



Analysis of Changes in Customers' Market Basket Across Different Branches of a Chain Store Using Association Rules Technique and Its Impact on Product Placement: A Case Study of a Chain Store in Various Areas of Tehran

Raheleh Ansari Ashlaghi, Parastoo Mohammadi*, Bakhtiar Ostadi

Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran

Email: r.ansariashlaghi@modares.ac.ir, *p.mohammadi@modares.ac.ir, bostadi@modares.ac.ir

How to cite this paper: Ashlaghi, R.A., Mohammadi, P. and Ostadi, B. (2024) Analysis of Changes in Customers' Market Basket Across Different Branches of a Chain Store Using Association Rules Technique and Its Impact on Product Placement: A Case Study of a Chain Store in Various Areas of Tehran. *Open Access Library Journal*, 11: e12151.

<https://doi.org/10.4236/oalib.1112151>

Received: August 22, 2024

Accepted: September 22, 2024

Published: September 25, 2024

Copyright © 2024 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Is the shopping basket composition the same across different branches of a multi-branch store? Is the association between products the same in all chain store branches? In this paper, we studied the market baskets of customers across three branches of a chain store in Tehran using association rule techniques. The product layout in all branches of this chain store is uniform and does not utilize any specific technique. A review of the literature and researchers' hypotheses reveals that various factors can influence customers' shopping baskets, including geographical location, season, etc. The main focus of the researchers was on a specific store; therefore, this paper aims to examine the impact of store location on product associations. This study used association rule techniques to examine the relationship between FMCG products, the RFM model for customer clustering, and the FP-Growth algorithm. The examination of customer market baskets in these branches shows that the difference in branches in various regions of Tehran leads to changes in the relationship between purchased products, so multi-branch stores should avoid using a uniform product layout in all branches.

Subject Areas

Business Analysis, Business Management, Marketing

Keywords

Data Mining, Market Basket Analysis, Association Rules, FP-Growth Algorithm,

Tehran

1. Introduction

To date, researchers have provided many definitions for market basket analysis. Market basket analysis is a practical method for finding customer purchase patterns by examining common events from the transaction databases of stores [1]. It provides insights into customer consumption patterns and industry trends. Information about customer buying habits can help the seller choose the products or services they want. The main goal of market basket analysis is to improve the situation [2]. Market basket analysis empowers marketing and sales organizations to make better-informed decisions about where and how to deploy their efforts and resources. Millions of transactions occur in retail businesses, and the need to analyze them for higher profits has driven the adoption of market basket analysis [3]. In addition, using the MBA method can make shopping easier for visitors because products usually purchased together will be placed near each other [4]. By analyzing the market basket, recurring patterns can be found for offering related products together, thus increasing sales. Also, related products are placed so that customers can logically find items they might buy together, increasing customer satisfaction [5]. Moreover, studying consumer behavior is a complex and challenging task that requires a deep understanding of the factors influencing customer decision-making. Customers are influenced by various factors, including cultural, social, economic, and personal factors, making their behavior often difficult to predict. Analyzing and understanding consumer behavior allows retailers to stay ahead of competitors, respond better to market changes, and remain competitive in the ever-changing retail sector [6]. According to research conducted by researchers, it has been shown that basket analysis depends on various factors, including the season [7], dates and occasions, geographical location [8], and type of goods [1]. According to the study by Chen *et al.*, in a multi-store environment, a product may sell out of one store due to geographical, political, or environmental issues.

A study conducted in 2022 by T Formánek, O Sokol indicates that the geographical location of a store is a crucial factor that affects the volume and structure of sales. Understanding the complexity of location effects on sales dynamics and using such information may be a key element to a company's success in a competitive market environment. Generally, the geographical location of a store can be characterized by geographical-spatial and socio-demographic features. Distinct geographical location factors can have different effects on the sales of various products [7]. Additionally, according to a study conducted in 2018 in urban and rural areas of Bandar Abbas, it was found that the factors influencing customer satisfaction and purchase in urban areas differ from those in rural areas [9]. In another study by researchers, a recommendation system based on a customer

