JARTE
Journal of Applied Research in Technology & Engineering

5(2): 63-68, 2024
https://doi.org/10.4995/jarte.2024.20753
Received: November 23, 2023; Accepted: February 06, 2024; Published: July 31, 2024

ASSOCIATION RULE MINING ALGORITHM IMPLEMENTATION FOR E-COMMERCE IN THE RETAIL SECTOR

Namatullah Wahidi [a1], Rita Ismailova [a2*]

Kyrgyz-Turkish Manas University, Bishkek, Kyrgyz Republic.

[a1] 2151Y01003@manas.edu.kg, [a2] rita.ismailova@manas.edu.kg

Abstract:

The growth of online trading platforms and the development of market technology have forced businesses to take part in the analysis of client behaviour. Therefore, this research aims to analyse customer behaviour in the Kyrgyz Republic to enhance supplier's revenue, service quality, and customer satisfaction. For that, a data consisting of 577730 invoices from a company in the retail sector is used. To find the association rules, the data was analysed using the Apriori algorithm. The minimum support threshold was set at 40% and the minimum confidence level at 80%. Results generated 118 rules which revealed strong connections between items and showed up to 61.06% relationship between the consumption of products, suggesting a connection among the considered items. Thus, the association rule highlights the significance of association rule mining in uncovering valuable insights within sales transaction data. These insights can inform targeted marketing efforts, inventory management, and the enhancement of customer experiences and optimize business strategies to meet customer preferences, ultimately fostering growth and competitiveness in the retail sector.

Keywords: association rule mining; incremental data mining; data mining; Apriori algorithm.

Cite as: Wahidi, N., Ismailova, R. (2024). Association rule mining algorithm implementation for e-commerce in the retail sector. *J Appl Res Eng Technol & Engineering, 5*(2), 63-68. https://doi.org/10.4995/jarte.2024.20753

1. Introduction

The recent development in Internet technology is very useful for the growth of enterprises. Almost every organization is either represented on the internet or has started their business through e-commerce. In addition, organizations collect information about customers and their business transactions, which are highly valuable for company development and increase of sales. However, there is always missing data in such datasets.

There are various data mining techniques available to extract valuable and useful information for enterprises. As for the missing values, to extract unknown pieces of information from a large database repository data mining set of techniques is used. One of the most significant and well-liked data mining methods for extracting unknown knowledge from transaction databases is Association Rule Mining (ARM). The ARM is a component of machine learning (ML), which is responsible for the identification of interesting connections between objects in large transaction data sets. Association rules specify how and why certain elements are associated and communicate the relationships (Fister et al. 2023).

The history of ARM began with a seminal paper by Agrawal. Agrawal provided the theoretical foundation for the ARM process. He also proposed the first algorithm, called the Apriori algorithm (Agrawal and Srikant 1994). Since then it has been one of the most popular research areas in the field of knowledge discovery (Iváncsy et al. 2004). As the field of ARM has evolved over time, its algorithms have seen a significant shift in their underlying methodologies. Initially, many of the first ARM algorithms relied on deterministic techniques such as Frequent Pattern (FP) growth and Eclat, as evidenced by works like (Borgelt et al., 2005; Han et al., 2000; Zaki et al., 2000). These deterministic approaches paved the way for early breakthroughs in discovering associations within categorical and numerical variables and left a lasting impact on the field.

Despite the large number of algorithms for data analysis and the availability of detailed sales data, it is remarkable that the retail sector in Kyrgyzstan has not been thoroughly investigated. Although Kyrgyzstan has a market of relatively modest size, its distinctive characteristics make it an interesting subject for research, especially since its consumer preferences are similar to those of neighboring countries in Central Asia. The main motivation for this study is to analyze consumer habits and buying patterns in Kyrgyzstan, which may be seen as a typical example for the wider Central Asian region. Thus, the purpose of our current study is to conduct a market basket analysis using association rules. To achieve this, a dataset covering the year 2020 from a department store located in Bishkek, Kyrgyzstan is utilized. This dataset comprises detailed information on 50 individual product purchases made by consumers. In our analysis, the Apriori algorithm is employed, implemented through the Python programming language.

