RECOMMENDATION SYSTEM

SAI GURRAM

INTRODUCTION

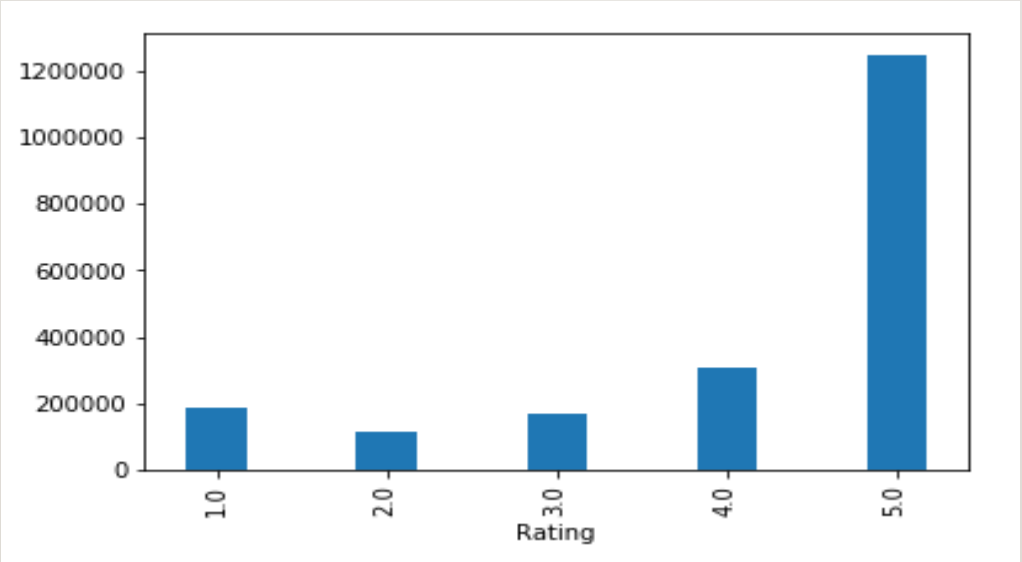
- Content-Based – Recommended based on content
- Item-Based – Recommended based on similarity between items
- User-Based – Recommended based on similarity between users
- Model-Based – Uses Matrix factorization and recommends based on users/items.
- Hybrid – Netflix, Amazon and other organizations are not relying on single approach and are combining the above approaches together a creating a hybrid system based on their business needs.

DATASET

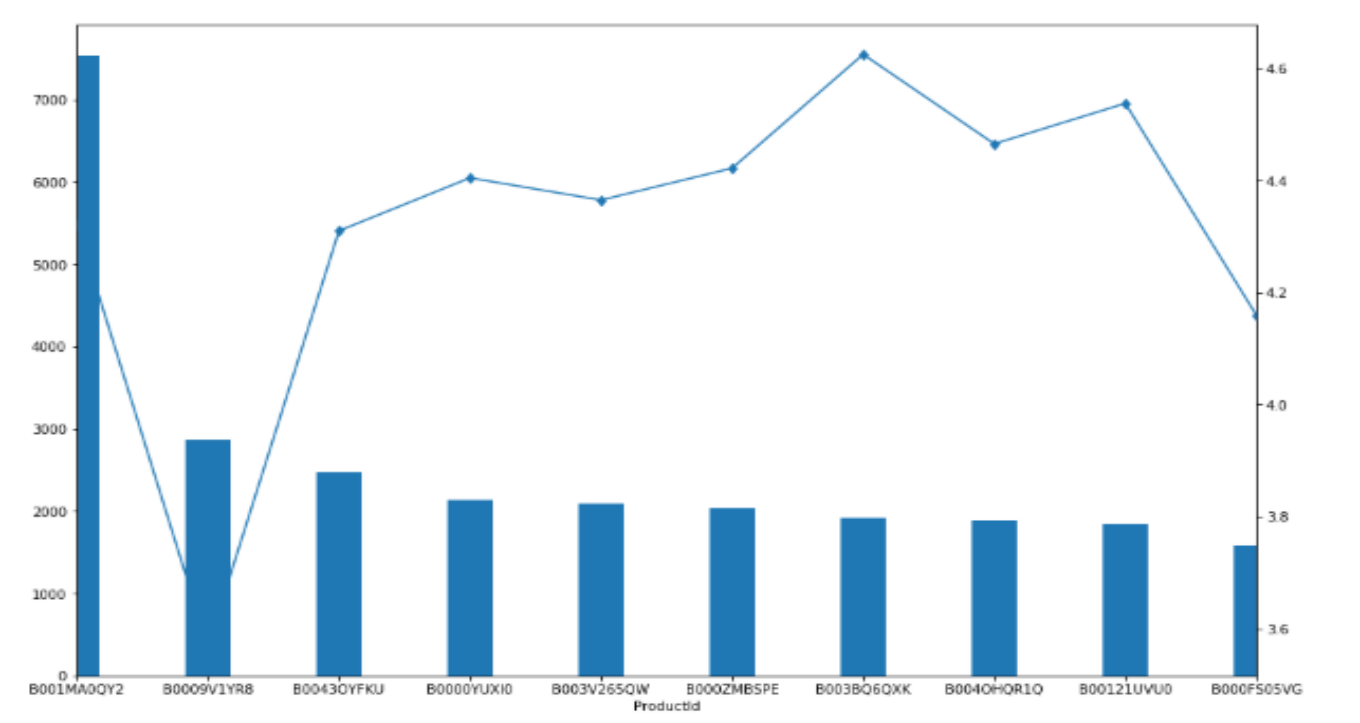
- Consists of Userid, productid, timestamp, ratings
- More than 2M records
- No null records
- No duplicates

DATA EXPLORATION

- The Average rating is 4.1 and more than 75% of the data has greater than 4 rating



TOP 10 PRODUCTS WITH USER COUNT AND AVERAGE RATING



Even though the user count for the first two products is higher the average rating is higher for the other products.

SUBSETTING THE DATA

- For computational reasons, subsetted the data to contain the only top 100 products.
- The new data contains 107098 records.

COSINE SIMILARITY

- Created an empty matrix and calculated cosine similarity between 100 products.
- There is a built in function in sklearn to calculate cosine similarity.
- Converted the similarity matrix into a dataframe.
- Obtained the top 4 products that can be recommended based on how similar they are to each other.
- Example: For product B0009FHJRS recommended products are B000UVZU1S, B003UH0528, B006L1DNWY.

SVD

- Converted the data into 10 columns using Singular Value Decomposition. This technique ensures all user-item interactions are captured and condensed into 10 columns.
- There is a built in function in sklearn –TruncatedSVD.
- Created a correlation matrix using the decomposed features and converted into a dataframe.
- Obtained the top 4 products that can be recommended based on how similar they are to each other.
- Example: For product B0009FHJRS recommended products are B0009OAGXI, B000L596FE, B00121UVU0 and B00639DLV2.

SUMMARY

- The model based recommendation system can now be used to recommend products.
- The sparsity and scalability issue of user based/item based systems has been resolved in model based system.

Areas of Improvement:

- Other Matrix factorization techniques can be used to improve model's performance.
- For item based approach adjusted cosine similarity and jaccard similarity techniques can be used.
- The model can be expanded to all the products in the dataset.
- A Hybrid system can be created by combining item based and model based system to improve the recommendations.