Journal of Engineering Science and Technology Review 16 (5) (2023) 108 - 122

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Performance Analysis of Association Rule Mining Algorithms: Evidence from the Retailing Industry

Bijayini Mohanty[1,*], Murchhana Tripathy[2] and Santilata Champati[3]

[1]*Centre for Data Science, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar – 751030, Odisha, India*
ORCID- 0009-0009-9313-0528, Email:)
[2]*Information Systems & Analytics, T A Pai Management Institute, Manipal Academy of Higher Education, Manipal-576104, Karnataka, India*
[3]*Department of Mathematics, FET, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar – 751030, Odisha, India*

Received 24 April 2023; Accepted 8 September 2023

Abstract

The Apriori and FP-growth algorithms have gained widespread popularity in various business applications. In the retailing industry they are widely used for market-basket data analysis and frequent pattern mining to gain valuable insights into customer purchasing behaviour. In this study, we conducted a comprehensive analysis of these two prominent association rules mining algorithms, utilizing six benchmark datasets from the UCI machine learning repository. Our investigation involved a thorough comparison of the execution time and the number of rules generated by both algorithms. Execution time is measured once by varying the support levels and next by varying the number of transactions and the support levels. Number of rules generated is estimated by varying the support levels of the rules. Through our rigorous experimentation, we derived insightful inferences that elucidated the utility of association rule mining in the retail industry. Moreover, we employed the Big-O method to compare the performance of the two algorithms and formulated a theorem that established FP-growth as Big-O of Apriori, substantiating the differences observed in their performance.

Keywords: f Association rule mining, Apriori algorithm, FP-growth algorithm, Retailing, Marketing, Big-O Method

1.Introduction

Data mining is the process of finding hidden information and useful patterns from large datasets. The obtained knowledge is used in practical applications to gain insight and improve the efficiency of the process involved. Several data mining techniques are used by business applications and to solve real-world problems. Depending on the nature of the dataset, often a technique is chosen. For example, if a dataset is labelled, then supervised learning is chosen and either classification or prediction is done. If the dataset has no labels, then unsupervised learning techniques are considered and clustering can be done. Association rule mining is an unsupervised learning technique because it works on an unlabeled dataset. It is used to find rules from a dataset typically containing transaction data. The rules help in discovering the interesting relations among the variables, and frequent patterns[1,2]. A transaction can be considered as a set of things occurring together at a particular instance. For example, the items occurring in one invoice during a customer's supermarket visit can be considered as a transaction. A dataset that contains several transactions can be referred to as a transaction dataset [3]. A transaction dataset typically contains transactions of the same type. For example, a retail transaction dataset contains transactions about the purchased items by the customers. In an online content provider like Netflix, a transaction consists of the movies and shows watched by a customer. Association rule mining has the ability to detect the association or relationship in a transaction dataset. This makes it an ideal method to be used for transactional data analysis or market basket analysis by retailers which in turn helps in understanding customer's buying behaviour, designing cross-selling techniques, customer segmentation and recommendation generation etc[4-8]. Retailing can be defined as the process of selling goods or services to the end consumer. It is the last stage of a supply chain [9] . As a result of the pandemic, the global retail sales fell by 2.9% in 2020, but soon in 2021 it bounced back with a growth rate of 9.7%.

Global retail sales are expected to hit around 31.7 trillion USD by 2025, up from the projected value of 27.34 trillion USD in 2022. With the discovery of new digital technologies such as web 2.0, social media and mobile computing, digitalization has influenced almost every business avenue and retailing is no exception [10,11]. Association rule mining has the potential to influence the last three factors positively and can be considered as one of the most influential data mining techniques in the retailing industry [12,13]. Analysis of the past sales records of a retail store through association rule mining show that customers' buying patterns are often the same and it is evident both in the case of an online store and a brick-and-mortar store[12]. The analysis also helps in formulating strategies for product placement and store layout in case of physical stores. Similarly, product pricing and bundling are done to attract more customers and to make the store profitable both for physical and on-line stores[13]. For example, in a supermarket, often milk, bread and eggs are bought together. There are two ways in which these associated products can be placed in a retail store. The first strategy is to place those in proximity so that it is easy for the customers to trace these products. The second strategy is to

*E-mail address: bijayinimohanty7@gmail.com
ISSN: 1791-2377
doi:10.25103/jestr.165.14

place the associated products far from each other so that a customer can indulge in some impulse purchasing while traversing from one product to the other. In both online and offline retailing, it is observed that certain products are sold in combo. For example, conditioners are combined with shampoos, different cleaning agents such as toilet cleaners, floor cleaner and Phenyl are sold in a combo. If you are buying a laptop from an electronic store, the store offers you to buy pen drives, mouse, card-readers, etc. at a discounted price. If you buy a branded cell phone, then the store offers you to buy a phone cover, a power bank or a memory card. This kind of product bundling helps in increasing sales, achieving the targeted sales and also gives some savings to the customers in the purchase they make. Furthermore, it helps the store managers to decide on the store layout or retail display which is an important retail decision to preserve sales and profits. It is also important because it affects the number of people visiting the store, the buying environment of the store and the purchasing behaviour of the customers and acts as a determinant of store reliability[13]. Association analysis is performed using two popular algorithms, the Apriori algorithm and the FP-growth algorithm. In this work, we have analysed the performance of both the algorithms on six retail datasets taken from the UCI machine learning repository and have compared their performance using execution time or run time of the algorithms and the number of rules generated by the algorithms. Execution time is calculated twice, once by varying the support level and next by varying the support level and number of transactions. Number of rules generated by the algorithms are calculated by varying the number of transactions and the support levels of the rules.

Through this research we aim to address the following three key questions.

1.To understand the behaviour of the Apriori and FP-growth algorithms on retail datasets in terms of key performance measures such as run time and time-complexity through Big-O analysis.
2. To understand the impact of varying support and confidence in identifying association rules
3. To identify a suitable algorithm for analysis of retail datasets which enables better decision making.

The rest of the article is organized as follows. Section two explains about the background of the work. Literature review is given in section three. Section four explains about the datasets used for the experiment. Results and discussions are given in section five. Section six covers analysis using Big-O method and followed by that section-7 enumerates the potential benefits of using the association rule mining in the retailing industry. The article has been concluded in section eight.

2. Background

This section briefly discusses about association rule mining and some of the important concepts related to it. In addition to that, the Apriori and the FP-growth algorithms are explained to set the context of the experiment.

2.1 Association Rule Mining

Association rule mining is a method of finding association and frequent patterns among the data objects in a dataset. The output of applying this method is implication rules of the form $A \rightarrow B$, where A and B are item sets and $A \cap B = \emptyset$. Association rule mining uses two rule interestingness measures such as 'support' and 'confidence' to decide the acceptability and strength of the rule. The support of a rule is defined as the percentage of transactions in which the rule is found [14-17]. The confidence of a rule is defined as the ratio between the number of transactions in which the rule is found and the number of transactions in which only the antecedent of the rule is found, expressed in percentage.

$$Support\ in\ Percentage = \frac{Number\ of\ observationscontaining\ the\ rule}{Total\ number\ of\ observations} \times 100 \tag{1}$$

$$\text{Confidence in Percentage} = \frac{Number\ of\ observations\ containing\ the\ rule}{Number\ of\ observations\ containing\ the\ antecedent\ of\ the\ rule} \times 100 \tag{2}$$

Support helps to remove infrequent rules and confidence helps to measure the reliability of the rule. Given a set of transactions, association rule mining finds rules which have the support and confidence values greater than the user defined threshold values.

2.1 The Apriori Algorithm

The Apriori algorithm was first proposed by Agrawal and Srikant in 1994 [18]. It works according to the apriori principle which states that all the item sets of a frequent item set are also frequent. In simple terms, if A, B, C is a frequent item set then, A, B, C ,A, B ,A, C ,B, C are also frequent. In contrast to this, if any of these subsets are infrequent then their subsequent supersets are also infrequent. For example, if A is infrequent then its supersets A, B ,A, C and A, B, C are also infrequent. It is mainly used for market basket analysis and helps to find those products that can be bought together. Due to large no of candidate generation, it requires large amount of storage. Also, the algorithm scans the raw data several times and hence it is a time-consuming process.

2.2 The FP-Growth Algorithm

The FP-growth algorithm was formulated by Agarwal in 1994 [18]. The algorithm represents the dataset as a tree like structure, known as frequent pattern tree or FP tree. It helps to find the frequent pattern without generating candidate item sets like the Apriori algorithm [19]. FP tree represents an item of the itemset in each node of the constructed FP tree. The root node always represents null.