*Corresponding author: R. Ismailova, rita.ismailova@manas.edu.kg

The paper is organized in the following way. The section two provides an overview of the related works. The third section outlines the study methodology, which is then followed by the research findings in the section four. Section five concludes the study.

2. Related Works

The Internet is one of the finest business platforms available today. Many businesses nowadays are implementing innovative business methods to grow their revenue. Prospective clients may learn about items, their characteristics, and comparisons between them thanks to online businesses. Comparatively speaking to traditional business, e-commerce eliminates the limitations of time and place (Soni et al., 2017). According to the study by (Chen, 2005) demonstrates that market basket analysis is a valuable technique for firms to get a deep understanding of their customers' purchasing patterns. However, current methodologies may prove inadequate in a multi-store setting, where the presence of products and client preferences may differ across different markets. The study found that the strength of relationships between products is valuable information that can be used for cross-selling, up-selling, offering coupons, and making other recommendations, and this information can be used to improve marketing, sales, service, and operation strategies in a variety of ways. For instance, the retail sector may use market basket research to find commonly co-purchased goods, create focused marketing campaigns, and improve product positioning in shops. In a separate investigation conducted by (Özçakir and Çamurcu, 2007), data mining techniques were used to analyze the sales data of a bread firm. They developed software using association rules. The implemented software used an Apriori algorithm to identify the correlations among goods that were bought together across multiple time periods, sales locations, and input values.

Another study, conducted by (Vujkovic et al., 2020) analyzed a dataset comprising 1641 transaction records. They set the minimum support threshold at 0.09 and the confidence threshold at 0.9. By employing the FP-Growth algorithm with these parameters, they generated 19 rules in approximately 6 seconds. These rules consisted of combinations of 3 items, demonstrating a rule strength of 112.66%. Remarkably, the reported accuracy of these rules was exceptionally high at 217%.

In medical research, Apriori is utilized to identify frequently occurring illnesses with related symptoms through frequent itemsets, as highlighted by (Rao et al., 2023). Another investigation showcased the advantages of employing a semi-supervised system for association rule filtering, (Araujo et al., 2022) demonstrated that this approach may yield superior results compared to both unsupervised and supervised systems, especially when dealing with limited training data.

Furthermore, Singh et al. (2021) presented a method for enhancing recommendation systems using a movie-lens dataset. By tailoring suggestions based on user profiles, the system offered highly effective personalized recommendations. Their findings suggested that items within the same category are commonly purchased together, contributing to the effectiveness of the recommendation system.

3. Method

3.1. Problem Statement

Basket analysis, referred to as market analysis of basket or affinity analysis, is a method used by businesses to figure out their consumers' purchasing behaviour. It includes analysing the things that consumers often buy together in order to detect patterns and relationships. This helps retailers in making well-informed decisions on product placement, cross-selling, and promotional techniques.

The primary objective of this study is to do a thorough examination and analysis of the Apriori algorithm on various datasets, taking into account the above-mentioned factors. The aim is to understand their capabilities in improving recommender systems. Recognizing the complexities involved in identifying frequently purchased products together and building accurate customer profiles in the e-commerce domain, researchers are aware of the challenges posed by categorizing products into various categories. As a result, this work aims to find algorithms that may successfully get around these difficulties and promote the creation of a recommender system with more flexibility.

