Institute of Research and Publication Indonesia (IRPI)
Public Research Journal of Engineering, Data Technology and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/predatecs
Vol. 1 Iss. 2 January 2024, pp: 95-101
ISSN(P): 3024-921X | ISSN(E): 3024-8043

Implementation of Association Rules Algorithm to Identify Popular Topping Combinations in Orders

Rizki Aulia Putra[1*], Margareta Amalia Miranti Putri[2], Sri Maharani Sinaga[3], Sania Fitri Octavia[4], Raihan Catur Rachman[5]

[1234]Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[4]The One Academy Petaling Jaya, Malaysia

E-Mail:[1]12050313356@students.uin-suska.ac.id,
[2]12050320348@students.uin-suska.ac.id, [3]12050326545@students.uin-suska.ac.id,
[4]saniafitrioctavia@gmail.com, [5]raihancaturr@gmail.com

Received Aug 10th 2023; Revised Dec 30th 2023; Accepted Apr 15th 2023
Corresponding Author: Rizki Aulia Putra

Abstract

Association rule is a data mining technique to find associative rules between a combination of items. This research aims to apply association rules algorithm in identifying popular topping combinations in food orders. This application aims to help restaurant owners or food businesses understand their customers' preferences and optimize their menu offerings. Data obtained from kaggle, the association rules algorithm is applied to this dataset to identify patterns or combinations of toppings that often appear together in orders. The results of this study show toppings with chocolate as a popular item in orders. These findings can provide valuable insights for food business owners in structuring their menus and determining attractive offers for customers. This study also applied a comparison between the apriori, fp- growth and eclat algorithms, with the result that the best item transaction rule was found: a combination of dill & unicorn toppings with chocolate with 60% confidence. Overall, the application of eclat algorithm in this study provides the best performance with higher execution speed, thus providing insight into customer preferences regarding topping combinations in food orders. Despite the shortcomings of the data form from this study, it is expected to help business owners in optimizing their offerings, increasing customer satisfaction, and improving their business performance.

Keyword: Apriori Algorithm, Association Rule Mining, Eclat Algorithm, FP Growth Algorithm, Market Basket Analysis

1. INTRODUCTION

Machine Learning (ML) and Artificial Intelligence (AI) have come a long way in data analysis and computation in recent years, which usually enables applications to work smart [1]. Data mining and its methods exist to achieve goals with the advancement of information technology and the need to extract useful information from datasets [2]. Due to the large amount of data stored in data warehouses, databases, online analytics processes, and other data storage places, one can try to search for information manually, which can take a lot of time [3].

Market basket analysis is a data mining technique that is widely used to discover associations between products. Market basket analysis gives the provider useful information about the products that its customers put together. This information can be used to influence customer purchases, rearrange retail store layouts, allocate products on shelves, etc. [4]. The results of market basket analysis allow companies to make easier decisions about how to bundle products and how to reorganize stores in a better way to increase sales and better market analysis results [5].

The association rule identification process performed on the market basket is a series of actions that include finding frequency itemsets, correlations, and relationships between data, as well as finding the relationship rules themselves using appropriate algorithms [6]. Association rules can play an important role in the management of business processes that require decision making, and thus can significantly affect the business of a particular company [7]. Association rule mining is the process of extracting valuable knowledge that describes the relationships between data items from large amounts of data [8]. As one of the most classic methods in data mining, it is widely used in various industries. For example, when applied to the commercial retail industry, it analyzes customer purchasing characteristics to guide shelf arrangement, pricing, and commodity purchasing [3].

As the amount of data generated by orders in distribution companies increases every day, attention needs to be paid to the selection of topping patterns for distribution. Because this selection can greatly affect company performance [7]. This research aims to find the best algorithm based on support and confidence and other parameters in finding the best frequent itemset pattern of topping combinations so that it helps and supports sellers in making decisions [9] regarding marketing strategies, so that customer satisfaction can be achieved [1].

