Optimal Product Ranking for Improving Customer's Decision-Making

Tarun Khanna tarun19339@iiitd. **Yugansh Sharu** yugansh19347@iiitd.ac.in **Ritik Aggarwal** ritik19320@iiitd.ac.in
Tanishq tanishq18108@iiitd.ac.in **Saurav Ranjan** saurav19387@iiitd.ac.in

Problem Formulation

E-commerce websites are known for offering an extensive range of products. However, this can make it challenging for shoppers to locate relevant information efficiently, even when using user-generated content. While simplistic ranking methods like price, rating, and relevance are commonly provided, they do not offer a personalized shopping experience. To overcome this limitation, a more individualized system can be implemented. Such a system would consider several factors, including the user's recent shopping and browsing history, item characteristics, user interaction with the product page, and micro/macro behaviors. By combining these factors, the system would produce more personalized recommendations for customers, enhancing their overall shopping experience.

Introduction

Our approach requires considerable background information about user interactions and items. This information needs to be cataloged and added to the system whenever there are new items or users. These interactions and behaviors by customers are called Micro Behaviors. The use of micro behavior data can be highly effective in content-based recommender systems, allowing for personalized recommendations that meet the unique needs of each user.
For Implementing our system, we are considering three basic user interactions, i.e., View, Cart, and Purchase, for each item and calculating an overall interaction score for each user item. One advantage of this approach is that it can overcome the "cold start" problem, which occurs when there are few or no user-item interactions. Due to limited data availability within a brief period of time, we opted to use content-based filtering to recommend products on e-commerce. Despite this, the recommender system can still provide somewhat accurate recommendations to the user.

Related Works

The traditional approach to recommender systems focuses on macro interactions between users and items but suffers from data sparsity problems.
However, recent studies have shown that micro-behaviors, such as complaints, review comments, and post-click behaviors, can offer a deep understanding of users and improve recommendation accuracy. Several studies have developed models incorporating micro-behaviors and item knowledge to address data sparsity issues in session-based recommendation systems. Multi-task learning for session-based recommendation: Including User Micro-behaviors and Item Knowledge (Meng et al., 2020) develop Session-based Recommendation (SR) to predict the next item based on a user's session. It shows that micro-behaviors offer deep insight into a user's preference, while item knowledge provides side information to alleviate data sparsity.

These models use clustering, multi-task learning, and neural graph networks to capture user preferences and transition patterns. Furthermore, micro-behaviors such as reading comments, wishlisting, and dwell time can alter the probability of an order getting placed. Reading patterns can be leveraged to estimate user preference levels. Overall, micro-behaviors provide a new perspective for improving the accuracy of e-commerce recommender systems. Enhanced Recommendation System for Post-Click Behaviors Liang et al. (Post-Click Behaviors Enhanced Recommendation System, 2020).introduce a recommendation system that leverages these patterns to estimate users' preference levels. We studied different scientific literature already published / available exploring this approach towards product recommendation and have incorporated suggestions and improvements from them to create a more suitable model.