3.2. Apriori Algorithm

The Apriori algorithm finds the most frequent item sets or elements in a transaction database and identifies association rules between the items. It was initially introduced by (Agrawal and Srikant ,1994). The algorithm is a stepwise technique that starts with the simplest rule and adds individual products to the k + 1 product set, where k sets of products are used. At the initial stage, the process involves setting a support threshold and the algorithm defines a subset of items, for which an individual value exceeds this threshold. Typically, the first selection includes single items. All products falling below the support threshold are dismissed. Products that go through the first step form two-item product clusters. The calculated support values of these product clusters are also compared with the support value initially determined, and the product clusters below are again ignored. But these ignored product clusters are candidates for two-product rules in the future (for the right-hand side of the rules). This process merges the products frequently until the specified support value is reached, and finally, no more curing clusters can be found. After the detection of frequent product clusters, the rule-finding process is started. There is a minimum predetermined value of support as well as association rules above a confidence value (Aggarwal, 2015; Giudici, 2005).

3.3. Procedure

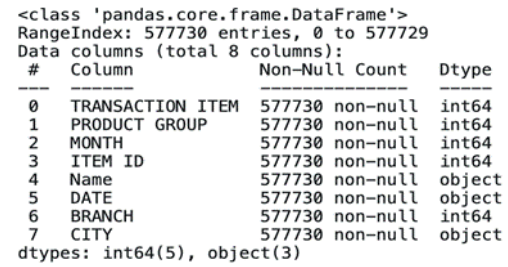
In the analytical phase of our research, we employed the Apriori algorithm, a widely recognized method for association rule mining, implemented using the Python

programming language. This algorithm is specifically designed to reveal underlying patterns and associations within transactional datasets. Meticulous parameterization ensured methodological rigor.

The minimum support threshold, influencing the identification of frequent itemsets, was judiciously set at 40%, signifying the minimum occurrence frequency required for an itemset to be considered significant. Simultaneously, the minimum confidence level, indicative of the strength of association between antecedent and consequent in a rule, was rigorously established at 80%. The purpose of this parameterization was to achieve a harmonious equilibrium between revealing significant correlations and reducing interference within the dataset.

Implementation was executed through the Python programming language, leveraging its versatile libraries and functionalities conducive to algorithmic intricacies. The use of Python facilitated efficient algorithm execution, ensuring reproducibility and transparency in the research methodology.

This comprehensive approach to algorithmic implementation and parameter specification underscores the rigor and precision in our research, enhancing the reliability and interpretability of the generated association rules.



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 577730 entries, 0 to 577729
Data columns (total 8 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   TRANSACTION ITEM  577730 non-null  int64
 1   PRODUCT GROUP     577730 non-null  int64
 2   MONTH             577730 non-null  int64
 3   ITEM ID           577730 non-null  int64
 4   Name              577730 non-null  object
 5   DATE              577730 non-null  object
 6   BRANCH            577730 non-null  int64
 7   CITY              577730 non-null  object
dtypes: int64(5), object(3)
```

(a)

Index	ANSACTION IT	ODUCT GROU	MONTH	ITEM ID	Name	DATE	BRANCH	CITY
504725	12001014064	1	11	10035000010000	Jsm Grn Tea 100Gr	27/11/20	1	BISHKEK
345306	12001022096	2	12	20023000050000	Coffee Cuba	4/12/20	2	OSH
367679	12001012497	1	3	10015960010000	Granul 100 GR	5/3/20	1	BISHKEK
140961	12001030663	1	7	10050500025000	pip fruit 250Gr	5/7/20	3	TALAS
310820	12001041781	1	8	10015960010000	Granul 100 GR	4/8/20	4	KARAKOL
429140	12001013062	1	1	10051500010000	pip fruit (spon)	3/1/20	1	BISHKEK
314981	12001051026	4	8	40000300035000	Chocolates 350Gr	12/8/20	5	NARYN
510066	12001014149	1	1	10010200010000	Protein 100 GR	4/1/20	1	BISHKEK
123081	12001010548	1	10	10050100010000	tropical fruit 100Gr	17/10/20	1	BISHKEK
354694	12001022339	1	10	10050100010000	tropical fruit 100Gr	16/10/20	2	OSH
282337	12001011554	1	10	10041300020000	Kids biscuit rfsyz)	28/10/20	1	BISHKEK
279047	12001011549	2	5	20024000050000	Coffee	5/5/20	1	BISHKEK
137083	12001060625	1	9	10010100025000	Opa 250Gr	2/9/20	6	JALALABAD
221696	12001011152	1	2	10050100025000	tropical fruit 250Gr	4/2/20	1	BISHKEK
111902	12001010508	1	4	10050100005000	tropical fruit 050Gr	13/4/20	1	BISHKEK
349143	12001022294	3	11	30012000025000	milk 250Gr	26/11/20	2	BATKEN
424934	12001013013	1	6	10015960010000	Granul 100 GR	11/6/20	1	BISHKEK
438523	12001013170	1	7	10050100050000	tropical fruit 500Gr	2/7/20	1	BISHKEK
384656	12001012853	1	12	10050100050000	tropical fruit 500Gr	10/12/20	1	BISHKEK
513798	12001014214	1	7	10050200010000	soybean 100Gr	7/7/20	1	BISHKEK

(b)

Figure 1: Sample retail sector transaction data for Market Basket Analysis a) field characteristics; b) data set view.

3.4. Dataset

For the current study, a dataset of 577 730 invoices from a company in the retail sector is used. In each invoice, the items purchased represent shopping transactions. An invoice is an indication of the products that go into the customer's shopping cart. In the invoice, each collection of lines is a representation of the purchase items (Figure 1). The current dataset contains purchase data for the year 2020. In this dataset, 'TRANSACTION ITEM' is the invoice or the identification number of the customer who purchased a product. There are about 7655 various customers in this data, 'PRODUCT GROUP' is the category of products such as tea, rice, chocolate, and milk, 'TEM ID' and 'Name' are the 'ID' and the name of the product, there are 242 unique products, 'DATE' is the date of purchase, 'BRANCH' and 'CITY' are the area of product purchase, there are 8 different areas and 9 different cities.

The dataset under consideration comprised sales transactions involving approximately 242 distinct items. These transactions were made by a total of 7652 customers, spanning eight different municipalities within the Kyrgyz Republic.

4. Results and discussion

4.1. Preliminary data analysis

In the initial phase of our investigation, preliminary data analysis was conducted by importing a dataset stored in a CSV file into our research program. This process was facilitated through the utilization of the Pandas library in conjunction with the Python programming language. To enhance the dataset's suitability for subsequent analysis employing the Apriori algorithm, redundant columns were removed.

The initial analysis sought to examine the distribution of sales among these municipalities. The findings reveal that the city of Bishkek exhibited the highest sales volume, accounting for 64% of the total sales transactions, while the Batken region reported the lowest sales volume, representing a mere 1% of the overall sales distribution (Figure 2). The data analysis provides unequivocal evidence that sales statistics in Bishkek significantly surpass those in other places, with an average disparity

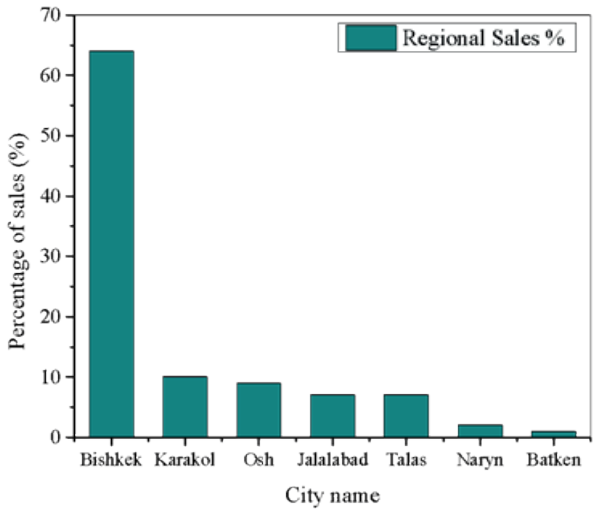


Figure 2: Regional sales distribution.

of around 50%. The evident divergence in consumer buying patterns and economic dynamics between the urban center of Bishkek and its surrounding regions is highlighted by this prominent variance.

As a next step, the analysis was extended to the distribution of sales based on product types. The primary aim of this phase was to identify association rules among items that exhibited substantial prevalence in sales transactions. To achieve this objective, items with notable representation in the dataset were highlighted.

The analysis of all products revealed that 'Tropical Fruit' emerged as the highest-selling product category, boasting approximately 44681 sales (Figure 3). This designation positions it as the most prominent product category among the 577730 available items. Furthermore, the examination of the top 20 products elucidated that these items collectively accounted for 62% of the total sales. Notably, the leading 5 products independently contributed to 26% of the overall sales volume.

This distinction is of paramount importance in our study, as it underpins our strategy to identify association rules exclusively among items that enjoy substantial purchase frequency. Based on these results, it was decided to narrow the focus of the study to items of significant popularity, thereby managing the overall number of item sets within a scope of 358192 invoices. Despite the a slightly small decrease in the number of invoices, the focus facilitates a more targeted and efficient exploration of association rules within our dataset.

4.2. Apriori algorithm results

In the context of ARM, two crucial parameters, namely the minimum support, and the minimum confidence, play pivotal roles in the calculation and discovery of associations (Li et al., 2019). These parameters offer analysts the means to control the precision of association rules. When aiming for highly accurate association rules, a minimum support threshold, and a high confidence percentage should be established. Conversely, if the objective is to derive association rules with a lower degree of accuracy, lower values can be set for both the minimum support and confidence parameters. To elaborate, support represents the percentage of item combinations that appear within the database, while confidence signifies the percentage of robust relationships between items.