2.3 Numerical Example of Association Rule formulation

In this section, we are presenting an example of association rule mining, where we show the discovery of frequent patterns and formulation of the final rules.

We consider the scenario of a supermarket where grocery items, fruits, vegetables, and daily need items are sold. For our example purpose, we consider five transactions as given in figure 1. The items considered for the transactions are Bread (B), Jam(J), Milk (M), Butter (Bu), Banana (Ba), Orange (O), Sweet Lime (SL), Cheese (C), Egg (E), Cornflakes (CO). Minimum support is 60% and minimum confidence is 80%.

Transaction id (Tid)	Itemset
1	B, J, M, CO, Bu, Bo
2	O, J, M, Co, Bu, Ba
3	B, SL, Co, Bu
4	B, C, E, Co, Ba
5	E, J, J, co, Bu

Fig. 1. Transaction dataset of a supermarket

Step-1: Find the support count of each item set and keep it in the candidate set table C_1.

Table 1. Candidate Set C_1

Item Set	**Support Count**
B	3
J	4
M	2
Bu	4
Ba	3
O	1
SL	1
C	1
E	2
Co	5

Since support is 60%, remove all those items for which the support count is less than 3 (60% of 5 is 3) and form a new table L_1.

Table 2. L_1

Item Set	**Support Count**
B	3
J	4
Bu	4
Ba	3
Co	5

Step-2: Find candidate item set C_2 from L_1

Table 3. Candidate Set C_2

Item Set	**Support Count**
B, J	1
B, Bu	2
B, Ba	2
B, Co	3
J, Bu	3
J, Ba	2
J, Co	3
Bu, Ba	2
Bu, Co	4
Ba, Co	3

Form L_2 from C_2 by removing the item sets for which the support count is less than 3.

Table 4. L_2

Item Set	**Support Count**
B, Co	3
J, Bu	3
J, Co	3
Bu, Co	4
Ba, Co	3

Step-3 Find candidate item set C_3 from L_2

Table 5. Candidate Set C_3

Item Set	**Support Count**
B, Co, J	0
B, Co, Bu	2
B, Co, Ba	2
Co, J, Bu	3
Co, J, Ba	2
J, Bu, Ba	1

Form L_3 from C_3 by removing the item sets for which the support count is less than 3

Table 6. L_3

Item Set	**Support Count**
Co, J, Bu	3

From table 6 we find only one frequent item-set and using it, the association rules can be formed. The association rules are presented in table 7.

Table 7. Association rules

Association Rules	**Support**	**Confidence**
Co ∧ J ⇒ Bu	**3**	**100%**
Co ∧ Bu ⇒ J	3	75%
J∧ Bu ⇒ Co	**3**	**100%**
Bu⇒ Co ∧ J	3	75%
J⇒ Co∧ Bu	3	75%
Co⇒ J ∧ Bu	3	60%

The final association rules considered are the ones which have a confidence greater than 80%. In this case the highlighted rules in table 7 only satisfy the criteria and hence only two association rules are found from the data considered.

2.4 Association Rule Mining example using the Orange Data Mining Tool

In this section, we show the extraction of frequent patterns and formulation of association rules by using the orange data mining tool. By using the inbuilt widgets and the food mart dataset, we find the frequent patterns and association rules presented in figure 3 and 4 respectively. Figure 1 shows the workflow for finding the frequent patterns and association rules and figure-2 show the dataset in a data table. The tool is easy to use, gives great data visualization and data mining capabilities and hence we used it to quickly conduct association rule mining. The tool can handle very large datasets as well and hence can be used by managers without diving deep into the technical details of association rule mining.

