International Journal of Applied Mathematics, Sciences, and Technology for National Defense
Volume 4, Issue 2, 141-154
e_ISSN: 2985-9352, p_ISSN: 2986-0776
https://www.journal.foundae.com/index.php/JAS-ND
DOI : https://doi.org/10.58524/app.sci.def.v4i2.1311

Sparsity-based memory scalability analysis of association rule mining algorithms using e-commerce heterogeneous multi-datasets for decision support systems

Esti Fadhilah
Telkom University
INDONESIA

Diandra Chika Fransisca*
Telkom University
INDONESIA

Article Info

Article history:

Received: May 8, 2026
Revised: June 30, 2026
Accepted: July 21, 2026

Keywords:

Association Rules Mining
Data mining
E-commerce
Heterogeneous multi-dataset
Sparsity

Abstract

Background: Many studies have compared the Apriori, FP-Growth, and ECLAT algorithms. Most of the previous literature focuses on runtime evaluation (execution speed) on homogeneous datasets. Research specifically mapping the visualization of memory curves against the level of data sparsity on common e-commerce heterogeneous.
Aims: This study aims to analyze the relationship between sparsity and memory usage of three classic Association Rule Mining algorithms such as Apriori, FP-Growth and ECLAT using a heterogeneous e-commerce dataset.
Method: A quantitative approach using Association Rule Mining was applied to different raw datasets from 721 to 30000 transactions. Transactions focused on various item types. The Apriori, FP-Growth, and ECLAT algorithms are implemented in Python, using grid search for minimum support and confidence adjustment to generate frequent item sets and association rules for performance comparison.
Result: The results show that the relationship between sparsity and memory usage differs significantly in the three classical association rule mining algorithms. The ECLAT algorithm shows low memory consumption because it uses a vertical TID-list, while the Apriori and FP-Growth algorithms show high memory consumption because they are related to their pattern search methods (Candidate Generation and FP-Tree).
Conclusion: This study shows that memory usage is not only influenced by sparsity, but also by the characteristics of the dataset. Then the results of this analysis provide an overview of heterogeneous e-commerce datasets when used in the classical association rule mining algorithm.

To cite this article: Fadhilah, E. & Fransisca, D. C. (2026). Sparsity-based memory scalability analysis of association rule mining algorithms using e-commerce heterogeneous multi-datasets for decision support systems. *International Journal of Applied Mathematics, Sciences, and Technology for National Defense,* 4(2), 141-154.

***Corresponding Author:**
Diandra Chika Fransisca, Telkom University, Indonesia, Email: diandraf@telkomuniversity.ac.id

INTRODUCTION

E-commerce is no longer just a side trend but has become a strong and inseparable part of the business or trade world (Criveanu, 2025). The growth of e-commerce, from its early days as a niche market to its current status as a dominant force in the global economy, has been marked by rapid technological advances and changing consumer preferences (Bao et al., 2025). E-commerce often collects a lot of information about consumer behaviour (Syah et al., 2025). Consumer behaviour that involves buying a few products, occurs due to impulsive behaviour toward a specific product (Moghddam et al., 2024). This behaviour is what allows e-commerce to create a highly popular flagship product. It's a common mistake for e-commerce companies to expand their product offerings without a clear strategy (Gumasing et al., 2023; Ma & Gu, 2024).

E-commerce businesses can leverage the transaction data they collect as their most important analytical tool for understanding consumer behaviour (Nosova & Guleva, 2021); the insights gained from this data are invaluable for presenting offers that are relevant and tailored to each consumer. To process this large volume of data, e-commerce businesses can implement techniques such as data mining (Stylianou & Pantelidou, 2025). Data mining is the process of discovering previously unknown knowledge from large datasets, which can specifically be used in commercial applications to identify customer purchasing patterns (Niu, 2023). One of the primary tasks in data mining for shopping cart analysis is Association Rule Mining (ARM) (Trasierras et al., 2024). This technique aims to identify all sets of items (products) that frequently appear together in a single transaction. There are several classic algorithms commonly used for this research, including Apriori, FP-Growth, and ECLAT have very different approaches and performance characteristics. The Apriori algorithm is the first and most classic algorithm in ARM, using a step-by-step iterative search method to discover patterns (Zhang & Zhang, 2023).

The Apriori algorithm is a classic join-based method that operates on horizontal data formats using a level-wise search strategy; however, it has a major limitation in the form of high time and I/O inefficiency because it requires the creation of complex “itemset” candidates and repeated database scans for each iteration. Apriori algorithm is only able to capture static (fixed) purchasing relationships or patterns and is less effective when used alone to analyze consumer behavior which is actually very dynamic and continues to change over time (Stylianou & Pantelidou, 2025). Meanwhile, the FP-Growth algorithm uses a compressed data structure called an FP-Tree (Frequent Pattern-Tree) for frequent pattern mining, which allows it to avoid explicitly generating item set candidates as in Apriori. The parameters used for the FP-Growth algorithm are compared based on execution time to measure the speed when generating itemsets in milliseconds (ms), measuring individual memory usage, number of transactions and assessing the sensitivity of the support threshold to changes (Srinadh, 2022). The ECLAT algorithm works by utilizing a vertical data format; the ECLAT (Equivalence Class Transformation) algorithm offers a different approach by using a vertical data format (storing items along with their transaction ID lists) and a depth-first search (DFS) method, although this method is very fast on sparse datasets because it reuses the same transaction IDs without rescanning the database (Kallay & Dan Mihoc, 2025).