The primary objective within the confines of current study was to devise precise association rules that illuminate robust connections among items. Employing the Apriori algorithm with parameter values set at a minimum support of 40% and a minimum confidence level of 80%, 118 rules were successfully generated. These rules delineate significant associations among items co-purchased, thereby providing insights into the patterns and relationships inherent in the dataset. (Figure 4) provides a visual representation of the top 20 rules characterized by the highest support values. Similar analysis of more than 900 transaction data from a grocery shop generated. A total of 145 association rules were created, with a minimum support value of 20%. This highlights the reliability of the results and demonstrates the effectiveness of association rule mining techniques in understanding consumer behaviour and purchasing patterns in retail environments (Dio et al., 2023).

The result of the association rule is given in the form of if x, then y. Based on the outcomes derived from one of the association rules generated by the Apriori algorithm, it can be determined that there exists a pattern between the purchase of 'Tropical fruit 100 gr' and 'Kids biscuit 100 gr' as shown in (Figure 5). Specifically, it is seen that there is a 61% support for the association, indicating that the two items are often bought together. Additionally, the confidence accuracy of this association is measured at 79 % suggesting a high likelihood that if a customer purchases 'Tropical fruit 100 gr', they will also purchase 'Kids biscuit 100 gr'.

The same likelihood was observed on items 'Tropical fruit 100 gr' and 'Protein 100 gr'. It is seen that there exists a 50% percentage of support for the relationship, the statistical accuracy of this connection has been quantified at 56%, meaning that it is likely that customers who buy 'Tropical fruit 100 gr' are also prepared to buy 'Protein 100 gr' with the highest support.

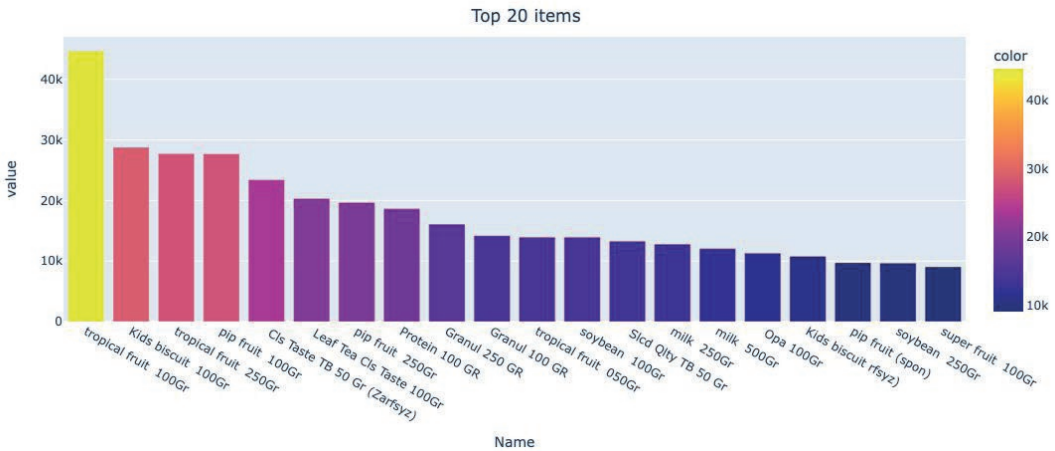
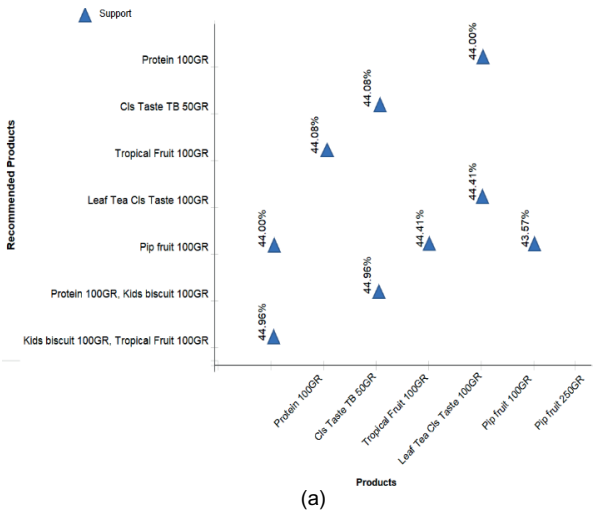
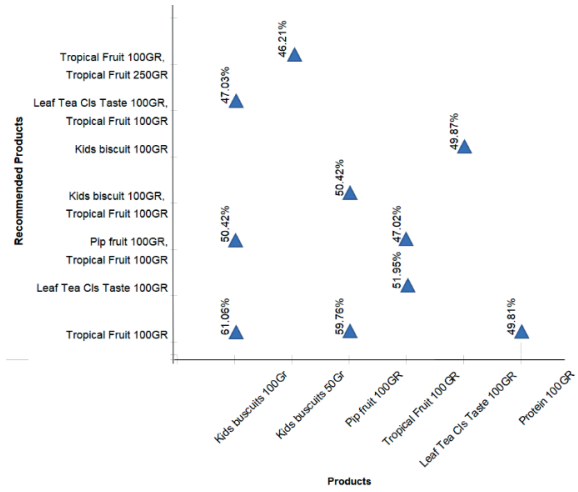


Figure 3: Top 20 sales items.



(a)



(b)

Figure 4: Top 20 items with highest support percentages.

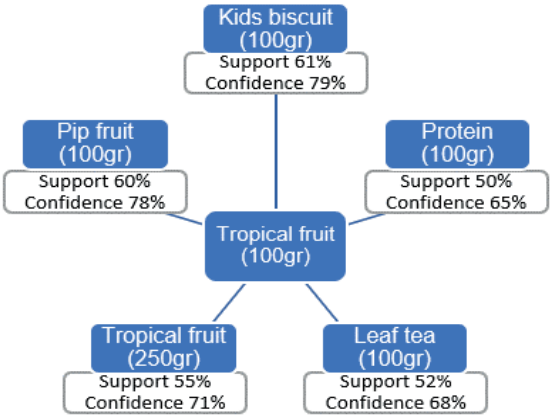


Figure 5: Relationship between products with highest supports.

5. Conclusion

In this study, an in-depth analysis of sales transaction data from various municipalities was conducted within the Kyrgyz Republic to investigate potential enhancements for the recommendation system used in retail sectors, intending to enhance its overall quality. Leveraging the Apriori algorithm, the aim was to uncover valuable association rules among items frequently purchased by customers. Our investigation yielded several noteworthy insights that contribute to both the understanding of consumer behaviour and the optimization of business strategies in the retail sector.

The analysis began by examining the distribution of sales across municipalities, revealing significant variations in purchase patterns. Notably, the city of Bishkek emerged as the dominant contributor to total sales, underlining the importance of tailoring marketing and inventory strategies to specific regional dynamics.

Further, our analysis delved into the distribution of sales by product types. Our finding highlighted the importance of focusing on high-frequency items when crafting association rules, thus improving the relevance and effectiveness of product recommendations.

