By Jesse Moore

Movie Recommendation System

Table of Contents

Overview

The goal of this project is to develop a Top 5 movie recommendation system to enhance our user's experience with our platform.

We seek to develop this recommendation system using both collaborative and content based filtering methods; providing our users with personalized recommendations based upon their user history.

This will utilize a hybrid approach that factors for problems with new users and new movies not having any history with our platform.

Business Problem

Our client is starting a new streaming service and is concerned about entering a saturated marketplace and wants to ensure that it’s users have the best user experience on their platform and feel the most engaged.

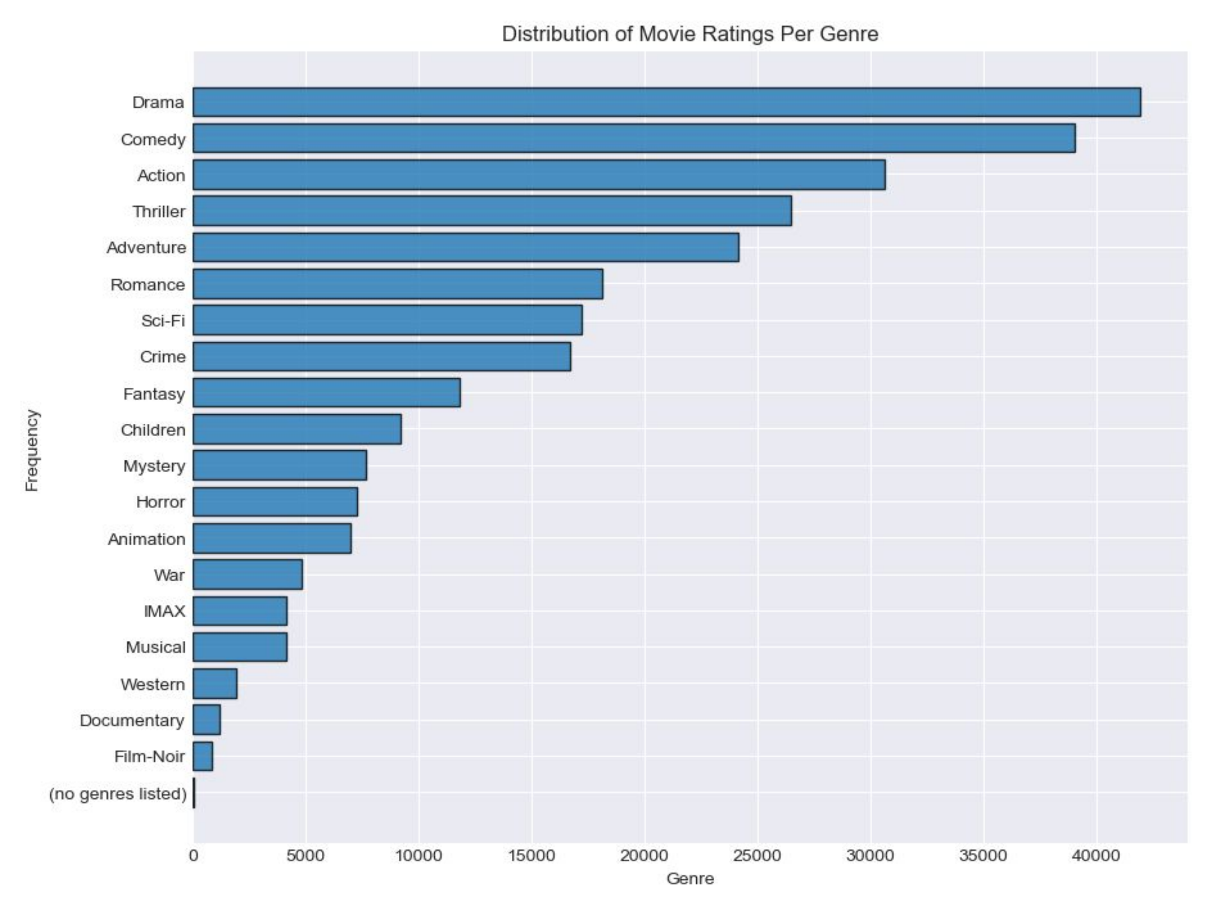
Using a collaborative filtering method that factors in users previous movie ratings we will provide users with top 5 recommendations based upon their previous ratings.