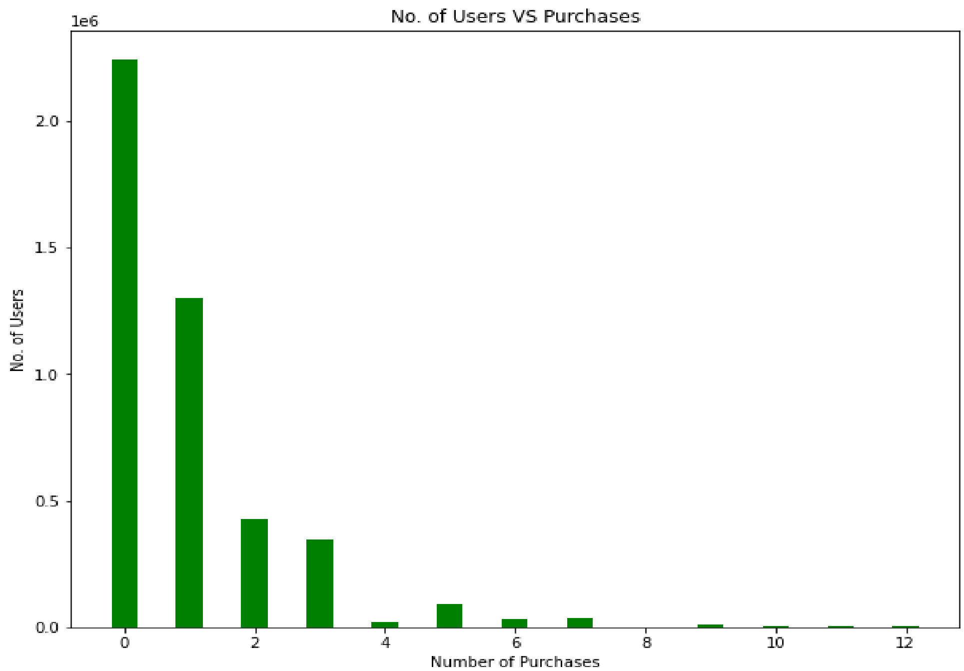
Preprocessing

To build a content-based recommender system for an e-commerce web store, we used the "event_type" column of the dataset as the target, which has three categories: View, Cart, and Purchase. User scores were assigned to these interactions, with a view scoring 1, a cart scoring 10, and a purchase scoring 50. The items were then categorized into five price categories relative to their item categories.

As multiple interactions can occur per user per item, we calculated the sum of user scores for each unique user-item interaction using a group by operation. We applied MinMaxScaler to ensure consistency and obtain interaction scores between 0 and 1. An interaction score above 0.5 indicates a high purchase probability, while scores below 0.5 suggest otherwise. These scores are from the ratings for the User-Item Matrix used for model building. The prices category column is created where we have labeled all products of each category from 1 to 5 on the basis of their price.

Choosing a Model

An important preliminary step in model selection and development is to visualize the target variable of the recommender system, which includes the categories of View, Cart, and Purchase. The analysis results show that the distribution of the target variable is highly imbalanced, with a significantly low conversion rate of only from view to purchasing.



- **Split the data**:

  To select the appropriate model for the recommender system, we began by dividing the entire dataset into separate training and test datasets using a Simple Validation approach and split into the ratio of 3:2 of training to test, respectively.

- **Generate User-Item matrix**

  Next, the training set was transformed into a sparse User-Item Matrix and assigned empty entries '0' for the user score.

- **Selected model**

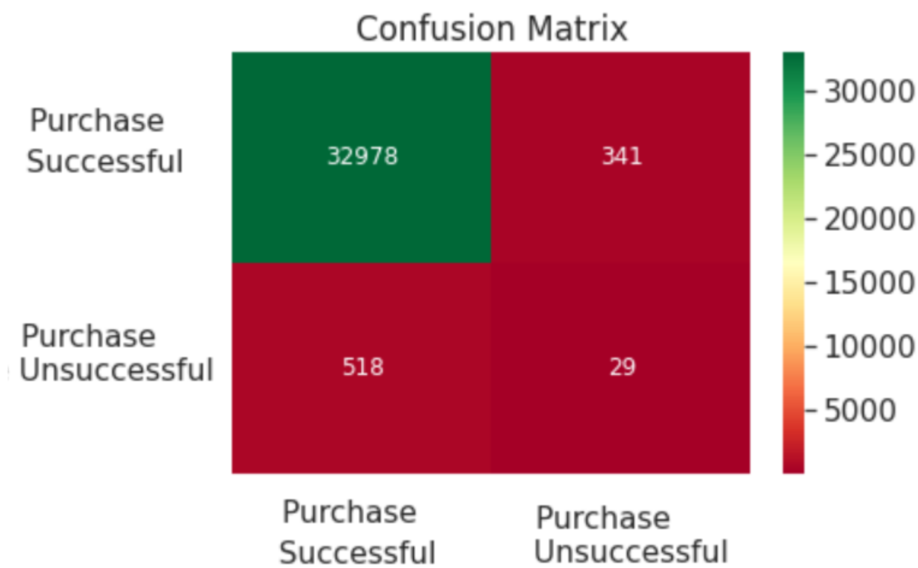
  Select Model for filtering by item, price, and brand similarity matrix:- We created three item-to-item similarity matrices based on price, category, and brand. The price categories were generated during data preprocessing, and we used this data to calculate the similarity between each item's price category using the inverse of the Euclidean method. We used cosine similarity for the similarity between each category and brand of items. Next, we multiplied all three similarity matrices of price, category, and brand and took their dot product with our user-item similarity matrix. Finally, we scaled the resulting output matrix between 0 and 1.

- **Prediction**

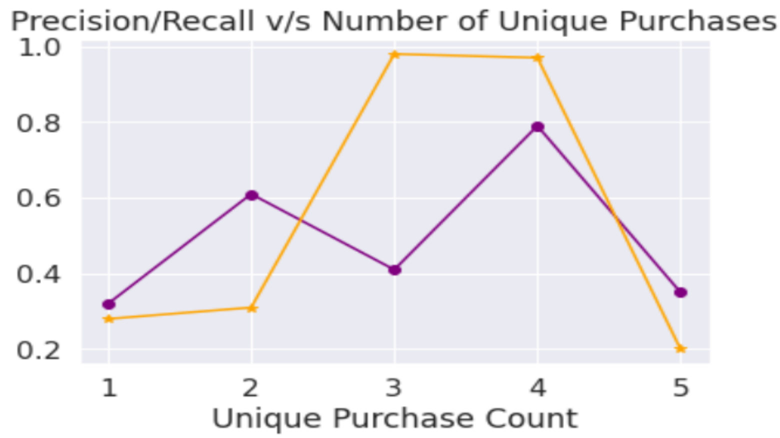
The final matrix is merged with the test dataset. Based on the final similarity score, purchase predictions are calculated. If the similarity score is above 0.5, then the item is considered purchasable.

