Illinois Institute of Technology

Department of Computer Science

CS584 - Section: 02

Machine Learning

Product Recommendation Model

Authors:	*Student Number:*
Xiaoting ZHOU	A20503290
Nevil Jack DENIS	A20474215
Hongbo WANG	A20524770

February 28, 2023

1 Problem Description

According to a report by Marist Poll - Digital Economy Poll [1], approximately 76% of adults in the United States have shopped online, a trend that has been growing steadily over the years. Despite the benefits of e-commerce, however, statistics from Baymard Institute [2] show that the average cart abandonment rate is a staggering 69.99%. One reason for this is that customers can become distracted by the sheer amount of options available to them, causing them to lose sight of their original intention. As a result, they may add unnecessary items to their cart, only to abandon them before finalizing their purchase and checking out.

The product recommendation model is a method to solve the problem of user cart abandonment. The e-commerce giants have unique models to help them sell their products, like Amazon, eBay, and Walmart. However, these models cannot fit well in small e-commerce platforms. Large companies (Giants) utilize vast amounts of data to train their models, with an abundance of features.

However, small e-commerce companies don't typically have access to data sets in the tens of billions necessary for model training. As a result, if they blindly apply a pre-existing model, over-fitting will show.

2 Work Different

Our model, Product Recommendation Model (PRM), is a lightweight product recommendation model that is suitable for small and medium-sized e-commerce platforms.

By prioritizing the individual customer's shopping journey over individual products, PRM is able to piece together a detailed history of each customer's preferences and provide them with personalized product recommendations. Additionally, PRM can help increase seller turnover by predicting best-selling items and providing personalized services to each customer. This mutually beneficial solution can be implemented through a simplified machine-learning model.

The goal of the PRM was to provide a mutually beneficial solution through the use of lightweight models. This solution leverages simplified machine learning models to increase sales and improve the overall shopping experience for customers.

3 Finished Work

Dataset Selecting (Kagge): The dataset is a relational set of files describing customers' orders over time. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, we provide between 4 and 100 of their orders, with the sequence of products purchased in each order. We also provide the week and hour of the day the order was placed and a relative measure of time between orders.

4 Preliminary Plan

4.1 Project Analyzing

Product Recommendation Model (PRM) is a process of identifying associations among entities and objects that frequently appear together, such as the collection of items in a shopper's Basket. When used appropriately, PRM can be an effective tool for Businesses/Companies in understanding consumer behavior better and influencing it. This paper will enable Instacart to enhance the user experience by suggesting the next likely product to purchase to the customer during the order process.

4.2 Data Preprocessing

The data sets were provided by Instacart Technology Company and were taken from Kaggle to perform the analysis. The data sets provided by Instacart had complete information on over 3 million grocery orders from more than 200,000 Instacart users. Both product data and customer data from Instacart includes 50,000 unique products, week and the time of purchase, different product aisle, and departments.

4.3 Model Selecting

We will use the Apriori algorithm for mining association rules and make a comparison with Frequent Pattern Growth Algorithm. XGBoost is the classifier that we have selected to implement.

4.4 Model Training and Testing

- Training: Enter the training set and hyper-parameters into the model to enable the model's adjustment and training process.
- Testing: Enter the testing set into the trained model, and get the predicted value for each customer.

4.5 Model Evaluation

The purpose of Model Evaluation is to gauge the accuracy of the model's performance. Once the testing data is fed into the model, the function assesses the discrepancies between the predicted and actual values to reflect the model's accuracy. This process enables the Model Evaluation to serve its intended function.

4.6 Report Writing

Using Latex and CVPR (or ACM) format to finish the final report.

References

[1] *NPR/Marist Poll of 1057 National Adults*. (June, 2020). URL: https://maristpoll.marist.edu/wp-content/misc/usapolls/us180423_NPR/NPR_Marist%20Poll_Tables%20of%20Questions_May%202018.pdf.

[2] *48 Cart Abandonment Rate Statistics 2023*. URL: https://baymard.com/lists/cart-abandonment-rate.