Journal of Computing Innovations and Emerging Technology

Volume 1 | Issue 1 | July (2025)

Published by **Novamind Press**

JCIET

Journal of Computing Innovations and Emerging Technology

Analyzing the Effectiveness of Apriori and ECLAT Algorithms in Frequent Itemset Mining

Dhina Puspasari Wijaya [1*], Pipit Febriana Dewi[2], Nurul Mega Saraswati[3]

[1] *Department of Informatic, Computer and Engineering Faculty, Universitas Alma Ata, Yogyakarta, Indonesia*
[2] *Universitas Nahdatul Ulama Yogyakarta, Indonesia*
[3] *Department of Informatic, Science and Technology Faculty, Universitas Peradaban, Brebes, Indonesia*
**Corresponding author: dhina.puspa@almaata.ac.id*

***Abstract* — Consumer transaction data recorded through the point of sale (POS) system contains purchase patterns that can be used to support marketing strategies. In the context of convenience stores that have high transaction volumes and large product variations, product association analysis is important to uncover customers' shopping habits. This study compares the effectiveness of a priori and ECLAT algorithms in conducting frequent itemset mining on transaction data of daily necessities. Both algorithms were evaluated based on the number of rules generated, support value, confidence, lift, and execution time efficiency. The dataset used is Groceries, which represents actual transactions in the retail environment. The results showed that although a priori and ECLAT produced 25 rules. ECLAT was superior in execution times four times faster than Apriori without compromising the quality of the rules. Most rules have a confidence between 30–50% and a lift above 1.5, signifying a meaningful association. This study concludes that ECLAT is more suitable for use in complex and dynamic minimarket transaction data scenarios, and is recommended as the basis for the development of product recommendation systems and association-based bundling strategies.**

Keywords—Apriori, Association, ECLAT, Rules, Transaction

Manuscript received 2 July 2025; revised 7 July 2025; accepted 10 July 2025. Date of publication 11 July 2025.

I. INTRODUCTION

Consumer transaction data that is automatically recorded in the point of sale (POS) system has become an important asset in business decision-making. This data stores hidden information that, if processed with the right method, can provide strategic insights such as consumer spending patterns, product purchasing tendencies, and seasonal preferences [1]. One popular approach in data mining that can dig into knowledge from such data is frequent itemset mining, or itemset search that often appears in transactions. Frequent itemset mining is the core of association rule mining, which is a technique that aims to find relationships between items in a transaction database. This technique is particularly useful in the context of market basket analysis, where organizations can find out which products are frequently purchased at the same time. This information is very valuable for designing marketing strategies such as product placement, bundling, and a more personalized recommendation system.

The two most commonly used algorithms in the itemset search process that appear frequently are a priori and ECLAT. A priori, introduced by Agrawal and Srikant in 1994, uses a breadth-first search approach and the principle of downward closure, in which large itemset is built iteratively from small itemsets. Although conceptually simple, a priori has a weakness in computational efficiency due to the large number of candidate search processes and checks of the entire dataset. On the other hand, ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal) is an alternative algorithm that uses a depth-first search approach and tidset-based data representation (transaction ID set). This approach significantly reduces the amount of access to databases and is more efficient for large, dense datasets. However, ECLAT has limitations in terms of scalability against very sparse datasets or very large number of transactions. Although both have been widely used, the comparison of effectiveness between a priori and ECLAT in the context of time performance, the number of rules generated, and the quality of the rules (seen from the value of support, confidence, and lift) are still relevant topics for further research. Real transaction data such as the purchase of daily necessities, the effectiveness of algorithms is key in determining the best approach to use to extract valuable purchase patterns. In the context of minimarkets, sales transactions that occur every day produce large amounts of data and are dynamic. Each purchase note records the combination of products purchased simultaneously by the consumer. If analyzed properly, these purchasing patterns can reveal hidden relationships between products, such as consumers' tendency to buy instant noodles with eggs, or coffee with bread. This kind of information can be used for various strategic purposes, such as more effective shelf arrangement, bundling product offerings, and the development of a recommendation system that increases sales.

However, the complex characteristics of minimarket transaction data, such as high transaction volumes, large item variations, and irregularities in purchasing patterns, demand the use of analysis methods that are not only accurate but also computationally efficient. Therefore, algorithms such as a priori and ECLAT are important to study in more depth to find out which methods are most suitable for exploring product associations in real convenience store scenarios. The use of the right algorithm not only improves the efficiency of the analysis process, but also has the potential to have a direct impact on improving operational performance and minimarket marketing strategies.

