Malaysian Journal of Computing, 9 (1): 1746-1758, 2024

eISSN: 2600-8238

EXPLORATORY ANALYSIS WITH ASSOCIATION RULE MINING ALGORITHMS IN THE RETAIL INDUSTRY

Alaa Amin Hashad[1], Khaw Khai Wah[2*], Alhamzah Alnoor[3] and XinYing Chew[4]

[1,2]School of Management, Universiti Sains Malaysia, Gelugor 11800, Malaysia*
[3]Management Technical College, Southern Technical University, Basrah, Iraq
[4]School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Malaysia
[1]alaa.hashad@student.usm.my, [2]khaiwah@usm.my, [3]alhamzah.malik@stu.edu.iq, [4]xinying@usm.my

ABSTRACT

Every year the retail sector expands quickly. These industries are becoming more competitive and difficult to operate in due to their expansion. Changing consumer buying habits, a decline in people's spending capacity and an increase in international retailers are a few of the difficulties that must be overcome. In the context of mining frequent item sets, many methods have been proposed to push various kinds of limitations inside the most well-known algorithms. This study presents an exploratory analysis for retail stores that uses market basket analysis as one of the data mining techniques to identify frequent patterns in customer purchases. The proposed method is based on comparing two algorithms: Apriori and Frequent Pattern Growth (FP- Growth). The study used a retail store dataset consisting of 522,064 rows and 7 variables. Data pre-processing was performed to clean and encode the data to be used in the model. The dataset limitation involves 25% null values in the ID column. To address this, missing customer IDs are filled with the last valid ID, assuming repeated purchases. The FP-Growth algorithm was found to be faster and more effective than the Apriori algorithm in extracting frequent item sets and generating association rules. The retail industry based on these frequent item sets is expected to increase sales by recommending highly associated items to customers.

Keywords*: Association Rule, Apriori, Data Mining, FP-Growth, and Market Basket Analysis.*

Received for review: 16-07-2023; Accepted: 25-03-2024; Published: 01-04-2024
DOI: 10.24191/mjoc.v9i1.21433

1. Introduction

There is intense competition in the trading and technology market, as more technology is used, leading to increased revenue and rapid growth. Every retail owner is searching for the best way to attract more customers through marketing campaigns and understanding the market. Studying the market means that you should know the behavior of your consumer and then build your business plan (Hidayat et al., 2019). To meet this competition, you need to be the best one in your field. This achievement needsa more selective and intelligent marketing strategy. The most effective factor which will help them more is data. Of course, all working companies have a transactional dataset about their customers.This data is considered the word of proficiency as it contains a lot of information about the customer. By using this data, they can make a Market Basket Analysis (MBA) which will help them to matchthe right customer with the right goods (Qisman et al., 2021).

The Apriori algorithm is ideal for use in MBA, also known as purchasing transaction analysis, to discover associations between purchased items. As a result, the Apriori algorithm is used as the basis for this investigation (Kurnia et al., 2019). Using a transactional dataset, MBA is a data mining technique for examining the associations of goods that regularly occur simultaneously from regular purchasing and selling. The basic MBA concept aids in identifying the related item pairings in the store (Izang et al., 2019). The conventional Apriori and FP-Growth algorithms are currently commonly employed in landslide deformation response analysis because of their straightforward design, understandable theory, and straightforward implementation. Incorrect association rules between triggers were mined in significant numbers throughout their use. This would make developing reliable association rules between triggering events and landslide deformation much more challenging and time-consuming (Linwei et al., 2023).

Retail stores have many products, which make the customer feel lost when need something specific. Also, the customer may need something and unfortunately didn't find the complementary goods. Which makes the customer disappear despite the existence of the goods and demand. Retailers often collect enormous amounts of transaction data, but it can be difficult to make sense of this data and extract valuable insights. One approach to analyzing transaction data is market basket analysis, which aims to identify patterns or associations among items that are frequently purchased together. In this study, a method for performing market basket analysis is proposed to determine the best model to design a recommendation system and the best way to assess the high correlation between products for each retailer to increase transactions. Python language and also, Tableau software will be used to visualize, build models and present the champion model output graphically.

2. Related Work

A study was executed for utilizing a mechanism for intelligent recommendations in an electronic-commerce recommender system. The proposed methodology has the following characteristics. Client preferences and customer demand associations are automatically learned by segmenting clients by nation, unlike previous recommendation systems that train simply from transaction records. The decrease in irrelevant elements in the dataset will make it possible to create better recommendations. To avoid providing suggestions that would disappoint customers, consumers who are most likely to buy suggested things are then identified with the aid of the Apriori Algorithm, association rules, aggregation, and pruning (Mechery et al., 2022).

Finding patterns in data can be done using data mining techniques. These patterns might offer some suggestions or insights that improve the capacity for decision-making (Shahidan et al., 2023). One such instance is one in which a client has shifted allegiances, as in circumstances where specific contactlenses are prescribed. The information produced as an output is predictions about whether a client will switch for a prediction or about the type of lens they may choose or be prescribed in specific situations. There are many data mining techniques, and the following are techniques used: Association, Classification, Prediction, Clustering, and Outlier Analysis (Dubey et al., 2021).

