JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 101 – 108 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3164
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

COMPARATIVE STUDY OF FP - GROWTH AND APRIORI IN GROCERY ANALYSIS

Ahmad Cahyono Adi[1*]
[1]Computer Science and Electronics, Universitas Gadjah Mada
email: *ahmadcahyonoadi@mail.ugm.ac.id

Abstract: The growing business environment causes the business world to create in order to continue to survive, one of which is by increasing sales. One way is to use the data approach. One method of data approach that is widely used is Market Basket Analysis. This study uses the Market Basket Analysis method with the a priori algorithm and FP-Growth. Grocery dataset analysis using two algorithms, Apriori and FP-Growth using a minimum support parameter of 0.45, has results sorted by the top 10 associations with the best confidence value. The association with the highest support value found is "Whole Milk -> Other Vegetables" with a support value of 0.0748347737. The analysis concludes that both algorithms produce the same association "Other Vegetables -> Whole Milk" with a Support value of 0.0748347737.

Keywords: apriori; FP-Growth; market basket analysis.

Abstrak : Lingkungan bisnis yang semakin berkembang menyebabkan dunia bisnis harus berkreasi agar dapat terus bertahan, salah satunya dengan cara meningkatkan penjualan. Salah satu caranya adalah dengan menggunakan pendekatan data. Salah satu metode pendekatan data yang banyak digunakan adalah Market Basket Analysis. Penelitian ini menggunakan metode Market Basket Analysis dengan algoritma apriori dan FP-Growth. Analisis dataset grosir menggunakan dua algoritma, Apriori dan FP-Growth dengan menggunakan parameter support minimum 0,45, memiliki hasil yang diurutkan berdasarkan 10 asosiasi teratas dengan nilai confidence terbaik. Asosiasi dengan nilai support tertinggi yang ditemukan adalah “Whole Milk -> Other Vegetables” dengan nilai support sebesar 0,0748347737. Analisis menyimpulkan bahwa kedua algoritma tersebut menghasilkan asosiasi yang sama “Other Vegetables -> Whole Milk” dengan nilai Support sebesar 0,0748347737.

Kata Kunci: apriori; FP-Growth; market basket analysis

INTRODUCTION

In the current competitive business landscape, companies must consistently innovate and adapt to maintain their competitiveness. Efforts to do this include enhancing product quality and boosting product diversity. A proficient approach is the analysis of transaction data, which facilitates decision-making for enterprises. By analyzing transaction data, firms can acquire significant insights into customer behavior and preferences, aiding in the optimization of marketing and sales tactics. Consequently, the analysis of transaction data is essential for firms to compete and expand in progressively competitive marketplaces.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 101 – 108 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3164
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Analyzing sales transaction data is challenging due to the substantial volume of data and the constraints of data processing techniques [1,2].

The difficulties of handling extensive datasets with restricted tools can be mitigated by employing data mining technology. Data mining entails the examination of particular datasets to reveal significant information through diverse methodologies. A prevalent data mining technique is association rules, referred to as Market Basket Analysis (MBA).

Market Basket Analysis is a technique extensively utilized in marketing to identify associations between items or product categories. The main objective is to ascertain the things acquired concurrently by buyers. This strategy enhances consumer spending by positioning frequently purchased items in proximity to one another. For example, MBA is often utilized in the analysis of shopping baskets at grocery stores. This analysis seeks to produce association rules or patterns within the data, emphasizing both precision and utility. Efficient data processing is a crucial element in resolving these situations [5].

A study was undertaken to identify trends of things purchased concurrently over the course of a year. This study utilized the Apriori and FP-Growth algorithms to discern patterns within the dataset, and their outcomes were juxtaposed. The results indicated that the Apriori Algorithm executed and presented outcomes more rapidly, although produced a lesser quantity of rules compared to FP-Growth. In contrast, FP-Growth required more time for data processing but generated a greater quantity of rules [6].

The Apriori Algorithm is a method extensively employed in numerous research projects concerning market basket analysis. The aim is to forecast the probability of product acquisition based on previous transactions or historical data [7]. The Apriori algorithm generates association rules through an iterative process that identifies itemsets with k occurrences based on itemsets with k-1 occurrences [8]. The first phase entails establishing threshold values for support and confidence. This approach focuses on the itemsets of commodities purchased in a single transaction by consumers [9].

The FP-Growth Algorithm detects patterns through frequency counts in a dataset [10]. Unlike the Apriori approach, which uses candidate generation, FP-Growth employs a tree-based structure (FP-Tree) to locate common itemsets [11]. This method enhances efficiency relative to Apriori, especially in contexts with extensive datasets [12].

This study intends to compare the performance of the FP-Growth and Apriori algorithms in Market Basket Analysis using supermarket datasets. The findings will be displayed in comparative tables that delineate the parameters produced by each method. These comparisons will elucidate the correlations produced by the two algorithms, finally offering insights for grocery store proprietors in picking the most appropriate algorithm for their sales analysis requirements. Research examined the utilization of the FP-Growth algorithm in data mining. This technique detects prevalent elements within a dataset and formulates association rules. For instance, when a consumer acquires a 3 kilogram LPG cylinder, they are also predisposed to purchase a pack of Djarum Super 12 cigarettes, exhibiting a support value of 0.02 and a confidence value of 1.00 [13].

Improve performed a comparative analysis of the Apriori and FP-Growth algorithms to ascertain the most common itemsets. Their analysis found that the

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 101 – 108 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3164
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Apriori Algorithm provided eight rules with a minimum support of 0.06 and confidence of 0.01, whereas the FP-Growth Algorithm produced 14 rules [14].

The Research employed the Apriori and FP-Growth algorithms in data mining to examine association patterns for product arrangement at Mohare Supermarket. Using a minimum support of 20% and confidence of 70%, the Apriori Algorithm created 10 rules with a support value of 0.32258605 and an accuracy of 12.8%. Concurrently, FP-Growth produced 78 rules with a support value of 2.51612903 and an accuracy of 78%. The association regulations enabled the supermarket to refine product arrangements and improve sales tactics [15].

This study does a comparative examination of the Apriori and FP-Growth algorithms in Market Basket examination utilizing supermarket data. The results, displayed in comparative tables, emphasize the performance of each method, offering insights into their advantages and disadvantages. The findings intend to assist grocery enterprises in selecting the most appropriate algorithm for sales analysis, providing relevant information for decision-making while addressing the industry's specific needs.

METHOD

This study was executed in multiple phases. The initial phase involves inputting the dataset, preprocessing the data, generating association rules, and evaluating the outcomes. The stages are depicted in Figure 1.

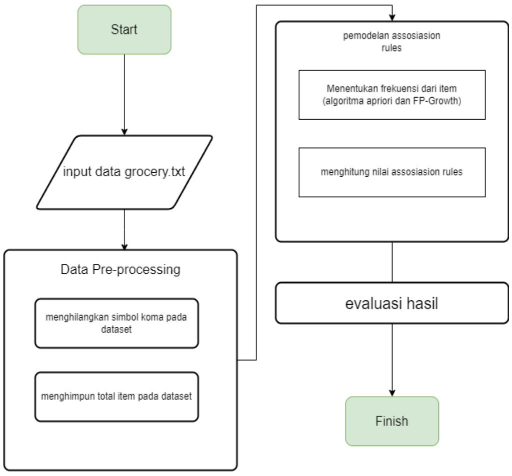

Figure 1. *Research Flowchart*

The research was conducted by inputting data acquired in .txt format. Processing was conducted utilizing the Google Colab platform. In the subsequent phase, the data underwent preprocessing to eliminate superfluous parts. Subsequently, association rule modeling was conducted utilizing the Apriori and FP-Growth algorithms. The final phase involves assessing the outcomes by juxtaposing the performance metrics of each algorithm.

Data Pre-Proccessing

Data preparation involves transforming or encoding data to enable accurate representation, facilitating rapid comprehension by machines. Data preparation can be defined as a procedure whereby an algorithmic model rapidly analyzes the characteristics of the data [16]. In this investigation, the dataset underwent preprocessing to exclude the comma symbols. Subsequently, the data was aggregated by quantifying each item to derive the top 10, representing the most often occurring entries. The data is displayed in Table 1 and illustrated in Figure 2.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 101 – 108 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3164
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

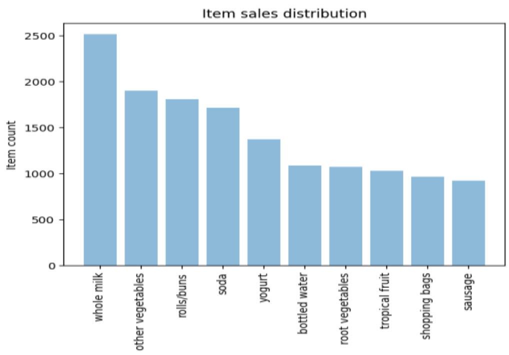

Figure 2. Research Flowchart

Tabel 1. Top 10 item with the highest total

No	Item Name	Total
1	Whole Milk	2513
2	Other Vegetables	1903
3	Rolls/Buns	1809
4	Soda	1715
5	Yogurt	1372
6	Bottled Water	1087
7	Root Vegetables	1072
8	Tropical Fruit	1032
9	Shopping Bags	969
10	Sausage	924

Assosiasion Rules

In association analysis, associations between combinations of items are discerned through a data mining method. A primary metric employed in this methodology is support, which quantifies the frequency of occurrence of an item or a combination of items within the entire dataset. The frequency is established by utilizing a formula to compute the ratio of occurrences within the transactions [17].

$$Support(A) = \frac{\text{Jumlah transaksi pada A}}{Total\ Transaksi} \quad (1)$$

Then to determine the support value of 2 items can be determined using the formula:

$$Support(A,B) = \frac{\sum TransaksimengandungAdanB}{\sum \text{Transaksi}} \quad (2)$$

Confidence is a metric that describes how often two items appear together in a transaction. Confidence is expressed as item A has appeared in a transaction. To calculate confidence, the following formula is used:

$$P(B|A) = \frac{\sum TransaksimengandungAdanB}{Totaltransaksi} \quad (3)$$

FP-Growth Algorithm

FP-Growth is employed to ascertain the likelihood of a rule manifesting across all transactions, predicated on the support value. FP-Growth consists of three stages: The initial phase, termed Conditional Pattern Base Generation, encompasses a subdatabase referred to as the Conditional Pattern Base. This Conditional Pattern Base comprises the original path and the resultant pattern derived from the FP-tree.

The second stage involves calculating the support parameters derived from the conditional pattern, after which items above the minimal support threshold will be incorporated into the FP-Tree. The Third Stage, known as the Frequent Itemset Search, can identify itemset rules when the Conditional FP-tree consists of a single path by amalgamating items for each Conditional FP-tree. Nonetheless, if not, the FP-Growth generation occurs recursively. [18].

Apriori Algorithm

The apriori algorithm is a method used to identify the highest frequency of a list of things. This algorithm comprises two phases. The first stage entails assessing the highest frequency pattern by looking for a combination of each item that meet the support value of an item.

Tabel 2. Nilai *Confidence* (Apriori)

No	Antecedent (Frozenset)	Consequent (Frozenset)	Support	Confidence	Lift
1	Whole Milk	Other Vegetables	0.0748347737	0.29287703939	1.5136340948
2	Other Vegetables	Whole Milk	0.0748347737	0.38675775091	1.5136340948
3	Yogurt	Whole Milk	0.0560244026	0.40160349854	1.5717351405
4	Whole Milk	Yogurt	0.0560244026	0.21925984878	1.5717351405
5	Root Vegetables	Whole Milk	0.0489069649	0.44869402985	1.7560309524
6	Whole Milk	Root Vegetables	0.0489069649	0.19140469558	1.7560309524
7	Rolls/Buns	Whole Milk	0.0566344687	0.30790491984	1.2050317893
8	Whole Milk	Rolls/Buns	0.0566344687	0.22164743334	1.2050317893
9	Other Vegetables	Root Vegetables	0.0473817996	0.24487651077	2.2466049285
10	Root Vegetables	Other Vegetables	0.0473817996	0.43470149253	2.2466049285

Tabel 3.Nilai Confidence (FP-Growth)

No	Antecedent (Frozenset)	Consequent (Frozenset)	Support	Confidence	Lift
1	Other Vegetables	Whole Milk	0.0748347737	0.38675775091	1.513634094
2	Whole Milk	Other Vegetables	0.0748347737	0.29287703939	1.513634094
3	Rolls/Buns	Whole Milk	0.0566344687	0.30790491984	1.205031789
4	Whole Milk	Rolls/Buns	0.0566344687	0.22164743334	1.205031789
5	Whole Milk	Yogurt	0.0560244026	0.21925984878	1.571735140
6	Yogurt	Whole Milk	0.0560244026	0.40160349854	1.571735140
7	Root Vegetables	Other Vegetables	0.0473817996	0.43470149253	2.246604928
8	Other Vegetables	Root Vegetables	0.0473817996	0.24487651077	2.246604928
9	Root Vegetables	Whole Milk	0.0489069649	0.44869402985	1.756030952
10	Whole Milk	Root Vegetables	0.0489069649	0.19140469558	1.756030952

This phase may employ the subsequent formula [19]:

$$Support(A,B)=\frac{\text{Jumlah transaksi pada variabel A dan B}}{TotalTransaksi}\,x\,100\% \quad (4)$$

After finding the highest frequency pattern, the next step is to find the association rule that meets the minimum requirements by calculating the existing association level. To calculate the Confidence value, the following formula is used:

$$Confidence\ P(B|A)=\frac{\sum TransaksimengandungAdanB}{totalteransaksi}\,x\,100\% \quad (5)$$

RESULT & DISCUSSION

The analyzed dataset employs a rules parameter with a minimum support threshold of 0.45. The associations are presented in order, showcasing the top 10 associations based on their Confidence values. The confidence values obtained from the apriori algorithm and the FP-Growth algorithm are displayed in Tables 2 and 3.

Table 2 indicates that the association with the highest support value is Whole milk -> Other vegetables, with a support value of 0.0748347737. The support value between whole milk and other vegetables is similar, yet the confidence value of each association varies despite having the same lift value.

Referring to table 3, it can be seen that the association with the highest support value is the other vegetables -> Whole milk association with a support value of 0.0748347737. The support value between whole milk and other vegetables is similar; however, the confidence value of each association varies despite having the same lift value. The analysis of the two algorithms reveals differences in the order of associations produced; however, both conclude the same association: other vegetables -> Whole milk, with a support value of 0.0748347737. The distinction in the order of association arises from the inherent characteristics of each algorithm. Apriori emphasizes exploring possibilities based on prior transactions, whereas FP-Growth directly seeks frequent itemsets, utilizing the FP-Tree structure. FP-Growth demonstrates greater efficiency than Apriori when evaluated based on speed. Thus, in the case of compiling sales goods, it can refer to the results of apriori because it is based on the transaction history of buyers who have more insight in seeing buyer purchasing patterns based on data. Thus, the analysis of the Grocery sales market basket is more flexible using apriori.

CONCLUSION

The analysis of the Grocery dataset employed two algorithms, Apriori and FP-Growth, with a minimum support threshold of 0.45. The results are organized according to the top 10 associations exhibiting the highest confidence values. The association with the highest support value found is "Whole Milk -> Other Vegetables" with a support value of 0.0748347737. The analysis indicates that both algorithms yield the same association "Other Vegetables -> Whole Milk" with a Support value of 0.0748347737. FP-Growth demonstrates superior efficiency compared to Apriori regarding processing speed.

The algorithms yield comparable association results; however, the sequence varies owing to their distinct characteristics. Apriori identifies potential associations derived from prior transactions, whereas FP-Growth conducts a direct search for frequent item-sets utilizing FP-Tree. The association ranking results provide valuable insights for grocery owners regarding optimal product combinations, potentially enhancing sales through data-driven marketing techniques derived from previously sold items. This permits the inclusion of multiple items or the formation of a package. The placement position of the item may indicate the outcomes of the conducted analysis.

BIBLIOGRAPHY

[1] A. Setiawan And and R. Mulyanti, "Market Basket Analysis Dengan Algoritma Apriori Pada Ecommerce Toko Busana Muslim Trendy (Market Basket Analysis With Apriori Algorithms In Ecommerce Trendy Muslim Clothing Stores)," *Juita: Jurnal Informatika*, vol. 8, pp. 11–18, 2020.

[2] H. Hernawati, "Analisis Market Basket Dengan Algoritma Apriori (Study Kasus Toko Alief)," *Ikraith Informatika*, vol. 2, no. 1, pp. 13–17, 2018.

