Customer Product Recommendations Using Association Rule Mining

Sandeep Batta
sabatta@iu.edu
Indiana University

Kamalesh Kumar MG
kmandak@iu.edu
Indiana University

project-sabatta-kmandak

Abstract

With the dynamic retail environment, where personalized product recommendations are essential for retailers aiming to increase their revenue, our project aims to leverage association rule mining, specifically the Apriori algorithm, to extract patterns from a retail dataset obtained from Kaggle. This algorithm enables us to uncover hidden patterns within a retail dataset sourced from Kaggle, which encompasses information about customers, products, orders, aisles, and departments. The primary objective is to derive associations among products purchased by customers and generate recommendations based on these associations. These insights will enhance the shopping experience for customers by suggesting items they are more likely to be interested in and boost profits for retailers. Throughout the project, we have explored various recommendation models, including association rule mining, item-based collaborative filtering, and order-based collaborative filtering, each offering unique insights and capabilities. The experimentation with different models allows us to provide a comprehensive and effective solution that aligns with the diverse needs of the retail environment.

Keywords: Data Mining, Association Rule Mining, Recommendation System, Apriori Algorithm, Machine Learning, Exploratory Data Analysis, Collaborative Filtering

1. Introduction

In today's highly competitive and data-driven retail landscape, understanding and predicting customer behavior is crucial for businesses seeking to enhance customer satisfaction and drive revenue. Customer product recommendations play a pivotal role in achieving these objectives. Recommender systems leverage the power of data mining and machine learning techniques to provide personalized product suggestions to customers, thereby increasing engagement and sales. The literature in this field has witnessed significant advancements and diversification. Notably, association rule mining techniques, as demonstrated in the paper [1] have become instrumental in extracting valuable insights from ever-expanding datasets. The authors emphasize the importance of precision and accuracy in mining association rules, addressing the complexities and execution times associated with traditional data mining techniques. Furthermore, the role of Association

Rule Mining (ARM) in recommendation systems is not limited to this paper alone. Another notable study [2], explores the intricacies of using ARM to enhance recommendation systems. This paper emphasizes how ARM is becoming increasingly crucial in dealing with the challenges of modern recommendation systems. It highlights how researchers are actively seeking innovative solutions to better align with the evolving preferences of consumers.

2. Methods

2.1 Dataset

The kaggle dataset we considered for our project revolves around a relational set of files capturing the dynamics of customer's orders. The dataset comprises anonymized information from over 3 million grocery orders by more than 200,000 users. Each user is represented by 4 to 100 orders. This comprehensive retail dataset is instrumental for analyzing customer behavior and understanding product relationships.

File Descriptions:
- **aisles.csv**: Maps aisle IDs to aisle names, offering insights into the types of products available in each aisle.
- **departments.csv**: Associates department IDs with department names, categorizing products based on their nature.
- **order_products_train.csv, order_products_prior.csv**: Specify products in each order.
- **products.csv**: Links product IDs to names, along with aisle and department IDs.

The dataset lends itself to the application of algorithms like Apriori for association rule mining. Analyzing customer purchase history can generate valuable recommendations based on observed patterns and associations between products.

2.2 Data Preparation and Exploration

This pivotal stage involves acquiring, cleaning, and structuring datasets to lay the groundwork for meaningful insights. In this section, we detail our approach to data preparation, starting with the loading of essential datasets and moving through the initial exploration and preprocessing steps. Subsequently, we embark on a comprehensive data exploration journey, uncovering patterns and nuances within the Market Basket dataset.

2.2.1 Data Preparation

- **Data Loading**:
  In this initial step, we utilize the Pandas library to load essential datasets for our analysis. The primary datasets include:
  - \`order_products__train.csv\`
  - \`order_products__prior.csv\`

- `products.csv`

- **Initial Data Exploration and Preprocessing**:
  Following data loading, we engage in initial exploratory steps to understand key elements related to order and product information. The focus is on streamlining the data by selecting relevant columns and merging datasets. This preprocessing phase sets the foundation for a cohesive and refined dataset.
- **Handling Missing or Inconsistent Data**:
  As part of data preparation, it is crucial to address any missing or inconsistent values. We thoroughly examine the dataset for such issues and implement strategies to handle them, ensuring the integrity and reliability of the data for subsequent analysis.