market basket was implemented. According to the suggestions and hypotheses stated in the study, implementing the algorithm in different geographical locations may affect the algorithm results [2]. Therefore, this paper proposes a comprehensive framework based on machine learning for analyzing consumer behavior trends, gaining insights into the data, and enabling data-driven decision-making. To create the framework, we analyzed the transactions of this store's customers and identified and studied the demands of the market baskets of loyal customers. By providing insights into consumer behavior to retailers, it is expected that the proposed framework will draw attention to the difference in customer market baskets across their branches. According to this information, for the upcoming research, we know that Tehran has 22 districts, and the chain store under study has branches in all of them. In addition, we assume that different stores can have different product combinations over different periods. This means that each store can have its own product combination, and the product combination in a store can dynamically change over time. It has been proven that when an organization can understand consumers' buying habits, it becomes easier to improve their business performance indicators [6].

Our research aimed to address this point, and the results and recommendations will improve marketing performance indicators such as customer satisfaction and loyalty. In this study, the impact of different geographical areas is only examined regarding the difference in product associations across various branches. The research also attempts to focus on a similar category of goods and a similar brand in all branches. This research is centered around applying machine learning technologies in Tehran's retail sector. This paper is organized as follows: Section 2 reviews and defines the problem in Tehran chain stores. Sections 3, respectively, discuss the types of algorithms proposed for examining customer market baskets and the methods of clustering loyal customers. Section 4 presents some case study results and discusses the findings. Finally, Section 5 provides conclusions, and in Section 6, we will talk about directions for future work and research.

2. Problem Definition

As explained in the previous section, the current retail environment is highly challenging and competitive due to market diversity, price change pressures from discounts, increased price transparency, and competition among companies. Traditional approaches for strategic pricing differentiation and product-related promotions are now more practical in the retail industry. In this competitive nature, treating customers as the company's main asset increases the organization's value. The retail sector is a complex and ever-evolving market that heavily relies on customer behavior. Studying consumer behavior is a complex and challenging task that requires a deep understanding of the factors influencing customer decision-making. Customers are influenced by various factors, including cultural, social, economic, and personal factors, which often make their behavior difficult to predict. Many of these product pairs that consumers purchase together are generally

known. However, given the fact that a typical supermarket contains hundreds of items bought by thousands of customers purchasing numerous products, understanding the less obvious related product pairs becomes difficult. Definition: If in branch A of the store, item x is placed next to item y, then in branches B and C, item x is also placed next to item y. Definition: The decision to place item x next to item y has been made entirely traditionally and without using association rules. Based on definitions 1 and 2, it has been identified in one of the chain stores in Tehran that all products on the shelves and floors in the branches are arranged uniformly, and no association rules are used in any of these branches.

3. Methods

3.1. FP-Growth Algorithm

There are many algorithms related to Frequent Itemset Mining (FIM). The Apriori and FP-Growth algorithms are the most fundamental FIM algorithms. Researchers have examined the fundamental differences between these algorithms in **Table 1**. According to studies [10] [11], the FP-Growth algorithm is recognized as a popular mining algorithm. It scans the database only twice and efficiently discovers all standard frequent item sets, particularly compared to the Apriori algorithm. FP-Growth has three strengths.

Firstly, FP-Growth compresses the entire database into a relatively small data structure (FP-tree), which results in scanning the database only twice. Secondly, it creates a frequent pattern growth formula to avoid generating many candidate itemsets. Thirdly, it generates detailed layers of the tree to discover frequent itemsets and reduces computational complexity. Experimental results show that FP-Growth is faster than the Apriori algorithm and several other frequent item mining methods [10] [12]. As explained in this paper, we have used the FP-Growth algorithm to analyze the transactional data of the chain store under study.

Table 1. Comparison of two algorithms [13].

Fpgrowth	Apriori
Scans the database only once, making it fast	Scans the database multiple times, making it slow
Used when database data is large	Used when database data is small
Stores a set of conditional FP-trees for each item in memory	Stores a transformed version of the database in memory
Creates a conditional FP-tree for each item	Generates frequent patterns by creating item sets through pairings like one-item sets, two-item sets, and three-item sets.