Previous research conducted [10] Through the analysis and research of the Apriori algorithm, it was found that when the execution of the improved Apriori algorithm was verified through specific examples, indicating that this method is feasible to use. As for FP Growth, it is shown that the effect of association rules and the effectiveness of identification algorithms in the field of energy saving greatly increase the energy saving potential of equipment, reducing the amount of time required to pass through the data set when creating frequent and complex item sets [11]. Therefore, both algorithms have fast execution times depending on the amount of data used (a priori is faster with less data used). The application of Eclat Algorithm, by handling large data in less time and documentation of association rules can help and support sellers in making decisions regarding marketing strategies. hence, customer satisfaction is achieved which will increase revenue [5].

In this research, the basic concepts of association rules are briefly presented, which is the starting point for detailed analysis and application of the above-mentioned on the presented real data set. After selecting a suitable algorithm and its application to find frequent itemsets [4][12], the last step is to identify and visualize the association rules. In the conclusion, the parameters used are support, which is the likelihood of a rule and confidence, the degree of trust in a rule obtained high results [13]. As well as analysis and guidelines that aim to assist sellers in providing the best combination of toppings to get customer satisfaction based on these research tests and provide future improvements and possible research [7].

2. MATERIAL AND METHOD

This research uses 3 algorithms in modelling, namely Apriori, FP-Growth and Eclat. Some important parts of this research consist of collecting data, pre-processing data, modelling process with the proposed algorithm, analysing and representing the results of the research. For more details, the flowchat of the methodology is shown in Figure 1.

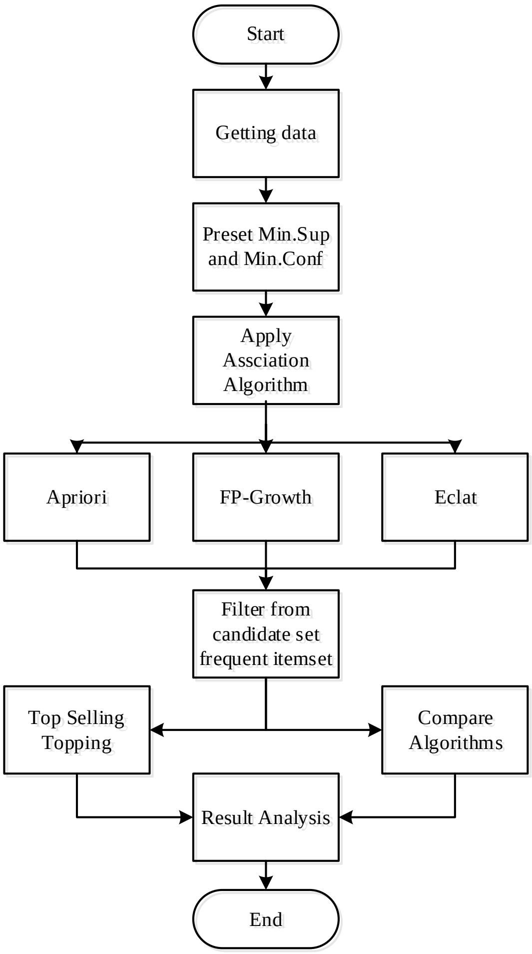

Figure 1. Research Methodology

2.1 Association Rules

Association rule mining is an important part of data mining technology. It is mainly used to find implicit relationships between multiple data, so as to describe the laws and patterns of multiple attributes simultaneously [11]. These rules can be expressed in the form X → Y, which means that when itemset X appears, itemset Y may also appear. The purpose of association rule learning is to find a collection of frequently occurring itemsets from those itemsets [6] [3].

Decision Assessment:

1. Support (S): The fraction of the number of transactions X and Y that belong to the set of transactions with the sum of all transactions.

$$Support\ (X \rightarrow Y) = Support\ (X \cup Y) = P(XY) \tag{1}$$

2. Confidence (C): tests the number of products in Y in transactions containing X, the number of transactions containing X and Y against the total number of transactions containing X in the transaction set.

$$Confidence\ (X \rightarrow Y) = \frac{Support\ (X \cup Y)}{Support\ X} = P(Y|X) \tag{2}$$

3. Itemset: one or more items obtained.
4. Frequent itemset: a collection of items where the support is greater than or equal to the min support limit.