2.2.2 Data Exploration

- **Products in Transactions**:
  We delve into the dataset to explore the count of transactions for each product. This analysis enables us to identify popular products based on transaction frequency, providing valuable insights into product distribution and customer preferences.
- **Merging Orders and Products**
  Our dataset is enriched by merging information from the orders and products datasets. This step involves appending product names to the order dataset, enhancing the interpretability of the data and enabling a more comprehensive understanding of customer purchasing patterns.
- **Aggregating Orders**
  To facilitate further analysis, we aggregate products within each order. This results in a refined dataset where each order is represented by a consolidated list of products. This aggregated format simplifies the data structure for subsequent exploration and analysis.
- **Exploring Transactions**
  We explore transactions by dissecting the aggregated product names. This step provides insights into the variety of unique items present in our dataset, contributing to a deeper understanding of product diversity and customer behavior.

2.3 Association Rule Mining using Apriori Algorithm

Our project utilizes the Apriori algorithm for association rule mining on the dataset. We calculate support and confidence, generating frequent itemsets and association rules. By setting minimum thresholds, we extract significant associations, revealing valuable insights for optimizing recommendations and improving the shopping experience. The resulting data frame captures detailed information about antecedents, consequents, and support values, offering a robust foundation for data-driven decision-making.

2.4 Recommendation Generation

Recommendation generation is a crucial aspect of recommendation systems, which are designed to assist users in discovering items or content that align with their preferences, interests, or needs. The goal of recommendation generation is to provide personalized and relevant suggestions, enhancing user experience and engagement. We have employed three distinct recommendation models to enhance the recommendation process. The first model utilizes association rules, extracting patterns from transaction data and considering subsets of associations to recommend related products. The second model, item-based collaborative filtering, leverages user-item interactions to suggest items based on the preferences of users with similar tastes. The third model, order-based collaborative filtering, takes into account the sequence of user actions or purchases to generate recommendations. This multifaceted approach aims to cater to different aspects of user preferences and behavior

2.4.1 Association Rule-based Recommendation:

For association rule-based recommendation, I employed the Apriori algorithm to mine frequent itemsets and generate association rules from the transaction data. The algorithm identifies items that often appear together in transactions, allowing the system to suggest related products based on these associations. Moreover, I considered all possible subsets of association rules to comprehensively capture nuanced patterns in customer purchasing behavior.

2.4.2 Item-based Collaborative Filtering:

To implement item-based collaborative filtering, I calculated the similarity between products using a matrix of user-item interactions. Specifically, I utilized the cosine similarity measure to identify items with comparable user engagement patterns. By examining the preferences of similar users and recommending items based on those preferences, this model delivers personalized product suggestions.

2.4.3 Order-based Collaborative Filtering:

For order-based collaborative filtering, I analyzed the sequential patterns in user actions. This involved understanding the order in which customers interacted with products and deriving patterns from these sequences. Leveraging the temporal aspects of user behavior, the system generates recommendations by considering the historical order patterns of users, providing personalized suggestions aligned with their purchasing sequences.

3. Results

To gauge the influence of different support and confidence thresholds on association rule mining, we conducted a sensitivity analysis. The objective was to understand how adjusting these thresholds influences the number of frequent itemsets and association rules generated by the Apriori algorithm. The table below summarizes our findings for various threshold combinations:

Support Threshold	Confidence Threshold	Number of Association Rules
0.005	0.005	174
0.0075	0.005	58
0.005	0.0075	174

3.1 Association Rule-based Recommendation:

Product List: ['Super Spinach! Baby Spinach', 'Sweet Baby Kale', 'Banana']
Recommendations: ['Baby Bok Choy']

3.2 Item-based Collaborative Filtering:

Product List: ['Super Spinach! Baby Spinach', 'Sweet Baby Kale', 'Banana']
Recommendations: ['Honeycrisp Apple', 'Baby Bok Choy', 'Ready Wipe Flushable Wipes', 'Strawberries', 'Pumpkin & Spinach Stage 2 Baby Food']