A study was performed to propose a Market Basket Analysis (MBA) with Apriori Algorithm to examine 2366 transactional data from a case-study company that offers fresh products in online commerce between July and December 2021. The regulations of the association make it clear which products customers will purchase simultaneously if they purchase this product. After using an MBA and organizing the items according to Association Rules, the location is changed to reduce the average distance between orders while adhering to the principle that the highest associations should be located closest to one another. From 31.8 meters to 14.5 meters or a 54.4% decrease, the average distance per order has decreased (Jirapatsil et al., 2022).

According to Samboteng et al., (2022), all data were analyzed by using the market-based analysis method, which develops patterns for each dataset. The association rule with an Apriori algorithm is one approach to market-based analysis that is under discussion. When using sales recommendations to help customers (owners) get recommendations when seeing details of the itemset purchased, this method generates sales transactions with significant correlations between the products in the transaction. According to the trials in this study's findings, it takes less time to develop suggestions and fewer recommendations are offered the higher the minimum support (min_sup) and minimum confidence (min_conf), but the recommendations given are from transactions that frequently arise.

The Apriori algorithm is an interpretation technique of mining association rules, which is implemented into web-based software. The program can analyze transaction data to find candidates and frequent itemsets, after which it can provide association rules that can be shown visually and textually. In the study of some data, it was shown that the program may create more association rules, the smaller the minimal support, and confidence are set. The test results show that creating the candidate 2-item set is typically the step that takes the longest amount of time (Listiawan et al., 2021).

A study conducted for a store by Pradana et al., (2022), showed that the buyer's interest in buying fashion products has decreased significantly. The store needs a system that can forecast customer purchasing patterns and preferences. Data mining techniques were used to process sales transaction data as the system's foundation. In this situation, the association method, also known as MBA, is the most appropriate approach. This technique operates by examining the buying behaviours of multiple customers at the same time. The FP-Growth algorithm is their choice since it is quicker and more effective. A system that can quantify the value of product associations in transactional data was the research output.

By considering the repeated patterns (items) in the frequent itemsets, we were able to create an algorithm for the FP-Growth technique. By effectively grouping the overlapping items, this algorithm can reduce the space and time complexity in the case and apply the strategies to the data in this format. Implementing FP-Growth using the tries data structure has thus increased the performance of association rule mining undoubtedly (Goel et al., 2017).

According to Linwei et al., (2023), the performance of the optimized Apriori algorithm, the conventional Apriori, and the FP-Growth algorithms were compared in three trials using monitoring data from the Baishuihe landslide. These experimental findings demonstrate that the optimized algorithm required a calculation time that was only 1/432 of the traditional Apriori algorithm and 1/80 of the FP-Growth algorithm when the same strong association rules with high factor dimensions and levels that geological engineers and scientists care about were obtained. This indicates that the suggested optimized algorithm has excellent potential for use in the analysis of enormous amounts of high-dimensional data for landslide dangers.

The Apriori algorithm and the FP-Growth algorithm were the main algorithms (Khedkar et al., 2021). The focus of this work is to examine the effectiveness of these two algorithms in terms of database size, time complexity, and space complexity. The Apriori technique required higher time complexity, but FP-Growth required more space complexity, according to their analysis of these two algorithms. In contrast to the FP-Growth Algorithm, which employs trees repeatedly to add new types of transactions to reduce time complexity, the Apriori Algorithm can be employed when there are no time limitations but little space available. They concluded that the Apriori algorithm has higher efficiency and can be used to work on larger databases whereas the FP-Growth algorithm is suitable for smaller databases to avoid space complexity.

According to Aldino et al., (2021), RapidMiner was used to process transaction data and execute association rules mining, also known as MBA, by contrasting the FP-Growth and Apriori algorithms. According to the findings of this investigation, which used data from 1641 transaction rows, the FP-Growth requires 6 seconds to develop 19 rules and creates a combination of 3 itemsets with a rule strength of 112.66% and an accuracy of 217%. When compared to this, the Apriori 30-second algorithm generates 6 rules and combines 3 items with a rule strength of 52.47% and an accuracy of 46%. FP-Growth algorithm is superior to the Apriori algorithm, according to the results of the algorithm comparison.

A study by Ünvan, (2021) was carried out to do an association rules-based MBA. The Association Rules algorithms Apriori and FP-Growth were tested in that sequence. Due to the categorical nature of the data collection, the Apriori algorithm produced no results. As a result, the top 10 rules were determined using the FP-Growth algorithm and the conviction value. In such research, Joshi et al., (2022), a method for decreasing the dataset items with the best-selling items has been developed to prevent this massive computation. Associative rule generation for the FP-Growth and the Apriori algorithms, together with common item combinations, begin with different percentages of trending products, ranging from 30 to 55%. According to the results, it is possible to recommend the same visit itemset and affiliation rules in a shorter amount of time than the yields obtained by computing everything separately. Additionally, it is discovered through time comparison that Distributed FP-Growth computation requires less time than the Apriori calculation.