Classic association rule mining algorithms have been extensively evaluated, with many studies emphasizing execution time, memory usage, and rule quality. However, in practice, different dataset characteristics exhibit heterogeneity that impacts their computational performance. Therefore, analysing memory usage on heterogeneous data is crucial for understanding algorithm behaviour on different datasets.

Previous studies by Srinadh (2022) have primarily evaluated the classic ARM algorithm based on computational performance metrics such as execution time and memory usage, without investigating the impact of dataset characteristics on these metrics. Although memory usage has been reported, as in the study by Hunyadi et al. (2025), it was treated only as an evaluation metric and was not analysed as a function of the dataset’s sparsity level. Research by Dwiputra et al. (2023) focused on execution time, memory usage, and the number of high-quality rules for generating contextual recommendations. This study measures memory usage directly compared to execution time and does not compare it based on sparsity; the evaluation was conducted on a single dataset (homogeneous dataset), so the results cannot yet be generalized to other datasets (heterogeneous dataset).

Most studies have evaluated classical association rule mining algorithms such as Apriori, FP-Growth, and ECLAT by comparing execution time, memory usage, and rule quality. However, memory usage is generally treated as a standard performance metric, and few similar studies have been conducted on homogeneous e-commerce data. Therefore, there is a gap in understanding the sparsity level in heterogeneous e-commerce data. Research on the relationship between sparsity and memory usage provides practical insights into selecting memory-efficient classical association rule mining algorithms for scalable and resource-efficient transaction analysis. The results of this research will contribute to the development of data processing systems to support the security, stability, and sustainability of e-commerce infrastructure. This study aims to investigate the relationship between the sparsity level of a dataset and the memory scalability of classical association rule mining algorithms using a heterogeneous dataset e-commerce transaction.

METHOD

This study employs a quantitative experimental research design, implemented in Python, to evaluate the scalability of the classic association rule mining algorithm on a heterogeneous e-commerce transaction dataset with varying levels of sparsity.

Four heterogeneous e-commerce transaction datasets were used in this study from publicly available datasets collected from Kaggle. These datasets differ in terms of transaction volume, comparing raw data and data with structured transaction formats (N), the number of unique items, average basket size, density, and sparsity, thereby providing diverse experimental conditions for analysing the memory scalability of Apriori, FP-Growth, and ECLAT. The characteristics of these datasets are summarized in Table 1. Dataset Summary.

Table 1. Dataset summary.

No	Dataset	Source	Transaction Raw/N	Unique Items	Avg. Basket Size	Density	Sparsity
1	D1	Kaggle	720/199	7	2.73	0.3905	0.6095
2	D2	Kaggle	1000/275	7	2.20	0.3143	0.6857
3	D3	Kaggle	10000/6016	7	1.54	0.2196	0.7804
4	D4	Kaggle	30000/7963	32	2.35	0.0735	0.9265

The research instruments consist of dataset characteristics and evaluation metrics used to analyse the performance of classical association rule mining algorithms. Prior to the mining process, each heterogeneous dataset is characterized based on the number of transactions, the number of unique items, the average basket size, density, and sparsity to describe its structural properties. These characteristics are then used to investigate their relationship with the memory scalability of the evaluated algorithms.

Table 2. Dataset characteristics and evaluation metrics.

Category	Variable	Description
Dataset Characteristics	Number of Transactions	Total number of transactions in each dataset
	Number of Unique Items	Total distinct items
	Average Basket Size	Average number of items per transactions
	Density	Ratio of existing items to all possible items
	Sparsity	Complement of density (1 – density)
Evaluation Metrics	Execution Time	Total execution time of the algorithm (second)
	Peak Memory	Memory consumption during execution (MB)
	Number of Frequent Itemset	Number of generated frequent itemset
	Number of Association Rules	Number of generated association rules
	Average Confidence	Average confidence of all generated rules
	Average Lift	Average lift of all generated rules

The performance of Apriori, FP-Growth, and ECLAT is evaluated using several computational metrics, including execution time, memory usage, the number of frequent item sets, the number of association rules, the average confidence level, and the average lift. Although execution time and rule quality metrics were included for comparison purposes, this study primarily focuses on memory usage as the key indicator for analysing memory scalability under different data set sparsity conditions. The dataset characteristics and evaluation metrics used in this study are summarized in Table 2 of dataset characteristics and evaluation metrics.

The experimental environment consists of hardware and software resources used to implement the proposed methodology and evaluate the performance of the Association Rule Mining (ARM) algorithm. The experiments were conducted using an ASUS TUF Gaming F15 laptop powered by an Intel® Core™ i5-10300H processor (2.50 GHz), 16 GB of RAM, and a 477 GB SSD, running the Windows 11 Home Single Language (64-bit) operating system. Google Collaboratory served as the primary computing platform for data preprocessing, algorithm implementation, and performance evaluation thanks to its flexible cloud-based computing environment.

The implementation was developed using Python, supported by several libraries. Pandas and NumPy are used for data manipulation and numerical computations; mlxtend is used to implement the Apriori, FP-Growth and ECLAT algorithm. In addition, 'trace malloc' are used to measure peak memory consumption during algorithm execution, while 'Matplotlib' is used to visualize memory scalability curves.