3.2. Customer Clustering

As reviewed in [14], several methods for customer segmentation exist, but most are based on customers' behavioral, psychological, geographic, and demographic

information. However, customer behavioral information based on RFM analysis is emphasized because it uses a small set of features for segmentation. To cluster store customers, the RFM model is used. RFM stands for Recency, Frequency, and Monetary Value, and it is becoming a prevalent form of clustering in the retail industry. This is particularly due to its simplicity of implementation with minimal help from data scientists and its straightforward interpretation due to the visual nature of its results. The three main factors of RFM can be explained as follows: Recency (R) Represents the time interval between the date of the last purchase and the most recent date in the statistical period. The smaller the time interval, the higher the R-value. Frequency (F): Indicates the number of times a customer has made purchases during the statistical period. The higher the F value, the more loyal the customer is to the company. Monetary Value (M): Represents the total value of transactions made by customers during the statistical period. The higher the M value, the more revenue for the company [15]. RFM analysis is a common clustering method for explaining customer purchasing behavior based on transaction data. Valuable customers have the highest frequency and monetary value and the lowest recency. These three variables belong to behavioral variables and can be used as clustering variables by observing customers' attitudes towards the product, brand, profit, or even loyalty from the database. The RFM scoring process uses quintiles to quantify customer behavior. The first quintile with the highest value (most minor for recency) is marked as 5. The next quintile is marked as 4, and so on. Finally, all customers are represented by 555, 554, 553, ..., 112, 111. The most valuable customer group is 555, while the worst is 111 [16].

3.3. Execution Steps of the Model

As shown in **Figure 1**, the framework development process begins with the implementation of the CRISP-DM model. To better understand this business, the received data was reviewed and analyzed. Each branch has approximately 700,000 transactions within a specified historical period. These transactions include customer purchases across all product categories of this chain store, including household items, health and beauty products, food items, and tools. Therefore, with the help of SQL software, the food product information for each branch was initially separated. In the second stage, refrigerated food items were also reviewed and cleaned. In the third stage, it was determined that all product names are fully specified along with the brand names, and according to the research hypothesis, product brands do not affect the result. At this stage, three refined Excel files were created, each containing columns for Branch Name Transaction Date Transaction Amount Customer Name and Code Purchased Product Clustering Loyal Customers for:

- Branch Name
- Transaction Date
- Transaction Amount
- Customer Name and

- Code
- Purchased Product

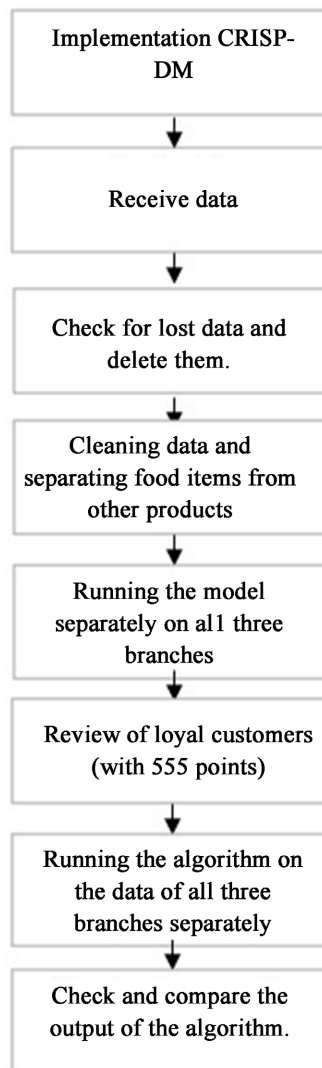


Figure 1. Execution steps of the model.

Clustering Loyal Customer

After determining the execution steps of the research and cleaning the transaction data, the next step is to find the store's loyal customers under study. One of the main reasons for selecting loyal customers for this research is to test the hypothesis that in-person shopping behavior in different geographical areas affects the relationship of purchased products in the customer shopping basket for each branch. Outlier data, such as a customer purchasing from a specific branch only once, can affect the research outcome. Therefore, to avoid this issue, the focus is on loyal customers.

As we explained in section 3.2 customer classification by RFM method, in this section, we classified customers according to the amount of shopping baskets, the number of times they purchased, and the last purchase of customers with the

customer data we had.

We have shown some of these data in **Table 2**. The first column is the unique ID of the customers in each store, the second column is the shopping frequency, and the third column Represents the time interval between the date of the last purchase and the most recent date in the statistical period. Loyal customers for each branch are identified and introduced according to **Table 2**. This research examines the last purchase date according to the available data over three months. Customers with high recency scores are more likely to make repeat purchases.