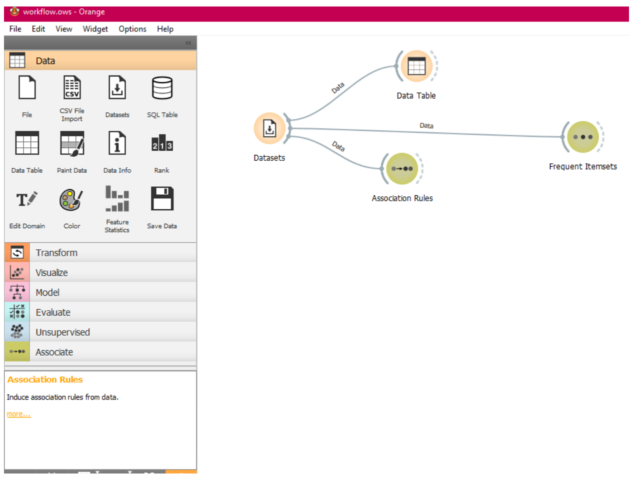

Fig. 1. Association Mining work flow

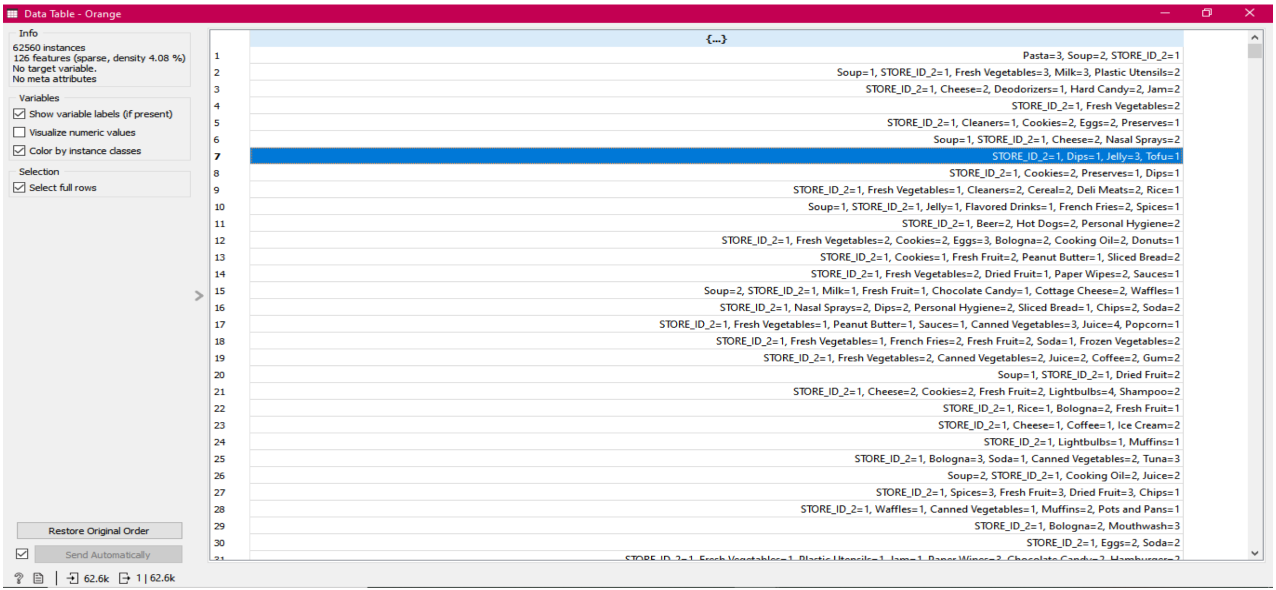

Fig. 2. Data table for Orange Food mart dataset

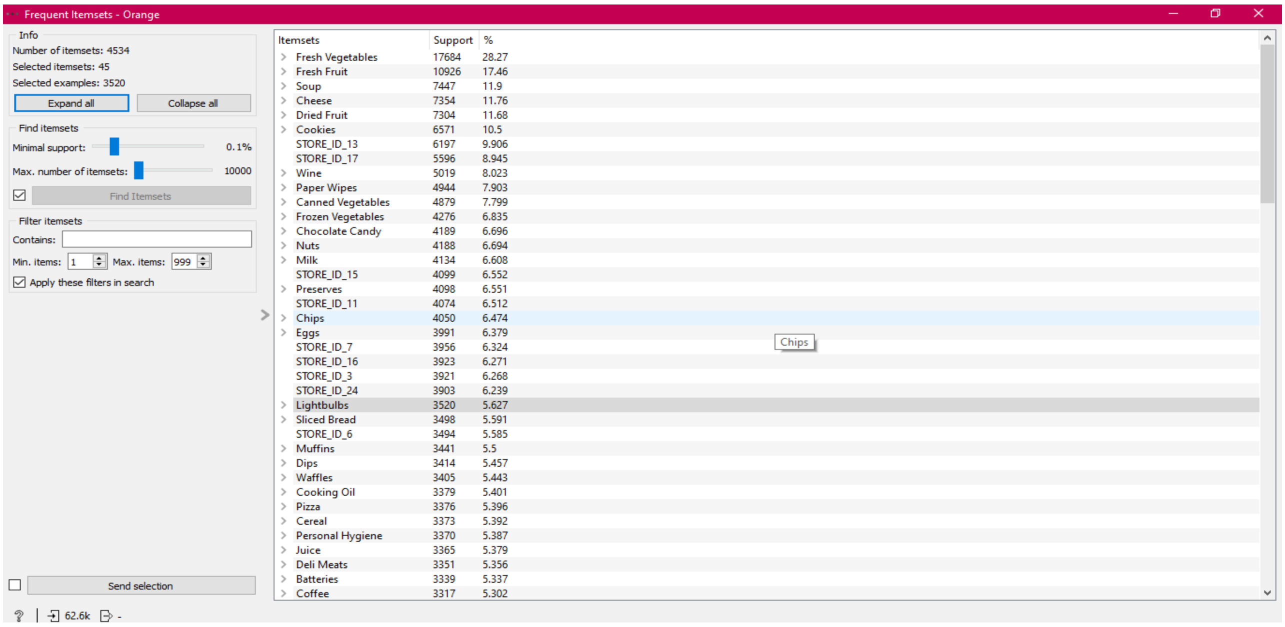

Fig. 3. Frequent Item sets

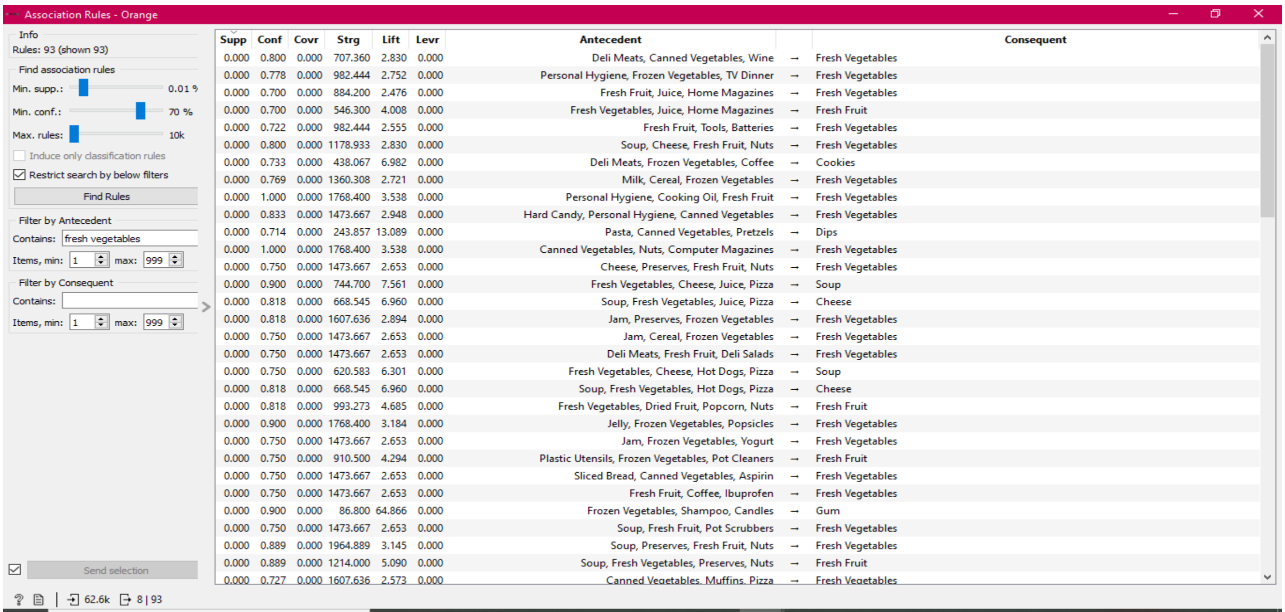

Fig. 4. Association Rules

3. Literature Review

This section presents some of the important contributions in the concerned field of research.
Arboleda et.al[20] proposed a novel technique which is very effective for stoke planning and future layouts. Hermina et.al [2] Investigate and Compare the respective run times of the apriori and FP growth algorithms using the neighbourhood grocery store dataset. Kulkarni et.al[21] Discuss how retail company uses market basket analysis technique to increase sales of their goods by analysing customers buying patten. Pradana et.al [22] tested 571 transections of a retail store to analysing the buying behaviour by using frequent-pattern growth algorithm. Aldino et.al [23] tested the apriori algorithm and f-p growth algorithm in Rapid-Miner software for the business databases. Dubey et.al [14]provides a comparative analysis between Association Rule Mining (ARM) and Collaborative filtering (CF) by understanding the frequent activities of buyers. Patwary et.al [24]determined the relationship between customers' purchase and their goods. It will be very helpful for supermarket managers to maintain CRM (customer relationship management). Kiani et.al [25]applied association rule mining on the Iran supermarket dataset and found that the number of generated frequent item sets increased significantly when the product exhibition time periods were taken into consideration. Pillai and Jolhe[1] provided some valuable insights based on a supermarket case study for better cross-selling, and up-selling of goods. Also, discuss integration tasks for newly launched products. Hossain et.al [26] proposed an approach by using the apriori and f-p growth algorithms concepts to avoid the computation of large-scale data by reducing the items of the considered dataset with top-selling products. Ahlers et.al[27] analysed the web data of five local shopping platforms in Germany to get frequent buying patterns. Atlal et.al [16] conducted a market basket survey by using the Apriori and Eclat algorithms and also discussed their implementation process and necessity. Griva et.al [3] used a business analytics approach that mined customer visit segments from market basket sales data. The knowledge obtained helped in forming market campaigns and design the store layout. Vanaja and Belwal[28] proposed an aspect-level sentiment analysis by using the amazon customer review data and focused on each review to get valuable aspects. A novel methodology of re-mining was introduced by Demiriz et.al [29] to enrich the traditional data mining process. They explored all the factors behind both positive and negative association rules which also predicted the class level of the data. Kaur and Kang [13] provided an association rule mining algorithm called ARM-Predictor that helped to examine the customer behaviour pattern and helped to increase the sales and profit. McDowell et.al [12]examined empirical associations between website features and online conversion rates through their analysis. They concluded that certain website features were used to convert the e-commerce visitors into their buyers. Tewari et.al[30] proposed a model for a book recommendation system with the help of combined features of content-based filtering, collaborative filtering and association rule mining. Abdulsalam et.al [8]provided a framework to know how market basket analysis could help business intelligence through association rule mining by using the apriori algorithm. Agarwal and Ranjan [18]used association rule mining to find rules between item sets in a large dataset of customer transactions. Ahn [31]used neighbourhood-based association rule mining on synthetic datasets to solve the product assignment problem and used lift to measure the effectiveness of cross-selling products. Cil [32] proposed a framework combining association rule mining and multi-dimensional scaling to improve the supermarket store layout. Liao et al. [33] used the apriori algorithm on the data of a Taiwan-based retailing company Carrefour to investigate the issues of brand extension and product line development. To reduce the execution time of the Apriori algorithm, Al-Zawaidin et al.[7] used the classical Apriori algorithm along with the features of the items and weights of candidate item sets to generate frequent item sets and association rules. Anderson et.al [34] purposed a model to understand how retailers used business intelligence and data mining tools to improve customer relationship management in retailing. Chen et.al [35] proposed a novel shelf space management scheme with the help of association analysis. Chen et.al [36] used association rule mining to identify the behaviour pattern of customers and used the association rules to find the relationship between customer's profiles and the products purchased.

4. Datasets used for Experiment

Six datasets for the experiment are considered from the UCI machine learning repository (https://archive.ics.uci.edu/ml/index.php) such as Wholesale customers data (440×8), Online-retail-2 (541904×8), Online-retail (1067371×8), Click stream data for online shopping data set (1655474×14), Bank marketing (45211×17), Online shopper purchasing intention data set (45211×17). Conducting experiments with different size datasets enables identifying the limitation of the experiment and decision making in terms of developing the solution. It also helps to validate the performance of the algorithm. The objective is to compare the performance of the Apriori and the FP-Growth algorithms on the datasets. The basic details about the datasets are given in table 8. We have performed the experiments using Python 3.8.

Table 8. Datasets used in the Experiment

Name of the data set	Number of instances	Number of attributes	Number of web hits
Wholesale customers data	440	8	440101
Online-retail-2	541904	8	159717
Online-retail	1067371	8	733856
Click stream data for online shopping data set	165474	14	60635
Bank marketing	45211	17	1784516
Online shopper purchasing intention data set	12330	18	187619

5. Results and Discussion

We consider three different criteria to compare the performance of the algorithms. Criteria-1 is based on the execution time or run time at varying support level. and criteria-2 is on the number of rules generated at varying support level. Criteria-3 is based on the execution time at different number of transactions and at varying support levels.

Support levels of 0.2, 0.4, 0.5, 0.6, 0.8 and transaction limits of 300, 600, 900, 1200, 1500 are considered to run the experiments. The computing requirements of association rule mining can be very high, particularly for market basket datasets. Where the number of transactions is very high. We randomly selected the support level and gradually increase interval up to 0.2 to see the results. Varying the support level is very helpful for the real-world retail dataset to achieve appropriate result. The support level can be adjusted to achieve a balance between the accuracy of the rules and their usefulness. By adjusting the support value, we will get the desire able result with better accuracy. We have implemented the Apriori and FP growth algorithm on six benchmarked retail data set such as Wholesale customers data, Online-retail-2, Online-retail, click stream data for online shopping data set, Bank marketing, Online shopper purchasing intention data set in Python. The Python program was executed on a HP Notebook (14s-dr1008tu) equipped with an Intel Core i5 processor and 8GB of RAM with 512GB NVMe M.2 solid-state drive (SSD). The program was developed by using Python 3.8 software also utilizing some popular libraries such as NumPy, Pandas, matplotlib and scikit-learn. First, we have pre-processed our dataset by using scikit-learn for generating the associations rules. To check the better accuracy level, we have use different support values for the experiment.

Table 9 shows the execution time for both the algorithm with different support level and table 10 shows the number of rules generated with various minimum support levels. Subsequently, table 11 shows the execution time of the algorithms with varying number of transactions and different support level. The Wholesale Customer's dataset has many 'NA' values for the execution time because it has only 440 objects and for transaction values greater than 440, the transaction time field gets 'NA'. Figure 5 through 10 show the execution time for both the algorithms. X-axis represents the number of transactions and Y-axis represents the run time of the algorithms. In the figures, blue line is used to represent the Apriori algorithm and orange line is used to represent the FP growth algorithm. For all the datasets a similar pattern is observed and that is when the number of input transactions is less than 200, the run time for both the algorithms is same. As the number of transactions increase, Apriori takes more time than FP growth.

To visualize the results properly, we present the bar chart for the online-retail data set in figure-11, showing execution time of Apriori and FP growth with different support levels. X-axis represents different minimum support level and Y-axis represents the run time of the algorithms in miliseconds. Figure-11 clearly represents that the FP growth algorithm consumes less time than the apriori algorithm.

To visualize the results properly, we present a bar chart for the online-retail data set in figure-12, showing the number of rules generated by Apriori and FP growth with different support levels. X-axis represents different minimum support level and Y-axis represents the total number of transactions for the entire dataset. It is evident that for lower support levels, Apriori produces a greater number of rules than FP growth.

Algorithm 1. Apriori Algorithm

1. pip install pandas mlxtend
2. import pandas as pd
3. from mlxtend.frequent_patterns import apriori
4. from mlxtend.frequent_patterns import association_rules
5. def read_wholesale_dataset():
6. Adjust the path accordingly if the dataset is in a different location
7. dataset_path = "wholesale.csv"
8. def get_association_rules(dataset, min_support=0.1, min_confidence=0.5):
9. #Apply the Apriori algorithm to find frequent itemsets
10. frequent_itemsets = apriori(one_hot_encoded, min_support=min_support, use_colnames=True)
11. # Generate association rules
12. rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=min_confidence)
13. return rules
14. start_time = time.time()
15. dataset = read_wholesale_dataset()
16. # Adjust the values for min_support and min_confidence as needed
17. min_support = 0.05
18. min_confidence = 0.5
19. # Get the association rules
20. rules = get_association_rules (dataset, min_support, min_confidence)
21. # Calculate the number of association rules
22. num_association_rules = len(rules)
23. end_time = time.time()
24. runtime = end_time - start_time
25. print ("Number of Association Rules:", num_association_rules)
26. print ("Runtime (seconds):", runtime)

Algorithm 2. FP growth Algorithm

1. pip install pandas mlxtend
2. import pandas as pd
3. from mlxtend.frequent_patterns import fpgrwoth
4. from mlxtend.frequent_patterns import association_rules
5. def read_wholesale_dataset():
6. Adjust the path accordingly if the dataset is in a different location
7. dataset_path = "wholesale.csv"
8. def get_association_rules (dataset, min_support=0.1, min_confidence=0.5):
9. #Apply the fpgrwoth algorithm to find frequent itemsets
10. frequent_itemsets = fpgrwoth (one_hot_encoded, min_support=min_support, use_colnames=True)
11. rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=min_confidence)
12. return rules
13. start_time = time.time()
14. dataset = read_wholesale_dataset()
15. # Adjust the values for min_support and min_confidence as needed
16. min_support = 0.05
17. min_confidence = 0.5
18. # Get the association rules
19. rules = get_association_rules (dataset, min_support, min_confidence)
20. # Calculate the number of association rules
21. num_association_rules = len(rules)

```
end_time = time.time()
runtime = end_time - start_time
print ("Number of Association Rules:", num_association_rules)
```

To visualize the results properly, we present the bar chart for the online-retail data set in figure-11, showing execution time of Apriori and FP-Growth with different support levels. X-axis represents different minimum support level and Y-axis represents the run time of the algorithms in miliseconds. Figure-11 clearly represents that the FP-growth algorithm consumes less time than the apriori algorithm.

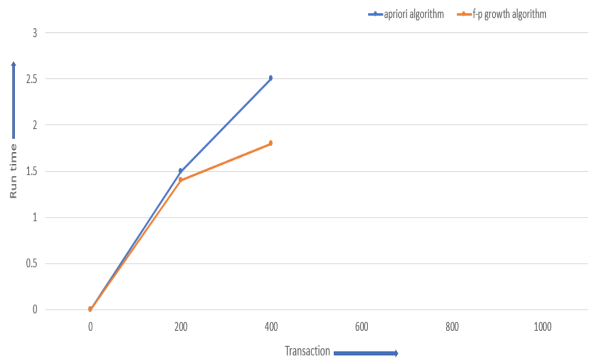

Fig 6. Run time for the Online customer data set

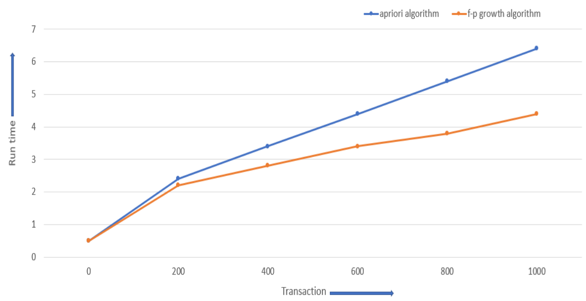

Fig. 5. Run time for the whole sale Retail – 2 data set

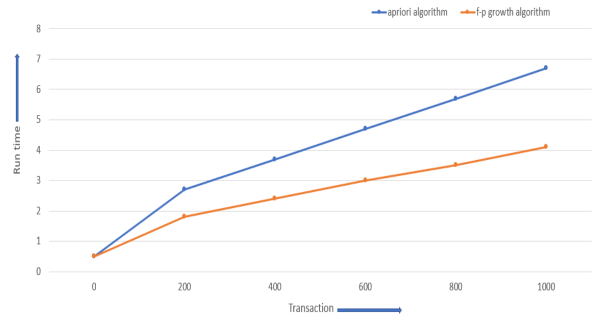

Fig. 7. Run time for the online-retail data set

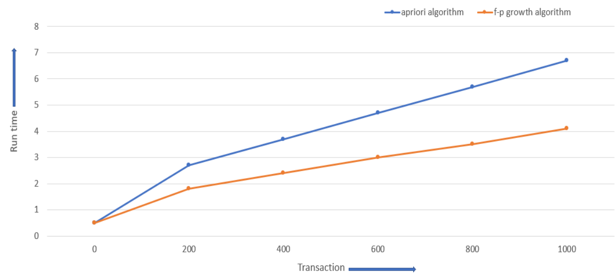

Fig. 8. Run time for the online shopping data set

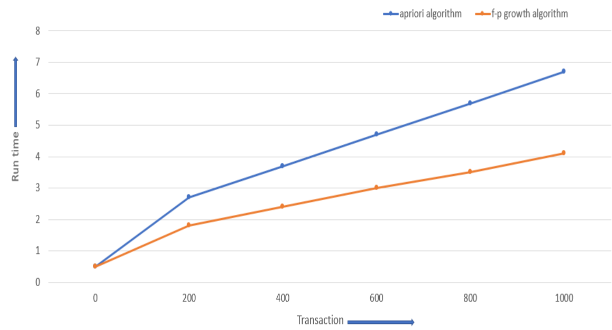

Fig. 9. Run time for the Bank Marketing data set

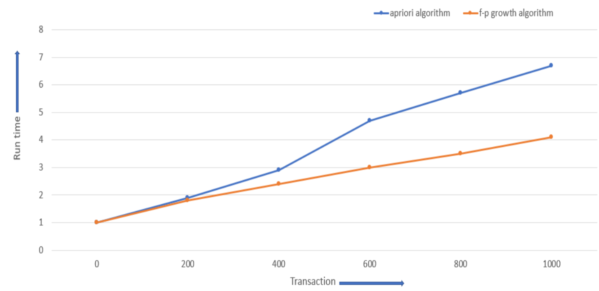

Fig. 10. Run time for the online shoppers purchasing intention data set

Table 9. Execution time for Apriori and FP-growth with various minimum support level