Support reflects the probability that items contained in X and Y appear in the transaction set at the same time. Confidence reflects the conditional probability of occurrence of Y in transactions containing X. That is, the proportion of (X ∪ Y) support to (X) support [3]. In this research, we use data from kaggle.com, which will be tested to determine the two results of our research, namely the best toping combination and the results of using and comparing the three association algorithms with predetermined support and confidence.

2.2 Apriori Algorithm

Apriori algorithm is used to find frequent item sets with iteration layers to get association rules [14] . The algorithm terminates when empty frequent item sets are found [15]. An iterative technique called a breadth-first search (level-wise search) with which search area is used to check (k+1) itemsets [6]. This algorithm is recursive and is based on the frequent set rule generation process and the goal is to find the largest frequent set of K items. Often, the set evaluation is based on support and confidence [3]. The Formula is as below[16]:

1. Support Formulas

$$\text{Support (A,B)} = \frac{Number\ Of\ Transactions\ A\ dan\ B}{\sum Transactions} \text{ x 100 \%} \tag{3}$$

2. Confidence Formula

$$\text{Confidence (A,B)} = \frac{Number\ Of\ Transactions\ A\ dan\ B}{\sum Transactions\ A} \text{ X 100\%} \tag{4}$$

3. Lifrt Ratio Formula

$$\text{Lift Rasio} = \frac{Support\ (A,B)}{Support\ A\ x\ Support\ B} \tag{5}$$

2.3 FP-Growth Algorithm

FP Growth follows a divide-and-conquer strategy. First, it builds a frequency pattern tree or FP- tree by retrieving frequently occurring items sorted in order of support count and then uses the FP-tree to obtain association information. The best advantage of FP Growth is that it scans the database only twice and does not generate a large number of candidate sets [4].

In the process of generating FP Tree, it only needs to scan the database twice to get frequent item sets and build FP Tree. The FP Growth algorithm changes the problem from finding long frequent patterns to multiple short patterns recursively, and then connecting suffixes. At the same time, the FP Growth algorithm greatly reduces the search time [8]. The data analysis technique is realized by applying the FP-Growth algorithm. This algorithm has three stages of work that is [17] :

1. Generating a conditional pattern base
2. Generating conditional FP-Tree
3. Search for frequent item sets

2.4 Eclat Algorithm

Eclat algorithm is an algorithm based on depth-first search (DFS). It uses a vertical database design i.e., rather than explicitly including all transactions; it can be stored along with their distribution & uses a crossing point based way to handle calculating the support of an itemset [6]. The Formula is as below [18]:

$$S(X) = \frac{N_x}{N} \tag{6}$$

$$C(X \rightarrow Y) = \frac{S(X \rightarrow Y)}{S(X)} \tag{7}$$

$$L(X \rightarrow Y) = \frac{S(X \rightarrow Y)}{S(X).S(Y)} \tag{8}$$

Where:

X = Antecedent
Y = Consequent
N_x = Number of incidents with pattern X
N = Total number of incidents
$S(X)$ = Support of pattern X
$S(X \rightarrow Y)$ = Support of frequent pattern formed by X and Y
$C(X \rightarrow Y)$ = Confidence of the association rule ðX ! YÞ
$L(X \rightarrow Y)$ = Lift of the association rule ðX ! YÞ

2.5 Dataset Description

In our research, we used a dataset obtained from Kaggle. With the selection of data features taken, namely <1000 customer transaction data orders, for view dataset at table 1.

Table 1. Dataset Description

Dataset Name	No. of Samples	No.of Attributes
Market Basket Analysis	999 (non-null)	16 (bool)

The data is used to find the best combination of items and comparative evaluation of association algorithms. Each attribute is binominal with no blank data.

