Grocery Market Transactions Analysis using Association Rules and Apriori Algorithm

[1] Riad Solh Anouar El Mahraoui Amal [2]

***Abstract*— understanding customer purchasing behavior is crucial for developing effective marketing strategies and improving overall sales performance. One powerful technique to uncover hidden patterns in transactional data is association rule mining, which is particularly useful in the context of grocery stores and e-commerce platforms. By identifying frequent product combinations, businesses can leverage these insights for cross-selling and personalized product recommendations [9].**

This project applies association rule mining and the Apriori algorithm to analyze grocery transactions and discover valuable insights about product relationships. The goal is to uncover frequent itemsets—combinations of items that are often bought together—and generate association rules that predict future purchases [1]. By exploring transaction data from a grocery store, a system is created that not only identifies common purchasing patterns but also supports decision-making for inventory management and marketing strategies [3].

Keywords: Association Rule Mining, Apriori Algorithm, Frequent Itemsets, Product Recommendations, Inventory Management, Marketing Strategies

I. INTRODUCTION

In recent years, the grocery retail industry has been experiencing an increasing reliance on data analytics to enhance customer satisfaction, boost sales, and optimize operations. The availability of large-scale transactional data provides an excellent opportunity for retailers to identify hidden patterns and gain actionable insights [2]. Among the various techniques used for analyzing transactional data, **Association Rule Mining** has emerged as one of the most effective approaches. This technique is based on finding associations between different items that are frequently purchased together, which helps businesses understand consumer behavior [9].

The **Apriori algorithm**, introduced by Agrawal et al. [1], is a widely used method for mining frequent itemsets and generating association rules. By identifying relationships between products, retailers can improve various aspects of their business, including cross-selling, targeted promotions, and inventory management [3]. Furthermore, data-driven approaches in retail are also being employed for building recommendation systems that predict the products that customers are most likely to purchase next based on historical purchase data [8].

In this project, **grocery transaction data** is analyzed using the Apriori algorithm to identify frequent itemsets and generate association rules that highlight relationships between products. These rules are particularly valuable for building recommendation systems, optimizing marketing campaigns, and improving inventory decisions [4]. The insights derived from these association rules can significantly enhance business strategies and provide a more personalized shopping experience for customers.

II. LITERATURE REVIEW

Several studies have explored the use of Association Rule Mining in analyzing transactional data to uncover meaningful patterns in customer purchasing behavior. Agrawal et al. [1] introduced the Apriori algorithm, which became a foundational technique for discovering frequent itemsets and association rules in large datasets. Subsequent research focused on optimizing Apriori's efficiency, leading to the development of FP-Growth [5], which reduces computational complexity by eliminating the need for candidate generation. Applications of these techniques in retail analytics have demonstrated their effectiveness in identifying product associations and enhancing recommendation systems [6]. Prior studies on market basket analysis have also explored generalizing association rules beyond simple co-occurrence patterns, allowing for a deeper understanding of customer behaviors [7]. Building upon these works, this study applies Association Rule Mining to the Groceries dataset, utilizing Apriori to extract valuable insights and develop a recommendation function tailored to real-world shopping patterns.

III. METHODOLOGY

A. Data Preprocessing

The dataset used for this study contains grocery transactions recorded over a certain period. The preprocessing steps included data cleaning, where missing values and duplicates were handled to ensure data consistency. Transaction identification was performed by assigning unique transaction IDs to group purchased items. Standardization techniques were applied to format product names uniformly and avoid duplicates caused by inconsistencies.

B. Exploratory Data Analysis (EDA)

To understand purchasing trends, various exploratory data analysis techniques were performed. General transaction statistics were computed to summarize the number of transactions and unique products. Monthly purchase trends were analyzed to identify fluctuations in consumer behavior over time. Frequent item analysis was conducted to determine the most commonly purchased products, and customer shopping behavior was examined to understand the distribution of

*This work was not supported by any organization

[1]A. Riad Solh is with Faculty of Sciences, Mohammed V University, 4 Ibn Batouta Avenue, P.O. Box 1014, Rabat, Morocco

[2]A. El Mahraoui is with Faculty of Sciences, Mohammed V University, 4 Ibn Batouta Avenue, P.O. Box 1014, Rabat, Morocco

items per transaction and identify top shoppers based on transaction frequency.