Name of the data set	Number of instances	Number of attributes	Minimum support level	Apriori algorithm	FP-growth algorithm
Whole sale Customer's data	440	8	0.2	2.52	1.66
			0.4	2.24	1.54
			0.5	2.20	1.50
			0.6	2.14	1.45
			0.8	2.07	1.33
Online-retail-2	541904	8	0.2	34.6	21.7
			0.4	33.1	21.1
			0.5	32.5	20.5
			0.6	31.7	19.8
			0.8	30.2	18.7
			0.2	39.1	21.7

Online-retail	1067371	8	0.4 0.5 0.6 0.8	38.2 36.7 35.9 34.4	21.2 20.5 19.8 18.7
Click stream data for online shopping data set	165474	14	0.2 0.4 0.5 0.6 0.8	47.1 46.5 45.1 44.6 43.9	39.8 38.2 37.5 36.9 36.2
Bank marketing	45211	17	0.2 0.4 0.5 0.6 0.8	46.1 44.9 43.1 42.9 41.5	34.9 34.1 33.3 32.9 31.1
Online shopper purchasing intention data set	12330	18	0.2 0.4 0.5 0.6 0.8	37.4 36.5 36 35.7 34.8	26.2 25.5 24.9 24.2 23.9

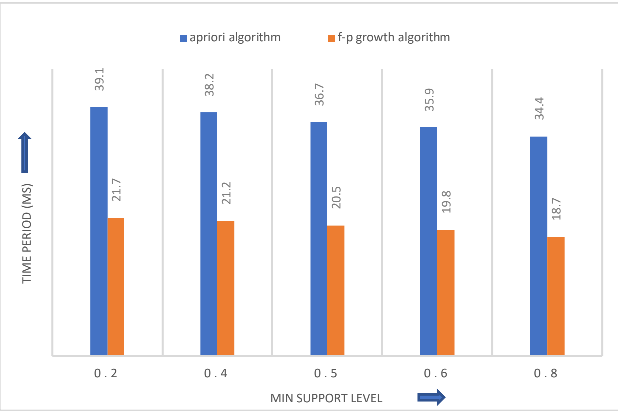

Fig. 11. Bar Chart showing the execution time of Apriori and FP-Growth with different support levels

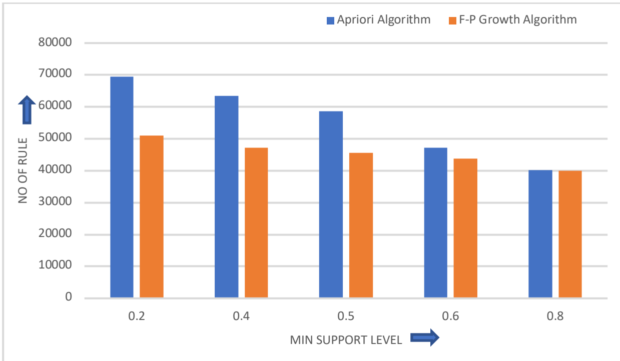

Fig. 12. Bar Chart showing the number of rules generated by Apriori and FP-growth with various support levels

Table 10. Number of rules generated by Apriori and FP-growth with various minimum support level

Name of the data set	Minimum support level	Number of Rules Generated in Apriori algorithm	Number of Rules Generated in FP-growth algorithm
Whole sale Customer's data	0.2 0.4 0.5 0.6 0.8	144 120 105 92 81	108 100 95 87 80
Online-retail-2	0.2 0.4 0.5 0.6 0.8	30501 28749 25309 21969 19801	24507 23105 21512 20899 19769
Online-retail	0.2 0.4 0.5 0.6 0.8	69432 63399 58600 47237 40108	50987 47107 45679 43837 40001
Click stream data for online shopping data set	0.2 0.4 0.5 0.6 0.8	30756 27800 24961 22638 20790	26709 24997 23152 21512 20659
Bank marketing	0.2 0.4	6708 5998	8102 7571

	0.5	5253	6698
	0.6	4617	4279
	0.8	3934	3856
Online shopper purchasing intention data set	0.2	3989	3607
	0.4	3541	3100
	0.5	2999	2722
	0.6	2457	2409
	0.8	2102	2089

To visualize the results properly, we present a bar chart for the online-retail data set in figure-12, showing the number of rules generated by Apriori and FP-Growth with different support levels. X-axis represents different minimum support level and Y-axis represents the total number of transactions for the entire dataset. It is evident that for lower support levels, Apriori produces more number of rules than FP-growth.

Table 11. Execution time for Apriori and FP-growth with different number of transactions with various minimum support level

Name of the data set	Minimum support level	Number of Transactions	Apriori Algorithm (In ms)	FP-growth Algorithm (In ms)
Wholesale Customer's data	0.2	300	2.01	1.3
		600	NA	NA
		900	NA	NA
		1200	NA	NA
		1500	NA	NA
	0.4	300	1.81	1.01
		600	NA	NA
		900	NA	NA
		1200	NA	NA
		1500	NA	NA
	0.6	300	1.62	0.98
		600	NA	NA
		900	NA	NA
		1200	NA	NA
		1500	NA	NA
	0.8	300	1.35	0.74
		600	NA	NA
		900	NA	NA
		1200	NA	NA
		1500	NA	NA
Online-retail-2	0.2	300	10.5	5.4
		600	11.3	6.3
		900	12.01	7.5
		1200	13.2	8.1
		1500	14.6	8.9
	0.4	300	10.6	4.1
		600	11.2	4.7
		900	11.9	5.3
		1200	12.5	5.9
		1500	13.2	6.5
	0.6	300	9.9	3.2
		600	10.3	3.8
		900	10.8	4.4
		1200	11.3	4.8
		1500	11.9	5.4
	0.8	300	7.6	2.4
		600	8.3	2.8
		900	8.9	3.3
		1200	9.6	3.7
		1500	10.01	4.1
Online-retail	0.2	300	8.9	1.7
		600	9.6	2.2
		900	10.2	2.6
		1200	10.9	3.1
		1500	11.7	3.4

	0.4	300	7.9	1.51
		600	8.6	1.76
		900	9.3	2.34
		1200	9.8	2.89
		1500	10.5	3.01
	0.6	300	5.9	1.22
		600	6.6	1.48
		900	7.1	1.76
		1200	7.6	2.01
		1500	8.3	2.4
	0.8	300	5.69	0.89
		600	6.32	1.15
		900	6.70	1.37
		1200	7.29	1.71
		1500	7.76	1.99
Click stream data for online shopping data set	0.2	300	8.98	5.59
		600	9.67	6.21
		900	10.27	6.94
		1200	10.84	7.69
		1500	11.09	8.07
	0.4	300	7.31	4.95
		600	7.98	5.63
		900	8.69	6.09
		1200	9.93	6.98
		1500	10.87	7.69
	0.6	300	6.69	4.33
		600	7.23	4.97
		900	7.95	5.73
		1200	8.68	6.05
		1500	9.34	6.86
	0.8	300	5.83	3.09
		600	6.47	3.86
		900	7.05	4.30
		1200	7.89	4.92
		1500	8.49	5.47
Bank marketing	0.2	300	1.89	0.89
		600	2.97	1.41
		900	3.63	2.03
		1200	4.17	2.66
		1500	4.98	3.19
	0.4	300	3.81	1.93
		600	4.55	2.79
		900	5.09	3.57
		1200	5.86	4.08
		1500	6.37	4.89
	0.6	300	5.91	2.69
		600	6.72	3.12
		900	7.13	3.88
		1200	7.98	4.47
		1500	8.53	5.01
	0.8	300	6.69	4.03
		600	7.59	4.99
		900	8.01	5.61
		1200	8.6	6.17
		1500	9.5	6.90
Online shopper purchasing intention data set	0.2	300	4.27	2.86
		600	4.91	3.49
		900	5.63	4.02
		1200	6.01	4.63
		1500	6.9	5.36
	0.4	300	2.56	1.97
		600	3.15	2.43
		900	3.99	3.01

		1200	4.65	3.69
		1500	5.2	4.19
	0.6	300	1.88	1.08
		600	2.63	1.76
		900	3.29	2.24
		1200	3.29	2.99
		1500	4.67	3.67
	0.8	300	1.27	0.12
		600	1.98	0.51
		900	2.64	0.87
		1200	3.11	1.15
		1500	3.89	1.77

To visualize the results properly, we present the corresponding bar chart for the retail data set. In this case a fixed support level of 0.2 has been considered and the algorithms are run by varying the number of transaction levels.