Table 2. An example of the output of the RFM model for surveying loyal customers.

Customer code	Frequency	Recency	Monetary
0040011139811600166	1	19	243,000
0040131398111600012	1	66	27,000
0040111398111600163	1	59	252,000
0040121398111600297	3	33	1,572,000
004011139811600335	3	9	162,000
....
0040111398111600177	1	0	27,000
0040111398091500045	1	67	54,000
0040081398809115009	2	9	81,000
0040111398091500133	4	15	243,000

4. Results and Discussion

In this study, we aimed to analyze market baskets to observe association rules for frequently purchased food items in each branch of grocery stores using the FP-Growth algorithm.

Initially, a minimum support value was introduced. The support threshold parameter specifies that the minimum coverage items must have to be confirmed as a frequent item set. This threshold can be determined as a percentage of the total transactions or as a specific number of transactions [17]. Based on statistical reports and analyses of each item's repetition frequency, the support threshold considered in this study was set at 7%.

The number of association rules generated is determined by the confidence threshold parameter. McLennan *et al.* suggest that in a sparse data set, such as a purchase transaction table, this threshold should be considered between 5-10% to derive reasonable rules.

Accordingly, a confidence threshold of 7% was chosen.

In basket analysis, we aim to find groups of items that often appear together and provide a recommendation based on the repetition of items in transactions containing other items. Afterward, the results of the three algorithms are evaluated to compare them. The expected outcome is to facilitate an understanding of

all consumer transactions at a specific grocery store branch and identify products that are frequently purchased together at that branch. The results of the FP-Growth algorithm, implemented in Python software, included 10,000 recommendations, a portion of which are shown in **Figures 2-4**. The first column contains the items for which the recommendations are provided, and the second column contains the items often purchased with the leading item.

4.2. Market Basket Analysis Results for Branch 1

In Branch 1, the association rule algorithms produce similar extraction rules with a maximum support of 0.1176, maximum confidence of 1, and a lift of 8.5.

Below are the details of the top 10 rules based on the highest confidence values (**Figure 2**). Transactions in the branch1 show the highest association among items like:

```
[ ] sort=df1.sort_values(by=['lift','confidence'],ascending=False)
sort
```

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning:
`should_run_async` will not call `transform_cell` automatically in the future. Please pass the result to `transformed_cell` argument and any exception that

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
503	(Eggs, wafers, Rani)	(bread, pasta)	0.117647	0.117647	0.117647	1.0	8.500000	0.103806	inf	1.000000
507	(bread, pasta)	(Eggs, wafers, Rani)	0.117647	0.117647	0.117647	1.0	8.500000	0.103806	inf	1.000000
567	(Eggs, wafers,Oil, Rani)	(bread, pasta)	0.117647	0.117647	0.117647	1.0	8.500000	0.103806	inf	1.000000
575	(Eggs, wafers, Rani)	(bread,Oil, pasta)	0.117647	0.117647	0.117647	1.0	8.500000	0.103806	inf	1.000000
578	(bread,Oil, pasta)	(Eggs, wafers, Rani)	0.117647	0.117647	0.117647	1.0	8.500000	0.103806	inf	1.000000
...
1631378	(Eggs, drinks, chips)	(Oil)	0.117647	0.764706	0.117647	1.0	1.307692	0.027682	inf	0.266667
1631382	(Sauce, drink)	(Oil)	0.117647	0.764706	0.117647	1.0	1.307692	0.027682	inf	0.266667
1631389	(Eggs, chips,drinks)	(Oil)	0.117647	0.764706	0.117647	1.0	1.307692	0.027682	inf	0.266667
1631394	(Sauce, drink,Chips)	(Oil)	0.117647	0.764706	0.117647	1.0	1.307692	0.027682	inf	0.266667
1631397	(Eggs,Sauce, drinks, chips)	(Oil)	0.117647	0.764706	0.117647	1.0	1.307692	0.027682	inf	0.266667

Figure 2. Output for Branch 1. (bread and pasta→, eggs, Rani drink, and wafers) And (bread and pasta→, eggs, oil, Rani drink, and wafers).