C. Apriori Algorithm & Association rules

The Apriori algorithm was applied to generate frequent itemsets and extract association rules. The process involved identifying itemsets that meet a predefined minimum support threshold, generating association rules that satisfy minimum confidence and lift criteria, and evaluating the rules using visualizations such as support-confidence graphs and lift heatmaps. Additionally, a recommendation function was developed to suggest products based on extracted association rules, allowing for personalized product recommendations based on shopping patterns.

IV. PROPOSED WORK

The proposed work focuses on analyzing customer purchasing patterns in the Groceries dataset using Association Rule Mining.

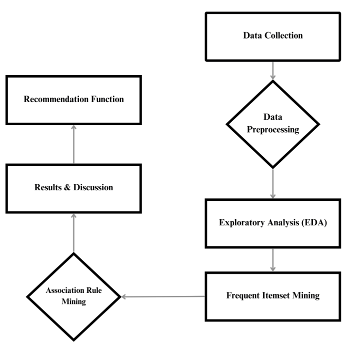

Fig. 1. System Design

The process begins with data collection and preprocessing to ensure a clean and structured dataset. Through Exploratory Data Analysis (EDA), key insights into purchasing behaviors are uncovered, highlighting trends in frequently bought items. The Apriori algorithm is then applied in the Frequent Itemset Mining step to identify strong relationships between products. The discovered associations are analyzed in the Results & Discussion section, providing meaningful interpretations of shopping habits. To enhance the practical value of these insights, a Recommendation Function is developed, allowing for personalized product suggestions.

V. RESULTS AND DISCUSSION

A. Frequent Itemsets

The support values indicate the proportion of transactions containing each itemset [I]. Among individual items, bottled beer (4.53%) and beef (3.39%) appear most frequently. Additionally, berries (2.17%) and beverages (1.66%) are commonly purchased items.

The high frequency of whole milk in multiple itemsets suggests it is a staple product, often bought alongside other grocery items. This finding indicates that promoting whole milk in conjunction with other frequently purchased goods could boost overall sales.

Support	Itemsets
0.0081	(baking powder)
0.0339	(beef)
0.0218	(berries)
0.0166	(beverages)
0.0453	(bottled beer)
⋮	⋮
0.0076	(root vegetables, whole milk)
0.0090	(sausage, whole milk)
0.0116	(whole milk, soda)
0.0082	(whole milk, tropical fruit)
0.0112	(yogurt, whole milk)

TABLE I

FREQUENT ITEMSETS AND THEIR SUPPORT VALUES

B. Frequent Multi-Itemsets

The most common multi-item purchases include (whole milk, other vegetables) with a support of 1.48%, followed by (rolls/buns, whole milk) with 1.40% as Table [II] highlights. These results indicate strong associations between dairy products, baked goods, and fresh produce.

A key insight from Table [II] is that whole milk is frequently purchased alongside multiple items, including other vegetables, soda, and yogurt.

This reinforces its role as a core grocery product that drives additional purchases. Additionally, the association between rolls/buns and whole milk suggests potential for targeted promotions on breakfast-related items.