3. RESULTS AND ANALYSIS

3.1 Rule Algorithm

In this study, we obtained rule results for each algorithm with the same results and we also compared the execution time of the Apriori, FP-Growth and eclat algorithms. With a minimum support of 10% and a minimum confidence of 60%. The results of modelling the Apriori, FP-Growth and Eclat algorithms based on predetermined parameters are with the best formulation at the support value of 10% and minimum confidence of 60%. Each is shown in table 2, table 3 and table 4.

Table 2. Results of Apriori Algorithm

Antecedents	Consequents	Support	Confidence	Lift
Dill, Unicorn	chocolate	0,101101	0,60119	1,426578
Dill, Milk	chocolate	0,114114	0,6	1,423753

Table 3. Results of FP-Growth Algorithm

Antecedents	Consequents	Support	Confidence	Lift
Dill, Unicorn	chocolate	0,101101	0,60119	1,426578
Dilk, Milk	chocolate	0,114114	0,6	1,423753

Table 4. Results of Eclat Algorithm

Antecedents	Consequents	Support	Confidence	Lift
Dill, Unicorn	chocolate	0,101101	0,60119	1,426578
Dilk, Milk	chocolate	0,114114	0,6	1,423753

The results of the three rules from each algorithm, resulted in the same combination or rule. The combination of Dill, Unicorn and Chocolate items is the best, so these results can be analyzed in the future. To get the algorithm in this study, it is used by comparing the execution time of each algorithm in carrying out the calculation process. Comparison of 3 association algorithms was done with minimum support and confidence kept at 10% & 60% for all experiments..

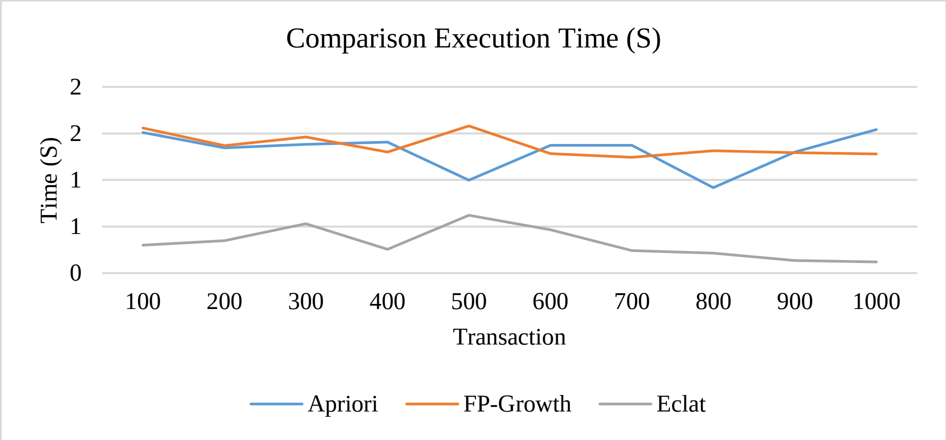

Figure 2. The Required Time for Apriori, FP Growth and Eclat Algorithms

Transactions come from a dataset that is broken into several sample data, with the execution speed of each algorithm application calculated. Eclat Algorithm was obtained with a speed of less than 1 second [19].

3.2 The Count of the Items in the Transactions

The data is processed to determine what toppings are popular among customers by calculating the highest number of transactions. The results show that the topping that is often purchased by customers is Chocolate with 421 transactions, then butter and yogurt have the same transaction results and Ice Cream and sugar have fewer transactions than the three previous toppings. By looking at the diagram above, it can be clearly seen which toppings are dominant and less attractive to buyers. The popular toping can view figure 3.