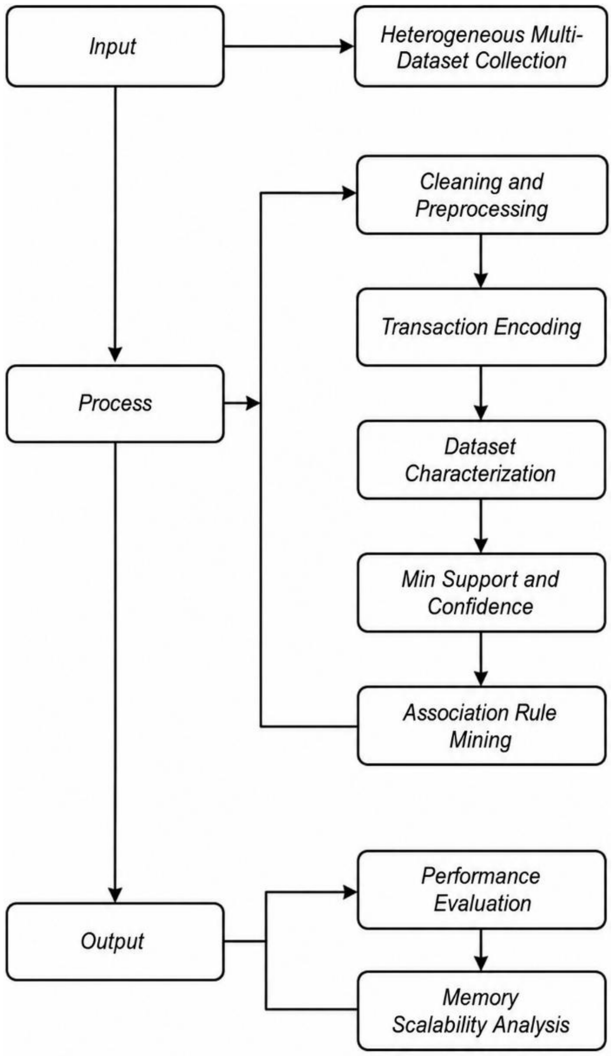

Figure 1. Flowchart of the method.

The Input stage collects datasets characterized by heterogeneous data. This is followed by the Process stage, which includes cleaning and preprocessing, transaction encoding, dataset characterization, determining minimum support and confidence, and modelling using the association rule mining algorithm. The Output stage performs performance evaluation and memory scalability analysis based on its relationship to sparsity. For further details, the steps are explained in sequence below.

Heterogenous Multi-Dataset Collection

The process of collecting datasets from various sources with non-uniform format and structure characteristics, also known as Heterogeneous Multi-Dataset Collection (Bhilwarawala et al., 2026). In this study, we focus on collecting datasets sourced from one open-source datasets on Kaggle, paying attention to key features such as Customer ID, Product Name (which indicates the type of product), the amount of data, and the shipping status. Shipping status is considered because it plays a role in ensuring alignment with the study's primary focus on data usage specifically, whether the product reaches the customer. The differences between the datasets are shown in Table 1 of dataset summary.

Cleaning and Preprocessing

The data preprocessing and cleaning stage plays a vital role in harmonizing data subsets drawn from various generators or different environments. Without structured preprocessing to address inconsistencies in the characteristics and formats of raw data, the model will struggle to recognize patterns objectively, which ultimately reduces its performance (Zhang et al., 2026). This step involves removing duplicate data (as shown in Figure 2), checking for null values (as shown in Figure 3). Next, one dataset was selected from each of the four datasets that had a shipping status of "delivered" (as show in Figure 4).

```
duplicates = df.duplicated().sum()
df = df.drop_duplicates()
print(duplicates)
```

Figure 2. Removing duplicate data.

```
df.isnull().sum()
```

Figure 3. Checking for null values.

```
df_diterima = df[df['order_status'].str.lower() == 'delivered']
```

Figure 4. Select shipping status of delivered.

Transaction Encoding

Association Rule Mining (ARM) is defined as one of the most important techniques in data mining, which is the core process of Knowledge Discovery in Databases (KDD) for identifying valid, novel, and meaningful patterns in a dataset. Originally proposed by Agrawal et al. in 1993 for market basket analysis, this technique aims to identify sets of items that frequently appear together in a transaction dataset (Aloisi et al., 2025). A market basket used to represent transaction data (a binary matrix), where rows represent transactions and columns represent items (Harini et al., 2024).

Data Characterization

After the three stages described earlier have been completed, each transaction dataset is analysed in this stage to identify its structural properties prior to the Association Rule Mining process. This stage looks at the characteristics of the heterogeneous dataset used, including the number of unique items, average basket size, density and spread. Additionally, dataset density is calculated as the ratio of the number of actual item occurrences to the total number of possible item occurrences in the transaction-item matrix (1) based on Equations (1). Dataset sparsity is then calculated as the complement of density, indicating the proportion of item values that do not appear in the transaction matrix (0) based on Equations (2). These structural characteristics are used to investigate the relationship between dataset sparsity and the memory scalability of Apriori, FP-Growth, and ECLAT across various heterogeneous e-commerce transaction datasets (Anastasov et al., 2026).

