Amazon Product Recommendation System

Project 3 - Recommendation Systems

Amanda Edwards
7/21/2025

Contents / Agenda

- Business Problem and Data Overview
- Exploratory Data Analysis
- Rank Based Model
- User-User Similarity-based Model
- Item-Item Similarity-based Model
- Matrix Factorization based Model
- Conclusion and Recommendations

Business Problem and Data Overview

- Business problem: This project aims to build a recommendation system for Amazon to recommend items to users based on their product history ratings or how similar users have rated products.
- Overview of the dataset:
    - There are 3 variables: userId, productID, and rating.
    - Due to the large dataset (7,824,482 observations), we have filtered it to include:
        - Users that have rated at least 50 products
        - Products that have at least 5 ratings

Exploratory Data Analysis

- Number of rows in the data set = 65290
- Number of columns in the data set = 3
- Data types:
    - user_id and prod_id are ‘object’ variables
    - rating is a ‘float’ since it has a decimal
- No missing values in any of the columns

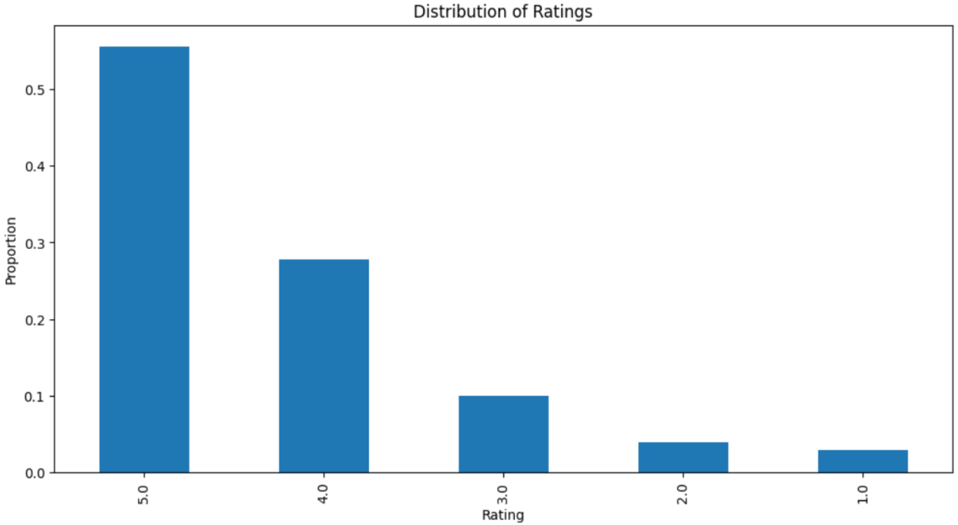
Exploratory Data Analysis

- Summary Statistics:
  - Count = 65290 – aligned and equivalent to the number of rows in dataset
  - Minimum and Maximum value are appropriate of 1 and 5, respectively, no abnormal values
  - 25th percentile value is 4, meaning only 25% of the data points are below this level; and 75th percentile value is 5.
  - Mean of 4.29 with standard deviation of 0.989
  - The data is skewed left due to more values in the higher ranges

	rating
count	65290.000000
mean	4.294808
std	0.988915
min	1.000000
25%	4.000000
50%	5.000000
75%	5.000000
max	5.000000

Exploratory Data Analysis

- The accompanying bar plot shows the distribution of 'rating’:
    - Rating of 5.0 comprises about 57% of the reviews
    - Rating of 4.0 makes up approximately 28% of the reviews
    - Ratings of 3.0, 2.0, 1.0 make up the remaining approximately 15% of reviews



Exploratory Data Analysis

- Number of observations in final data = 65290
- Number of unique users = 1540
- Number of unique items = 5689
- We filtered the data to only include users who rated over 50 items, so this aligns with the fact that there are less number of users compared to items rated. We can also infer that items are being rated by multiple users.
- Looking at individual users, we grouped the data and sorted by the user with the most reviews – one user rating 295 different products

Rank Based Model

- Approach taken to build the Rank-based Model:
    - We grouped the data by product ID and then found the mean and count of the ratings
    - Created a dataframe with these columns and values
    - Defined a function to get the top n products based on the highest average rating and minimum interactions based on popularity

Rank Based Model

- Using the Rank-based Model the following top 5 recommendations were made based on popularity:
  - With minimum interactions of 50:
    - 'B001TH7GUU', **'B003ES5ZUU'**, 'B0019EHU8G', 'B006W8U2MU', 'B000QUUFRW'
  - Minimum minimum interactions of100:
    - **'B003ES5ZUU',** 'B000N99BBC', 'B007WTAJTO', 'B002V88HFE', 'B004CLYEDC'

- One of the products, **B003ES5ZUU,** is included in both groups, indicating it consistently is rated well among users

User-User Similarity-based Model Performance

- RMSE = 1.0012:
  - RMSE shows how close predicted ratings are to actual ratings. This score means that predictions are about 1 point variation from actual score
- Precision = 0.855:
  - 85.5% of the items that were predicted as relevant to the user, were actually relevant. This percentage is good and the model isn't recommending many false positives
- Recall = 0.858
  - Recall means that of the relevant items to a user, 85.8% were recommended by the model (of the users relevant items, the model recommended 85.8% of them)
- F_1 score = 0.856:
  - Harmonic mean between precision and recall – shows good accuracy of the model in recommendation

User-User Similarity-based Model – Hyperparameter tuning

- After hyperparameter tuning, the user-user similarity model had the following results:
  - RMSE = 0.9526
  - Precision = 0.847
  - Recall = 0.894
  - F1-Score = 0.87

- RMSE, F1-score and recall all improved over the first user-user model. Precision reduced from 0.855 to 0.847.
- Since RMSE and F1 are better in the tuned model, this is the better model because predicted ratings have less error and accuracy is better.

