DELTA CORNER PRODUCTIONS

HULU MOVIE RECOMMENDATIONS

This project built a Recommendation System model that can be adopted to make movie recommendations for a user based on the ratings given by other different users.

INTRODUCTION

Personalized movie suggestions are vital to customer engagement and retention in today's fiercely competitive streaming market. With 11% of the streaming market, Hulu wants to improve the algorithm it uses to suggest movies to its users. The objective of this project is to create a machine learning model that recommends the best 5 movies to users based on their past evaluations of other films.

The Hulu Technology Team, which is led by the company's CTO, is our target audience for this project.The goal of this research is to investigate how to improve the algorithm that generates suggestions for Hulu users.

BUSINESS OBJECTIVES

The main objective is to build a movie recommendation model that can provide the top 5 movie recommendations to a user based on their previous ratings. The model intends to enhance Hulu's current recommendation system.

DATA UNDERSTANDING

The dataset used in this project was sourced from [Grouplens] (https://grouplens.org/datasets/movielens/100k/), utilizing information from IMDb and TMDb. The data contains over 100,000 movie ratings provided by 610 different users for about 9,700 movies, collected over a timespan from 1996 to 2018. Each user has reviewed at least 20 different movies.

While the full dataset consists of approximately 1.9 million ratings, this project focuses on a subset of 100,000 ratings due to time constraints.

KEY FEATURES

- User ID: A unique identifier for each user.
- Movie ID: A unique identifier for each movie.
- Ratings: A score from 0.5 to 5.0 given by a user for a movie.
- Timestamp: The time when the rating was provided.
- Movie Metadata: Information about the movies, such as title and genres.

This data is crucial for building a collaborative filtering model that can predict which movies a user is likely to rate highly based on the behavior of similar users.

EXPLORATORY DATA ANALYSIS

EDA was thoroughly conducted to ensure tashowed us that the movies dataset has 9,742 entries with columns: movieId, title, and genres.

There are no missing values or duplicate movieId entries, ensuring each movie is uniquely identified.

The title column has a few duplicates, potentially due to different versions of the same movie.

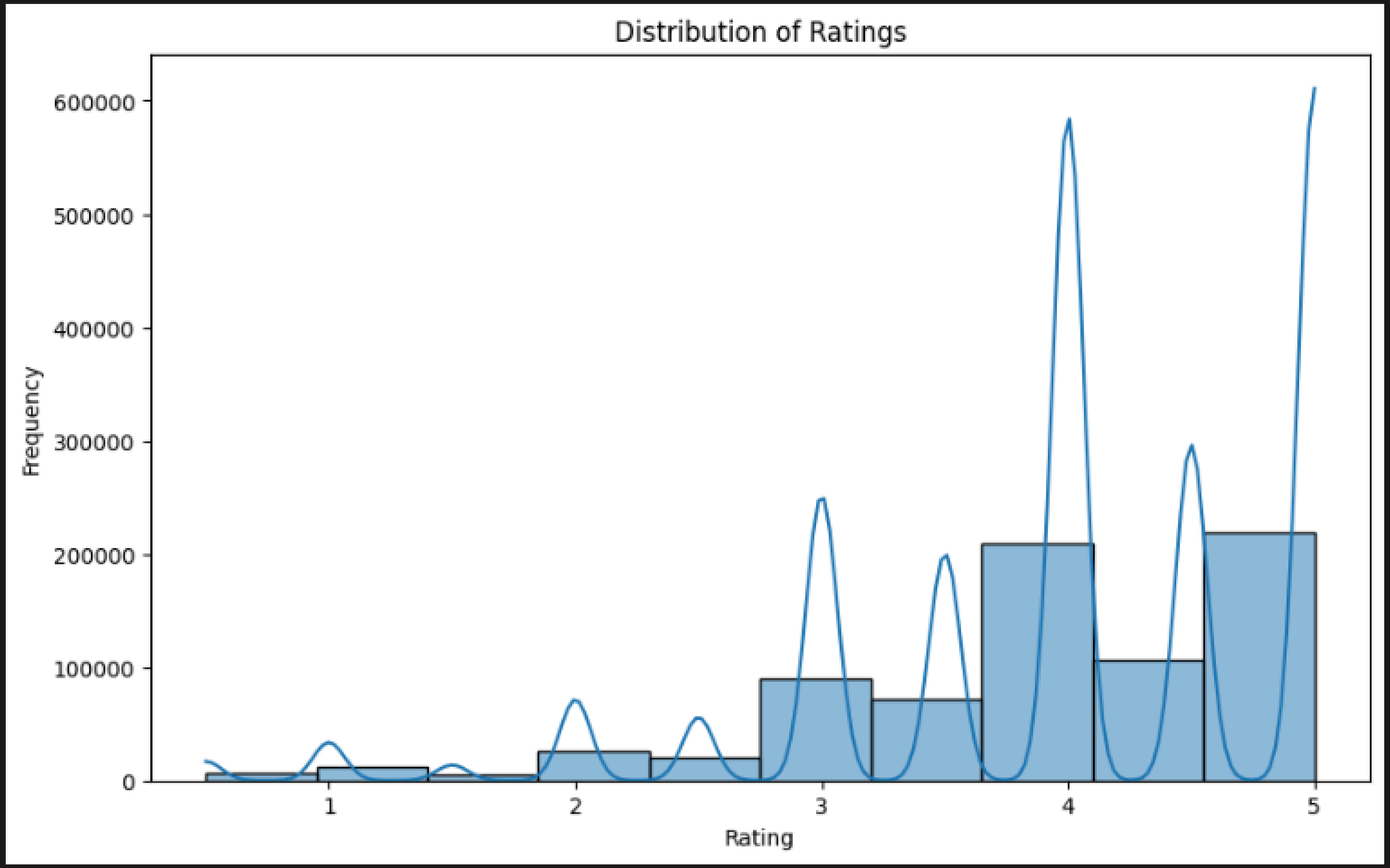
The genres column contains 951 unique genre combinations, with "Drama" being the most common.