3.3 Order-based Collaborative Filtering:

Items in Order ID: ['2% Reduced Fat Milk', 'American Slices Cheese', 'Apricot Preserves', 'Artichokes', 'Bag of Organic Bananas', 'Bag of Organic Lemons', "Biscuits Orange Pim's", 'Boneless Skinless Chicken Breast Fillets', 'Chocolate', 'Clementines', 'Dairy Milk Fruit & Nut Chocolate Bar', 'Everyday Facial Tissues', 'French Lavender Hand Wash', 'Fresh Fruit Salad', 'Just Crisp', 'Macaroni And Cheese', 'Matzos', 'Meyer Lemon', 'Mini & Mobile', 'Mini Original Babybel Cheese', 'Natural Artesian Water', 'One Ply Choose A Size Big Roll Paper Towel Rolls', 'Organic Hass Avocado', 'Organic Raspberries', 'Original Black Box Tablewater Cracker', 'Parmesan', 'Sensitive Toilet Paper', 'Spaghetti Pasta', 'Tea', 'Thin', 'Wafer']

Recommendations based on Order: ['Simply Clean Fragrance Free Wipes', 'No Sugar Added Muesli Cereal', 'Light & Fit Greek Crunch Coconut Chocolate Bliss', 'Activia Mixed Berry/Black Cherry Lowfat Yogurt']

4. Discussion

The results obtained from each recommendation model reveal distinct patterns in suggesting additional products based on different aspects of customer behavior. The association rule-based recommendation suggests 'Baby Bok Choy' in association with 'Super Spinach! Baby Spinach', 'Sweet Baby Kale', and 'Banana.' In contrast, item-based collaborative filtering recommends a diverse set of products, including 'Honeycrisp Apple' and 'Pumpkin & Spinach Stage 2 Baby

Food.' Order-based collaborative filtering, focusing on sequential patterns, suggests items such as 'Simply Clean Fragrance Free Wipes' and 'Activia Mixed Berry/Black Cherry Lowfat Yogurt' based on specific order sequences.

Selecting the most suitable recommendation model depends on the specific requirements and objectives of the recommendation system. The association rule-based recommendation model is advantageous for its simplicity and interpretability, making it effective for scenarios where items are frequently purchased together. Item-based collaborative filtering excels in providing personalized suggestions based on user-item interactions, making it suitable for situations where user preferences are a key factor. On the other hand, order-based collaborative filtering considers sequential patterns and temporal dependencies, making it appropriate for scenarios where the order of item selection matters. A combination approach, leveraging different models for various aspects of recommendation, may offer a balanced solution.

5. Author Contribution

In terms of author contributions, the project benefited significantly from a collaborative effort.

Sandeep Batta involved in implementing association rule mining. handled data loading, preprocessing, and integration, focusing on key datasets like "order_products__train.csv" and "order_products__prior.csv." The Apriori algorithm was efficiently implemented for generating frequent itemsets and association rules. Also explored unique items, calculated support and confidence values, and assessed the impact of different thresholds on rule generation. The resulting association rules were stored for further analysis, contributing to the project's objective of generating insightful recommendations based on customer behavior.

Kamalesh Kumar MG took charge of implementing three distinct recommendation models: association rule-based recommendation, item-based collaborative filtering, and order-based collaborative filtering.

References

- [1] An improved association rule mining algorithm for large data by Zhao et al. https://www.degruyter.com/document/doi/10.1515/jisys-2020-0121/html
- [2] Applicability of Association Rule Mining in Recommendation System for Big Data Analysis by Vishakha et al. https://ieeexplore.ieee.org/abstract/document/9823695
- https://en.wikipedia.org/wiki/Apriori_algorithm
- https://medium.com/mlearning-ai/recommendation-systems-arl-association-rule-learning-bed1a07b5d9a
- Processing distributed transactions in a predefined order - https://dl.acm.org/doi/10.1145/3427796.3427819
- Shukla, A. (2021). https://towardsdatascience.com/comprehensive-guide-on-item-based-recommendation-systems-d67e40e2b75d