From the experiment, we can infer the following.

1. Apriori takes more time than FP-growth on the same dataset irrespective of the support level or the number of transactions.
2. Apriori produces more rules than FP-growth for lower support levels. But as the support increases the difference between the number of rules produces by both the algorithms reduce and at a higher value of support such as 0.8, the rules generated by both the algorithms are almost the same.

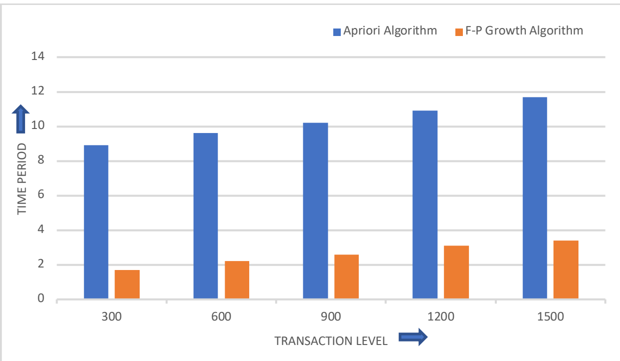

Fig. 13. Bar Chart showing the execution time of Apriori and FP-Growth with 0.2 support level

5.1 Decision on Adapting an algorithm

Both Apriori and FP growth are classical association rule mining algorithms. Both of them take a transactional dataset as the input and produce frequent patterns and association rules as the output. The experiments show that FP growth is less time consuming than Apriori, but the difference between the time taken by both the algorithms is very marginal. In today's digital age when there is no scarcity on the availability of computing power, software and hardware resources and internet speed, such small difference on run time does not matter much. Thus, for low and medium size datasets which have only thousands of rows and very few columns, any of these two algorithms can be used. For datasets having millions of rows and hundreds of columns it may be required to make a choice between Apriori and FP growth. Further, if all possible rules are needed from the dataset, then Apriori can be chosen because it always produces a greater number of rules than FP growth. Additionally, depending on the requirement of a retailer the algorithms can be used. For example, if product bundling is the target, then FP growth can be used because frequent patterns will be more relevant for the application. On the other hand, if product placement and redesigning the store layout are concerned then Apriori algorithm can be used, because a greater number of rules can be helpful in placing the products which have the probability of being bought together**.**

6. Analysis using Big-O method

Big O notation is used to evaluate the performance of algorithms, and to distinguish the efficiency of one algorithm from another. It provides an upper bound of the time taken by an algorithm as the input size increases. Given two functions, it formalizes the notion that two functions either "grow at the same rate," or one function "grows faster than the other one". Here, we have used the Big-O method to examine the performance of the association rule mining algorithm as the size of the data set increases that is mathematically the size of the dataset moves towards infinity.

Let f and g be function from the set of real numbers to the set of real numbers. We say that $f(x)$ is O $(g(x))$ [read as "$f(x)$ is Big-O of $g(x)$"] if there are constant C and k such that

$$|f(x) \leq C\,|(g(x))| \quad \forall\ x \leq x_0 \text{, whenever } x > k \qquad (3)$$

In other words, the absolute value of $f(x)$ is at most a positive constant multiple of $g(x)$ for all sufficiently large x in the domain.

Throughout this analysis, $f(x)$represents the FP-growth algorithm and $g(x)$ represents the Apriori algorithm. It describes the limiting behaviour of the given function concerning the data set. A close look at figures 14 through 19 explains well the behaviour of the algorithms. We observe that when the number of input transactions was less than 100, the efficiency of both algorithms was the same. At exactly 100 transactions, the lines corresponding to $f(x)$ and $g(x)$ coincide reflecting that both the functions take the same time. Further, we observe that as the number of transaction increases, $g(x)$ takes more time. Hence, we conclude that $f(x)$ performs better than $g(x)$. This behaviour of the algorithms has been observed across all the datasets.

The equations of the graphed lines are generated by using Microsoft excel. E in the equations represents Euler number (E=2.718281828459045…). Table 12 gives the equations $f(x)$ & $g(x)$ and the value of C and k for all the data sets.

Based on the above discussion, we write the theorem 1.

Theorem 1: If the runtime of the FP-growth algorithm is $f(x)$, and the runtime of the Apriori algorithm is $g(x)$, then FP-growth is Big-O of Apriori.

Proof: We prove this by considering the equations from Wholesale customers dataset