The following are some relevant research references to strengthen the research position where research shows an increased focus on the efficiency and scalability of frequent itemset mining algorithms, especially a priori and ECLAT. Several studies, highlight the classical limitations of A-priori algorithms in terms of memory efficiency and speed in large datasets [2] [3], encouraging the development of ECLAT-based solutions [4]. The ECLAT approach is considered computationally superior because it uses depth-first search and tidset intersection techniques, as evidenced by the implementation of PartECLAT and RDD-ECLAT on the Spark platform, which consistently results in faster execution times than a priori, especially on dense and distributed data [5]. Although ECLAT has execution times for large datasets [6], however, ECLAT has advantages in reading data vertically [7]. Comparison of the ECLAT, A Priori and FP-Growth methods shows the advantages of Eclat over a priori and Fpgrowth [8]. Another study showed that a priori it provided more rule options than ECLAT with a ratio of 9 rules and 11 rules, but program execution showed an advantage of ECLAT for 0.2 seconds [9]. A priori comparison of ECLAT and a priori in groceries data shows that the constant value of confidence and apriori tends to decrease [10]. The combination of Bloom Filters in Eclat is able to significantly increase the associative ability of Eclat [11]. In addition to the advantages of ECLAT, a priori it basically also still has a positive value or good performance as in the combination for data protection [12] as well as student performance data [13]. Other results also show significance when combined with the K-Means algorithm [14] or when with the SVM classification algorithm [15].

Overall it confirms that while a priori still has practical value on small or medium datasets, ECLAT and its variants are more feasible for use on large-scale data due to their improved efficiency and scalability. Research trends are also starting to shift towards algorithm optimization in the context of data distribution, parallel processing, and adaptation to real-time consumer behavior recommendation and analysis systems.

II. METHODOLOGY

The methodology and stages used in this study are as follows:

a. Experimental Design

The study uses a quantitative approach with computational experiments. The aim is to measure the efficiency of execution time and memory usage of a priori algorithm on datasets of varying sizes and characteristics. The results of a priori are also compared with other more modern algorithms such as FP-Growth to get an idea of their relative performance.

b. Dataset

The dataset used was obtained from an open source, namely Groceries data. The dataset was chosen because it has the number of transactions and the variety of items that reflect real-world scenarios. The dataset is divided into subsets to observe the impact of data scale on algorithm performance.

c. Data Pre-processing

Before the analysis is performed, the data goes through a cleanup process, including the removal of blank values, duplication, and normalization of items. The transaction format is then converted into a form that is appropriate for the association mining process, namely a basket representation (list of items per transaction).

d. Algorithm Implementation

Apriori algorithm is implemented compared to the results on ECLAT

e. Performance Measurement

The metric used to evaluate performance is execution time: the total time it takes for the algorithm to generate frequent itemsets and association rules, as well as the results of their rules.

f. Analysis and Visualization

The measurement results are analyzed descriptively and visualized in the form of a graph.

III. RESULT AND DISCUSSION

The data used in this study is data from open sources by containing transaction data as many as 3500 data. The example data is shown in Table *I*.

Table I. Example Dataset

StockCode	Desc	Qty	UnitPrice	Cust_ID
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2,55	17850

71053	WHITE METAL LANTERN	6	3,39	17850
84406B	CREAM CUPID HEARTS COAT HANGER	8	2,75	17850
84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	3,39	17850
84029E	RED WOOLLY HOTTIE WHITE HEART	6	3,39	17850

From this data, descriptive analysis was carried out first to identify the characteristics and distribution of the data before being processed at the next stage of research. The results show an average value of 4.41 items per transaction with a minimum value of 1 and a maximum value of 32. Visualization shown on Fig *1*

