International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 03, Issue 04, October 2022
e-ISSN : 2745-9659
https://ijcis.net/index.php/ijcis/index

Comparison of Apriori Algorithm and FP-Growth in Managing Store Transaction Data

1st Syukron Anas, 2nd Nelson Rumui, 3rd Andi Roy, 4th Pujo Hari Saputro
Informatics Manajemen Politeknik Negeri FakFak
FakFak, Indonesia
[1]ukowoeish@gmail.com, [2]nelsonrumui@gmail.com, [3]andiroy@polinef.id, [4]pujoharisaputro@unsrat.ac.id

Abstract*—The role and position of data in today's digital era are very important, data can be likened to a resource that can be explored further to produce new information or knowledge. Seeing the importance of data position, several solutions can be offered in getting more value from data, one of which is the use of Data Mining techniques with association techniques, several types of association techniques are a priori algorithms and FP-Growth algorithms. Based on the research results, the a priori algorithm produces a combination of goods with a confidence value of 98.4 and a support value of 98.4, and the algorithm produces a combination of goods with a support value of 95.2 and a confidence value of 95.2. The comparison of these two algorithms in making associations results in a faster execution time of the FP-Growth algorithm than Apriori, and the Apriori algorithm produces more varied itemset combinations.

Keywords: Data Mining, Apriori, FP-Growth

I. INTRODUCTION

Data is one of the most vital things in today's world. Data is currently one of the things that can support various aspects of decisions to be taken. Decision-making based on data is a wise step because the data reflects the facts that occur in the field, so it can describe the real conditions, and will affect the quality of the decisions to be taken. One aspect that uses a lot and involves a lot of data is the sales process, in this case,s more specific to retail stores that provide a variety of products. Transactions that occur in stores in a day can reach hundreds of transactions, it can be estimated how much data is in sales transactions. With a lot of data in the sales data, the data should not just be a record that has no added value and only fills data storage from transaction data storage which is sure to increase over time.

Seeing the problem of data management, the role that can be taken in providing more value to the data is to perform or apply data mining techniques. Data Mining is the process of looking for interesting patterns in data sets by applying certain techniques [1]. In data mining, there are several types and techniques of data processing, such as data grouping techniques or classification and clustering, or performing data association processes. This process is used to find relationships from many data in a large database. Association rules are useful for finding important relationships between items in each transaction, these relationships can indicate whether or not a rule is strong in the association [2]. The data that is processed using the association technique can be used as a reference for the next strategic step. This is because consumer buying interest is influenced by several factors, and one of the important factors is the convenience obtained when making transactions. To increase the convenience of transactions, the application of association rules can be used to find the associative relationship of item combinations. The combination of items here is intended to be able to find items that have a high relationship if other items are purchased.

Several methods of association rule include Apriori and FP-Growth. In this study, we will present a comparison of these two association methods related to the performance of the association process. To strengthen the research, the following literature reviews were carried out.

As a research reinforcement, the first research reference was from Erwin who wrote Market Basket Analysis with Apriori Algorithm and FP-Growth. This study concludes that the Apriori algorithm requires a long computational time to get frequent item sets, while FP-Growth takes a shorter time, so it becomes more efficient [3]. Domi Sepri wrote a priori and FP-Growth comparison in his writings with better results on the FP-Growth algorithm in generating frequent itemset because it searches for frequent item-sets using the FP-Tree system so there is no need to scan the database repeatedly if a combination of itemset has been found previously [4]. Furthermore, Luki Henando wrote a comparison of these two association methods and got the results of the association rule system that can be used in decision-making to find alternative laptop brands that are widely enjoyed by consumers. [5]. Muhammad Mariko researched the same thing by comparing these two algorithms by getting the results that the greater the support value and the confidence value set, the shorter the algorithm processing time, and the smaller the support and confidence value, the longer it will take to process the algorithm. [6]. The article published in 2018 written by Abu Salam concluded that FP-Growth can find more association rules than Apriori because the process in FP-Growth does not require several iterations in the process so more association rules can be obtained, between the two algorithms. FP-Growth algorithm rule association has a higher level of accuracy than apriori [7]. Hita Maulidiya wrote a comparison of the method with the results of the FP-Growth algorithm which is better at forming association rules. The results of this study can be used as consideration in making food packages at Kopkartex [8]. Similar results were also obtained in Sita Anggraeni's research, that the FP-Growth algorithm has a better accuracy rate than Apriori [9]. Research from

International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 03, Issue 04, October 2022
e-ISSN : 2745-9659
https://ijcis.net/index.php/ijcis/index

Anthony Anggrawan also got results similar to the final results of the a priori algorithm getting a total of 4 rules, and the FP-Growth algorithm as many as 6 rules with the best results on the FP-Growth algorithm with a lift ratio of 1.27908 [10]. Maliha Hossain wrote a comparison of these two algorithms with the final result that the FP-Growth algorithm requires a faster time than the Apriori algorithm. [11]. Apart from the association algorithm, the follow-up of data association can be done by making decisions by applying decision-making algorithms with one of the SAW and WP algorithms [12]. On the other hand, data mining processing can also increase customer satisfaction which can be open to data processing with other data mining techniques such as clustering with one of the techniques such as K-Means to get customer satisfaction.[13]. The latest research reference from Mustakim also produces thesame conclusions as several reference studies, namely the results of this study, the FP-Growth algorithm produces a better level of effectiveness and time efficient a priori algorithm [14].

II. RESEARCH METHODS

2.1. Stage Research

The stages in this research begin with the study literature and continue with data collection. The data collected was applied using the association technique using a priori algorithm and FP-Growth. The results of the application of these two association algorithms will be compared with each other. Flow details are shown in **Error! Reference source not found.**

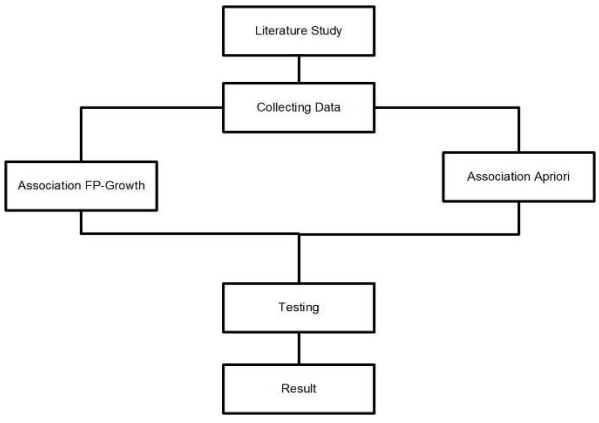

Figure 1. Research Stage

2.2. Apriori

Apriori is the first and basic algorithm for finding frequent itemsets proposed by R. Agrawal and R. Srikant in 1994 [15]. Apriori involves an approach known as a level-wise search, where k-itemsets are used to explore (k + 1)-itemsets. Here, at first frequent 1-itemsets are found by scanning the database which satisfies the minimum support. Again, frequent 2-itemsets are found by using frequent 1-itemsets. So this process continues until frequent item sets can be found [16]. Apriori follows an anti-monotonic property that states that every subset of a frequent itemset must also be frequent and it uses a breadth-first search to count the candidate items frequently. this algorithm has two main steps-

Joining step: To find LK , a set of candidate k-itemsets is generated by joining (LK 1) with itself [16].

Pruning step: Any (k 1)-itemset that is not frequent cannot be a subset of the frequent k-itemset [16].

2.3. FP-Growth

The Apriori algorithm has two major demerits it generates a huge number of candidate sets and scans the database a lot of time. To overcome the disadvantages of the Apriori algorithm, the FP Growth algorithm is used. FP Growth follows a divide-and-conquer strategy. At first it constructs a frequent pattern tree or FP-tree by taking the frequent items which are sorted in the order of descending support count and then it uses that FP-tree to obtain the association information [16]. The best advantage of FP Growth is it scans the database only two times and does not generate a huge number of candidate sets.

III. RESULT AND ANALYSIS

The data used is buying and selling transaction data at the store. The data used are 125 data based on sales. Transaction data is shown in Table 1.