From table 9, we have the following equations

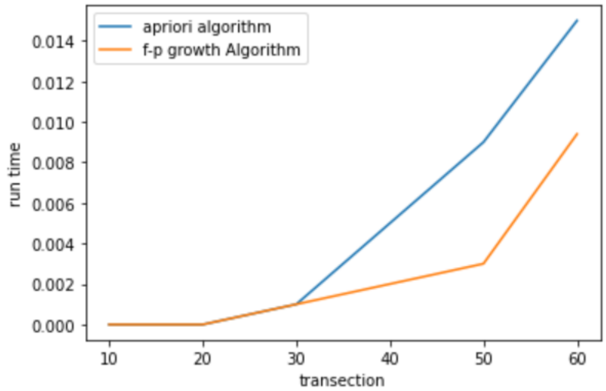

Fig. 14. Big-o graph for wholesale dataset

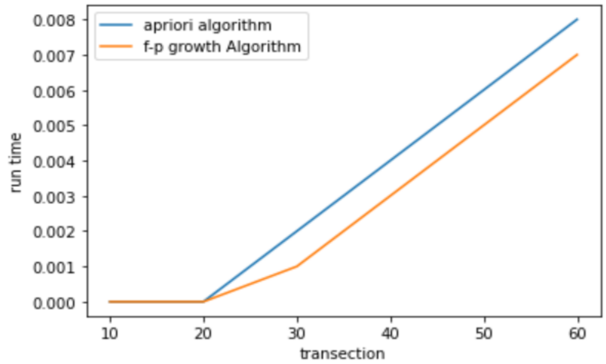

Fig.15. Big-o graph for Online-retail-2 dataset

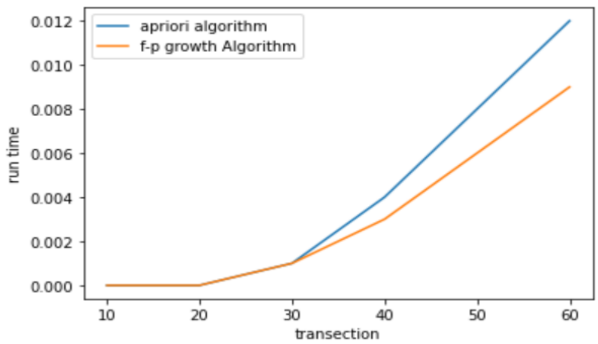

Fig. 16. Big-o graph for Online-retail online shopping data set

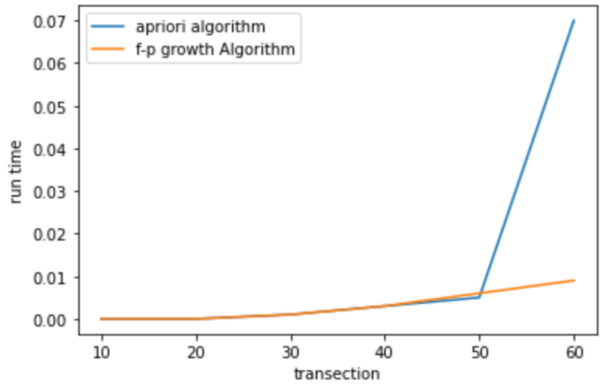

Fig. 17. Big-o graph Click stream data for data set

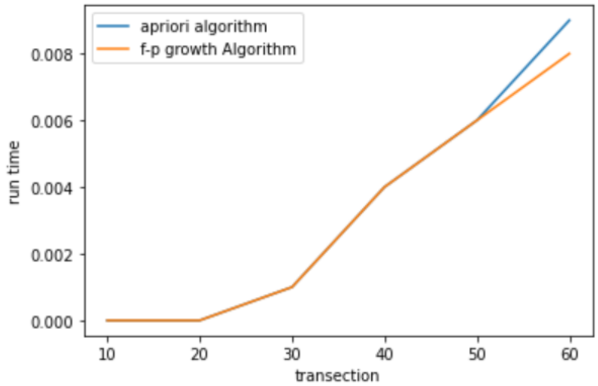

Fig. 18. Big-o graph for Bank marketing data set

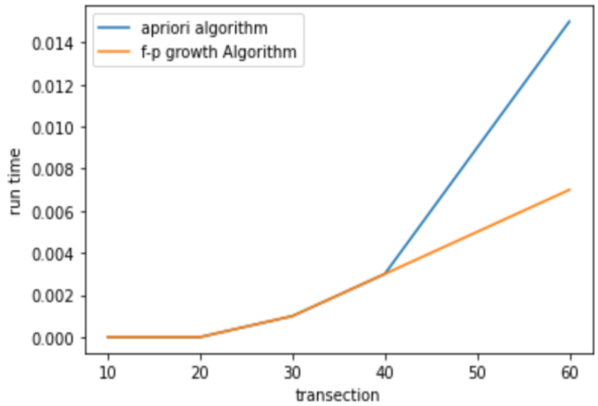

Fig. 19. Big-o graph for Online shopper purchasing intention data set

$$f(x) = 4E - 05x^5 - 0.0005x^4 + 0.0025x^3 - 0.0053x^2 + 0.0046x - 0.0014 \tag{4}$$

$$g(x) = 8E - 05x^5 - 0.0015x^4 + 0.0095x^3 - 0.0275x^2 + 0.0354x - 0.016 \tag{5}$$

Table 12. Value for $f(x), g(x), C$ *and* k from Big-O graph

Name of Data sets	Fig. No	$f(x)$	$g(x)$	C	k	$f(x) \leq C(g(x))$
Wholesale customers data	14	$4E - 05x^5 - 0.0005x^4 + 0.0025x^3 - 0.0053x^2 + 0.0046x - 0.0014$	$8E - 05x^5 - 0.0015x^4 + 0.0095x^3 - 0.0275x^2 + 0.0354x - 0.016$	1	1	Satisfied
Online-retail-2	15	$-2E - 05x^5 - 0.0003x^4 + 0.0018x^3 - 0.0047x^2 + 0.0051x - 0.002$	$2E - 05x^5 + 0.0003x^4 - 0.0026x^3 + 0.0097x^2 - 0.0154x + 0.008$	1	1	Satisfied

Online-retail	16	$-8E-06x^5+0.0001x^4-0.0007x^3+0.0024x^2-0.0038x+0.002$	$2E-05x^5-0.0003x^4+0.0024x^3-0.0072x^2+0.0091x-0.004$	1	1	Satisfied
Click stream data for online shopping data set	17	$-8E-06x^5+0.0001x^4-0.0007x^3+0.0024x^2-0.0038x+0.036$	$0.0005x^5-0.0082x^4+0.0465x^3-0.1228x^2+0.149x-0.065$	1	1	Satisfied
Bank marketing	18	$7E-05x^5-0.0012x^4+0.0075x^3-0.0213x^2+0.0269x-0.012$	$7E-05x^5-0.0013x^4+0.0082x^3-0.0232x^2+0.0292x-0.013$	1	1	Satisfied
Online shopper purchasing intention data set	19	$2E-05x^5-0.0003x^4+0.0018x^3-0.0047x^2+0.0051x-0.002$	$8E-05x^5+0.0014x^4-0.0083x^3+0.0236x^2-0.0306x+0.014$	1	1	Satisfied

Let us verify $|f(x)| \leq C|g(x)|$

$$|4E-05x^5-0.0005x^4+0.0025x^3-0.0053x^2+0.0046x-0.0014|$$

$$\leq C\,|(8E-05x^5-0.0015x^4+0.0095x^3-0.0275x^2+0.0354x-0.016)|$$

Let $C=1$ & $k=1$

$$\Rightarrow |4E-05(1)^5-0.0005(1)^4+0.0025(1)^3-0.0053(1)^2+0.0046(1)-0.0014|$$

$$\leq |1(8E-05(1)^5-0.0015(1)^4+0.0095(1)^3-0.0275(1)^2+0.0354(1)-0.016)|$$

$$\Rightarrow 4E-05-0.0005+0.0025-0.0053+0.0046-0.0014 \leq 1(8E-05-0.0015+0.0095-0.0275+0.0354-0.016)$$

Let, us substitute the value E we considering E value up to four decimal places.

$$\Rightarrow 4(2.7182)-4.9999 \leq 1(8(2.7182)-4.9999)$$
$$\Longrightarrow 10.8728-4.9999 \leq 1(21.7456-4.9999)$$
$$\Longrightarrow 5.8729 \leq 16.7457$$

From the above calculation we can say that $f(x) \leq C(g(x))$istrue i.e FP-growth is Big-O of Apriori.

The same result has been found for all the other datasets and hence it can be concluded that FP-growth is Big-O of Apriori.

7. Potential Benefits of using association rule mining in retailing industry

The retail industry has the potential to boost the economy of a country by creating new job opportunities, increasing labour income and GDP and enhancing both local business and export. The amalgamation of data mining technology with retailing helps realize these goals. Several data mining techniques such as classification, clustering, dimensionality reduction and association rule mining have their own roles to play as far as retail data is concerned. For example, clustering is widely used for customer segmentation and Classification is used for customer's loyalty analysis. Similarly, association rule mining has got many potential applications in the retailing industry. Since, this paper focuses on association rule mining, we list out some of the potential benefits of using association rule mining in the retailing sector.

1. Association rule mining can help retailers to foresee the requirements of their business and thus helps them to maintain adequate stock and inventory management. This in turn helps manufacturing units to produce goods in appropriate amount and prevents supply chain disruption.
2. The analysis result improves the efficiency of the business by enabling selling of large portfolio of products. This helps not only in range selling but also helps to sell products in the long tail.
3. Improves efficiency and productivity of the retail managers. The frequent patterns and association rules discovered by the algorithms only require to be verified by the managers. However, they do not have to scan through the large transactional datasets to find the buying patterns of the customers.
4. Increases in staff productivity. For example, just by looking at the rules the store layout can be reconfigured by the staff.
5. Increases human capital in the retail industry.
6. Enhances customer experience.
7. Helps in floor space utilisation.

8. Limitations and Future Work

The Apriori algorithm needs to generate and store a large number of candidate item sets during its execution, which can lead to high memory consumption, especially for datasets with a large number transaction. Apriori needs more iteration over the data to find frequent item sets, which can be computationally expensive for large datasets. As the number of items in the dataset increases, the number of potential item sets grows exponentially. Consequently, the algorithm's performance degrades rapidly with an increasing number of items. Researchers can explore parallel versions of the Apriori algorithm to make it more efficient in handling large-scale datasets [11, 22] Additionally, optimization techniques can be applied to reduce the number of candidate item sets generated during the execution.

The initial construction of the FP growth tree involves scanning the dataset to determine the frequency of single items. This step can be time-consuming for large datasets with many unique items. The recursive nature of the FP growth algorithm can lead to deep recursive calls and may result in stack overflow errors or increased memory consumption for datasets with long item sets. Research can focus on developing parallel versions of the FP growth algorithm using multi-core processors and distribute the workload efficiently, thereby improving its performance on large datasets [6, 37]. Optimizing memory usage during the construction and traversal of the FP growth tree can enhance the algorithm's efficiency and make it more applicable to datasets with a large number of unique items. Investigating techniques to dynamically determine the minimum support threshold for different stages of the FP growth algorithm can lead to better adaptability and improved performance on datasets with varying sparsity levels.

Challenges were more in data preparation and run time when we worked on a real dataset. More infrastructure was required for conducting experiments on real datasets, which may be needed additional investments.

9. Conclusion

This study examines the effectiveness of the Apriori and FP growth algorithms in the retail industry, utilizing six retail datasets from the UCI machine learning repository. The performance of these algorithms is evaluated based on their execution time and the number of rules generated. To assess execution time, two estimations are made for each algorithm: one by varying only the support level, and another by varying both the support level and the number of transactions. Rules are generated by varying the support levels of both algorithms. The results of the experiments demonstrate that FP growth requires less time than Apriori, indicating that FP growth has a lower computational complexity than Apriori and thus FP growth is Big-O of Apriori.