$$\text{Density} = \frac{\text{Total Transactions}}{\text{Number of Customers} \times \text{Number of Products}} \tag{1}$$

$$\text{Sparsity} = 1 - \text{Density} \tag{2}$$

Min Support and Confidence

In this stage, the minimum support and minimum confidence parameters were set as the minimum thresholds for identifying relevant patterns based on Equations (1) through (3). The data mining process was then carried out by comparing the performance of the Apriori algorithm, FP-Growth algorithm and the ECLAT algorithm. The model configuration was optimized using a grid search (Dendy & Tanamal, 2026), which comprehensively evaluated the parameter space and determined that a minimum support of 0.005 or 0.05% and a minimum confidence of 0.1 or 10% yielded the most statistically significant rules. Although the number of transactions in each dataset varies, a minimum support value of 0.05% and a minimum confidence level of 10% were applied uniformly across all experiments. This experimental design eliminates parameter bias and ensures that performance comparisons reflect the influence of dataset characteristics, including transaction volume, the number of unique items, average basket size, density, and scarcity. By using the same parameters, the influence of dataset characteristics on algorithm performance can be observed more objectively.

$$\text{Support}\ (A) = \frac{\text{Number of Transaction Containing } A}{\text{Total Transactions}} \tag{3}$$

Support for two items, such as item A and B, using Equation 4.

$$\text{Support}\ (A, B) = \frac{\text{Number of Transaction Containing } A \cap B}{\text{Total Transactions}} \tag{4}$$

Confidence is a measure of the strong association between two products in the total demand for either product, equation (5) is used to calculate the confidence for products A and B (Wahyuningsih & Suharsono, 2023).

$$\text{Confidence}\ (A, B) = \frac{\text{Number of Transaction Containing } A \cap B}{\text{Number of Transaction Containing } A} \tag{5}$$

Association Rule Mining

Based on the data that has been prepared, the next step is to formulate a method and implement a data mining model focused on the application of Association Rule Mining (ARM) algorithms such as Apriori, FP-Growth, and ECLAT. The minimum support and confidence thresholds were applied as previously specified during the determination of these thresholds; these values were chosen because they were capable of generating item sets and rules across all four types of datasets. These values were chosen because this study aims to make a fair comparison using these minimum thresholds; the results are presented in Tables 4 through 7.

Performance Evaluation

The research process ultimately focused on evaluating and understanding the results obtained. The evaluation results included comparisons of model execution time, the number of item sets formed, the number of rules, and peak memory (MB). The generated association rules were then validated using average confidence and lift ratios (Felix & Tjen, 2025) to measure the strength of the correlation between items and ensure that the rules were valid and not merely coincidental, across the three association rule mining algorithms used: Apriori, FP-Growth, and ECLAT. To evaluate the memory efficiency of the implemented algorithms, we used a built-in Python library called trace malloc (Mudumba & Kabir, 2024). This library tracks memory allocations and records peak memory usage during the execution of the Apriori, FP-Growth, and Eclat algorithms.

Memory Scalability Analysis

The final stage of this study involves analysing the memory scalability of the Association Rule Mining (ARM) algorithms under evaluation. Memory scalability refers to an algorithm's ability to maintain efficient memory utilization as the structural characteristics of a dataset change, particularly with regard to the datasets sparsity level. Unlike conventional performance evaluations, which primarily emphasize execution time, this study focuses on memory consumption as the primary indicator of algorithmic scalability. Peak memory usage was recorded during the execution of the Apriori, FP-Growth, and ECLAT algorithms on each heterogeneous dataset. The collected memory measurements were then compared across datasets with different sparsity levels to identify variations in memory behaviour.

The performance of Apriori, FP-Growth, and ECLAT will be analysed based on execution time, memory usage, the number of frequently occurring item sets, the number of association rules, the average confidence level, and the average lift. Data set characteristics, including transaction volume, average basket size, density, and sparsity, will be calculated in advance for each heterogeneous data set. A comparative analysis will then be conducted to evaluate the memory scalability of each algorithm under different sparsity conditions. The results will be presented using comparison tables and memory scalability curves to illustrate the behaviour of each algorithm across various heterogeneous datasets.

RESULTS AND DISCUSSION

The experimental results obtained from evaluating the memory scalability of the Apriori, FP-Growth, and ECLAT algorithms using heterogeneous e-commerce transaction datasets. The results include dataset characteristics, algorithm performance, and the relationship between dataset sparsity and peak memory usage. To facilitate interpretation, the findings are presented through descriptive tables, comparative analyse, and visualization using curves. These datasets differ in terms of transaction volume, comparing raw data and data with structured transaction formats (N), the number of unique items, average basket size, density, and sparsity.

Table 3. Dataset summary.

No	Dataset	Transaction Raw/N	Unique Items	Avg. Basket Size	Density	Sparsity
1	D1	720/199	7	2.73	0.3905	0.6095
2	D2	1000/275	7	2.20	0.3143	0.6857
3	D3	10000/6016	7	1.54	0.2196	0.7804
4	D4	30000/7963	32	2.35	0.0735	0.9265

The research results from using the minimum support equation of 0.005 and minimum confidence of 0.1, produce the number of frequent itemset and rules that match the characteristics of the dataset in Table 4 up to Table 7.

Table 4. D1 Performance evaluation.

Dataset/N	Algorithm	Execution Time (second)	Peak Memory (MB)	Frequent Itemset	Rules	Avg. Confidence	Avg. Lift
D1/199	Apriori	0.00663	0.4916	109	644	0.294048	1.200281
	FP-Growth	0.03968	0.5154	109	644	0.294048	1.200281
	ECLAT	0.00056	0.4869	109	644	0.294048	1.200281

Table 5. D2 Performance evaluation.

Dataset/N	Algorithm	Execution Time (second)	Peak Memory (MB)	Frequent Itemset	Rules	Avg. Confidence	Avg. Lift
D2/275	Apriori	0.00961	0.3396	94	439	0.274594	1.265244
	FP-Growth	0.05198	0.3645	94	439	0.274594	1.265244
	ECLAT	0.00352	0.3356	94	439	0. 274594	1.265244

Table 6. D3 Performance evaluation.

Dataset/N	Algorithm	Execution Time (second)	Peak Memory (MB)	Frequent Itemset	Rules	Avg. Confidence	Avg. Lift
D3/6016	Apriori	0.00546	0.4199	37	68	0.157669	0.69185
	FP-Growth	0.03552	0.7934	37	68	0.157669	0.69185
	ECLAT	0.00139	0.0336	37	68	0.157669	0.69185

Table 7. D4 Performance evaluation.

Dataset/N	Algorithm	Execution Time (second)	Peak Memory (MB)	Frequent Itemset	Rules	Avg. Confidence	Avg. Lift
D4/7963	Apriori	0.01266	4.8609	209	74	0.110357	1.023972
	FP-Growth	0.06195	4.4449	209	74	0.110357	1.023972
	ECLAT	0.00901	0.0365	209	74	0.110357	1.023972

From the results of the performance evaluation in Table 4 to Table 7, we continue by looking at the comparison of data characteristics with memory usage in Table 8 to Table 11.

Table 8. D1 Dataset characteristics and peak memory usage of association rule mining algorithms.

Dataset/N	Unique Items	Avg. Basket Size	Density	Sparsity	Algorithm	Peak Memory (MB)
D1/199	7	2.73	0.3905	0.6095	Apriori	0.4916
					FP-Growth	0.5154
					ECLAT	0.4869

Table 9. D2 Dataset characteristics and peak memory usage of association rule mining algorithms.

Dataset/N	Unique Items	Avg. Basket Size	Density	Sparsity	Algorithm	Peak Memory (MB)
D2/275	7	2.20	0.3143	0.6857	Apriori	0.3396
					FP-Growth	0.3645
					ECLAT	0.3356

Table 10. D3 Dataset characteristics and peak memory usage of association rule mining algorithms.

Dataset/N	Unique Items	Avg. Basket Size	Density	Sparsity	Algorithm	Peak Memory (MB)
D3/6016	7	1.54	0.2196	0.7804	Apriori	0.4199
					FP-Growth	0.7934
					ECLAT	0.0336

Table 11. D4 Dataset characteristics and peak memory usage of association rule mining algorithms.

Dataset/N	Unique Items	Avg. Basket Size	Density	Sparsity	Algorithm	Peak Memory (MB)
D4/7963	32	2.35	0.0735	0.9265	Apriori	4.8609
					FP-Growth	4.4449
					ECLAT	0.0365

Then, from the quantitative comparison of Table 4 up to Table 11, visual analysis is needed to see the pattern in the relationship between sparsity and memory in each classic association rule mining algorithm. The research results are further illustrated with memory scalability curves, an analysis of the relationship between sparsity and memory.

The analysis of the relationship between sparsity and peak memory consumption using four heterogeneous datasets shows the relationship between the two depending on the association rule mining strategy used. Figure 5 shows that the Apriori algorithm consumes relatively stable memory between datasets D1, D2, and D3. Then, on dataset D4, it experiences a high increase in large datasets with an increase of up to 4.8 MB because Apriori is wasteful on large datasets after storing many itemset combinations. From these results, the Apriori algorithm is not only determined by the sparsity of the dataset but also by the process of forming the itemset. Figure 6 shows that the FP-Growth algorithm has relatively low memory consumption on datasets D1, D2, and D3. Then, on dataset D4, memory consumption increases to 4.4 MB. FP-Growth experiences waste because it has to create a tree structure and its pointers on large data. In the FP-Growth algorithm, it can be associated with a substantial increase in the number of unique items, which enlarges the FP-tree structure and requires additional memory for node storage and tree traversal, and its memory usage is still affected by the complexity of the compressed tree representation. Figure 7 shows that the ECLAT algorithm is more efficient on the D4 dataset, in contrast to the conditions of the Apriori and FP-Growth algorithms which experience increased memory. ECLAT works using a vertical TID-list where when the dataset has high sparsity, it indicates that many items appear in few transactions. When looking at the size of the sparsity which touches 92% of empty matrix cells, this benefits ECLAT. As the dataset becomes sparser, each item appears in fewer transactions, resulting in shorter TID lists and more efficient slicing operations. Consequently, the memory required to store intermediate data structures is significantly reduced. This explains why the ECLAT algorithm exhibits superior memory usage scalability when processing very sparse transaction datasets.

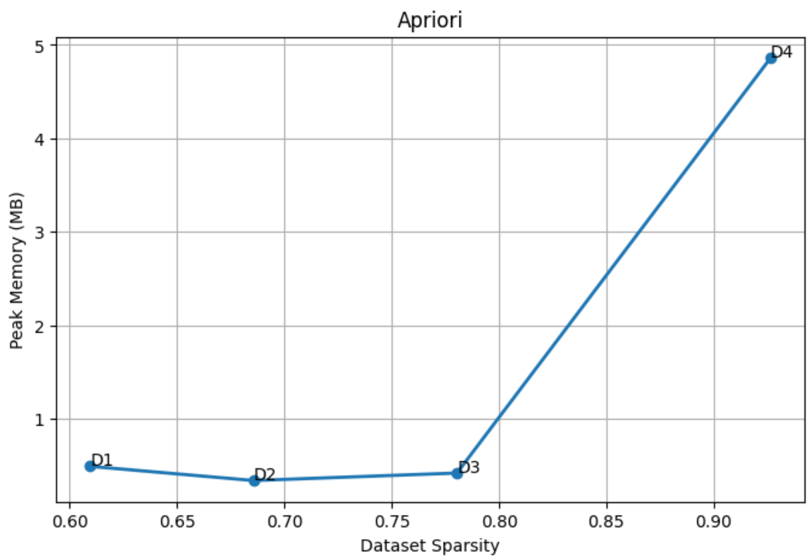

Figure 5. Sparsity–memory relationship of apriori.

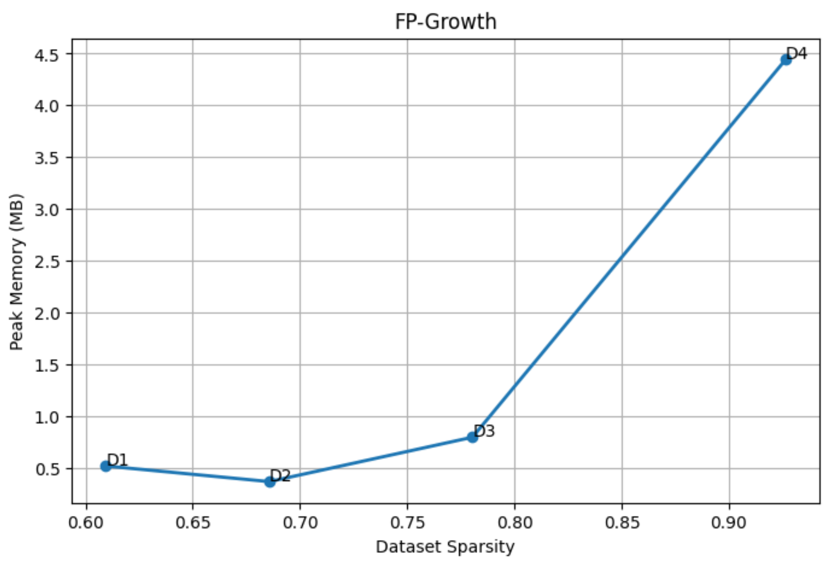

Figure 6. Sparsity–memory relationship of fp-growth.

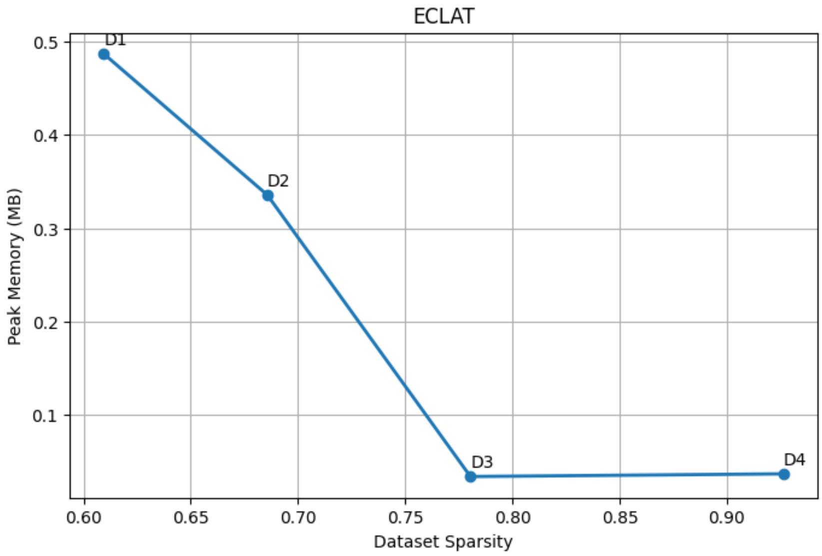

Figure 7. Sparsity–memory relationship of eclat.

From the research results, dataset sparsity alone does not fully explain the strong relationship on memory usage, in the D4 dataset which has the highest sparsity at 0.9265, memory behaviour is seen to have differences in each association rule mining algorithm. The ECLAT algorithm shows a reduction in memory usage due to vertical representation which benefits from shorter TID-lists in high sparsity datasets. Then, the Apriori and FP-Growth algorithms consume more memory caused not only by sparsity but also by increasing the number of unique items from 7 items to 32 items, expanding the search space for candidate generation in Apriori and the complexity of the FP-Tree structure in FP-Growth. Therefore, these findings indicate that the size of memory usage is influenced by dataset sparsity and other characteristics of the dataset structure and not by sparsity alone.

