

# A REPORT ON BUILDING A MOVIE RECOMMENDATION SYSTEM USING COLLABORATIVE FILTERING

## 1. BUSINESS UNDERSTANDING

### 1.1 Business overview

XMovies is a digital cinema company providing movies and tv shows on demand through their subscription platform. The company is seeking to develop a recommendation system based on user ratings that can enhance user satisfaction, engagement, and revenue generation for the platform. The system should leverage data science techniques to analyze user preferences and generate accurate and personalized recommendations, thereby increasing user retention, driving user interaction, and boosting overall platform usage. This will also grant a competitive advantage to other platforms offering the same services.

### 1.2 Objectives

#### *Main Objective*

The primary goal of the project is to build a movie recommendation system that provides top 5 movie recommendations to a user based on their ratings of other movies.

#### *Specific Objectives*

To enhance user satisfaction by providing accurate and personalized movie recommendations that align with each user's individual tastes and interests. Improve user retention by delivering a highly engaging and satisfying user experience through personalized recommendations.

### 1.3 Business criteria

To build a recommendation system that employs data science techniques to analyze user preferences and generate accurate and personalized recommendations, thereby increasing user retention, driving user interaction, and boosting overall platform usage.

### 1.4 Situation Assessment

#### *Datasets*

This project uses the MovieLens dataset that describes 5-star rating and free-text tagging activity from MovieLens, a movie recommendation service. This dataset has the following files;

- Movie dataset
- Ratings dataset
- Tags dataset
- Links dataset

#### *Softwares used*

- Jupyter Notebook
- Git
- Github
- Numpy Pandas
- scikit learn
- Canva

### *Assumptions*

Data provided is correct and upto date

## **2. DATA UNDERSTANDING**

For this project we will use the MovieLens dataset (ml-latest-small) as provided by the GroupLens research lab at the University of Minnesota. The dataset 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.

## **3. DATA PREPARATION**

This are the steps in preparing data

### *LOADING DATA*

load data from the excel sheets downloaded from Movielens as a csv file

### *EXPLORATION*

Loading the data and viewing the data using the `.shape()` , `.head()` , and `.info()` functions

### *ANALYSIS*

We performed various analyses in our dataset such as Exploratory Data Analysis for example generating the top 10 genres using the `ratings` data

## **4. MODELLING**

### *Preprocessing*

we Converted the DataFrame into a Surprise Dataset then Split the data into training and testing sets

### *model building and evaluation*

We used the following models to build the recommendation system

1. The SVD algorithm using KNNBaseline and KNNBasic for collaborative filtering

The models were evaluated using the RMSE and the MAE

We also tried the following methods to improve the model performance

- Ensemble methods
- Matrix factorization
- Collaborative filtering using baseline only
- Hybrid system using both collaborative and content filtering

## CONCLUSION

- SVD and BaselineOnly had the lowest RMSE values indicating better prediction accuracy compared to the other models.
- Top 10 most produced genres were the top 10 most rated genres

## RECOMMENDATIONS

- As SVD and BaselineOnly model had the lowest RMSE and relatively low MAE values, meaning they are both a good option for the collaborative filtering recommendation system, we recommend comparing these two models based on other metrics, such as coverage and diversity, to make a final decision.
- The ensemble model also performs well with a slightly higher RMSE compared to SVD. As this model allows for diversity and variety in recommendations, we recommend applying the ensemble model and seeing how it performs on the recommendation system.
- The Hybrid strategy that combines collaborative filtering and content-based filtering is a good alternative option although it has a higher RMSE compared to SVD, we recommend further investigation and fine-tuning this strategy to improve its performance.

In [ ]: