Product Recommendation

Department of Computer Science
New York Institute of Technology

Big Data Analytics
Prof:- Xueqing Huang
CSCI:-636

Abighna Allu
Ruchitra Kumaresan
Vineesha Chowdary
Sai Dheeraj Reddy Yapala

Table of Context

Abstract

Product Recommendation System is a machine learning-based project that provides personalized product recommendations to users based on their browsing and purchase history. The system utilizes collaborative filtering and content-based filtering algorithms to analyze user behavior and generate relevant recommendations. This project aims to improve the overall shopping experience for users, increase sales for e-commerce businesses

Introduction

In the expansive realm of global e-commerce, Amazon emerges as an unparalleled giant, captivating consumers with its unparalleled convenience, vast product selection, and a distinctive feature that significantly influences purchasing decisions—the wealth of product reviews contributed by fellow shoppers. These reviews form a virtual marketplace within the digital platform, serving as a rich tapestry of shared experiences, opinions, and insights. Beyond the transactional nature of online shopping, Amazon's product reviews create a dynamic community where customers engage in a collective dialogue about various products. This communal exchange fosters a sense of trust and transparency as potential buyers navigate the digital aisles armed with real-world feedback from their peers. The impact of these reviews is pivotal, shaping the online shopping landscape by empowering customers to make informed decisions. Buyers, armed with a plethora of perspectives, can assess products comprehensively, leading to more satisfying purchases. As a result, Amazon's review system functions as a valuable informational resource. It cultivates a sense of belonging within the vast e-commerce space, where shoppers are connected through shared experiences and a common pursuit of quality and satisfaction. Essentially, Amazon's product reviews are a cornerstone of the platform's success, transforming online shopping into a collaborative and community-driven endeavor.

Data Collection

Data Set Link:-

https://www.kaggle.com/datasets/arhamrumi/amazon-product-reviews

The dataset containing over 568,000 consumer reviews on various Amazon products is a valuable resource for our project, primarily due to its comprehensive set of variables that provide diverse insights into customer feedback. Including essential fields such as ID, ProductId, UserId, ProfileName, Helpfulness Numerator, Helpfulness Denominator, Score, Time, Summary, and Text ensures a rich and multifaceted perspective on consumer experiences. Our analysis and recommendation system specifically focused on leveraging ProductId, UserId, and Score. ProductId and UserId enable us to identify specific products and users, facilitating a more personalized and targeted approach to recommendations. The Score variable, representing the rating given by users, serves as a crucial indicator of customer satisfaction. By utilizing these key variables, we aim to extract meaningful patterns and trends from the dataset, enhancing the accuracy and relevance of our recommendations. Overall, the dataset's richness in relevant information makes it a pertinent and beneficial asset for our project, allowing us to build a robust recommendation system based on the experiences and opinions of a large and diverse consumer base.

Code
Link to our Code:-

https://colab.research.google.com/drive/1uKfqyS0HIxMrRCdhf7NUJcQNSQVKBPtr?usp=sharing

Here we have used three different approaches

1. **Rank Based Product Recommendation**
   Objective :
   - Recommend products with highest number of ratings
   - Target new customers with most popular products

   The result will be the top 5 products with 20 and 50 as maximum ratings or interactions

So we approached this method by
- calculating the average rating for each product
- calculating total number of ratings for each product
- created a dataframe using these values and sort it by average
- Writing a function to get n top products with specific minimum number of interactions.

2. **Similarity Based Collaborative Filtering**
   Objective:
   Provide personalized and relevant recommendations to users

The output of this will be the top 10 products based on interactions of similar users
   Approach:
   - We wrote similarity score of the desired user with each user in the interaction matrix uses cosine similarity score and append to an empty ;ist and sort it
   - Extract the similar user and similarity scores from the sorted list
   - Remove original user and its similarity score and return the rest
   - For each similar user we find n products with which the similar user has interacted with but not the actual user
   - And return the number of products

3. **Model Based Collaborative Filtering**

   Objective:

   Provide personalized recommendations to users based on their past behavior and preferences, while also addressing the challenges of sparsity and scalability that can arise in other collaborative filtering technologies. The output will be the recommended top 5 products for a particular user

   Approach:

   - Taking the matrix of product ratings and converting it to a CSR(compressed sparse row) matrix. This is done to save memory and computation time, since only the non-zero values need to be stored.
   - Performing singular value decomposition(SVD) on the sparse or csr matrix. SVD is a matrix decomposition technique that can be used to reduce the dimensionality of a matrix of product ratings to 50 latent features
   - Calculating the predicted ratings for all users using SVD. the predicted ratings are calculated by multiplying the U matrix, the sigma matrix and Vt matrix.
   - Storing the predicted ratings in a dataframe. The dataframe has the same columns as the original matrix of product ratings. The rows of the dataframe corresponds to the users. The values in the dataframe are the predicted ratings for each user

Result

We finally conclude that each model has their own pros and cons. And rank based is easy as it does not involve any complex algorithms. But each model gives different results based on the user, ratings and similarity, so it depends on the user which product they want to pick in the end.

References

- https://towardsdatascience.com/what-are-product-recommendation-engines-and-the-various-versions-of-them-9dcab4ee26d5
- https://www.itransition.com/machine-learning/recommendation-systems
- https://gritglobal.io/blog/product-recommendation-system/
- https://www.javatpoint.com/product-recommendation-machine-learning