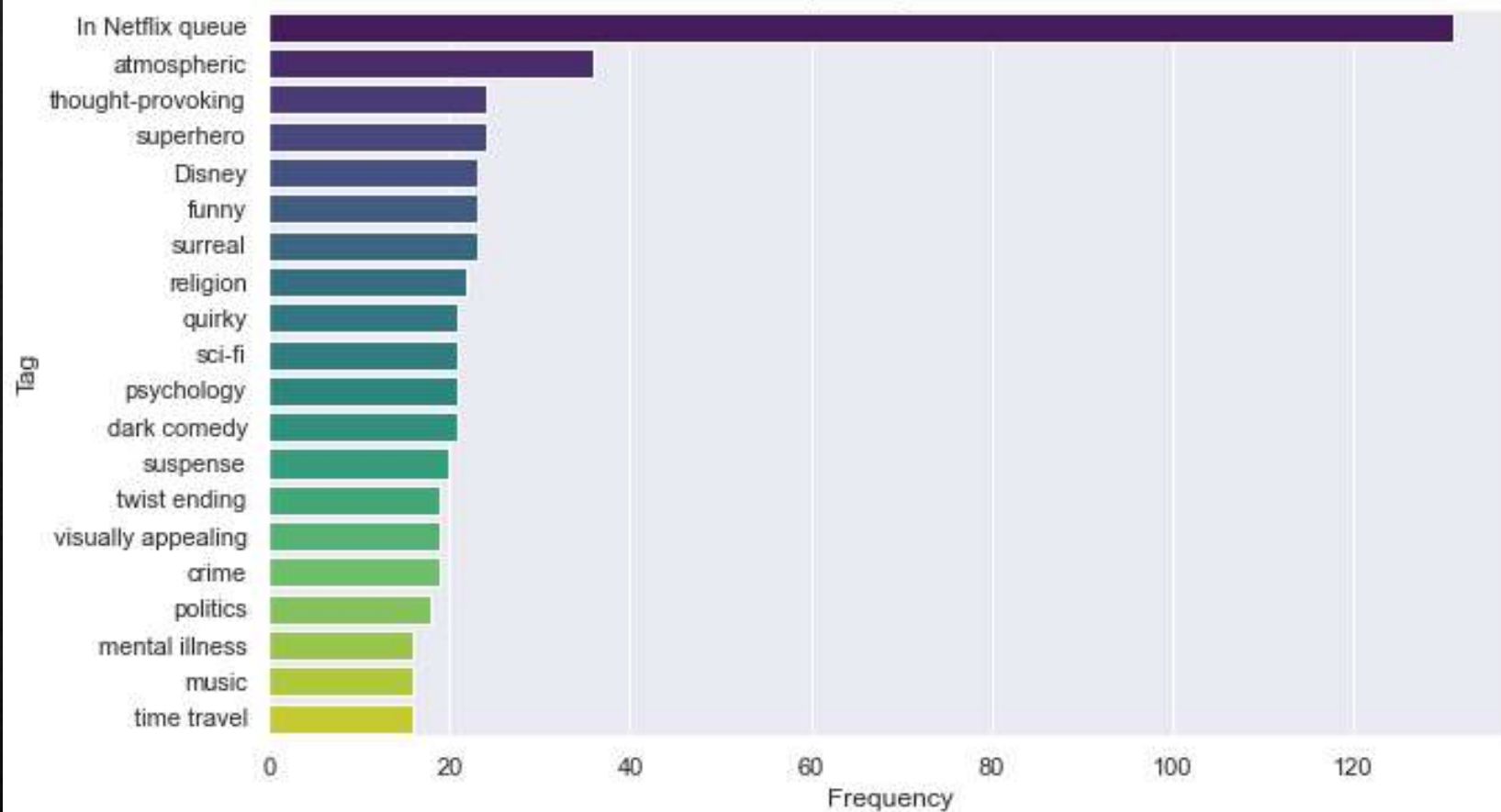
rating distribution



Number of Movies per Genre



Top 20 Tags Used

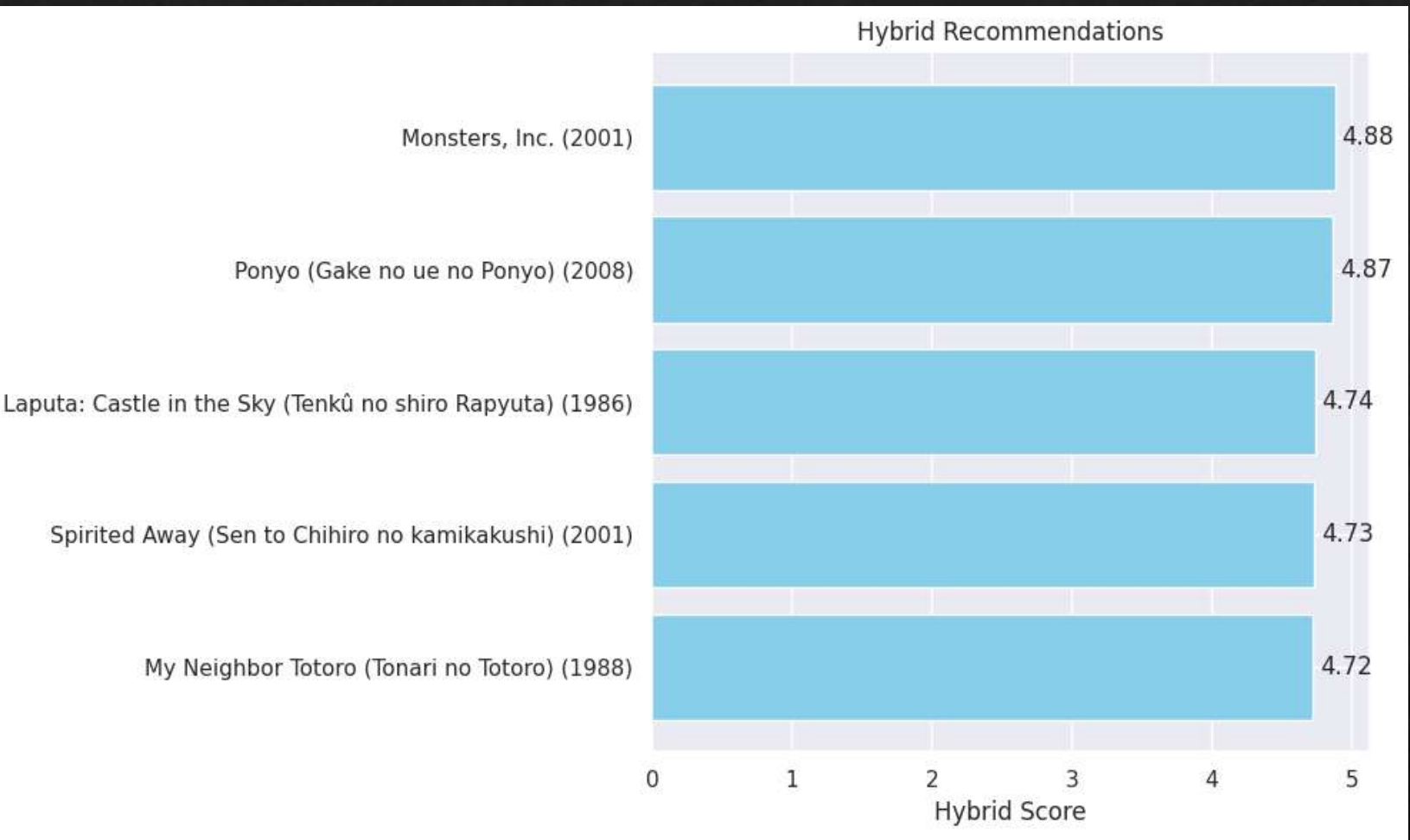


MODELING.

- We started with the KNN model as our baseline. The idea behind KNN is that it assumes that similar things are usually close together. Since recommendation systems are all about finding and using those similarities
- Singular Value Decomposition (SVD). SVD is better equipped to handle sparsity and scale, so we hope it will improve performance, especially in terms of lower RMSE and MAE scores.
- Collaborative Filtering(Matrix Factorization) - This will recommend movies based on user preferences and similarities with other users
- Content-Based Filtering (TF-IDF + Cosine Similarity) – This will recommend to a user the movies they would love in reference to those the user liked.
- Hybrid Approach: Combines multiple techniques for better results.

MODEL PERFORMANCE.

- **KNN PERFORMANCE.** – We evaluated our KNN model using RMSE, MAE, and cross-validation. The results were as follows: 1. RMSE - 0.9829. 2. MAE - 0.7621. 3. CV - The predictions from the model are about 0.99 stars off using RMSE and 0.77 stars off using MAE. We can say that the model is on average one star off. This could be due to KNN's struggles with very sparse datasets and its limitations when working with larger ones.
- SVD.- 1, Average RMSE_SVD: 0.8634. 2. Average MAE_SVD: 0.6602
- **Collaborative Filtering(Matrix Factorization).** - The model performed fairly well, with an RMSE of 0.88 and an MAE of 0.68; the prediction is less than one point from the actual ratings. F1 score of 0.73, indicating a good balance between correctly recommended items and coverage.



Conclusions

- The initial analysis and implementation of the recommendation system have shown promising results, particularly with SVD& COLLABORATIVE FILTERING. The evaluation metrics indicate that the model is capable of providing relevant recommendations based on user interactions.
- The preprocessing steps undertaken were crucial in preparing the data for analysis. Cleaning and transforming the data helped in reducing noise and improving the model's accuracy.
- While the current model performs well, there is room for improvement. Future iterations should focus on incorporating more diverse data sources and refining the algorithms used.
- Overall, this project lays a solid foundation for developing a sophisticated recommendation system. With further enhancements and user engagement, it has the potential to significantly improve user experience and satisfaction.