



MOVIE RECOMMENDATION SYSTEM PROJECT



CONTENT

1. BUSINESS UNDERSTANDING
2. DATA UNDERSTANDING
3. MODELING
4. CONCLUSION
5. RECOMMENDATION



BUSINESS UNDERSTANDING



OVERVIEW

The entertainment industry is rapidly evolving, and streaming services generate vast amounts of user interaction data. Personalized recommendations have become crucial for improving user engagement and retention. The MovieLens dataset, provided by the GroupLens research lab at the University of Minnesota, offers a rich set of user ratings that can be leveraged to build a robust movie recommendation system. This project aims to develop a personalized recommendation model that provides users with highly relevant movie suggestions based on their past preferences. By utilizing collaborative filtering techniques, we can enhance the user experience by helping them discover movies they are likely to enjoy, thereby increasing platform engagement and reducing decision fatigue.

PROBLEM STATEMENT

Users often struggle with finding movies they will enjoy due to the overwhelming number of choices available on streaming platforms.

Traditional browsing methods, such as genre-based filtering or trending lists, may not always reflect an individual's unique preferences.

To address this challenge, we aim to build a recommendation system that suggests the top 5 movies for a user based on their past ratings. The system will leverage collaborative filtering techniques to analyze user interactions and provide personalized movie recommendations.

By implementing an efficient recommendation system, we can improve user satisfaction, optimize content discovery, and create a seamless entertainment experience.

OBJECTIVES

- Develop a movie recommendation model using the MovieLens dataset (100K ratings).
- Generate top 5 personalized movie recommendations for each user.



DATA UNDERSTANDING



The MovieLens dataset contains 100,000 ratings, as well as demographic information and movie metadata, collected from 9,000 movies by 600 users. The dataset was collected and made available by GroupLens, a research lab at the University of Minnesota. The purpose of the dataset is to enable research in recommendation systems and related fields.

The dataset can be downloaded from the GroupLens website (<https://grouplens.org/datasets/movielens/latest/>).

DATA EXPLORATION

The data was individually checked for null values and for duplicates in the data. The datatypes of each of their columns was checked. Different tables were merged and their contents also checked for null values and for duplicates. Columns that were not important were dropped like the timestamp column.

EXTERNAL DATA SOURCE VALIDATION

GroupLens, a research lab at the University of Minnesota made available the MovieLens dataset. The dataset has been used in various academic publications and competitions, which further validates its credibility. GroupLens is a reputable source for research in the field of recommendation system.



MODELING



We trained three models.

The models include:

- SVD(Singular Value Decomposition)
- KNNBasic(Simple nearest neighbours)
- KNNWithMeans(Neighbours but with mean adjustment)

We did hyperparameter tuning with GridSearchCV on the best model.

We then trained the best model with pipeline.



CONCLUSION

- Our recommendation system successfully generates top-5 movie recommendations for users based on past ratings.
- SVD (Singular Value Decomposition) was the best-performing algorithm, demonstrating strong predictive accuracy.
- GridSearchCV and Pipelines optimized model performance, ensuring robust parameter selection and efficient training.
- Popular movies and genres receive disproportionately high ratings, leading to potential bias in recommendations

RECOMMENDATION

- Introduce a diversity factor, ensuring recommendations include different genres rather than just high-rated films.
- Allow users to explicitly specify their preferences (favorite genres, actors, etc.). Provide personalized dashboards with insights into user movie-watching patterns.
- Increase the promotion of less popular movies. By promoting these movies, it may lead to an increase in their popularity and more positive ratings.
- Increase the supply of drama, comedy, and action movies: These genres were the most popular among the users, and therefore, there is a higher chance of success if more movies in these genres are pushed in the platform.
- Consider user-generated tags: Although the tags were not included in this analysis, they can provide valuable insight into how users perceive movies. By analyzing user-generated tags, it may help in understanding the users' preferences and improving the movie recommendation system.



THANK YOU