4.3. Market Basket Analysis Results for Branch 2

In Branch 2, the association rule algorithms produce similar extraction rules with a maximum support of 0.11 and a maximum confidence of 1. Below are the details of the top 10 rules based on the highest confidence values (**Figure 3**).

Transactions in the Railway branch show the highest association among items like:

4.4. Market Basket Analysis Results for Branch 3

In Branch 3, the association rule algorithms produce similar extraction rules with a maximum support of 0.111 and a maximum confidence of 1. Below are the details of the top 10 rules based on the highest confidence and lift values (**Figure 4**).

Transactions in the Behrood branch show the highest association among items like:

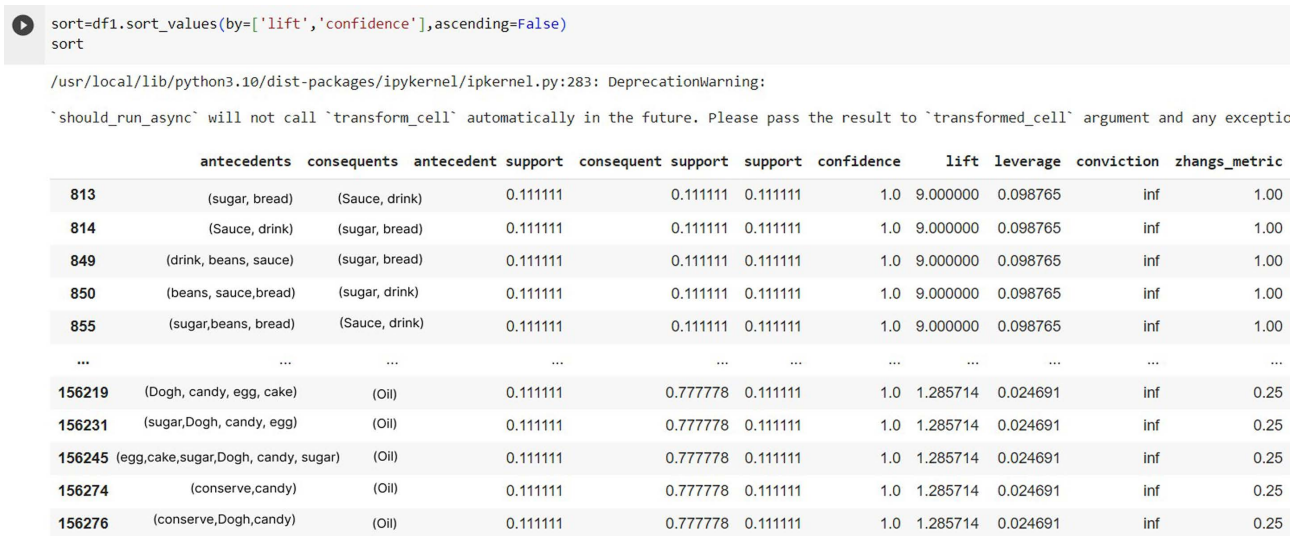


Figure 3. Output for Branch 2. (sauce and drinks → sugar and bread) And (sugar and drinks → sauce, beans, and bread).