Past studies revealed that the most well-known association algorithms the Apriori and the FP-growth, are frequently employed to assist in identifying relationships in transactional data. Comparing the Apriori and the FP-tree algorithms has been used in some studies to assess the differences between them. This research provided details on how these algorithms work. Because these algorithms could determine which things were frequently sold together, the analysis's findings demonstrated that they were advantageous for business. Therefore, to further support our claims, we investigated these two algorithms by evaluating their performance in terms of execution time, the number of itemsets produced, and the correspondence between the results of each algorithm and the real transaction records (Christian et al., 2021).

3. Methodology

This paper aims to conduct an exploratory analysis utilizing association rule mining algorithms within the retail industry. The objective is to determine the champion algorithm for identifying frequent patterns in customer purchases.

3.1 Dataset Description

The dataset was obtained from the Kaggle website (https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis). It comprises transactional records covering all transactions that occurred over a defined period.

	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
...	...	...	...	...	...	...	...
522059	581587	PACK OF 20 SPACEBOY NAPKINS	12	2011-12-09 12:50:00	0.85	12680.0	France
522060	581587	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.1	12680.0	France
522061	581587	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	France
522062	581587	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	France
522063	581587	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	France

Figure 1. The dataset consists of 522,064 customer data

Figure 1 shows the sample of dataset which contains information about 552,064 transactions from 37 countries. The dataset contains 7 features derived from customer data and transactions that the customer processed. Table 1 provides more details on these features.

Table 1. Dataset Description

Feature Name	Feature Description
BillNo	a 6-digit number assigned to each transaction. Nominal.
Itemname	Product name. Nominal.
Quantity	The quantities of each product per transaction. Numeric.
Date	The day and time when each transaction was generated. Numeric
Price	Product price. Numeric.
CustomerID	a 5-digit number assigned to each customer. Nominal.
Country	Name of the country where each customer resides. Nominal.

3.2 Data Preprocessing

Data preprocessing is to make the dataset applicable to implement a model from it using machine learning algorithms. Data pre-processing for this work include handling missing values and encoding categorical variables. Figure 2 shows the steps in cleaning up the row dataset. After cleaning up, one hot encoding function is applied to convert categorical variables into a numerical representation.

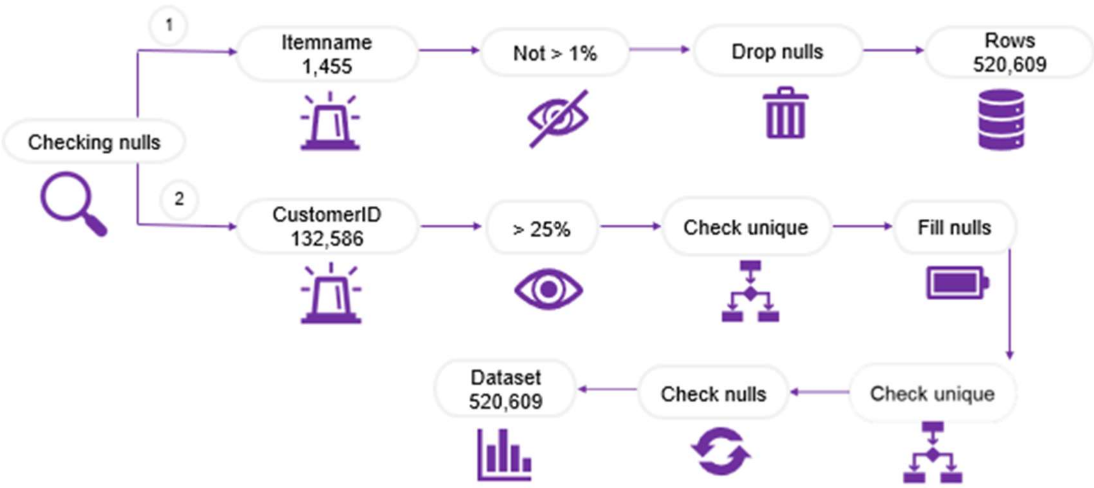

Figure 2. Handling missing values steps

The dataset has two columns with null values, Itemname and CustomerID as it represents in Figure 2. In Itemname column, the amount of the missing value is 1,455 which is less than 1% of the dataset, then this amount is dropped and the remaining amount will be 520,609. In the CustomerID column, the missing value amount is 132,586, since this subset accounts for 25% of the dataset, simply dropping these entries is not a viable option. To address this using Python, each missing CustomerID is imputed with the nearest valid unique ID based on the sorted order of bill numbers. To validate the correctness of the imputation, the unique values in all datasets, both before and after filling in the missing CustomerID values, are compared, and unique values remain consistent before and after the imputation.

3.3 Exploratory Data Analysis

This is a crucial step in the data science process because it helps to understand the characteristics of the data, identify any potential problems or issues, and determine the best approach for modeling and analyzing the data.

