

10.02 Exercise

Ross L. Kim-Schreck

DSC630 Predictive Analytics

Professor Hua

2024.05.16

10.02 Exercise

Write up

I used two datasets: ratings.csv as dt01 and movies.csv as dt02. I began by confirming the column names and dataset dimensions. On 10.02.03-01, I then merged the datasets. I then renamed the columns for convenience in future steps. On 10.02.04-01, I started building a model to calculate RMSE and MAE. I built a singular value decomposition algorithm (SDV) to break down the amount of data and to make predictions on unrated data. RMSE Mean is 0.89 and MAE is 0.68. RMSE is quite high, but MAE which is low indicating high model accuracy.

I then created a function to return movie title input, rate input, predict movie rating of users, sort ratings, and recommend movie titles. I then tested the function using a random movie. It seems that the function works as it returned ten selected movies based on my input.

In 10.02.11-02, I then created functions for users to be able to search movie titles. Could use fuzzy matching if user input is inaccurate as the current function requires a high level of input accuracy to return results. These functions require further development.

References

<https://www.geeksforgeeks.org>

<https://www.kaggle.com>

<https://365datascience.com>