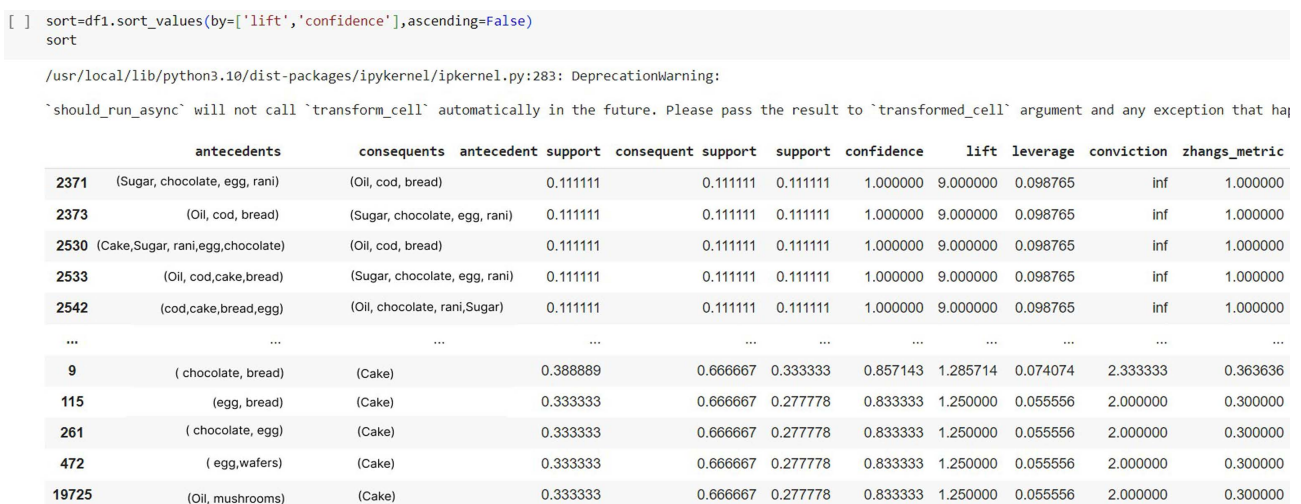


Figure 4. Output for Branch 4. (oil, bread, split peas →, chocolate, eggs, Rani drink, and sugar) And (chocolate, eggs, Rani drink, and sugar → oil, cake, and split peas).

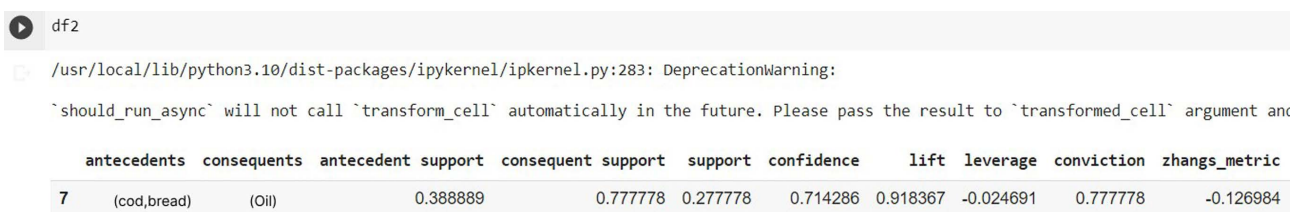


Figure 5. Independent items display.

Independent Items Analysis Items displayed in Figure 5 do not show any association with each other in Branch 2, considering minimum support of 0.7%. As shown, the lift and confidence values are less than one, and Zhang’s criterion is negative, indicating the independence of the items.

As seen in Figure 5, in Branch 2, oil and bread are independent of each other,

considering a minimum support of 0.7%. This is while in Branches 1 and 3, bread and oil are among the items that are associated with each other. Therefore, the shopping basket for each branch is still different from the other branches.

5. Conclusion

As the algorithm results indicate, the association of items purchased by customers in each branch of a chain store is different, implying that branch variation impacts the market basket. Therefore, chain stores should pay attention to these branch differences and conduct separate market basket analyses for each branch. The empirical evaluation in this study shows that the proposed method is computationally efficient. Additionally, we assume that different stores may have different product combinations over different periods. This means that each store can have its product combination, and the product combination in a store can dynamically change over time.

6. Future Recommendations

The following recommendations are made for future market basket analysis using data mining approaches:

- Implement the algorithm across different chain stores with larger data volumes.
- Consider factors such as seasonality or product brands.
- Investigate the relationship between various items, such as digital products.
- Analyze customer basket correlations in online stores.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Chen, Y., Tang, K., Shen, R. and Hu, Y. (2005) Market Basket Analysis in a Multiple Store Environment. *Decision Support Systems*, **40**, 339-354. <https://doi.org/10.1016/j.dss.2004.04.009>
- [2] Shahrabi, J.-N., et al. (2012) Analyzing Customers' Shopping Carts Using Dependency Rules in Citizen Chain Stores. *Iran Data Mining Conference*, Shahrod.
- [3] Caldas, G.M.I., Rubio, X.D.M., Lopez, Y.J.G. and Gutiérrez, J.A.T. (2022) A Product Net-work Analysis Using a Priori Algorithm for Extending the Market Basket in Retail. *Proceedings of the First Australian International Conference on Industrial Engineering and Operations Management*, Sydney, 20-21 December 2022, 2025-2034.
- [4] Sussolaikah, K. (2021) Market Basket Analysis for Determination of Consumer Behavior at XYZ Stores Using R Programming. *Advance Sustainable Science, Engineering and Technology*, **3**, Article 0210206. <https://doi.org/10.26877/asset.v3i2.8547>
- [5] Loraine Charlet Annie, M.C. and Ashok Kumar, D. (2012) Market Basket Analysis for a Supermarket Based on Frequent Itemset Mining. *International Journal of Computer Science Issues*, **9**, 257-264. <https://www.ijcsi.org/>
- [6] Alawadh, M. and Barnawi, A. (2024) A Consumer Behavior Analysis Framework toward Improving Market Performance Indicators: Saudi's Retail Sector as a Case Study. *Journal of Theoretical and Applied Electronic Commerce Research*, **19**,