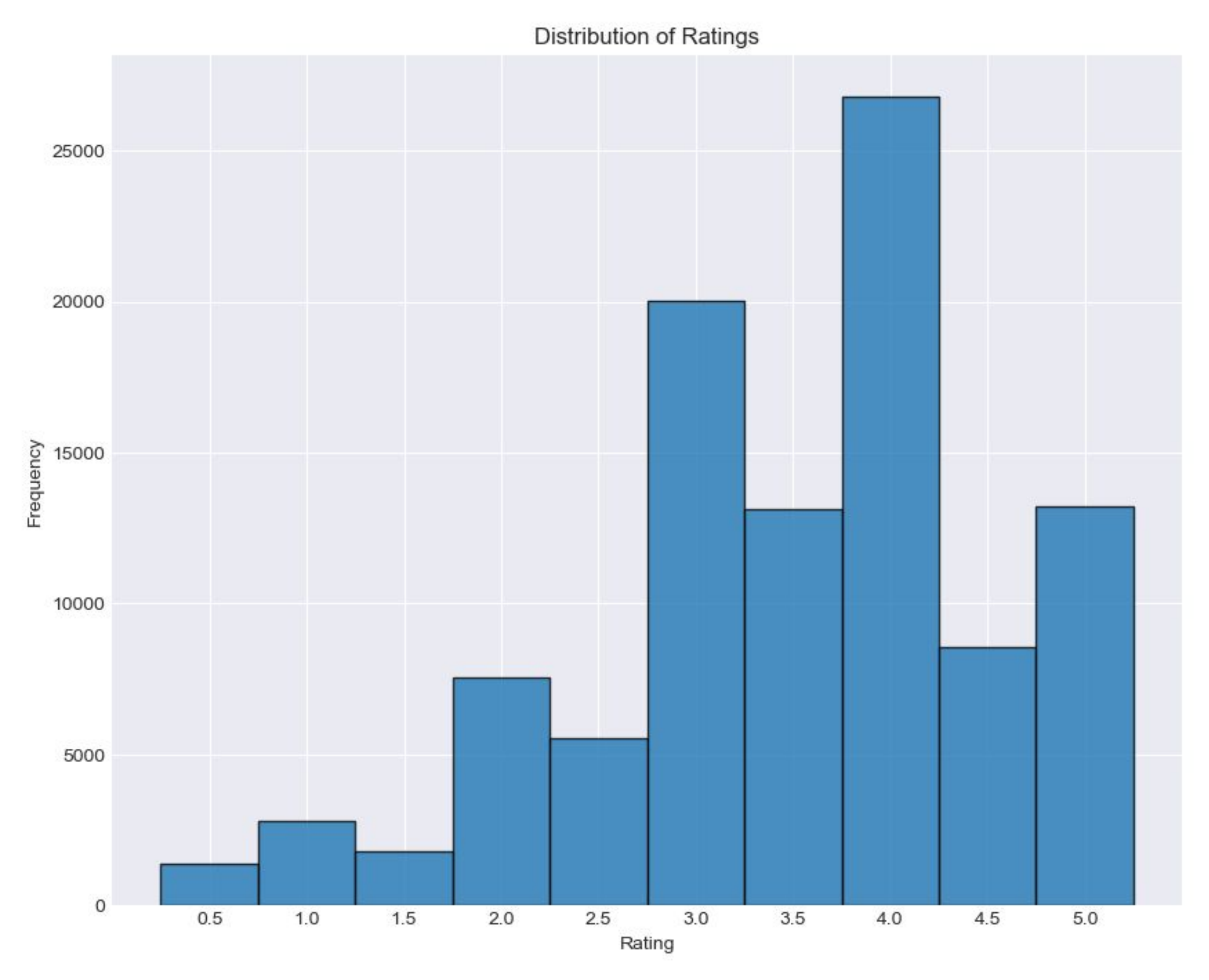
Data Understanding

The MovieLens dataset has been created by the GroupLens research lab at the University of Minnesota contains 100,000 ratings, 3,600 tags and was last updated on 9/2018.

While the dataset contains several csv files of data, we will focus on the 'movies.csv' and 'ratings.csv' datafiles to create our models. These files contain user ratings for movies, the user id, the movie id, and the movie title.

Movie Rating Freq. by Genre





Plot of Freq.
Of Movie Ratings

Modeling

We initially implemented an Alternating Least Squares (ALS) model using Spark for collaborative filtering.
ALS decomposes the user-item matrix to uncover latent factors, optimized via hyperparameter tuning (rank, maxIter, and RegParam).

Modeling

We then introduced a K-Nearest Neighbors (KNN) baseline model using the Surprise package, tuning for optimal k and min_k values. This approach strategizes to combat 'cold-start' problems where users have little to no rating history, or when movies have little to no ratings.

Conclusions

Implementing a hybrid recommendation system, we combined the strengths of ALS and KNN Baseline to deliver personalized movie suggestions while addressing the cold-start problem. Our grid search optimization identified KNN as the best-performing model (RMSE = 0.88), outperforming the ALS baseline (RMSE = 0.888). While KNN excels for users with rich rating histories, ALS remains essential for recommending movies to new users or sparsely rated items.

Recommendations:

Further improvements could be made by incorporating additional user behavior data—such as watch duration, rewatch frequency, and unrated views—to refine predictions. With this hybrid approach, our client can maximize user engagement and deliver high-quality recommendations as their streaming platform grows.

Thank you!!! LinkedIn - Github - Email