References

[1] A. R. Pillai and D. A. Jolhe, “Market basket analysis: Case study of a supermarket,” in *Advances in Mechanical Engineering: Select Proceedings of ICAME 2020*, Springer, Jan. 2020, pp. 727–734.

[2] D. Komarasay, M. Gupta, M. Murugesan, J. Jenita Hermina, and M. Gokuldhev, “Frequent Pattern Mining in Big Data Using Integrated Frequent Pattern (IFP) Growth Algorithm,” in *Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering: ICACIE 2021*, Sep. 2022, pp. 423–431.

[3] A. Griva, C. Bardaki, K. Pramatari, and D. Papakiriakopoulos, “Retail business analytics: Customer visit segmentation using market basket data,” *Exp. Syst. Appl.*, vol. 100, pp. 1–16, Jun. 2018, doi: 10.1016/j.eswa.2018.01.029

[4] A. N. Gun and B. Badur, “Assortment planning using data mining algorithms,” in *PICMET’08-2008 Portland International Conference on Management of Engineering \& Technology*, Jul. 2008, pp. 2312–2322.

[5] B. Berman, J. R. Evans, and P. Chatterjee, “Retail Management: A Strategic Approach (Vol.). Harlow.” United Kingdom: Pearson Education Limited, Jun. 2018.

[6] R. M. D. Datta M.P. Dutta, “Improvisation in Efficiency of Apriori Algorithm for Mining Frequent Itemsets,” *Int. J. Comput. Sci. Eng.*, vol. 6, no. 8, pp. 424–428, Aug. 2018,

[7] F. H. AL-Zawaidah, Y. H. Jbara, and A. L. Marwan, “An improved algorithm for mining association rules in large databases,” *World Comput. Sci. Inf. Technol. J.*, vol. 1, no. 7, pp. 311–316, 2011.

[8] J. L. Ruiz-Real, J. Uribe-Toril, J. C. Gázquez-Abad, and J. de Pablo Valenciano, “Sustainability and retail: analysis of global research,” *Sustainability*, vol. 11, no. 1, p. 14, Dec. 2018.

[9] M. A. Valle, G. A. Ruz, and R. Morrás, “Market basket analysis: Complementing association rules with minimum spanning trees,” *Expert Syst. Appl.*, vol. 97, pp. 146–162, May. 2018.

[10] J. Chen, C. Miller, and G. G. Dagher, “Product recommendation system for small online retailers using association rules mining,” in *Proceedings of the 2014 International Conference on Innovative Design and Manufacturing (ICIDM)*, Aug. 2014, pp. 71–77.

[11] S. Lian, Y. Xu, and C. Zhang, “Family profile mining in retailing,” *Decis. Support Syst.*, vol. 118, pp. 102–114, Mar. 2019.

[12] W. C. McDowell, R. C. Wilson, and C. O. Kile Jr, “An examination of retail website design and conversion rate,” *J. Bus. Res.*, vol. 69, no. 11, pp. 4837–4842, Jan. 2016, doi: 10.1016/j.jbusres.2016.04.040.

[13] M. Kaur and S. Kang, “Market Basket Analysis: Identify the changing trends of market data using association rule mining,” *Procedia Comput. Sci.*, vol. 85, pp. 78–85, Dec. 2016, doi: 10.1016/j.procs.2016.05.180.

[14] D. S. Pranav, D. Punj, T. Dubey, and P. Chawla, “Data mining in cloud computing,” in *2021 5th international conference on computing methodologies and communication (ICCMC)*, Jun. 2021, pp. 71–78.

[15] M. J. Zaki, “Scalable algorithms for association mining,” *IEEE Trans. Knowl. Data Eng.*, vol. 12, no. 3, pp. 372–390, Jun. 2000, doi: 10.1109/69.846291

[16] M. N. Dalal Al Attal, N. AlBaghli, N. Al Muhaimeed, and S. Al, “Redesigning a Retail Store Based on Association Rule Mining”Jul. 2018.

[17] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, Jun.1993, pp. 207–216.

[18] R. Agrawal, R. Srikant, and others, “Fast algorithms for mining association rules,” in *Proc. 20th int. conf. very large data bases, VLDB*, Sep. 1994, pp. 487–499.

[19] J. Wang, J. Han, Y. Lu, and P. Tzvetkov, “TFP: An efficient algorithm for mining top-k frequent closed itemsets,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 5, pp. 652–663, May. 2005.

[20] F. J. M. Arboleda, G. P. Ortega, and J. A. G. Luna, “Temporal visual profiling of market basket analysis,” *IAENG Int. J. Comput. Sci.*, vol. 49, no. 2, Jun. 2022.

[21] A. Kinge, P. B. Hrithik, Y. Oswal, and N. Kulkarni, “Customer Analytics Research: Utilizing Unsupervised Machine Learning Techniques,” in *Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2022*, Springer, Dec. 2022, pp. 501–515.

[22] M. R. Pradana, M. Syafrullah, H. Irawan, J. C. Chandra, and A. Solichin, “Market Basket Analysis Using FP-Growth Algorithm On Retail Sales Data,” in *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Oct. 2022, pp. 86–89.

[23] A. A. Aldino, E. D. Pratiwi, S. Sintaro, A. D. Putra, and others, “Comparison of market basket analysis to determine consumer purchasing patterns using fp-growth and apriori algorithm,” in *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, Oct. 2021 pp. 29–34.doi:10.1109/ICOMITEE53461.2021.9650317

[24] F. Tasnim, T. Ahmed, K. Ahmed, M. F. K. Patwary, and M. K. Newaz, “Potential Risk Factors and Association of Significant Factors of Blood Cancer in Bangladesh Using Data Mining Techniques,” *J. Eng. Sci.*, vol. 13, no. 1, pp. 9–20, Sep. 2022.

[25] G. H. Kiani, “Determining profitable products in the retail market with consideration of cash limitation and exhibition periods,” *J. Retail. Consum. Serv.*, vol. 55, p. 102079, July. 2020. doi : 10.1016/j.jretconser.2020.102079.

[26] M. Hossain, A. H. M. S. Sattar, and M. K. Paul, “Market basket analysis using apriori and FP growth algorithm,” in *2019 22nd*

international conference on computer and information technology (ICCIT), Dec. 2019, pp. 1–6.

[27] R. Ahlers *et al.*, “Are local retailers conquering the long tail? A web usage and association rule mining approach on local shopping platforms,” *Proc. Multikonferenz Wirtschaftsinformatik, MKWI'18*, Mar. 2018.

[28] S. Vanaja and M. Belwal, “Aspect-level sentiment analysis on e-commerce data,” in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, Jul. 2018, pp. 1275–1279.

[29] A. Demiriz, G. Ertek, T. Atan, and U. Kula, “Re-mining positive and negative association mining results,” in *Advances in Data Mining. Applications and Theoretical Aspects: 10th Industrial Conference, ICDM 2010, Berlin, Germany, July 12-14, 2010. Proceedings 10*, 2010, pp. 101–114.

[30] J. L. Tiwari and P. I. Terasaki, *HLA and disease associations*. Springer Science \& Business Media, Dec. 2012.

[31] K.I. Ahn, “Effective product assignment based on association rule mining in retail,” *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12551–12556, Nov. 2012. DOI:10.1016/j.eswa.2012.04.086

[32] I. Cil, “Consumption universes based supermarket layout through association rule mining and multidimensional scaling,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 8611–8625, Aug. 2012. DOI:10.1016/j.eswa.2012.01.192

[33] S. H. Liao, C.M. Chen, and C.H. Wu, “Mining customer knowledge for product line and brand extension in retailing,” *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1763–1776, Apr. 2008.

[34] J. L. Anderson, L. D. Jolly, and A. E. Fairhurst, “Customer relationship management in retailing: A content analysis of retail trade journals,” *J. Retail. Consum. Serv.*, vol. 14, no. 6, pp. 394–399, Nov. 2007, doi:10.1016/j.jretconser.2007.02.009

[35] S.J. Yen and A. Chen, “A graph-based approach for discovering various types of association rules,” *IEEE Trans. Knowl. Data Eng.*, vol. 13, no. 5, pp. 839–845, Sep. 2001. doi: 10.1109/69.956106.

[36] M. Chen, A. Chiu, and H. Chang, “Mining changes in customer behavior in retail marketing,” *Exp. Sys. with Applic.*, vol. 28, no. 4, pp. 773–781, May. 2005, doi: 10.1016/j.eswa.2004.12.033.

[37] W. Gan, J. Lin, P. Fournier Viger, H.-C. Chao, and P. Yu, “A Survey of Parallel Sequential Pattern Mining.” Aug. 2018.