[3] M. Prabukusumo And and N. Azhar, "Pemodelan Pola Belanja Pelanggan Produk Infrastruktur Dan Security Menggunakan Algoritma Fp-Growth," *Jurnal Ilmiah Komputasi*, vol. 21, pp. 305–316,

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 101 – 108 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3164
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

2022.

[4] I. W. Anwar Arifin, "PERBAIKAN TATA LETAK PASAR INDUK TRADISIONAL DI SANGATTA DENGAN METODE MARKET BASKET ANALYSIS (MBA)," *Res. J. Account. Bus. Manag.*, vol. 5, no. 2, p. 113, 2022.

[5] A. I. Idris Et, "Comparison Of Apriori, Apriori-Tid And Fp-Growth Algorithms In Market Basket Analysis At Grocery Stores," *Ijics*, vol. 6, no. 2, 2022.

[6] H. Harianto And and H. Eddy, "Analisa Data Transaksi Penjualan Barang Menggunakan Algoritme Apriori Dan Fp-Growth," *Jnanaloka*, pp. 35–43, 2020.

[7] A. N. Sagin And and B. Ayvaz, "Determination Of Association Rules With Market Basket Analysis: Application In The Retail Sector," *Southeast Eur. J. Soft Comput*, vol. 7, no. 1, 2018.

[8] H. Patron, "A Market Basket Analysis Of The Usauto-Repair Industry," *Journal Of Business Analytics Taylor & Francis*, vol. 3, no. 2, pp. 79–92.

[9] G. Shmueli, *Data Mining For Bussiness Analysis*. Hoboken: John Wiley & Sons, 2018.

[10] A. Aldi And and D. Fitrianah, "Penerapan Algoritma Fp-Growth Rekomendasi Trend Penjualan Atk Pada Cv. Fajar Sukses Abadi," *Jurnal Telekomunikasi Dan Komputer*, vol. 9, no. 1, pp. 49–60, 2019.

[11] A. H. Nasyuha Et, "Frequent Pattern Growth Algorithm For Maximizing Display Items," *Telkomnika*, vol. 19, no. 2, 2020.

[12] S. Hu, Q. Liang, H. Qian, J. Weng, W. Zhou, and P. Lin, "Frequent-pattern growth algorithm based association rule mining method of public transport travel stability," *Int. J. Sustain. Transp.*, vol. 15, no. 11, pp. 879–892, 2021.

[13] A. Wijaya and A. Raka, *Mencari Pola Pembelian Konsumen Menggunakan Algoritma Fp-Growth*. 2018.

[14] H. Mauliiya And and A. Jananto, *Asosiasi Data Mining Menggunakan Algoritma Apriori Dan Fpgrowth Sebagai Dasar Pertimbangan Penentuan Paket Sembako*. 2020.

[15] A. Harahap, A. L. R. Perangin-Angin, K. Kumar, and S. P. Parsaoran, "ANALISIS PENERAPAN DATA MINING DALAM PENENTUAN TATA LETAK BARANG MENGGUNAKAN ALGORITMA APRIORI DAN FP-GROWTH," *Tekinkom*, vol. 5, no. 2, p. 291, 2022.

[16] K. Maharana, S. Mondal, and B. Nemade, "A review: Data preprocessing and data augmentation techniques," *Global Transitions Proceedings*, 2022.

[17] R. Ramadhan And and E. I. Setiawan, "Market Basket Analysis Untuk Swalayan Ksu Sumber Makmur Dengan Algoritma Fp Growth," *Journal Of Intelligent System And Computation*, vol. 2, no. 1, pp. 34–39, 2021.

[18] I. Musdalifah And and A. Jananto, "Analisis Perbandingan Algoritma Apriori Dan Fp-Growth Dalam Pembentukan Pola Asosiasi

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 101 – 108 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3164
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Keranjang Belanja Pelanggan," *Progresif J. Ilmi. Kom*, vol. 18, no. 2, 2022.

[19] A. R. Wibowo And and A. Jananto, "Implementasi Data Mining Metode Asosiasi Algoritma Fp Growth Pada Perusahaan Ritel"," *Jurnal Teknologi Informasi Dan Komunikasi*, vol. 10, pp. 200–212, 2020.