

[Code](#)
[Issues](#)
[Pull requests](#)
[Actions](#)
[Projects](#)
[Wiki](#)
[Security](#)
[Insights](#)
[Settings](#)

main

 b-irungu Add files via upload ... 2 minutes ago ⌚ 12

[View code](#)

☰ README.md



This repository contains the following

- A Jupyter notebook
- A non-technical presentation.
- A CRISP-DM report
- A README file
- MOVIELENS data

BUSINESS OVERVIEW

Developing a recommendation system based on user ratings that can enhance user satisfaction, engagement, and revenue generation for the XMOVIES 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.

DATA The project uses the MovieLens dataset from [MovieLens](#) , which contains the following files;

- links data
- tags data
- movies data
- ratings data

The MovieLens dataset provides a rich and reliable source of movie ratings and associated information, making it a suitable choice for building and evaluating movie recommendation systems.

COLLABORATIVE FILTERING

collaborative filtering was used in the Recommendation system and it involved the following steps

1. Converting the DataFrame into a Surprise Dataset
2. Splitting the data into training and testing sets
3. Hypertunning Dataset to get the best parameters by Performing a gridsearch then fitting the SVD model
4. Cross-validation using KNNBasic and KNNBaseline; Evaluation metrics: RSME and MAE
5. Making predictions
6. Testing the recommendation systems

COLLABORATIVE FILTERING RESULTS

The result shows the best scores achieved by the SVD algorithm are RMSE OF 0.8693804004379402 and MAE of 0.6684190070988877

Comparison results between KNNBasic Algorithm and KNNBaseline Algorithms result:

- The KNNBaseline algorithm indicated a lower RMSE of 0.8774070745386995, which suggests better prediction accuracy compared to KNNBasic.
- The KNNBasic algorithm indicated a higher RMSE of 0.9730103781790413, indicating a slightly higher prediction error.

Based on the RMSE values provided, KNNBaseline outperformed KNNBasic in terms of prediction accuracy and recommendation quality.

IMPROVING THE RECOMMENDATION SYSTEM

The following techniques were adopted to try and improve the RMSE and MAE of the model; 1.

Ensemble methods

From this method, the 'RMSE' Increased by 0.01665019 from the results of the previous model, giving us an 'RMSE' of 0.8860305908074744 and MAE: 0.6800762378176585

2. Matrix factorization

From this method, the 'RMSE' Increased by 0.1084196 from the results of the previous model, giving us an 'RMSE' of 0.9778 and Average MAE: 0.7535

3. Collaborative filtering using BaselineOnly

From this method, the 'RMSE' Increased by 0.0050196 from the results of the previous model, giving us an 'RMSE' of 0.8744

4. A hybrid system using both collaborative and content filtering

From this method, the 'RMSE' Increased by 0.0739196 from the results of the previous model, giving us an 'RMSE' of 0.9433

CONCLUSION

- SVD and BaselineOnly had the lowest RMSE values indicating better prediction accuracy compared to the other models.
- The Ensemble model performs slightly worse than SVD in terms of RMSE but is still considered to have good accuracy.
- The ALS-based Matrix Factorization model and MF algorithm with ALS have higher RMSE values, suggesting lower accuracy.
- The Hybrid strategy also shows a higher RMSE value compared to SVD, indicating comparatively lower accuracy.

RECOMMENDATIONS

- Implement the SVD model to develop the movie recommendations system as it had the lowest RMSE and relatively low MAE values giving better prediction accuracy in movie recommendations
- As SVD and BaselineOnly models 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.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%