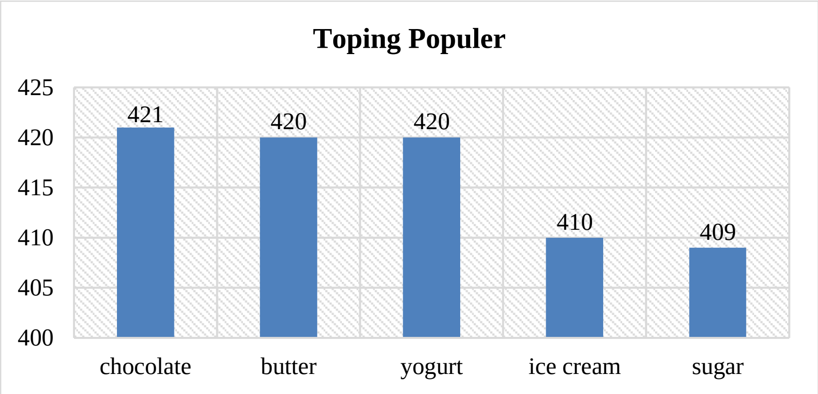

Figure 3. Top 5 Items

3.3 Rule Analysis

In the collection of processed data, all experiments were carried out on a Laptop with an Intel(R) Core(™) i3-10110U CPU @2.10GHz (4CPUs) ~2.6GHz main processor, 4 GB main memory and running the Microsoft Windows 11 operating system. All analysis was carried out using the python programming language. Minimum Support 10% and Minimum Confidence 60%, the results of the association rule algorithm implementation are obtained, table 5.

Table 5. Final Rule

Antecedents	Consequents	Antecedents of support	Consequents support	Support	Confidence
Dill, Unicorn	Chocolate	0,168	0,421	0,101	0,601
Dill, Milk	Chocolate	0,190	0,421	0,114	0,600

Rule 1 shows that when a customer buys a combination of Dill & Unicorn toppings, there is a 60% chance that they will buy Chocolate toppings, where Dill & Unicorn toppings are believed to appear in the same transaction set with 16% support. Rule 2 also explains that the Dill & Milk topping is 60% likely to buy the Chocolate topping. It is believed that Chocolate topping will be a key transaction item in the order with high confidence and support compared to other items, and it is explained that this item has a high occurrence in transactions [20].

4. CONCLUSION

From the results of the research analysis, it was found that chocolate topping is the most favorite topping with a total of 421 transactions, so chocolate topping is key in the combination of items in each topping purchase. With the combination of items purchased, namely dill, unicorn, there is a 60% chance of buying chocolate and the second item combination dill, milk with a 60% chance of buying chocolate. For the implementation of association algorithms used in this research shows, on the dataset tested, the best algorithms, namely, apriori, fpgrowth and eclat execute data quickly. However, the eclat algorithm has the advantage of faster execution time than other algorithms. Based on our view, that the data used has not been maximized in managing the algorithm so that for further research can use better and more data in terms of transactions and attributes or items of each transaction.

REFERENCES

[1] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Computer Science*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[2] M. H. Santoso, “Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom,” *Brilliance: Research of Artificial Intelligence*, vol. 1, no. 2, pp. 54–66, 2021, doi: 10.47709/brilliance.v1i2.1228.

[3] X. Zhang and J. Zhang, “Analysis and research on library user behavior based on apriori algorithm,” *Measurement: Sensors*, vol. 27, no. February, p. 100802, 2023, doi: 10.1016/j.measen.2023.100802.

[4] M. Hossain, A. H. M. S. Sattar, and M. K. Paul, “Market basket analysis using apriori and FP growth algorithm,” *2019 22nd International Conference on Computer and Information Technology, ICCIT 2019*, pp. 18–20, 2019, doi: 10.1109/ICCIT48885.2019.9038197.

[5] D. Al Attal, M. Naser, N. AlBaghli, N. Al Muhaimeed, and S. A. Al Awadh, “Redesigning a retail store based on association rule mining,” *Proceedings of the International Conference on Industrial Engineering and Operations Management*, vol. 2018, no. JUL, pp. 1948–1965, 2018.

[6] P. Matapurkar and S. Shrivastava, “Comparative Analysis for Mining Fuzzified Dataset Using Association Rule Mining Approach,” *ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, pp. 383–387, 2020, doi: 10.1109/ICRITO48877.2020.9198028.