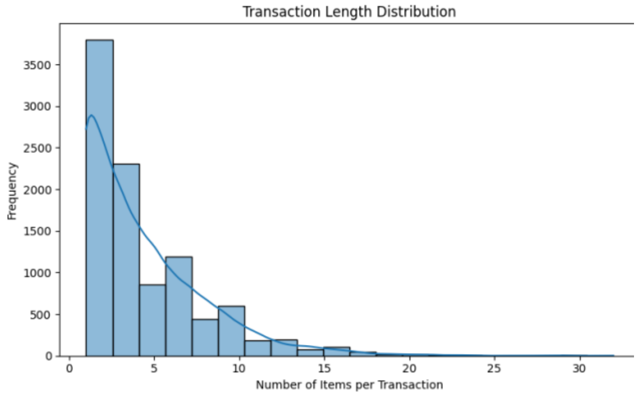

Fig 1. Data Distribution

The interpretation of the data states that the average consumer buys 4-5 items per transaction, indicating that the purchase of daily necessities is relatively small. The distribution is skewed to the right, as there are transactions of up to 32 items (possible monthly purchases or large stocks). The median (50%) is 3, meaning that more than half of the transaction is made up of $\leq$3 items. The fairly high variability (std 3.59) indicates that there is a segment of consumers with very different shopping patterns (small vs large). This is strengthened by the value of the results of apriori rules in Table *II*.

Table II. Rules Apriori

Antecedents	**Consequents**	**Support**	**Confidence**	**Lift**
beef	other vegetables	0.019725	0.375969	1.943066
beef	root vegetables	0.017387	0.331395	3.040367
beef	whole milk	0.021251	0.405039	1.585180
berries	other vegetables	0.010269	0.308869	1.596280
berries	whole milk	0.011795	0.354740	1.388328

The results of further analysis show that the results of the number of rules generated are the same as from a priori and Eclat with data of 120 rules in Fig *2*

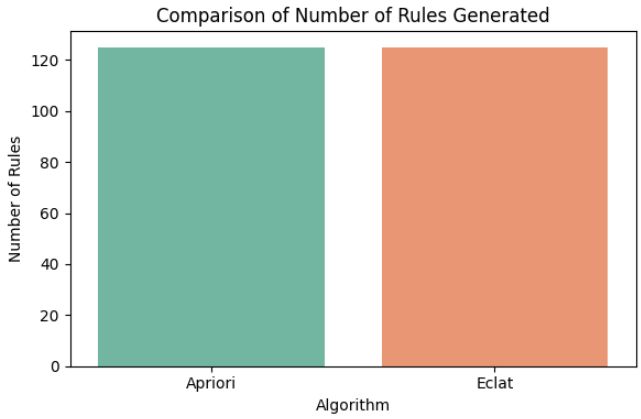

Fig 2. Number of Rules

Meanwhile, from the execution time aspect, it shows that the eclat algorithm is much faster than a priori with a priori execution time of 0.3560 seconds and Eclat of 0.0901 seconds or at Fig *3*

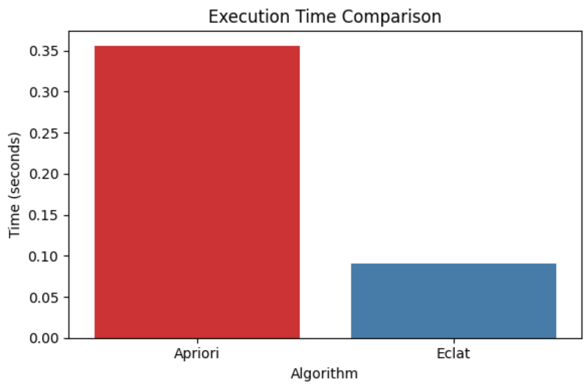

Fig 3. Execution Time

This analysis was continued with support, confidence and lift distribution with the results of each on Fig *4* Fig *5* Fig *6*.