The parameters of minimum support and minimum confidence proved crucial in our analysis, enabling us to fine-tune the accuracy of association rules. By setting these parameters at 40% and 80%, respectively, 118 rules was generated that revealed strong connections between items. Notably, these association were illustrated through specific examples, such as the relationship between ‘Tropical fruit 100 gr’ and ‘Kids biscuit 100 gr’ at 61.06% support, meaning, ‘Tropical fruit 100 gr’ and ‘Kids biscuit 100 gr’ may be sold together. As a result, a customer purchasing ‘Tropical fruit 100 gr’ will purchase ‘Kids biscuit 100 gr’ with a high probability.

The findings of our study demonstrate the potential to improve the accuracy of the recommendations system via the integration of diverse and different data sources. These results allow the company to make product planning to increase sales by placing the products that customers often purchased together nearby on shelves, while at the same time contributing to customer satisfaction by saving time for the customers.

In conclusion, our research underscores the significance of ARM in uncovering valuable insights within sales transaction data. These insights can inform targeted marketing efforts, inventory management, and the enhancement of customer experiences. By understanding and leveraging the relationships among frequently purchased items, businesses can optimize their strategies to meet customer preferences, ultimately fostering growth and competitiveness in the retail sector.

However, the datasets that were chosen had significant dissimilarities, hence impeding the achievement of the expected results. The primary finding of the study is that there is a need for more investigation to identify more appropriate datasets and enhance the methodologies and models used. Therefore it can be conclude that the use of the apriori algorithm in data mining has been established with the goal of generating valuable information patterns, specifically in the context of sales trends analysis, the utilization of the apriori data mining algorithm enables the identification of item association rules, specifically in consumer purchase patterns, which can be employed by the store to improve sales through the development of promotional strategies cantered around items that are

frequently purchased together, this program generates item association rules or list combinations based on customer purchase habits, with the aim of developing promotion tactics that are very precise and effective, as opposed to the manual implementation of promotional methods.

The study's conclusions show that using data mining methods in the field of science has aided in the discovery of new information patterns. Moreover, a comprehensive examination has been undertaken to evaluate the usage of these methodologies to comprehend the sales trends, particularly the item association rules.'

References

Aggarwal, C.C., Aggarwal, C.C. (2015). *Data classification*. Springer International Publishing. pp. 285-344. https://doi.org/10.1007/978-3-319-14142-8_10

Agrawal, R., Srikant, R. (1994), September. Fast algorithms for mining association rules. *In Proc. 20th Int. Conf. Very Large Data Bases, VLDB*. Vol. 1215, pp. 487-499.