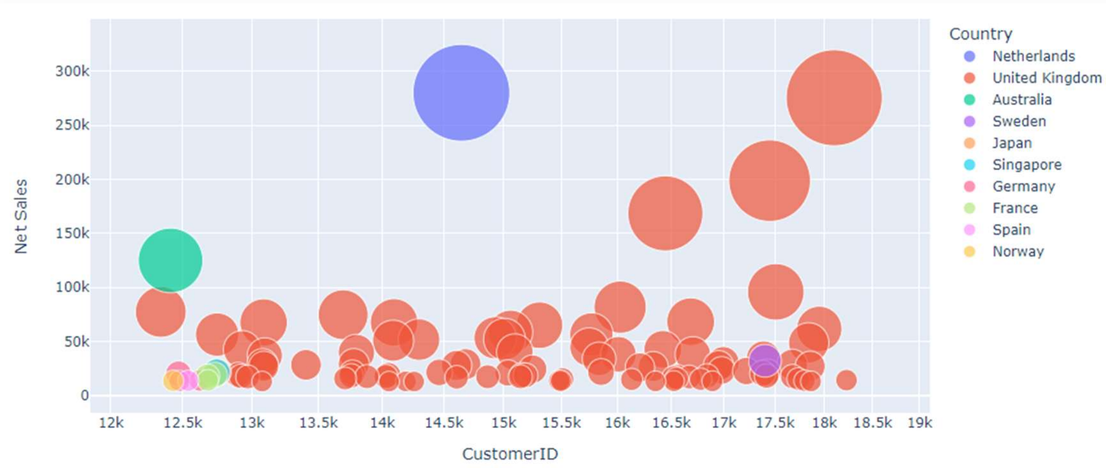

Figure 3. Top 100 Shoppers and Their Countries

Figure 3 shows the Top 100 shoppers and their countries by amounts of net sales accessed by the CustomerID. United Kingdom takes the long part as it appears by most customers and most sales for products which makes it the best market to invest in more branches.

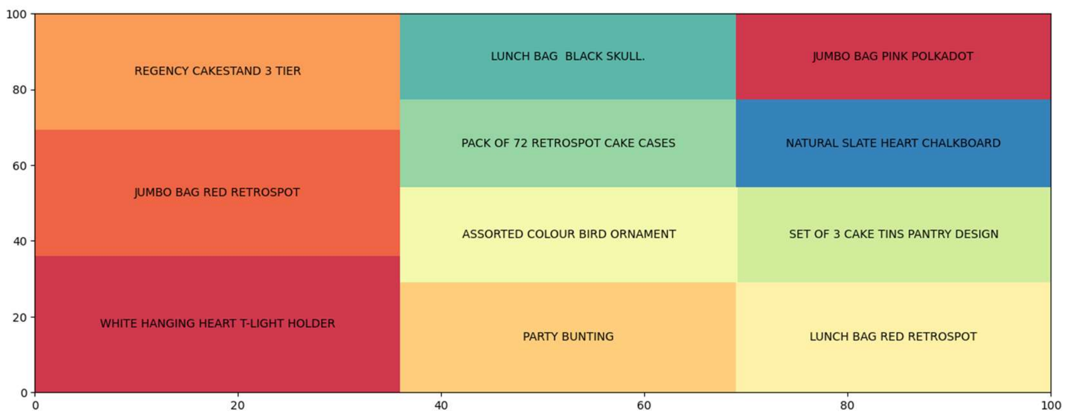

Figure 4. Treemap of Top-Sold Products

The treemap above in Figure 4 provides a visual representation of the top 10 products based on sales or quantities in the dataset. Each rectangle corresponds to a specific product, with the size of the rectangle indicating the relative sales volume or quantity. Darker shades represent higher sales, while lighter shades represent lower sales. The top 10 products listed above, such as 'WHITE HANGING HEART T-LIGHT HOLDER' and 'JUMBO BAG RED RETROSPOT', are contributing significantly to the overall sales.

Table 2 presents the top 10 products based on their frequency in the dataset. These products have the highest occurrence in the transaction records. The frequency column indicates how often each product appears, with higher frequencies indicating greater popularity. This aligns with the previous treemap visualization in Figure 4, where it is visually represented the sales or quantities of these top products. Understanding the frequency of these items is crucial for recognizing customer preferences and optimizing business strategies.

Table 2. The top 10 items frequency

Itemname	Frequency
WHITE HANGING HEART T-LIGHT HOLDER	2269
JUMBO BAG RED RETROSPOT	2087
REGENCY CAKESTAND 3 TIER	1930
PARTY BUNTING	1677
LUNCH BAG RED RETROSPOT	1570
ASSORTED COLOUR BIRD ORNAMENT	1465
SET OF 3 CAKE TINS PANTRY DESIGN	1360
PACK OF 72 RETROSPOT CAKE CASES	1328
LUNCH BAG BLACK SKULL.	1315
NATURAL SLATE HEART CHALKBOARD	1246

The line plot above in Figure 5, it is visualizing the period of the most frequently purchased product which is 'WHITE HANGING HEART T-LIGHT HOLDER' throughout the year 2011. The Net Sales, representing the monetary value of transactions, exhibit interesting patterns and fluctuations. These observations highlight the dynamic nature of sales for the most frequently purchased product, with fluctuations influenced by various factors throughout the year.