User-User Prediction Ratings

- Using the original model and after hyperparameter tuning, we predicted how a user would rate a product.
- Prediction for product #1400501466:
  - User A3LDPF5FMB782Z:
    - Estimated rating = 3.4
    - After tuning estimated rating = 4.29 (results indicate not enough nearest neighbors)
  - User: A34BZM6S9L7QI4:
    - Estimated rating = 4.29 (results indicate not enough nearest neighbors)
    - After tuning – estimated rating = 4.29 (results indicate not enough nearest neighbors
- Prior to tuning, the first user had an estimated rating of 3.4 for the above item, however after tuning the estimation showed that there were not enough nearest neighbors. This was seen for user A34BZM6S9L7QI4, as well. This may indicate that we need to run a different model.
- Since the value returned is 4.29 when lacking enough data points, the model may be providing the average ranking score (observed in summary statistics)

Item-Item Similarity-based Model

- Item-item model results:
  - RMSE: 0.9950; Precision: 0.838; Recall: 0.845; F_1 score: 0.841
- Predicting product #1400501466 rating:
  - User: A3LDPF5FMB782Z
    - Predicted rating = 4.27
  - User: A34BZM6S9L7QI4
    - Predicted rating = 4.29 (lacking enough nearest neighbors)
- The model predicted the above ratings for these two users. For the second user, the result indicated that the prediction was lacking enough nearest neighbors to make the estimate.

Item-Item After Hyperparameter Tuning

- Item-item model results (after tuning):
  - RMSE: 0.9576; Precision: 0.839; Recall: 0.88; F_1 score: 0.859
- Predicting product #1400501466 rating:
  - User: A3LDPF5FMB782Z
    - Predicted rating = 4.67
  - User: A34BZM6S9L7QI4
    - Predicted rating = 4.29 (not enough nearest neighbors)
- After hyperparameter tuning, we have a higher predicted rating for user A3LDPF5FMB782Z but the same rating for the second user, most likely due to lack of nearest neighbors.
- Additionally, the model result show improved results for RMSE, precision, recall and F1_score

Matrix Factorization based Model

- Original model results:
  - RMSE: 0.8882; Precision: 0.853; Recall: 0.88; F_1 score: 0.866
- Results after hyperparameter tuning:
  - RMSE: 0.8808; Precision: 0.854; Recall: 0.878; F_1 score: 0.866
- Predicting product #1400501466 rating:
  - User: A3LDPF5FMB782Z
    - Predicted rating = 4.08
    - After hyperparameter tuning, rating = 4.13
  - User: A34BZM6S9L7QI4
    - Predicted rating = 4.40
    - After hyperparameter tuning = 4.22

Matrix Factorization based Model

- Comparing original model and after tuning, the results are very similar between RMSE, precision, recall, and F1_score.
- The predicted rating was slightly higher for user A3LDPF5FMB782Z after hyperparameter tuning (4.08 versus 4.13)
- The predicted rating was lower for user A34BZM6S9L7QI4 after hyperparameter tuning (4.40 versus 4.22)

Conclusion and Recommendations

- Results summary/conclusion:
    - Since we want **RMSE** to be low, indicating most accurate results for predicing product rating, the matrix factorization model after hyperparameter tuning was best at 0.8808. When assessing the model, we want precision, recall and F1_score to be high as these metrics relate to relevance of a recommendation for the user. **Precision** was highest with user-user model at 0.855, though matrix factorization tuned model was very close at 0.84. Additionally, **recall** was 0.894 with the user-user model after hyperparameter tuning. The F1_score is the harmonic mean between precision and recal and was highest in the user-user model after tuning at 0.87.

Conclusion and Recommendations

- Model choice recommendation:
  - Based on these results, the user-user model after hyperparameter tuning showed best F1_score and recall, however when we predicted product ratings, there were not enough nearest neighbors. This suggests that there may be sparcity in the data, making it difficult to gain information from similar users or items to make predictions. The matrix factorization model had no indication of lacking nearest neighbors, and had best RMSE of all models. Precision was also only 0.01 lower than the best score.
  - **Therefore, I would recommend utilizing the matrix factorization tuned model for recommending products to users.**

- Additional recommendations:
  - From the rank based model, we determined 5 top ranked products that received a minimum of 50 reviews and 100 reviews. These items can be used as general recommendations based on popularity or if there are scarce data points for certain users.
  - B003ES5ZUU is consistently rated high among users. It might be beneficial to provide this item as a recommendation and look into this product to determine why it is so popular.

APPENDIX

Happy Learning !