- 152-171. <https://doi.org/10.3390/jtaer19010009>
- [7] Formánek, T. and Sokol, O. (2022) Location Effects: Geo-Spatial and Socio-Demographic Determinants of Sales Dynamics in Brick-and-Mortar Retail Stores. *Journal of Retailing and Consumer Services*, **66**, Article 102902. <https://doi.org/10.1016/j.jretconser.2021.102902>
- [8] Malik, M.H., Ghous, H., Ismail, M., Jamshaid, S. and Altaf, J. (2024) Market Basket Analysis for Next Basket Item Prediction Using Data Mining and Machine Learning. *Journal of Computing & Biomedical Informatics*, in Press.
- [9] Paslari, P., Asghar, M., Anvari, A.A. and Sharifi, M. (2018) The Effect of Regional Features on the Relationship between the Store Image, the Service Quality and WOM: The Case Study of Urban and Rural Areas in Bandar-Abbas. *Geography*, **8**, 267-287.
- [10] Sudirman, I.D., Bahri, R.S., Utama, I.D. and Ratnapuri, C.I. (2021) Using Association Rule to Analyze Hypermarket Customer Purchase Patterns. *Proceedings of the Second Asia Pacific International Conference on Industrial Engineering and Operations Management*, Surakarta, 14-16 September 2021, 12-23.
- [11] Yue, X. and Shi, F. (2017) Stock Pattern Mining and Correspondence Analysis Based on Association Rules. *Journal of Data Analysis and Information Processing*, **5**, 77-86. <https://doi.org/10.4236/jdaip.2017.53006>
- [12] Calvo, P. and Egea-Moreno, R. (2021) Ethics Lines and Machine Learning: A Design and Simulation of an Association Rules Algorithm for Exploiting the Data. *Journal of Computer and Communications*, **9**, 17-37. <https://doi.org/10.4236/jcc.2021.912002>
- [13] Sarath, U. and Nair, N.S. (2023) Apriori versus FP-Growth for Recommendation System. In: Tripathy, S., Samantaray, S., Ramkumar, J. and Mahapatra, S.S., Eds., *Recent Advances in Mechanical Engineering*, Springer, 155-162. https://doi.org/10.1007/978-981-19-9493-7_16
- [14] Rungruang, C., Riyapan, P., Intarasit, A., Chuarkham, K. and Muangprathub, J. (2024) RFM Model Customer Segmentation Based on Hierarchical Approach Using FCA. *Expert Systems with Applications*, **237**, Article 121449. <https://doi.org/10.1016/j.eswa.2023.121449>
- [15] Laksono, F.A., Rachmat, B. and Sutasrso, Y. (2024) B2B Customer Segmentation Based on Customer Lifetime Value Concept and RFM Modeling. *International Journal of Economics Development Research*, **5**, 539-337.
- [16] Ernawati, E., Baharin, S.S.K. and Kasmin, F. (2021) A Review of Data Mining Methods in RFM-Based Customer Segmentation. *Journal of Physics: Conference Series*, **1869**, Article 012085. <https://doi.org/10.1088/1742-6596/1869/1/012085>
- [17] Ghafari Ashtiani, P. and Davoudi, M. (2017) Review and Analysis of Market Baskets and the Order of Goods in Chain Stores. *Journal of Business Administration Research*, **8**, 161-184.