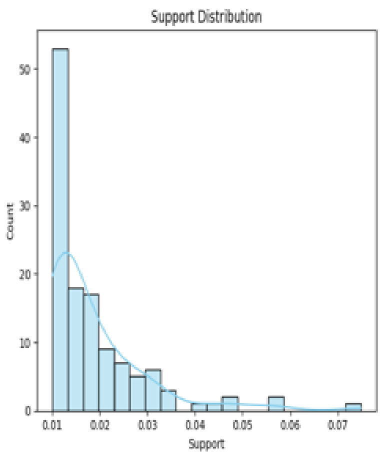

Fig 4. Support Distribution

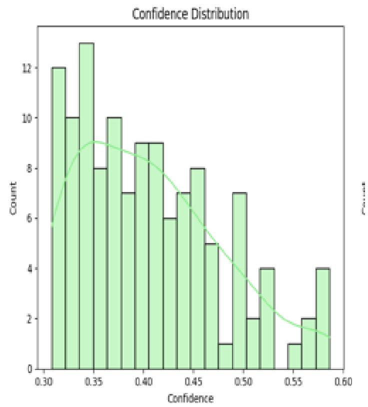

Fig 5. Confidence Distribution

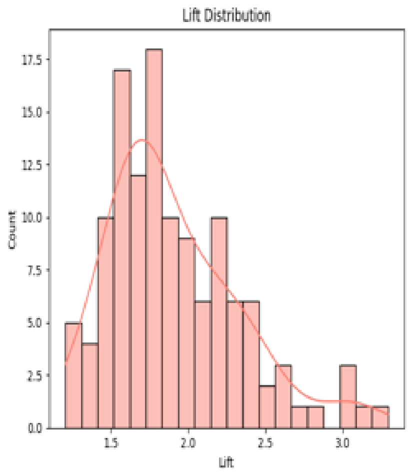

Fig 6. Lift Distribution

The results of the interpretation of the three charts are as follows:

Support Distribution

The majority of rules have low support (around 0.01–0.02). The distribution is right-skewed, indicating that few rules appear frequently in the transaction. This shows that the pattern of buying product combinations does not occur too often together in transactions. It is natural in the minimarket domain, because the variety of customer shopping is very high. However, rules with low support can still be valuable, especially if the confidence and lift are high.

Confidence Distribution

Most confidence is between 0.30–0.50, with some rules reaching ~0.58. The distribution is fairly even (not too skewed), indicating that the rules formed are quite consistent in their probability of occurrence. The rule is relatively well-established: if A is bought, the chances of B being bought are at 30–50%. This provides the basis for a less speculative recommendation system.

Lift Distribution

Elevators are mostly between 1.3–2.2, with peaks around 1.7–1.9. Some rules have lifts up to >3, which is very high. The whole >1 lift shows that all rules are positive associations not coincidences. A high lift signifies that the purchase of item A increases the probability of purchasing item B significantly compared to choosing randomly. Suitable for bundling analysis, promos, or product cross-recommendations.

The evaluation analysis uses the Kulczynski jaccard certainty zhangs_metric matrix where it will be paired with the Antecedents that have been formed with details 0 = beef, 1 = beef, 2 = beef, 3 = berries, 4 = berries. So the results are presented in Table *III*

Table III. Evaluation Model

Antecedents	Zhangs_Metric	Jaccard	Certainty	Kulczynski
0	0.512224	0.087191	0.226255	0.238957
1	0.708251	0.120677	0.249603	0.245455
2	0.389597	0.074113	0.200841	0.244103
3	0.386391	0.047440	0.143056	0.180971
4	0.289329	0.042584	0.133279	0.200450

Each matrix has a different result from Zhang's Metric & Jaccard showing that the higher this value gives the picture of a specific relationship, the better the correlation, so it can be concluded that the following points can be concluded on the rule:

1. Pair beef with root vegetables in promotions (e.g. 10% discount if you buy both)
2. Place whole milk and beef in close proximity on the shelf to reinforce the spontaneous buying association.
3. Use rules such as berries → whole milk with caution, as associations are weak and can be accidental.

The network visualization of each item is shown in the Fig *7*

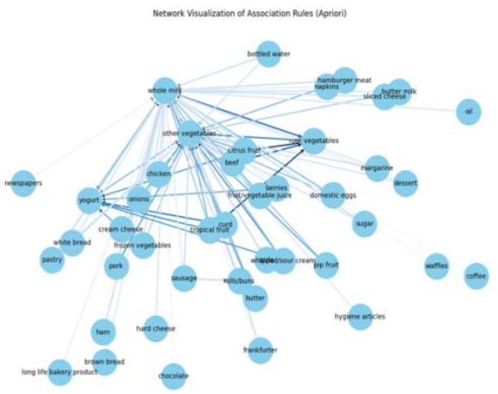

Fig 7. Network Visualization

IV. CONCLUSION

The results of the experiment in the research and analysis show the following points:

a. Statistical analysis shows that most convenience store customers make small transactions, with an average of only 4–5 products per transaction. This reflects a daily or weekly buying pattern.

b. The results of mining the association rules with the A priori algorithm yielded 125 rules, with an average confidence value of 40% and an average lift of 1.9, indicating that the rules were quite strong and did not occur by chance.

c. The histogram shows support is mostly low (1–2%), but it is reasonable for retail datasets. Confidence is evenly distributed between 30–50%, indicating medium to high reliability. Lifts are mostly >1.5, with peaks of around 1.7–1.9, which indicates a positive association worthy of action.

d. ECLAT shows a much faster execution time (about 4x) than a priori with quality-identical rule results. This makes ECLAT superior in efficiency for large datasets.