Overall, the dataset is clean, with appropriate data types, making it ready for use in the recommendation system.

EXPLORATORY DATA ANALYSIS

The dataset contains 765,271 movie entries, with movieIds ranging from 1 to 193,565, showing a wide variety of movies. User IDs span up to 610, indicating a diverse user base. Most interactions occurred around 2014-2015, with timestamps ranging from 1996 to 2018. The average movie rating is 3.99, suggesting generally positive user sentiment. Movies have release years between 1921 and 2018, with the majority released around 1995.

UNIVARIATE ANALYSIS



The distribution of ratings in the dataset is visualized in the histogram. It shows that the majority of ratings are clustered around the higher end of the scale, specifically from 3 to 5. The most frequent rating is around 4, indicating that users generally rate movies positively. There are fewer ratings on the lower end (1 to 2), suggesting that users might be inclined to rate only the movies they enjoyed. The Kernel Density Estimate (KDE) line reinforces the peak around higher ratings, emphasizing this positive bias in user ratings. This skewness is a common pattern in movie rating datasets.

BUILDING THE RECOMMENDATION SYSTEM

1. Collaborative Filtering (SVD)
Singular Value Decomposition (SVD) was used to capture user-item interactions based on ratings.
Data: Trained on user ratings (rt_df), with a test-train split to assess performance.
Cross-Validation: Applied 5-fold cross-validation to evaluate model accuracy.
Performance Metrics:
RMSE (Root Mean Squared Error): 0.874 (mean from cross-validation), indicating reasonable prediction accuracy.
MAE (Mean Absolute Error): 0.672, showing a relatively small average error between predicted and actual ratings.
Advantages: Learns user preferences for personalized recommendations.

BUILDING THE RECOMMENDATION SYSTEM

Hybrid Recommendation & Evaluation

1. Hybrid Model

Integration: Combined collaborative filtering (SVD) and content-based filtering to leverage both user preferences and movie attributes.

Scoring: Adjusted scores using a weighted combination of collaborative and content-based predictions.

Flexibility: Provides recommendations based on either specific movie titles or genres.

2. Evaluation Metrics

Metrics Used: RMSE and MAE to assess model accuracy on the test set.

Final RMSE: 0.6434

Final MAE: 0.4992

User Testing: Presented top movie recommendations for different genres and titles (e.g., "Drama," "Fight Club"), validating the model's ability to suggest relevant content.

Interpretation: The hybrid model successfully combines personalization with content similarity, providing tailored and accurate recommendations.

INTERPRETATION

BEST MODEL

The hybrid model successfully combines personalization with content similarity, providing tailored and accurate recommendations.

RECOMMENDATIONS

- Implement a hybrid recommendation system
- Incorporate User Feedback Loop
- Periodically Retrain the Model
- Include Diversity and Novelty
- Enhance Genre-Specific Recommendation

CONCLUSIONS

This model has revealed valuable insights and methodologies that can be generalized to Multiple Recommendation Systems for multiple business enterprises.
Our team looks forward to continually improving the model to yield desirable results.

THANK YOU