Araujo, L., Martinez-Romo, J., Bisbal, O., Sanchez-de-Madariaga, R., The Cohort of the National AIDS Network (CoRIS), Portilla, J., Portilla, I., Merino, E., García, G., Agea, I., Sánchez-Payá, J., Rodríguez, J. C., Giner, L., Reus, S., Boix, V., Torrus, D., Pérez, V., Portilla, J., Gómez, J. L., … Telleria, P. (2022). Discovering HIV related information by means of association rules and machine learning. *Scientific Reports, 12*(1), 18208. https://doi.org/10.1038/s41598-022-22695-y

Borgelt, C, 2005, August. An Implementation of the FP-growth Algorithm. In *Proceedings of the 1st International Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations* (pp. 1-5). https://doi.org/10.1145/1133905.1133907

Chen, Y.L., Tang, K., Shen, R.J., Hu, Y.H., 2005. Market basket analysis in a multiple store environment. *Decision support systems, 40*(2), 339-354. https://doi.org/10.1016/j.dss.2004.04.009

Dio, R., Dermawan, A. A., Putera, D. A. (2023). Application of Market Basket Analysis on Beauty Clinic to Increasing Customer's Buying Decision. Sinkron, 8(3), 1348–1356. https://doi.org/10.33395/sinkron.v8i3.12421

Fister Jr, I., Fister, I., Fister, D., Podgorelec, V., Salcedo-Sanz, S. (2023). A comprehensive review of visualization methods for association rule mining: Taxonomy, Challenges, Open problems and Future ideas. *arXiv preprint* arXiv:2302.12594. https://doi.org/10.1016/j.eswa.2023.120901

Giudici, P. (2005). *Applied data mining: statistical methods for business and industry*. John Wiley & Sons.

Han, E.H., Karypis, G., Kumar, V. (2000). Scalable parallel data mining for association rules. *IEEE Transactions on Knowledge and Data Engineering, 12*(3), 337-352. https://doi.org/10.1109/69.846289

Iváncsy, R., Kovács, F., Vajk, I. (2004). An Analysis of Association Rule Mining Algorithms. In *CDROM Proc. of Fourth International ICSC Symposium on Engineering of Intelligent Systems* (EIS 2004).

Li, K., Liu, L., Wang, F., Wang, T., Duić, N., Shafie-khah, M., Catalão, J.P.S. (2019). Impact factors analysis on the probability characterized effects of time of use demand response tariffs using association rule mining method. *Energy Conversion and Management, 197*, 111891. https://doi.org/10.1016/j.enconman.2019.111891

Özçakir, F.C., ÇAMURCU, A.Y. (2007). Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 6*(12), pp.21-37.

Rao, A.B., Kiran, J.S., Poornalatha G. (2023). Application of market–basket analysis on healthcare. *International Journal of System Assurance Engineering and Management, 14*(S4), 924–929. https://doi.org/10.1007/s13198-021-01298-2

Singh, P.K., Othman, E., Ahmed, R., Mahmood, A., Dhahri, H., Choudhury, P. (2021). Optimized recommendations by user profiling using apriori algorithm. *Applied Soft Computing, 106*, 107272. https://doi.org/10.1016/j.asoc.2021.107272

Soni, H.K., Sharma, S., Jain, M., 2017, February. Plausible characteristics of association rule mining algorithms for e-commerce. In *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics* (AEEICB) (pp. 36-39). IEEE. https://doi.org/10.1109/AEEICB.2017.7972379

Vujkovic, M., Keaton, J. M., Lynch, J. A., Miller, D. R., Zhou, J., Tcheandjieu, C., Huffman, J. E., Assimes, T. L., Lorenz, K., Zhu, X., Hilliard, A. T., Judy, R. L., Huang, J., Lee, K. M., Klarin, D., Pyarajan, S., Danesh, J., Melander, O., Rasheed, A., … Saleheen, D. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature Genetics, 52*(7), 680–691. https://doi.org/10.1038/s41588-020-0637-y

Zaki, M.J. (2000). Scalable algorithms for association mining. *IEEE transactions on knowledge and data engineering, 12*(3), 372-390. https://doi.org/10.1109/69.846291