The results obtained in this study show different structural characteristics, then in previous studies mainly focused on homogeneous datasets or datasets with relatively similar transaction structures, this study uses a heterogeneous e-commerce transaction dataset with substantial variations in the number of transactions, unique items, average basket size, density, and sparsity. As a result, the observed memory behaviour reflects not only the effect of dataset scarcity but also the interaction between several structural characteristics of the dataset.

The results of this study provide theoretical and practical implications for evaluating the application of classical association rule mining algorithms. From a theoretical perspective, the results show that sparsity alone is not sufficient to explain memory usage. Instead, memory usage is influenced by sparsity and other dataset characteristics, such as the number of unique items and the average bucket size. These findings contribute to the growing understanding that the performance of classical association rule mining algorithms should be evaluated using more heterogeneous datasets rather than solely using metrics such as execution time. From a practical perspective, the proposed memory usage analysis provides useful guidance in selecting an association rule mining algorithm tailored to the characteristics of the dataset. The ECLAT algorithm is proven to be efficient in memory usage for sparse datasets because it uses vertical data, and the Apriori and FP-Growth algorithms require additional consideration when applied to datasets with a large number of unique items because of their memory requirements related to candidate generation and FP-Tree.

Despite the contributions of this research, there are several limitations that need to be acknowledged. First, the testing used four heterogeneous e-commerce datasets, which may not fully represent the diversity of real-world transaction data mining. Second, the evaluation was limited to three classic association rule mining algorithms: Apriori, FP-Growth, and ECLAT, without considering more recent approaches. Third, the testing relied on the same minimum support and confidence and did not examine the influence of different parameters. Finally, while dataset sparsity was analysed as the primary test, other factors such as frequent item sets, transaction length, and dataset size can affect memory usage and require further research.

The results of this study have practical implications, specifically, the proposed memory scalability analysis can serve as a knowledge base Kayvanfar et al. (2024) for selecting the appropriate Association Rule Mining (ARM) algorithm according to the structural characteristics of Decision Support Systems (DSS) transaction datasets (Alahmadi & Jamjoom, 2022). Instead of selecting an algorithm solely based on execution time, decision makers can also consider dataset characteristics such as sparsity, number of unique items, and average bucket size to achieve more efficient memory utilization.

CONCLUSION

Based on the results of this study, sparsity-based memory scalability analysis of association rule mining using e-commerce heterogeneous multi-dataset helps provide an overview for researchers and practitioners in selecting the right classical algorithm evaluation apart from sparsity on heterogeneous transaction data. Dataset sparsity is not the only factor in assessing the memory size of the algorithm, the effect of sparsity depends on the search pattern used in each algorithm such as the Apriori algorithm that uses candidate generation, the FP-Growth algorithm that uses FP-Tree and ECLAT that uses vertical TID-list.

In further research, this research can be expanded by using heterogeneous datasets from various sources and the characteristics of the datasets in relation to memory usage. The development of a recommendation algorithm framework with dataset characteristics can be further research beyond sparsity.

AUTHOR CONTRIBUTIONS

E.F.: Supervision, resources, data curation, methodology, software, formal analysis, visualization, writing –original draft. D.C.F.: Conceptualization, methodology, funding acquisition, validation.

ACKNOWLEDGMENT

The author expresses sincere gratitude to the academic advisor for their invaluable guidance and feedback, to the Research and Community Service Institute (LPPM) of Telkom University for financial support, and to all parties who contributed directly or indirectly to the successful completion of this research.

CONFLICT OF INTEREST

The authors declare that have no conflict of interest.

REFERENCES

Alahmadi, D. H., & Jamjoom, A. A. (2022). Decision support system for handling control decisions and decision-maker related to supply chain. *Journal of Big Data*, 9(114). https://doi.org/10.1186/s40537-022-00653-9

Aloisi, V., Noce, S., Epicoco, I., Cipriano, C., Cafaro, M., Brundu, G., Arcidiaco, L., Spano, D., Aloisio, G., & Mereu, S. (2025). Application of association rule mining to assess forest species distribution in Italy considering abiotic and biotic factors. *Ecological Informatics,* 92, 103514. https://doi.org/10.1016/j.ecoinf.2025.103514

Anastasov, G., Kocaleva Vitanova, M., Zlatanovska, B., & Miteva, M. (2026). A review of mathematical approaches to recommendation algorithms: from collaborative filtering to deep learning. *International Journal of Scientific Development and Research,* 11(6), c92-c101. www.ijsdr.org

Bao, M., Latif, S., Bao, D., Latif, Z., & Lu, J. (2025). Overcoming e-commerce barriers in developing markets: A review of data-driven strategies for sustainable growth. *Sustainable Futures*, 10, 101408. https://doi.org/10.1016/j.sftr.2025.101408

Bhilwarawala, A., Rongmei, L., Sharma, H., Jena, A., Singh, K., Piri, J., & Dey, R. (2026). *BRIDGE and TCH-Net: Heterogeneous Benchmark and Multi-Branch Baseline for Cross-Domain IoT Botnet Detection*. arXiv. arXiv: 2604.11324v1. http://arxiv.org/abs/2604.11324

Criveanu, M. M. (2025). The impact of digital technology on e-commerce and sustainable performance in the EU. *Economies*, 14(1), 5. https://doi.org/10.3390/economies14010005