The suggestions for this study are

a. Product Promotion Optimization can use rules with high confidence and > 1.5 elevator to make purchase recommendations, such as buy beef promos, root vegetables discounts.

b. Arrange the placement of products that appear frequently at the same time (high lift) to make it easier to find them together on the shelf.

REFERENCES

[1] Musthofa Galih Pradana, *Information Retrieval*. Penamuda, 2024.

[2] H. Li and P. C.-Y. Sheu, "A scalable association rule learning heuristic for large datasets," *J Big Data*, vol. 8, no. 1, p. 86, Dec. 2021, doi: 10.1186/s40537-021-00473-3.

[3] B. A. Alrahwan and M. Farouk, "ASCF: Optimization of the Apriori Algorithm Using Spark-Based Cuckoo Filter Structure," *International Journal of Intelligent Systems*, vol. 2024, pp. 1–16, Jan. 2024, doi: 10.1155/2024/8781318.

[4] M. G. Pradana and K. Khoironi, "Market Basket Analysis Data Groceries Menggunakan Equivalence Class Transformation," *Journal Computer Science and Information Systems : J-Cosys*, vol. 5, no. 1, Mar. 2025, doi: 10.53514/jco.v5i1.593.

[5] M. Man, M. A. Jalil, and W. A. W. A. Bakar, "Fi-Eclat: An enhancement of incremental Eclat algorithm," 2023, p. 060006. doi: 10.1063/5.0110230.

[6] N. Lin and T. Win, "Optimizing the Performance of the ECLAT Algorithm using K-means Clustering for Frequently Itemset Mining for Large Datasets," in *2024 5th International Conference on Advanced Information Technologies (ICAIT)*, IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/ICAIT65209.2024.10754926.

[7] O. M. Ben Omran, "Using ECLAT Algorithm for Privacy Preserving on Publishing Data," in *2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, IEEE, May 2023, pp. 407–412. doi: 10.1109/MI-STA57575.2023.10169272.

[8] V. Srinadh, "Evaluation of Apriori, FP growth and Eclat association rule mining algorithms," *Int J Health Sci (Qassim)*, pp. 7475–7485, Apr. 2022, doi: 10.53730/ijhs.v6nS2.6729.

[9] M. S. Krishnan, A. S. Nair, and J. Sebastian, "Comparative Analysis of Apriori and ECLAT Algorithm for Frequent Itemset Data Mining," 2022, pp. 489–497. doi: 10.1007/978-981-16-3675-2_37.

[10] D. Mohapatra, J. Tripathy, K. K. Mohanty, and D. S. K. Nayak, "Interpretation of Optimized Hyper Parameters in Associative Rule Learning using Eclat and Apriori," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Apr. 2021, pp. 879–882. doi: 10.1109/ICCMC51019.2021.9418049.

[11] S. Abbasi and A. Moeini, "BloomEclat: Efficient Eclat Algorithm based on Bloom filter," 2021. [Online]. Available: http://jac.ut.ac.ir

[12] D. Dhinakaran and P. M. J. Prathap, "Protection of data privacy from vulnerability using two-fish technique with Apriori algorithm in data mining," *J Supercomput*, vol. 78, no. 16, pp. 17559–17593, Nov. 2022, doi: 10.1007/s11227-022-04517-0.

[13] R. M. Sumacot, "Association Analysis Using Apriori Algorithm of GANs-Expanded Student Performance Dataset," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 15, no. 2, pp. 101–108, Nov. 2024, doi: 10.21512/comtech.v15i2.11948.

[14] L. Kusak, F. B. Unel, A. Alptekin, M. O. Celik, and M. Yakar, "Apriori association rule and K-Means clustering algorithms for interpretation of pre-event landslide areas and landslide inventory mapping," *Open Geosciences*, vol. 13, no. 1, pp. 1226–1244, Oct. 2021, doi: 10.1515/geo-2020-0299.

[15] Z. Zixian, L. Xuning, L. Zhixiang, and H. Hongqiang, "Outburst prediction and influencing factors analysis based on Boruta-Apriori and BO-SVM algorithms," *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 2, pp. 3201–3218, Sep. 2021, doi: 10.3233/JIFS-210466.