Name: Yamini Aalla SJSU Id:014331018

RMSE Score: 1.55739

Goal: To Build a recommender system for a given book dataset.

Dataset:
There are two input files
1.train.csv – The Dataset has three features user_id, book_id and rating with total non-null values of 700000.
2.test.csv- The Dataset has two features user_id and book_id with total non-null values of 299606.

Implementation:

Reading the csv files to pandas data frame, invoking reader instance from surprise library and loading the dataset to surprise data structure using load_from_df()

Distribution of Ratings:
Ratings for the given user and books are in range from 0 to 5.
Where 239522 of the ratings are 0, 4448 of the ratings are 1.
Rating Distribution by User:
User with id 20755 gave highest number of 4287 ratings.
Rating Distribution by Book:
The most rated book has received 11213 ratings.

Cross-Validation:
For cross-validation I have used K-fold cross validation with number of splits = 5.

Algorithm:
From scikit-surprise chosen Matrix Factorization based SVD algorithm to predict the ratings for given user and book id.
SVD algorithm represents the rating matrix as the product of matrices representing user's factors and book factors.

Implemented SVD algorithm with following parameters:

- Learning rate for all parameters(lr_all) = 0.01,
- Regularization term for all parameters(reg_all) = 0.1,
- Number of factors(n_factors) = 200 and
- Number of iterations n_epochs = 105

For predicting the ratings for given test.csv file used predict() method.

Output File:

As per requirements the output file should be a csv file to achieve this:

- Created a dictionary with a two key value pairs, where first pair formed by appending both user and book ids with '-'and other pair is predicted ratings.
- Converted the dictionary to pandas data frame.
- By using .to_csv() method converted the data frame to csv file.

References:

https://surprise.readthedocs.io/en/stable/matrix_factorization.html
https://surprise.readthedocs.io/en/stable/getting_started.html#use-cross-validation-iterators

https://surprise.readthedocs.io/en/stable/getting_started.html#train-on-a-whole-trainset-and-the-predict-method
https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_csv.html