- **Training the Model**

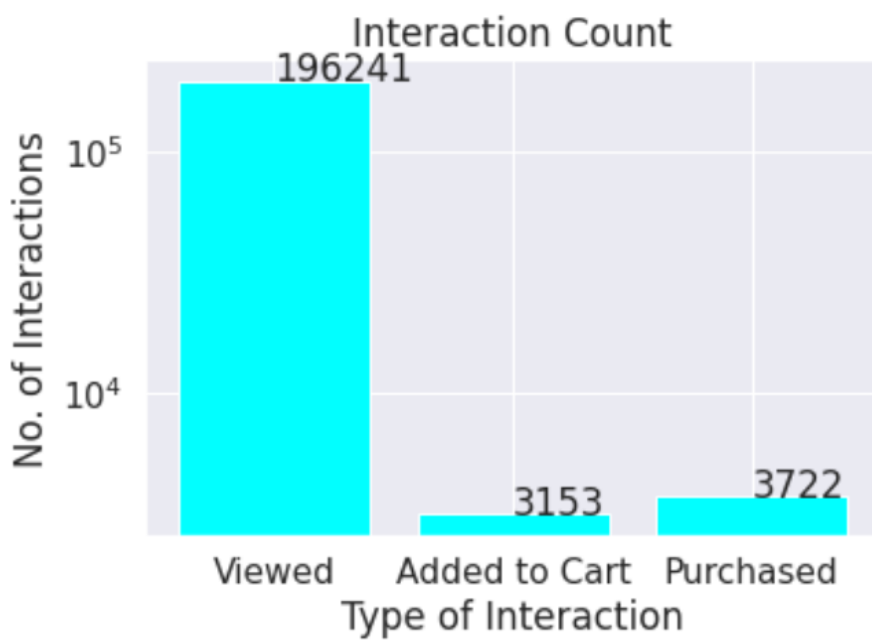
An intriguing trend emerged during the evaluation of Precision and Recall based on the number of unique purchases made.



The Precision of the model is significantly higher for customers who have previously made at least one purchase.



This indicates that when the model suggests a product to a buying customer, there is a greater likelihood that the customer will make the purchase.



However, false positives of predicted purchases of customers who did not actually buy a product bring down the overall precision of the model.

Limitations

I. Due to a sparsity of quality of data on micro behaviors, the accuracy of the content-based filtering is Inadequate as it is difficult to develop an accurate and effective recommendation.

II. The precision of the model decreases when it incorrectly predicts customer purchases who did not actually make a purchase, resulting in false positives.

III. The three micro-behaviors by which we calculate the interaction score do not fully capture the complexity and richness of the user preferences and may lead to oversimplified recommendations.

Conclusion

I. A content-based recommender has been created using multiple categories. To improve the accuracy of product recommendations on an e-commerce website, a content-based recommender was developed based on multiple categories such as price, brand, category, and product features. This approach involves analyzing the characteristics of the items that users have shown interest in and then recommending similar products to them.

II. The precision of the model is higher for eventual buyers. The precision of the model refers to the accuracy of the recommendations made by the system. In this case, the content-based recommender has a higher precision for users who are eventually buyers. This means that the system is better at suggesting products that are likely to be purchased by users who have a history of buying items from the website.

III. The recall of the model increases as the customer makes more purchases. A recall is a measure of the effectiveness of the recommender system to identify all relevant products for a particular user. As a customer purchases more items, the system can use this information to improve the recall of the model.

Future Works

I. More micro behaviors can be included to create a more accurate prediction.
II. Future work could involve using a data set covering a longer period of time.
III. A hybrid recommender could be developed using neural networks.

References

- *eCommerce behavior data from the multi-category store*. (2019, December 9). Kaggle. https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store
- Meng, W., Yang, D., & Xiao, Y. (2020). Incorporating User Micro-behaviors and Item Knowledge into Multi-task Learning for Session-based Recommendation. *ArXiv (Cornell University)*. https://doi.org/10.1145/3397271.3401098
- *Post-Click Behaviors Enhanced Recommendation System*. (2020, August 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9191633