Support	Multi-Itemsets
0.0106	(rolls/buns, other vegetables)
0.0097	(soda, other vegetables)
0.0148	(whole milk, other vegetables)
0.0081	(yogurt, other vegetables)
0.0081	(rolls/buns, soda)
0.0140	(rolls/buns, whole milk)
0.0078	(yogurt, rolls/buns)
0.0076	(root vegetables, whole milk)
0.0090	(sausage, whole milk)
0.0116	(whole milk, soda)
0.0082	(whole milk, tropical fruit)
0.0112	(yogurt, whole milk)

TABLE II

FREQUENT MULTI-ITEMSETS AND THEIR SUPPORT VALUES

C. Association Rules

Our analysis generated strong association rules based on confidence and lift values. The most significant rules include:

- (Whole Milk $\rightarrow$ Other Vegetables) with a confidence of 9% and a lift of 0.77.
- (Whole Milk $\rightarrow$ Rolls/Buns) with a confidence of 9% and a lift of 0.80.
- (Whole Milk $\rightarrow$ Yogurt) with a confidence of 7% and a lift of 0.82.

- (Whole Milk → Sausage) with a confidence of 6% and a lift of 0.94.

These findings confirm that customers who purchase whole milk are more likely to buy vegetables, baked goods, and yogurt. The rule with the highest lift (Whole Milk → Sausage) suggests a particularly strong relationship, meaning that sausage sales may benefit from promotional strategies that pair it with whole milk.

D. Itemset Network Analysis

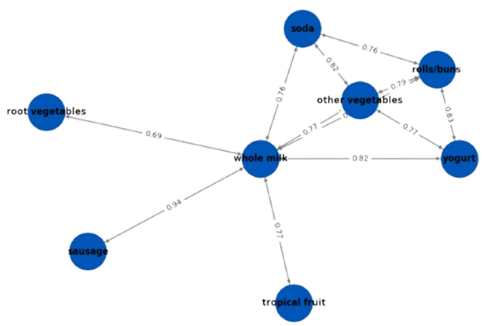

Fig. 2. Itemsets Network

To further understand the relationships between items, we constructed an itemset network visualization. The visualization revealed that whole milk acts as a central hub, frequently co-occurring with other staple products. The strongest association, Whole Milk → Sausage (lift: 0.94), suggests that placing these products near each other or offering bundle discounts could increase sales.

VI. CONCLUSIONS

Market Basket Analysis provides valuable insights into customer purchasing behavior. By identifying frequent itemsets (Table I) and strong multi-item associations (Table II), businesses can optimize store layouts, create targeted marketing campaigns, and increase overall sales. Our results highlight the central role of whole milk in driving multiple-item purchases, reinforcing its importance in promotional strategies.

APPENDIX

To maintain clarity and focus on the primary objectives, not all result images and code snippets are included in this report. However, the full set of results, including intermediate images, and the complete source code are available in the corresponding GitHub repository

https://github.com/aelmah/Data-Mining

REFERENCES

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2):207–216, 1993.

[2] Ravi S. Bose. Big data and analytics in the retail industry. *Retail Management Review*, 29(3):21–34, 2016.

[3] Shuo Chen, Weiwei Liu, and Ruijun Li. Mining association rules for retail analytics: A data-driven approach. *International Journal of Retail and Distribution Management*, 45(8):909–925, 2017.

[4] Sandeep Ghoshal and Renu Agarwal. Improving retail business performance with data mining. *Journal of Retail Analytics*, 9(4):251–262, 2015.

[5] Jian Pei Jiawei Han and Yiwen Yin. Mining frequent patterns without candidate generation. *ACM SIGMOD Record*, 29(2):1–12, 2000.

[6] Michael Steinbach Pang-Ning Tan and Vipin Kumar. Introduction to data mining. *Addison-Wesley*, 2006.

[7] Jeffrey D. Ullman Sergey Brin, Rajeev Motwani and Shalom Tsur. Beyond market baskets: Generalizing association rules to correlations. *ACM SIGMOD Record*, 26(2):265–276, 1997.

[8] Sandeep Shah and Umesh Kumar. Mining frequent itemsets for product recommendation in e-commerce websites. *Journal of Computer Science*, 3(6):383–391, 2007.

[9] Jai Srivastava, Rajeev Rastogi, and Kyuseok Shim. Mining association rules with item constraints. *ACM SIGMOD Record*, 29(2):60–72, 2000.