International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Market Basket Analysis Using Association Rule Mining for Enhanced E-Commerce Recommendation

Varun D, Kiran Kumar N.R, Pravin Kumar R, Raghul S, Mrs. C. Chinnima, AP/CSE

Department of Computer Science and Engineering, Kingston Engineering College, Vellore
Guide, Kingston Engineering College, Vellore Email: chinni16.engineering@kingston.ac.in,
Email id: [varunvicky2021@gmail.com, nrukmangadakiran@gmail.com, usraghul824@gmail.com, pravinpeabash@gmail.com]

ABSTRACT

Market Basket Analysis (MBA), also known as affinity analysis, is a vital data mining technique in e-commerce that uncovers purchasing patterns by identifying products frequently bought together. This project implements MBA using the Apriori algorithm and integrates it into an e-commerce platform to enhance the recommendation system. Through this implementation, the system suggests relevant product combinations, improving user experience and increasing average transaction value. Developed with Python, Django, and Streamlit, the system demonstrates how data-driven insights can inform product placements, optimize inventory management, and personalize marketing strategies.

KEYWORDS: Market Basket Analysis, Apriori Algorithm, Association Rule Mining, Recommendation System, E-commerce, Python, Django

1. INTRODUCTION

E-commerce platforms aim to provide personalized user experiences that drive sales and increase customer satisfaction. A powerful method to achieve this is by analyzing customer purchasing behavior using Market Basket Analysis (MBA). MBA reveals relationships between items by examining historical transaction data and identifying product combinations that frequently appear together. The insights gained can power features such as “Frequently Bought Together” and “Customers Also Bought,” improving product visibility and streamlining the shopping process. Through this project, we implement a full-stack recommendation system using MBA and the Apriori algorithm to extract valuable patterns, visualize them interactively, and apply them in real-time to enhance the customer journey.

2. RELATED WORK

Over the years, several algorithms and frameworks have been explored to mine association rules effectively. The classical Apriori algorithm is well-suited for discovering frequent itemsets in structured transaction data. FP-Growth has also emerged as an efficient alternative, especially for large datasets, due to its compressed representation of itemsets. Genetic algorithms and soft computing techniques have been proposed to improve the accuracy of discovered rules. Several studies show that combining these algorithms with optimization techniques enhances performance in terms of rule relevance and computational cost. While most prior work focused on backend efficiency, this project also addresses frontend usability and real-world integration through Django and Streamlit.

3. METHODOLOGY

Our methodology begins with the collection of transactional data containing details like customer ID, product names, and timestamps. This raw data undergoes preprocessing to remove null values, duplicates, and inconsistencies. We then group transactions into item baskets. The Apriori algorithm is applied to these baskets to identify frequent itemsets based on a predefined support threshold. From these frequent itemsets, we derive association rules, filtered further using confidence and lift values to ensure significance. Finally, these rules are integrated into a recommendation engine deployed through Django (backend) and Streamlit (frontend). The system displays intelligent product suggestions based on customer cart contents or previous purchase history.

4. SYSTEM ARCHITECTURE AND IMPLEMENTATION

The architecture comprises multiple components:

- **Data Preprocessing:** Cleans raw transaction logs.
- **Analysis Engine:** Applies Apriori to extract frequent itemsets and generate rules.
- **Rule Filtering Module:** Retains only rules with high business impact.
- **Recommendation Engine:** Matches real-time cart data with association rules.
- **Frontend (Streamlit):** Displays "Frequently Bought Together" and "Customers Also Bought" suggestions.

5. RESULTS AND DISCUSSION

The system was tested using sample transaction data that mimics real-world e-commerce purchases. For example, transactions involving smartphones often included screen protectors and cases, a pattern successfully identified by the model. Generated rules such as {Smartphone} → {Charger, Case} with a confidence of 78% and lift of 1.35 proved actionable for upselling. These recommendations, when shown through the user interface, led to better user engagement and increased basket size. Limitations observed include challenges with sparse data and cold-start products that had not been previously purchased. These can be addressed in the future by integrating collaborative filtering or content-based models.

6. CONCLUSION AND FUTURE WORK

Market Basket Analysis, through association rule mining, offers a powerful approach to transforming raw purchase data into actionable business intelligence. Our implementation demonstrates the value of Apriori-based recommendations in improving product discovery and increasing user satisfaction. This paper not only proves the technical feasibility of such systems but also underlines their practical impact when integrated into full-stack platforms like Django and Streamlit. For future work, we propose incorporating machine learning models that consider user preferences, temporal trends, and external variables like seasonality. We also aim to deploy the system in a production environment and monitor key performance indicators such as conversion rate and user retention.

7. EVALUATION METRICS

The system's effectiveness was measured using standard association rule mining metrics:

- **Support:** Percentage of transactions containing the itemset.
- **Confidence:** Likelihood that the consequent is purchased when the antecedent is.
- **Lift:** Indicates whether the rule is statistically significant or a result of random chance. We also observed user behavior metrics like average cart size and click-through rates to evaluate the practical impact of recommendations.

8. ETHICAL CONSIDERATIONS

All customer data used in the study was anonymized to ensure privacy. The system is designed in compliance with data protection regulations such as GDPR. Recommendations are generated based solely on purchase behavior and not on personal identifiers. Future enhancements will include explainability features to help users understand why certain products are being suggested.

9. APPENDIX

The appendix includes sample transaction data used for testing the association rule mining algorithm, along with code excerpts that illustrate the implementation of the Apriori algorithm in Python. Screenshots of the deployed recommendation system are also provided, showcasing real-time product suggestions generated based on the derived association rules.

Key configurations such as the minimum support and confidence thresholds used for rule generation are documented to facilitate reproducibility. The structure of the Django-based backend and Streamlit frontend has been outlined to help developers understand the full-stack integration approach. Additionally, a brief overview of the libraries and development tools employed in the project—such as Pandas, NumPy, mlxtend, and WAMP Server—is included to support future enhancements or adaptations of the system.

10. CASE STUDY

To validate our system, we simulated a series of customer sessions using mock transactional data. When a user added "Laptop" to the cart, the system recommended accessories like "Mouse," "Laptop Bag," and "External Storage," based on prior shopping behavior patterns. These recommendations aligned well with known industry trends and showcased the practical effectiveness of the model. Businesses implementing this system can expect increased user engagement, reduced bounce rates, and higher conversion ratios.

11. COMPARATIVE ANALYSIS

We compared Apriori against FP-Growth and basic collaborative filtering techniques. Apriori, while slower on large datasets, offered interpretable rules ideal for business stakeholders. FP-Growth provided faster processing but required more memory. Collaborative filtering worked better in scenarios with rich user profiles but struggled with sparse datasets. Thus, Apriori remains a suitable choice for association mining in small-to-medium datasets where rule transparency is essential.

12. LIMITATIONS AND RECOMMENDATIONS

Although our system performs well in extracting meaningful associations, it is limited by the static nature of MBA. The cold-start problem for new users and products, lack of time-awareness, and inability to capture causality are notable limitations. To overcome these, hybrid models that incorporate machine learning or contextual metadata can be explored. Real-time data pipelines and A/B testing frameworks can also be added for continuous improvement.