

Data Report on Movie Recommendation System Project

Authors:

- George Oponi
- Donnah Mwaniki
- Vincent Kiplang'at
- Brian Gikonyo
- Jimmy Mumo
- Cynthia Gitonga
- Ian Odhiambo
- Jonquil Phelan

Overview

The objective of this project is to build a movie recommendation system that provides personalized recommendations to users based on their ratings of other movies. The system will utilize machine learning algorithms to analyze user ratings and similarities between movies to generate a list of top 5 movie recommendations for each user.

Business Understanding

In today's era of streaming services and a plethora of movie options, users often struggle to find movies that align with their preferences. A movie recommendation system addresses this challenge by leveraging user ratings to deliver tailored movie suggestions. Companies in the movie streaming industry, such as Netflix, Amazon Prime Video, and Hulu, can utilize recommendation systems to enhance customer engagement and retention. By offering personalized movie recommendations, these businesses can cater to diverse audience preferences, leading to growth and competitive advantage in the entertainment sector.

Objectives.

- Develop a customized movie recommendation system using cutting edge machine learning algorithms.
- Provide personalized movie recommendations for each user based on their preferences and viewing history.
- Enhance user experience and engagement by offering relevant and tailored movie suggestions.

Data Understanding.

Data Source.

The data set was obtained from grouplens .

Data Description.

This dataset (ml-latest-small) describes 5-star rating and free-text tagging activity from MovieLens, a movie recommendation service. It contains 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users between March 29, 1996 and September 24, 2018. This dataset was generated on September 26, 2018.

Users were selected at random for inclusion. All selected users had rated at least 20 movies. Each user is represented by an id, and no other information is provided.

The data are contained in the following files: **links.csv**, **movies.csv**, **ratings.csv** and **tags.csv**.

Data Preprocessing

1. Handle missing values, duplicates, and inconsistencies: in this step of the data analysis, the quality of the dataset is investigated and addressed. That requires

that missing values, duplicates and other anomalies present within the dataset are handled accordingly.

Missing values: In handling missing values, techniques like imputation and deletion are utilized. This step maintains the completeness and robustness of the analysis.

Duplicates: Duplicate values bring about redundancy within the dataset. In removing these values the dataset is much more streamlined and thus much more meaningful insights can be obtained.

Inconsistencies: Conflicts and erroneous entries are addressed in this step. In this stage of data cleansing and validation, the reliability of the dataset as well as its coherence is upheld.

2. Exploratory data analysis (EDA): in this stage, a deeper understanding to the idiosyncrasies of the dataset is established to understand data distribution and patterns.

Patterns and Relationships: in the given dataset, the common feature is “movie id” from which all other data manipulation is performed. Other important features singled out are user ratings and genres.

Model Selection and Development

The project utilized two main models: SVD and KNN. SVD accuracy beat out KNN, with an error rate of 0.8734 for SVD. In order to boost SVD's accuracy, GridSearchCV fine tuning is utilized. Upon fine-tuning the accuracy rises to 0.8687, with the best settings being $n_factors = 50$ and $reg_all = 0.05$.

Collaborative filtering algorithms were implemented as a means of recommending movies to users.

Results

Utilizing User ratings to feed into the SVD model allows for the generation of personalized movie recommendations. This is accomplished through the use of collaborative filtering.

These techniques have remarkable efficiency in predicting user preferences as well as generating personalized recommendations.

Top Recommended Genres:

Top genres being recommended are Drama, Comedy, Thriller and Action. Classic movies such as “The Godfather” and cult classics like “Fight Club” are among top recommended.

Highlight Recommendations:

The recommendation system excels in identifying and promoting classic movies with enduring appeal as well as cult classics cherished by enthusiasts. This approach resonates highly with users seeking timeless cinematic experiences as well as those seeking a casual viewing experience. This approach towards movie recommendations keeps the users hooked and thus greater loyalty and larger revenue streams for the streaming platforms.

Conclusion

The successful development of a movie recommendation system using collaborative filtering techniques represents a significant achievement with far-reaching implications for the streaming industry. By effectively leveraging user ratings data, the system delivers personalized movie suggestions that enhance user engagement and satisfaction. Moving forward, continued refinement and optimization of the recommendation system will further solidify its role as a cornerstone of the streaming platform, driving continued growth and success in the competitive entertainment landscape.