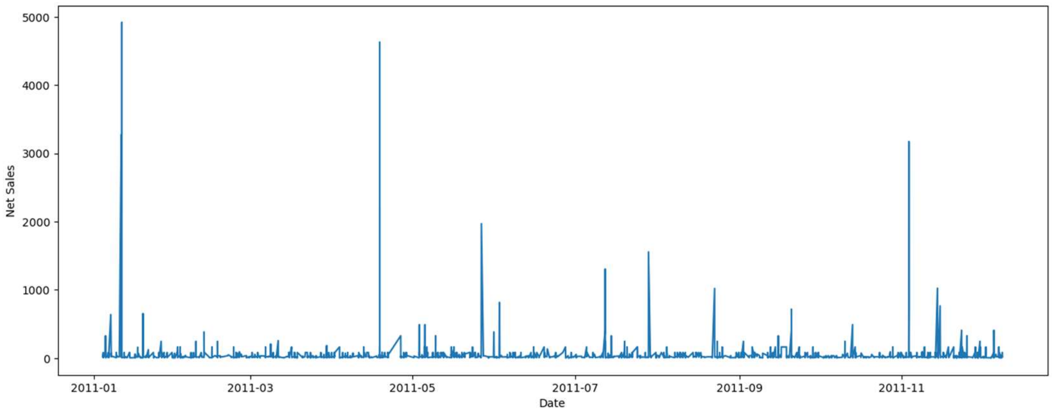

Figure 5. The time period of the frequently purchased product

3.4 Data Modelling

Modelling was implemented by Python to choose the champion model to identify frequent patterns in customer purchases. The Apriori algorithm and FP-Growth were employed to generate frequent itemsets. These algorithms were selected based on a review of past studies, where they demonstrated superior performance and efficacy in similar contexts, thereby ensuring robust and effective association rule mining for the retail dataset. Figure 6 displays the Apriori algorithm's flowchart. It looks for things that commonly show up together.

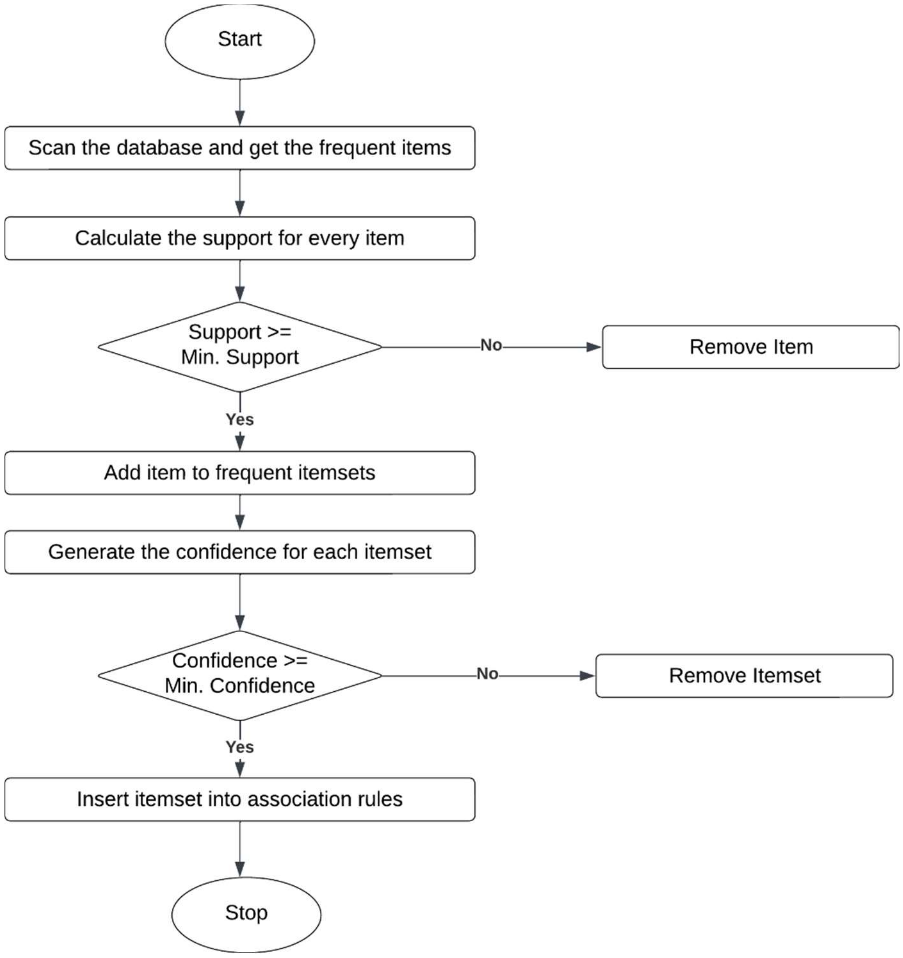

Figure 6. Apriori algorithm flowchart (Christian et al., 2021)

The criteria support, confidence, and lift are used to develop association rules to identify the most meaningful relationship. The support shows the degree of dominance an item or itemset has over the full set of data. Equation (1), where N is the quantity of data, is used to determine the support if a pair of items, X and Y, commonly appear together. Confidence, meanwhile, denotes the importance of certainty or the degree of correlation between the components of the association rule. Equation (2) is used to determine the percentage of Y in X-containing data. Additionally, equation (3) can be used to calculate a lift ratio, which represents how independent X and Y must be for the observed support to match the predicted one (Christian et al., 2021).

$$Support = \frac{Frequency\ (Y \times Y)}{N} \tag{1}$$

$$Confidence = \frac{Frequency\ (Y\ \times\ Y)}{Frequency(X)} \tag{2}$$

$$Lift = \frac{Support}{Support(X)\ .\ Support(Y)} \tag{3}$$

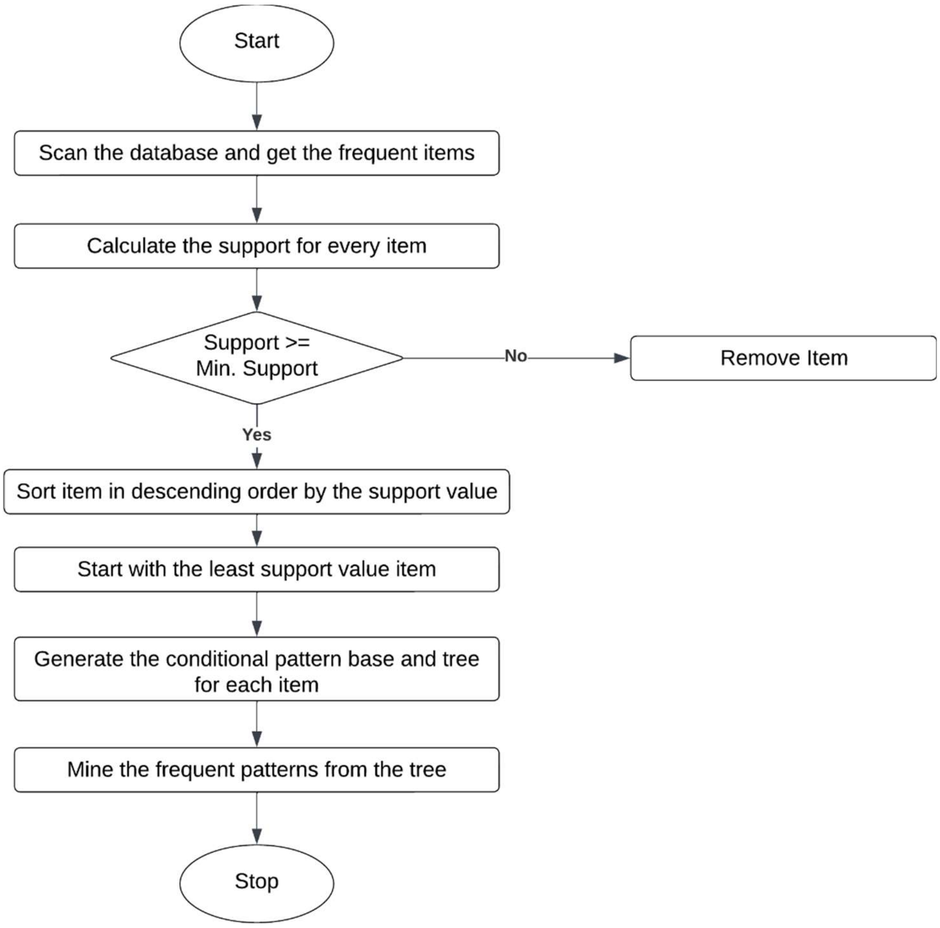

Figure 7. FP-Growth flowchart (Christian et al., 2021)

The FP-Growth algorithm, illustrated in Figure 7, organizes dataset transactions in a tree structure, linking each item to transactions containing it. In contrast to Apriori, FP-Growth efficiently discovers itemsets with minimal support. The "min_support" parameter sets the threshold for an item to be considered frequent based on its occurrence fraction. The "use_colnames" parameter, when True, uses dataset column names as item names. This study used a code to demonstrate this by finding frequent itemsets with a minimum support of 0.02, utilizing dataset column names for item names. The results are stored in a data frame for further analysis in the next chapter.

4. Results and Discussions

Table 3 shows that the Apriori algorithm took a significantly longer time to run than the FP-Growth algorithm on the given dataset. This suggests the FP-Growth algorithm may be more efficient in runtime performance compared to the Apriori algorithm on this dataset. In addition to the time taken with each algorithm, which gives an enormous difference, the number of frequent itemsets discovered using these two different algorithms isn't different as shown in Table 1.

Both the Apriori and the FP-growth algorithms returned 358 frequent item sets from the same dataset. This was done by setting the minimum support to be 0.02, which means it is returning items or itemsets that are included in at least 2% of the entire baskets.

Table 3. Comparison between the output of Apriori and FP-Growth

Model	**Computational Time**	**Frequent Itemsets**
Apriori	68.18 Sec.	358
FP-Growth	3.52 Sec.	358

Association rules using the Apriori Algorithm and FP-Growth Algorithm are giving the same values using the minimum support of 0.02 as in Figure 8 and Figure 9.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.036421	0.048248	0.029939	0.822011	17.037124	0.028181	5.347247
1	(ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY...	(GREEN REGENCY TEACUP AND SAUCER)	0.028207	0.048248	0.025485	0.903509	18.726262	0.024124	9.863609
2	(GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY...	(ROSES REGENCY TEACUP AND SAUCER)	0.029939	0.050129	0.025485	0.851240	16.981097	0.023984	6.385246
3	(JUMBO STORAGE BAG SUKI, JUMBO BAG PINK POLKADOT)	(JUMBO BAG RED RETROSPOT)	0.025435	0.102138	0.020388	0.801556	7.847797	0.017790	4.524521

Figure 8. Association rules using Apriori Algorithm

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(JUMBO STORAGE BAG SUKI, JUMBO BAG PINK POLKADOT)	(JUMBO BAG RED RETROSPOT)	0.025435	0.102138	0.020388	0.801556	7.847797	0.017790	4.524521
1	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.036421	0.048248	0.029939	0.822011	17.037124	0.028181	5.347247
2	(ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY...	(GREEN REGENCY TEACUP AND SAUCER)	0.028207	0.048248	0.025485	0.903509	18.726262	0.024124	9.863609
3	(GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY...	(ROSES REGENCY TEACUP AND SAUCER)	0.029939	0.050129	0.025485	0.851240	16.981097	0.023984	6.385246

Figure 9. Association rules using FP-Growth algorithm

The parallel coordinates plot will allow us to visualize whether a relationship exists between an antecedent and a consequent. We can think of it as a directed network diagram. The plotshows connections between 2 objects that are related and indicates the direction of the relationship. Figure 6 displays a parallel coordinates plot showing the 'antecedent' and 'consequent' values for each of the rules in the input data frame. Each line in the plot represents a different rule,and the values for the 'antecedent' and 'consequent' variables are plotted along the x-axis. The y-axis displays the names of the variables, and a legend showing the different rules is also displayed.

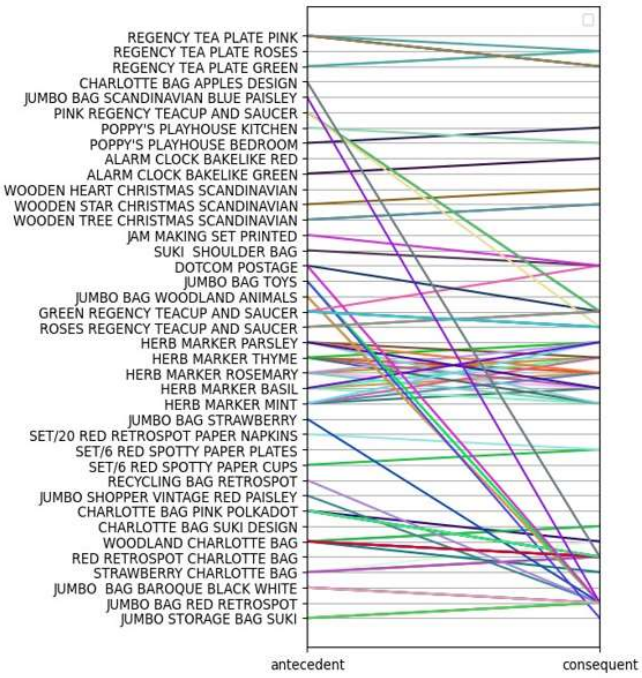

Figure 7. Parallel Coordinates

5. Conclusion

In conclusion, this study used a retail store dataset which consists of 522,064 rows of data and 7 variables. Before beginning the data modelling phase, data preprocessing is performed to delete null values which are not significant, handle missing values, and encode categorical variables as machine learning algorithms deal with numeric data only. As a result, for this cleaning, there are no nulls in the dataset, and it is encoded to be used in the model.

For modeling, two algorithms are proposed based on the past work reviewed, it begins with comparing the two models in terms of time. It showed that the FP-Growth algorithm is faster than the Apriori algorithm, the first one takes 3.525 seconds and the second one takes 68.189 seconds. Then, another comparison between several frequent itemsets with each algorithm was implemented. Both resulted in the same number, 358 frequent itemsets with a minimum support of 0.02. Finally, another comparison in terms of association rules values. After implementing the code by Python with the same minimum support = 0.02, each algorithm resulted in the same values of antecedents, consequents, antecedent support, consequent support, support, confidence, lift, leverage, and conviction. So, the FP-Growth is the champion model here as it is faster than the Apriori algorithm in terms of extracting frequent item sets.

One of the limitations faced in this project is the dataset, having many null values at the ID column, the contribution of null values is 25% of all the dataset. All of them cannot be deleted, as it will be not significant. Every missed value was filled of the customer ID with the last valid ID, supposing that this customer purchased this item also. Sometimes, this will work effectively, but other times it may result in a bias in the results. If the dataset includes many missing values, it may fail to conclude more reliable results.

Acknowledgement

Special thanks were extended to the School of Management, Universiti Sains Malaysia and Digital Egypt Builders Initiative.

Author Contribution

All authors have involved themselves equally in every section when writing this paper.

Conflict of Interest

The author has no conflicts of interest to declare.

References

Aldino, A. A., Pratiwi, E. D., Setiawansyah, Sintaro, S., & Putra, A. D. (2021). Comparison of Market Basket Analysis to Determine Consumer Purchasing Patterns Using Fp-Growth and Apriori Algorithm. *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering, ICOMITEE 2021*, 29–34. https://doi.org/10.1109/ICOMITEE53461.2021.9650317

Christian, M. A., Nathanael, N., Mauliani, A., Indrawati, A., Manik, L. P., & Akbar, Z. (2021). Real Market Basket Analysis using Apriori and Frequent Pattern Tree Algorithm. *ACM International Conference Proceeding Series*, 161–165. https://doi.org/10.1145/3489088.3489133

Dubey, S. K., Mittal, S., Chattani, S., & Shukla, V. K. (2021). Comparative Analysis of Market Basket Analysis through Data Mining Techniques. *Proceedings of 2nd IEEE International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2021*, 239–243. https://doi.org/10.1109/ICCIKE51210.2021.9410737

Goel, M., & Goel, K. (2017). FP-growth implementation using tries for association rule mining. *Advances in Intelligent Systems and Computing*, *547*, 21–29. https://doi.org/10.1007/978-981-10-3325-4_3

Hidayat, A. A., Rahman, A., Wangi, R. M., Abidin, R. J., Fuadi, R. S., & Budiawan, W. (2019). Implementation and comparison analysis of apriori and fp-growth algorithm performance to determine market basket analysis in Breiliant shop. *Journal of Physics: Conference Series*, *1402*(7). https://doi.org/10.1088/1742-6596/1402/7/077031

Izang, A. A., Goga, N., Kuyoro, S. O., Alao, O. D., Omotunde, A. A., & Adio, A. K. (2019). Scalable Data Analytics Market Basket Model for Transactional Data Streams. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 10, Issue 10). www.ijacsa.thesai.org

Jirapatsil, P., & Phumchusri, N. (2022). Market Basket Analysis for Fresh Products location improvement: A case study of E-Commerce Business Warehouse. *ACM International Conference Proceeding Series*, 23–28. https://doi.org/10.1145/3535782.3535786

Joshi, K., Rathod, D., Parekh, M., & Dave, J. (2022). Customer Behavioral Market Basket Approach for Most Frequent Pattern Identification Using Data Mining Techniques. *Smart Innovation, Systems and Technologies*, *248*, 663–670. https://doi.org/10.1007/978-981-16-4177-0_66

Khedkar, S. S., & Kumari, S. (2021). Market Basket Analysis using A-Priori Algorithm and FP-Tree Algorithm. *Proceedings - 2021 1st IEEE International Conference on Artificial Intelligence and Machine Vision, AIMV 2021*. https://doi.org/10.1109/AIMV53313.2021.9670981

Kurnia, Y., Isharianto, Y., Giap, Y. C., Hermawan, A., & Riki. (2019). Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm. *Journal of Physics: Conference Series*, *1175*(1). https://doi.org/10.1088/1742-6596/1175/1/012047

Linwei, L., Yiping, W., Yepiao, H., Bo, L., Fasheng, M., & Ziqiang, D. (2023). Optimized Apriori algorithm for deformation response analysis of landslide hazards. *Computers & Geosciences*, *170*, 105261. https://doi.org/10.1016/j.cageo.2022.105261

Listiawan, T., & Hudha, M. N. (2021). Design and implementation software for mining association rules (market basket analysis) to design product layout desicions. *Journal of Physics: Conference Series*, *1869*(1). https://doi.org/10.1088/1742-6596/1869/1/012121

Mechery, S., & Preethi, N. (2022). An Intelligent Recommendation System Using Market Segmentation. *Smart Innovation, Systems and Technologies*, *303 SIST*, 468–476. https://doi.org/10.1007/978-981-19-2719-5_44

Pradana, M. R., Syafrullah, M., Irawan, H., Chandra, J. C., & Solichin, A. (2022). Market Basket Analysis Using FP-Growth Algorithm on Retail Sales Data. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, *2022-October*, 86–89. https://doi.org/10.23919/EECSI56542.2022.9946478

Qisman, M., Rosadi, R., & Abdullah, A. S. (2021). Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores). *Journal of Physics: Conference Series*, *1722*(1). https://doi.org/10.1088/1742-6596/1722/1/012020

Robert C. Blattberg, B.-D. K. and S. A. N. (n.d.). *Blattberg, R. C., Kim, B.-D., & Neslin, S. A. (2008). Database marketing. International Series in Quantitative Marketing. https://doi.org/10.1007/978-0-387-72579-6*

Samboteng, L., Rulinawaty, Kasmad, M. R., Basit, M., & Rahim, R. (2022). MARKET BASKET ANALYSIS OF ADMINISTRATIVE PATTERNS DATA OF CONSUMER PURCHASES USING DATA MINING TECHNOLOGY. *Journal of Applied Engineering Science*, *20*(2), 339–345. https://doi.org/10.5937/jaes0-32019

Shahidan, S., Ooi, S. P. & Khaw, K. W. (2023). Prediction of employee promotion using hybrid sampling method with machine learning architecture. *Malaysian Journal of Computing*, 8(1), 1264-1286. https://doi.org/10.24191/mjoc.v8i

Ünvan, Y. A. (2021). Market basket analysis with association rules. *Communications in Statistics - Theory and Methods*, *50*(7), 1615–1628. https://doi.org/10.1080/03610926.2020.1716255