

Proposal Report: Customer Purchase Behavior Analysis using Association Rule Mining

Title: Customer Purchase Behavior Analysis using Association Rule Mining

Abstract

This project aims to analyze customer transaction data from a retail store using association rule mining to uncover hidden patterns and associations between purchased items. By leveraging the Apriori algorithm, this project seeks to identify frequent itemsets and generate strong association rules, providing valuable insights into customer purchase behavior. These insights can be used to optimize inventory management and plan targeted marketing campaigns, ultimately enhancing customer satisfaction and business performance.

Introduction

Retail businesses are constantly seeking ways to understand and predict customer purchase behavior to improve inventory management and develop targeted marketing strategies. Traditional methods often fall short in revealing complex patterns and associations hidden within large transaction datasets. This project addresses this problem by applying association rule mining to analyze customer transaction data, uncovering meaningful relationships between items that customers frequently purchase together. The findings will help in making informed decisions that align with customer preferences and behaviors.

Related Work

Association rule mining is a well-established data mining technique widely used in market basket analysis. Agrawal et al. (1993) introduced the Apriori algorithm, which efficiently identifies frequent itemsets and generates association rules. Subsequent research has refined and optimized these techniques for various applications, including retail, healthcare, and e-commerce. Studies have shown that understanding customer purchase patterns can significantly enhance inventory management, reduce stockouts, and improve customer targeting through personalized marketing campaigns (Agrawal et al., 1993; Han et al., 2004). This project builds on these foundations by applying association rule mining to a new dataset and exploring its practical applications in a specific retail context.

Proposed Work

Hypotheses/Questions:

- What are the frequent itemsets in the transaction data?
- What strong association rules can be derived from these frequent itemsets?
- How can these association rules be utilized to improve inventory management and marketing strategies?

Methodology:

Data Collection:

- Transaction data from a retail store, including transaction IDs and items purchased.

Data Preprocessing:

- Convert transaction data into a binary matrix format where each row represents a transaction, and each column represents an item.

Frequent Itemset Mining:

- Apply the Apriori algorithm to identify frequent itemsets meeting a minimum support threshold.

Association Rule Mining:

- Generate association rules from the frequent itemsets.
- Evaluate rules using support, confidence, and lift metrics to identify significant associations.

Tools and Technologies:

- Programming Languages: Python or R
- Libraries: pandas, mlxtend (Python) or arules (R)
- Data Visualization: Matplotlib, Seaborn (Python) or ggplot2 (R)

Evaluation Plan

Evaluation Metrics:

- **Support:** The proportion of transactions containing the itemset.
- **Confidence:** The likelihood that item Y is purchased when item X is purchased.
- **Lift:** The increase in the probability of purchasing item Y when item X is purchased.

Success Criteria:

- Discovery of high-support and high-confidence association rules.
- Insights that lead to measurable improvements in inventory management and targeted marketing effectiveness.
- Positive feedback from retail stakeholders on the applicability of the findings.

Timeline

- **Week 1-2:** Data collection and preprocessing.
- **Week 3-4:** Implementation of the Apriori algorithm for frequent itemset mining.
- **Week 5-6:** Generation and evaluation of association rules.
- **Week 7-8:** Analysis of results and preparation of the final report.

References

- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (pp. 207-216). ACM.
- Han, J., Pei, J., & Kamber, M. (2004). *Data mining: Concepts and techniques*. Elsevier.
- Lecture notes on Frequent Itemsets Mining and Association Rule Mining.