Table 1. Data Transaction

No	Items
1	BERAS OSING 25 KG, SABUN SABUN GIV, SEDAP SOTO, NIU, SAMSU, SAMPOERNA, TELUR
2	MINYAK GORENG TANGGUNG, TEH SARIWANGI, ROTI LEGIT, BISKUIT ROMA KELAPA
3	AQUA BESAR, BISKUIT ROMA KELAPA, YAKULT, JELYDRINK
4	SENTRAT BABI, BERAS OSING 5 KG, BOTOL MINYAK GORENG BESAR, PEPSODENT BESAR
5	MINYAK TELON BESAR, KAPAS, TISSU PASSEO, BISKUIT MARRIE SUSU, SEDAP SOTO, SEDAP GORENG
6	BUNGKUS SAMPOERNA, WALANG MASS, FIX MILD, TEH PUCUK, ROTI
7	MARLBORO, NIU, SUSU KENTAL MANIS, SEDAP SOTO, TEPUNG BUMBU SAJIKU, TELUR
8	TEPUNG TERIGU,TEPUNG BERAS ROSE BRAND, TEPUNG KETAN ROSE BRAND, KACANG TANAH, TELUR
9	MASAKO BESAR, MIWON, MINYAK GORENG BIMOLI, TERASI, TELUR,
10	AQUA GELAS, KERTAS MINYAK KOTAK, MINYAK GORENG BESAR, TERASI, MIWON, TERASI, GARAM
…..	
125	TEPUNG KUE, WINGS, BERAS OSING 5KG, MINYAK GORENG FITRI

The existing transaction data is then carried out by the mapping process based on the item items by mapping each transaction. So that it will produce tabular transaction data as shown in Table 2.

Table 2. Tabular Data

ABC SUSU	AQUA BESAR	JOS
Y	N	N
N	Y	N
N	N	N
N	N	N
N	N	N

After mapping, the next step is to arrange experimental scenarios for the two algorithms, namely by carrying out several test scenarios. The first test is setting the minimum support value to see the results of the rules generated by the two algorithms, the next is testing the execution time of the two algorithms, which is the fastest time to generate association rules.

The Apriori experimental table is shown in Table 3.

Table 3. Scenario Minimum Support Apriori

No	Nilai Minimum Support	Number of Rule
1	0,1	10
2	0,2	10
3	0,3	10

The FP-Growth experimental table is shown in Table 4.

Table 4. Scenario Minimum Support FP-Growth

No	Nilai Minimum Support	Number of Rule
1	0,1	10
2	0,2	7
3	0,3	7

In experiments using the two algorithms, different results were obtained. In the a priori algorithm, the resulting rules remain consistent even though the minimum support value is changed. While in the FP-Growth algorithm, the value changes at the minimum support value to 2.

From the experiment above, the same rule produced as many as 10 rules at a minimum support value of 0.1. The detailed data of the Apriori rule is shown in Table 5.

Table 5. Rule Apriori Algorithm

No	Rule	Support	Confidence
1	Blaster Neaplt --> Cup Ngetop	98,4	98,4
2	Big Sprite --> Wafer Tanggo	98,4	98,4
3	Minyak Kayu Putih --> Hit	97,6	97,6
4	Asatu --> Blaster Neaplt	97,6	97,6
5	Asatu --> Big Sprite	97,6	97,6
6	Marlboro Putih --> Blaster Neaplt	97,6	97,6
7	Marlboro Putih --> Big Sprite	97,6	97,6
8	Citra HBL Pearl White UV --> Blaster Neaplt	97,6	97,6
9	Citra HBL Pearl White UV --> Big Sprite	97,6	97,6
10	Blaster Neaplt --> Minyak GPU	97,6	97,6

The data on the FP-Growth rule is shown in Table 6.

