

# 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.

## Keywords

Data Mining, Association Rule Mining, Recommendation System, Apriori Algorithm, Machine Learning, Exploratory Data Analysis

## 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.

## Methods

- **Dataset:** The Kaggle dataset contains retail information like aisles, departments, customer orders, and product details. It includes records of orders, products purchased, and order specifics like order ID, user ID, and order set classification. This data facilitates the analysis of customer behavior and product relationships, enabling the application of algorithms like Apriori for association rule mining to generate recommendations based on customer purchase history.
- **Data Preparation and Exploration:** Exploratory Data Analysis (EDA) to understand the dataset, distribution of products, orders, users, etc and Data preprocessing for Merging relevant data, handling missing values, encoding categorical variables, etc.
- **Association Rule Mining using Apriori Algorithm:** Apply the Apriori algorithm to discover frequent itemsets and association rules based on the products purchased by customers. Set thresholds for min support and min confidence to extract significant rules. Generate association rules indicating which products are frequently purchased together.
- **Recommendation Generation:** Utilize the association rules to suggest product recommendations based on the discovered patterns. Develop a recommendation system that provides suggestions for additional products based on a customer's current selection or purchase history. Implement rules to suggest related products that customers might be interested in buying together.
- **Evaluation:** Evaluate the generated association rules and recommendations by varying the min support and min confidence values.

## 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://en.wikipedia.org/wiki/Apriori_algorithm)
- <https://cloud.google.com/blog/topics/developers-practitioners/bigquery-explained-working-joins-nested-repeated-data>
- <https://cloud.google.com/bigquery/docs/reference/libraries>
- <https://medium.com/mlearning-ai/recommendation-systems-arl-association-rule-learning-bed1a07b5d9a>