Amazon

Recommendation System

Diego Morales
November 2024

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

Project Overview

Recommendation systems have become essential for businesses to enhance user engagement and drive sales. By providing personalized and relevant product suggestions, these systems help consumers navigate the overwhelming number of choices available online.

This project focused on building a product recommendation system for Amazon using a dataset of user ratings for electronic products. The goal was to predict the ratings of products a user has not yet interacted with, allowing the system to offer personalized recommendations.

Data Overview

The dataset is a collection of labeled data of Amazon reviews of products. It contains the following attributes:

**userId:** Every user identified with a unique id
productId: Every product identified with a unique id
Rating: The rating of the corresponding product by the corresponding user
timestamp: Time of the rating. We will not use this column on this project

Exploratory Data Analysis

▶ The original dataset was reduced from 7,824,482 to **65,290 observations** by taking **users who have given at least 50 ratings**, and the **products that have at least 5 ratings**.

▶ Ratings range from **1 to 5**. The **average rating is 4.29**.

Distribution of Ratings
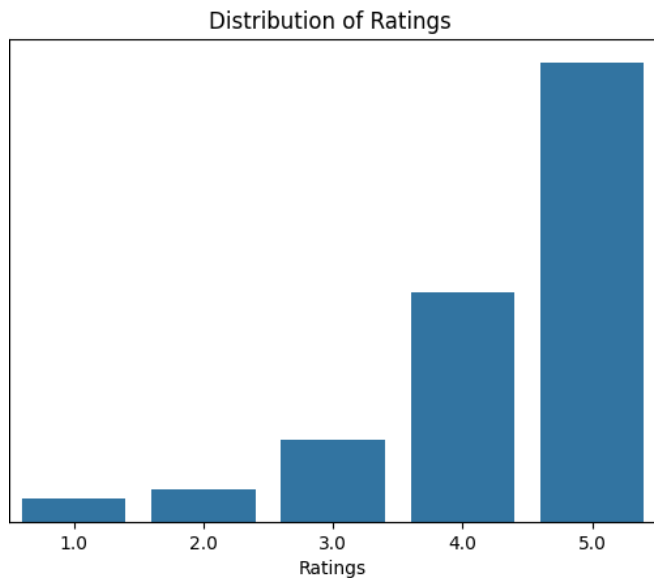
Rank-based Model

The model was created by calculating the average rating of products with at least 100 ratings. Below are the top 5 products based on popularity:

prod_id	avg_rating	rating_count
B003ES5ZUU	4.864130	184
B000N99BBC	4.772455	167
B002WE6D44	4.770000	100
B007WTAJTO	4.701220	164
B002V88HFE	4.698113	106

User-User Similarity-based Model

This model predicts ratings using data from users similar to the user of interest.

Results:

- 84% of recommended products are relevant
- 81% of relevant products are being recommended
- F1 score = 82%

Item-Item Similarity-based Model

This model predicts a user's rating for a product by analyzing ratings of items similar to the product of interest. Similarity between products is calculated based on user rating patterns.

Results:

- 83% of recommended products are relevant
- 80% of relevant products are being recommended
- F1 score = 82%

Matrix Factorization based Model

This model identifies latent features of users and items. It predicts ratings by decomposing the user-item interaction matrix to capture underlying patterns in the data.

Results:

- 85% of recommended products are relevant
- 80% of relevant products are being recommended
- F1 score = 82.5%

Conclusions

- The user-based and item-based models had a very similar performance. There was a slight difference in precision and recall but they achieved the same F1 score.

- The model based on Matrix Factorization achieved the highest F1 score and provides the best trade-off between precision, and recall. Therefore, my recommendation to the company is to implement this model for their recommendation system.