Table 6. Rule FP-Growth

No	Rule	Support	Confidence
1	Dupa --> ABC Susu	95,2	95,2
2	Sedap Soto --> ABC Susu	86,4	86,4
3	Sampoerna --> ABC Susu	10,4	10,4
4	Sampoerna --> Dupa	10,4	10,4
5	Sampoerna --> Sedap Soto	10,4	10,4
6	Dupa --> Sedap Soto	10,4	10,4
7	Sampoerna --> ABC Susu	10,4	10,4
8	ABC Susu --> Sampoerna	10,4	10,4
9	Dupa --> Sampoerna	10,4	10,4
10	Sampoerna --> ABC Susu	10,4	10,4

From the results of the two rules produced by the two algorithms, the results shown by the FP-Growth algorithm experienced several times the similarity of item combinations, while the a priori algorithm had more variations.

The results of the visualization of the results of the confidence and support values of the two algorithms are shown in Figure 2.

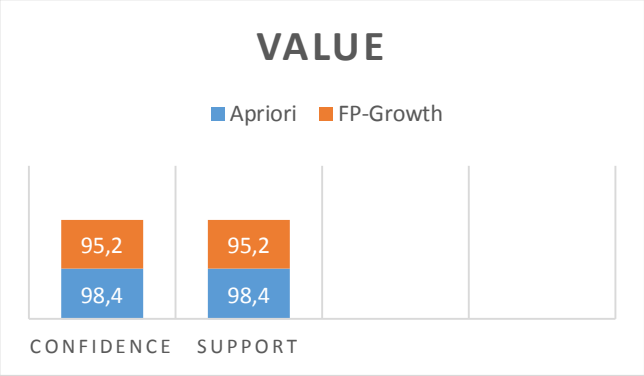

Figure 2. Value Comparison

The next test is seen from the execution time of the two algorithms in generating association rules. The results of the execution times of the two algorithms are shown in Figure 3.

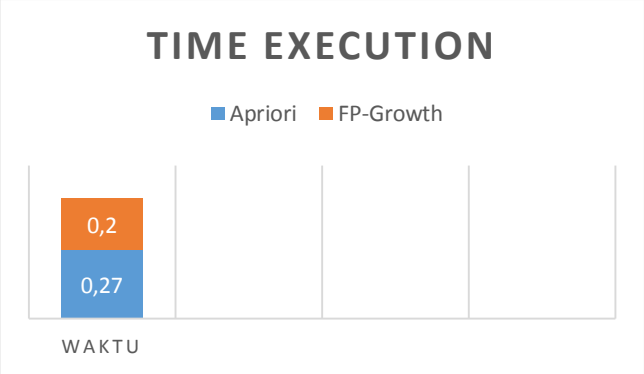

Figure 3. Time Execution

Analysis :
Based on several test scenarios, both algorithms have their respective advantages and disadvantages. The Apriori algorithm has a longer execution time, because the scanning process is more in-depth while using FP-Growth scanning is not too deep so it has a relatively faster time. However, the item combination results obtained by the Apriori algorithm are more varied than the FP-Growth algorithm.

IV. CONCLUSION

The conclusions that can be drawn from this research are:
1. The Apriori algorithm has a support and confidence value of 98.4.
2. The FP-Growth algorithm has a support and confidence value of 95.2.
3. The Apriori algorithm has a relatively long execution time, due to a deeper scanning process compared to the FP-Growth algorithm.
4. The FP-Growth algorithm has a faster execution time, because the scanning process is not too deep, and the results are less varied.
5. The results of the FP-Growth algorithm are 10 rules, with 5 rules having the same itemset.
6. The results of the Apriori algorithm are 10 rules, without any rules that have the same item set.

The suggestions in this research are:
1. Comparison with other association algorithms is possible for further research.
2. Trial with more varied and more data to get a better picture of the results of the association algorithm.

REFERENCES

[1] V. Kale, "Enterprise performance intelligence and decision patterns," *Enterp. Perform. Intell. Decis. Patterns*, pp. 1–262, 2017, doi: 10.4324/9781351228428.

[2] D. Nofriansyah, "Algoritma Data Mining dan Pengujiannya," *Andi Offset*, vol. 2, pp. 1–140, 2014.