[7] A. Nogo, E. Žunić, and D. Donko, “Identification of association rules in orders of distribution companies’ clients,” *EUROCON 2019 - 18th International Conference on Smart Technologies*, vol. 1, no. 1, pp. 1–4, 2019, doi: 10.1109/EUROCON.2019.8861951.

[8] Lingxizhu, Yufeiguo, and Jingyiwang, “Application of FP_Growth Algorithm of Sequential Pattern Mining on Container Maintenance Components Association,” *Proceedings - 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2020*, pp. 1026–1031, 2020, doi: 10.1109/CISP-BMEI51763.2020.9263649.

[9] S. Halim, T. Octavia, and C. Alianto, “Designing facility layout of an amusement arcade using market basket analysis,” *Procedia Computer Science*, vol. 161, pp. 623–629, 2019, doi: 10.1016/j.procs.2019.11.165.

[10] H. Bin Wang and Y. J. Gao, “Research on parallelization of Apriori algorithm in association rule mining,” *Procedia Computer Science*, vol. 183, pp. 641–647, 2021, doi: 10.1016/j.procs.2021.02.109.

[11] R. Qi and X. Guo, “Analysis of Intelligent Energy Saving Strategy of 4G/5G Network Based on FP-Tree,” *Procedia Computer Science*, vol. 198, no. 2021, pp. 486–492, 2021, doi: 10.1016/j.procs.2021.12.274.

[12] S. Kumar and K. K. Mohbey, “A review on big data based parallel and distributed approaches of pattern mining,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1639–1662, 2022, doi: 10.1016/j.jksuci.2019.09.006.

[13] M. Tian *et al.*, “Data Dependence Analysis for Defects Data of Relay Protection Devices Based on Apriori Algorithm,” *IEEE Access*, vol. 8, pp. 120647–120653, 2020, doi: 10.1109/ACCESS.2020.3006345.

[14] X. Wang, C. Song, W. Xiong, and X. Lv, “Evaluation of Flotation Working Condition Recognition Based on An Improved Apriori Algorithm,” Elsevier B.V., Jan. 2018, pp. 129–134. doi: 10.1016/j.ifacol.2018.09.404.

[15] Y. Zhang and L. Wang, “A optimization algorithm for association rule based on spark platform,” *Proceedings - 2020 International Conference on Computer Network, Electronic and Automation, ICCNEA 2020*, no. 2, pp. 82–86, 2020, doi: 10.1109/ICCNEA50255.2020.00026.

[16] Moch. Syahrir and L. Z. A. Mardedi, “Determination of the best rule-based analysis results from the comparison of the Fp-Growth, Apriori, and TPQ-Apriori Algorithms for recommendation systems,” *MATRIX : Jurnal Manajemen Teknologi dan Informatika*, vol. 13, no. 2, pp. 52–67, Jul. 2023, doi: 10.31940/matrix.v13i2.52-67.

[17] S. Bagui, K. Devulapalli, and J. Coffey, “A heuristic approach for load balancing the FP-growth algorithm on MapReduce,” *Array*, vol. 7, p. 100035, Sep. 2020, doi: 10.1016/j.array.2020.100035.

[18] S. Das, A. Dutta, M. Jalayer, A. Bibeka, and L. Wu, “Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using ‘Eclat’ association rules to promote safety,” *International Journal of Transportation Science and Technology*, vol. 7, no. 2, pp. 114–123, Jun. 2018, doi: 10.1016/j.ijtst.2018.02.001.

[19] T. Rahman, M. M. J. Kabir, and M. Kabir, “Performance Evaluation of Fuzzy Association Rule Mining Algorithms,” *2019 4th International Conference on Electrical Information and Communication Technology, EICT 2019*, no. December, pp. 20–22, 2019, doi: 10.1109/EICT48899.2019.9068771.

[20] 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,” *International Journal of Sustainable Transportation*, vol. 15, no. 11, pp. 879–892, 2021, doi: 10.1080/15568318.2020.1827318.