Dendy, L. E., & Tanamal, R. (2026). Association rule mining for truck body damage pattern analysis using apriori and CRISP-DM. *Journal of Applied Informatics and Computing*, 10(1), 417-428.. http://jurnal.polibatam.ac.id/index.php/JAIC

Dwiputra, D., Mulyo Widodo, A., Akbar, H., & Firmansyah, G. (2023). Evaluating the performance of association rules in apriori and fp-growth algorithms: market basket analysis to discover rules of item combinations. *Journal of World Science*, 2(8), 1229–1248. https://doi.org/10.58344/jws.v2i8.403

Felix, J., & Tjen, J. (2025). Machine learning in business: Product bundling strategy and customer segmentation via market basket analysis algorithm. *MABIS*, 16(01), 29-37.. https://doi.org/10.66003/mabis.v16i01.10550

Gumasing, M. J. J., Ong, A. K. S., Sy, M. A. P. C., Prasetyo, Y. T., & Persada, S. F. (2023). A machine learning ensemble approach to predicting factors affecting the intention and usage behavior towards online groceries applications in the Philippines. *Heliyon,* 9(10), e20644. https://doi.org/10.1016/j.heliyon.2023.e20644

Harini, J. V., Venu, G., Reddy, G. V. K., Datta, B. V., Goud, M. V., Mohammed, T. K. M. K., & Ashok, Prof. (2024). Market basket analysis. *International Journal for Research in Applied Science and Engineering Technology*, 12(5), 5224–5229. https://doi.org/10.22214/ijraset.2024.61662

Hunyadi, I. D., Constantinescu, N., & Ţicleanu, O. A. (2025). Efficient discovery of association rules in e-commerce: comparing candidate generation and pattern growth techniques. *Applied Sciences*, 15(10), 5498. https://doi.org/10.3390/app15105498

Kallay, P., & Dan Mihoc, T. (2025). Comparative analysis of frequent pattern mining algorithms. *Acta Universitatis Sapientiae, Informatica*, 17(1), 1-17. https://doi.org/10.1007/s44427-025-00008-1

Kayvanfar, V., Elomri, A., Kerbache, L., Vandchali, H. R., & El Omri, A. (2024). A review of decision support systems in the internet of things and supply chain and logistics using web content mining. *Supply Chain Analytics,* 6, 100063. https://doi.org/10.1016/j.sca.2024.100063

Ma, X., & Gu, X. (2024). New marketing strategy model of E-commerce enterprises in the era of digital economy. *Heliyon*, 10(8), e29038. https://doi.org/10.1016/j.heliyon.2024.e29038

Moghddam, H. A., Carlson, J., Wyllie, J., & Rahman, S. R. (2024). Scroll, Stop, Shop: Decoding impulsive buying in social commerce. *Journal of Business Research*, 182(2024), 114776. https://doi.org/10.1016/j.jbusres.2024.114776

Mudumba, B., & Kabir, M. F. (2024). Mine-first association rule mining: An integration of independent frequent patterns in distributed environments. *Decision Analytics Journal*, 10, 100434. https://doi.org/10.1016/j.dajour.2024.100434

Niu, M. (2023). Sports training strategies based on data mining technology. *Procedia Computer Science*, 228, 847–856. https://doi.org/10.1016/j.procs.2023.11.112

Nosova, V., & Guleva, V. (2021). Purchasing patterns in luxury consumption: transitional transaction categories. *Procedia Computer Science*, 193, 351–360. https://doi.org/10.1016/j.procs.2021.10.036

Srinadh, V. (2022). Evaluation of apriori, fp growth and eclat association rule mining algorithms. *International Journal of Health Sciences*, 6(S2), 7475–7485. https://doi.org/10.53730/ijhs.v6ns2.6729

Stylianou, T., & Pantelidou, A. (2025). A machine learning approach to consumer behavior in supermarket analytics. *Decision Analytics Journal*, 16, 100600.. https://doi.org/10.1016/j.dajour.2025.100600

Syah, R. B. Y., Elveny, M., Darmansyah, S., & Silviana, L. (2025). Enhanced detection of consumer behavioral shifts in e-commerce platforms with transformer-based algorithms. *Journal of Applied Data Sciences*, 6(4), 2582–2593. https://doi.org/10.47738/jads.v6i4.907

Trasierras, A. M., Luna, J. M., Fournier-Viger, P., & Ventura, S. (2024). Data heterogeneity's impact on the performance of frequent itemset mining algorithms. *Information Sciences*, 678., 120981. https://doi.org/10.1016/j.ins.2024.120981

Wahyuningsih, R., & Suharsono, A. (2023). Comparison of market basket analysis method using apriori algorithm, frequent pattern growth (fp-growth), and equivalence class transformation (eclat) (Case study: Supermarket "X" transaction data for 2021). *Business and Finance Journal,* 8(2), 192-201. https://doi.org/10.33086/bfj.v8i2.5226

Zhang, X., & Zhang, J. (2023). Analysis and research on library user behavior based on apriori algorithm. *Measurement: Sensors*, 27, 100802. https://doi.org/10.1016/j.measen.2023.100802

Zhang, Y., Zhuo, Z., Xu, S., Lv, S., Liu, Z., Qiu, J., Wang, Q., Liu, Y., & Zhou, S. K. (2026). *Towards personalized multi-modal mri synthesis across heterogeneous datasets*. arXiv. arXiv: 2602.19723. http://arxiv.org/abs/2602.19723