[3] U. S. A. Erwin (Jurusan Teknik Informatika, Fakultas Ilmu Komputer, "Analisis Market Basket Dengan Algoritma," *J. Generic*, vol. 4, pp. 26–30, 2009.

[4] D. Sepri and M. Afdal, "Analisa Dan Perbandingan Metode Algoritma Apriori Dan Fp-Growth Untuk Mencari Pola Daerah Strategis Pengenalan Kampus Studi Kasus Di Stkip Adzkia Padang," *J. Sist. Inf. Kaputama*, vol. 1, no. 1, pp. 47–55, 2017, [Online]. Available: https://www.jurnal.kaputama.ac.id/index.php/JSIK/article/view/27.

[5] L. Henando, "Algoritma Apriori Dan Fp-Growth Untuk Analisa Perbandingan Data Penjualan Leptop Berdasarkan Merk Yang Diminati Konsumen (Studi Kasus : Indocomputer Payakumbuh)," *J-Click*, vol. 6, no. 2, pp. 201–207, 2019.

[6] M. Mariko, "Perbandingan Algoritma Apriori Dan Algoritma Fp-Growth Untuk Rekomendasi Item Paket Pada Konten Promosi," *Explore*, vol. 11, no. 2, p. 24, 2021, doi: 10.35200/explore.v11i2.438.

[7] A. Salam, J. Zeniarja, W. Wicaksono, and L. Kharisma, "Pencarian Pola Asosiasi Untuk Penataan Barang Dengan Menggunakan Perbandingan Algoritma Apriori Dan Fp-Growth (Study Kasus Distro Epo Store Pemalang)," *Dinamik*, vol. 23, no. 2, pp. 57–65, 2019, doi: 10.35315/dinamik.v23i2.7178.

[8] H. Maulidiya and A. Jananto, "Asosiasi Data Mining Menggunakan Algoritma Apriori dan FP-Growth sebagai Dasar Pertimbangan Penentuan Paket Sembako," *Proceeding SENDIU 2020*, vol. 6, pp. 36–42, 2020.

[9] S. Anggraeni, M. A. Iha, W. Erawati, and S. Khairunnas, "The Analysis of Sales by Using Apriori and FP-Growth at PT. Panca Putra Solusindo," *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 3, no. 2, pp. 41–46, 2019, [Online]. Available: Riset dan E-Jurnal Manajemen Informatika Komputer.

[10] A. Anggrawan, M. Mayadi, and C. Satria, "Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma FP-Growth," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 125–138,

2021, doi: 10.30812/matrik.v21i1.1260.

[11] M. Hossain, A. H. M. S. Sattar, and M. K. Paul, "Market basket analysis using apriori and FP growth algorithm," *2019 22nd Int. Conf. Comput. Inf. Technol. ICCIT 2019*, pp. 18–20, 2019, doi: 10.1109/ICCIT48885.2019.9038197.

[12] E. T. L. Musthofa Galih Pradana, Kusrini, "PERBANDINGAN METODE WEIGHTED PRODUCT DAN SIMPLE ADDITIVE WEIGHTING DALAM SELEKSI PENGURUS FORUM ASISTEN (STUDI KASUS : UNIVERSITAS AMIKOM YOGYAKARTA)," vol. 4, no. 2, 2019.

[13] M. G. Pradana, A. C. Nurcahyo, P. H. Saputro, U. A. Ata, and U. A. Ata, "Penerapan Metode K-Means Klustering untuk Menentukan Kepuasan Pelanggan," *Citec J.*, vol. 7, no. 1, pp. 42–50, 2020.

[14] Mustakim *et al.*, "Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau," *J. Phys. Conf. Ser.*, vol. 1114, no. 1, 2018, doi: 10.1088/1742-6596/1114/1/012131.

[15] R. Agrawal, T. Imieliński, and A. Swami, *Mining Association Rules Between Sets of Items in Large Databases*, vol. 22, no. 2. ACM, 1993.

[16] J. Han, M. Kamber, and J. Pei, *Data Mining (Third Edition) The Morgan Kaufmann Series in Data Management Systems*. 2012.