

Application of Apriori and FP-Growth Algorithms in Analyzing Drug Purchasing Patterns

Gemma Tahmid Alfaridzi^{1*}, Febi Nur Salisah², Inggih Permana³

^{1,2,3}Fakultas Sains dan Teknologi, Sistem Informasi, UIN Sultan Syarif Kasim, Pekanbaru, Indonesia

Email: ^{1*}12250314082@students.uin-suska.ac.id, ²febinursalisah@uin-suska.ac.id, ³inggihipermana@uin-suska.ac.id

(* Email Corresponding Author: 12250314082@students.uin-suska.ac.id)

Received: January 5, 2026 | Revision: January 7, 2025 | Accepted: March 4, 2026

Abstract

Pharmacy sales transaction data contain valuable information on customer purchasing patterns; however, in practice, such data are often used merely as operational records, making relationships between purchased drugs difficult to identify. This study analyzes drug purchasing patterns using the Apriori and FP-Growth algorithms based on sales transaction data from Apotek Gadi Lamba Condét for the period January to June 2025. The transaction data were processed through data cleaning, drug name standardization, and transformation into transaction format, resulting in 7,038 transactions with 1,495 drug items. Association rule mining was performed using a minimum support of 0.01 and a minimum confidence of 0.17. The results show that the Apriori and FP-Growth algorithms generate ten identical association rules with the same support, confidence, and lift values, and all rules have lift values greater than one. Paracetamol 500 MG emerges as the most frequently involved drug in the association rules. These findings demonstrate that, for medium-scale pharmacy transaction datasets, Apriori and FP-Growth have equivalent capability in identifying drug purchasing patterns, with the primary difference lying in computational efficiency rather than the quality of the generated patterns.

Keywords: Data Mining, Drug Purchasing Patterns, Association Rule, Apriori, FP-Growth

1. INTRODUCTION

Pharmacies as healthcare service providers generate large volumes of drug sales transaction data that reflect customers' needs and purchasing behavior. These transaction records inherently contain valuable information regarding inter-drug purchasing relationships that frequently occur within a single transaction. However, in many pharmacies, including those in Indonesia, sales transaction data are still treated primarily as administrative records and have not been systematically utilized as analytical resources to understand customer purchasing patterns [1]. As a result, pharmacies often lack adequate insight into inter-drug relationships and recurring purchasing patterns derived from daily transaction activities.

A similar condition is observed at Apotek Gadi Lamba Condét, located in East Jakarta. Daily drug sales transactions recorded on a monthly and annual basis are generally stored as archives without further analysis. Manual and reactive recording practices limit the identification of meaningful drug associations, while anomalous transactions are difficult to detect due to the absence of baseline purchasing patterns for comparison [2]. This situation emphasizes the need for data-driven analytical approaches capable of transforming transaction data into structured, objective, and measurable information.

In this context, data mining provides an effective approach for converting transaction data into valuable knowledge. Data mining enables the extraction of hidden patterns from large datasets that cannot be identified through manual observation [3]. One data mining technique particularly relevant to pharmacy transaction analysis is association rule mining, which aims to discover relationships among items that frequently appear together within a transaction [4]. This technique has been widely applied in market basket analysis to gain deeper insight into consumer purchasing behavior [5].

Various algorithms have been developed to support association rule mining, among which Apriori and FP-Growth are the most commonly used. The Apriori algorithm operates by iteratively generating candidate itemsets based on the frequent itemset principle, whereas FP-Growth employs an FP-Tree structure to mine frequent patterns without explicitly generating candidate itemsets [6]. FP-Growth is generally recognized for its computational efficiency due to fewer database scans, while Apriori remains widely applied because of its simplicity and adaptability to small- and medium-sized datasets [7].

The quality of association rules generated by both algorithms is strongly influenced by dataset characteristics and the selection of minimum support and confidence thresholds. Inappropriate parameter settings may produce irrelevant rules and reduce the validity of analytical results [5]. Dataset size also affects algorithm performance, where FP-Growth tends to perform better in terms of processing speed on large datasets, while Apriori may be more memory-efficient for smaller datasets [8].

Previous studies have demonstrated that the application of Apriori and FP-Growth can generate meaningful purchasing patterns across various domains, including retail and pharmaceutical sectors. Pabendon and Purnomo reported that combining both algorithms can produce broader and more efficient association rules [9]. Idris et al. found that Apriori and FP-Growth yield comparable results in terms of rule quality, despite significant differences in

computational efficiency [10]. Other studies in the pharmaceutical domain highlight that drug transaction data exhibit inherent complexity due to variations in drug types, dosages, and purchasing frequencies, which require empirical investigation [11] [12] [13].

Nevertheless, studies that directly compare the results of Apriori and FP-Growth using pharmacy transaction data in Indonesia remain limited. Understanding whether both algorithms produce equivalent association patterns on medium-scale pharmacy transaction datasets is essential for determining appropriate analytical approaches without relying solely on theoretical assumptions [14].

Based on this background, this study examines the application of Apriori and FP-Growth algorithms to drug sales transaction data from Apotek Gadi Lamba Condet using a minimum support threshold of 0.01 and a minimum confidence threshold of 0.17. The results indicate that both algorithms generate identical association rules in terms of drug combinations as well as support, confidence, and lift values. These findings suggest that, for medium-scale pharmacy transaction datasets, the primary difference between Apriori and FP-Growth lies in computational efficiency rather than in the quality of the extracted drug purchasing patterns.

2. RESEARCH METHODOLOGY

This study employs a quantitative approach with an applied experimental method to analyze drug purchasing patterns based on pharmacy sales transaction data. This approach was selected because transaction data contain historical information that can be processed to objectively identify inter-drug relationship patterns using data mining techniques [3]. The research methodology was designed to ensure that the analysis process is reproducible and that the resulting findings can be systematically evaluated.



Figure 1. Research Methodology

2.1 Dataset

The dataset used in this study consists of drug sales transaction data collected from Apotek Gadi Lamba Condet, East Jakarta, covering the period from January to June 2025. The data were obtained from the pharmacy's internal digital recording system and represent real-world customer purchasing activities. Each transaction contains a set of drugs purchased together within a single purchase event. After an initial selection process, the final dataset comprised 7,038 valid transactions. The use of real transaction data enables the analysis to reflect actual purchasing behavior and healthcare needs of pharmacy customers [5].

2.2 Data Preprocessing

The data preprocessing stage was conducted to improve data quality and consistency prior to the pattern mining process. Preprocessing included the removal of irrelevant entries, standardization of drug name representations to prevent semantic duplication, and transformation of the data into a transaction-based itemset format. This stage is critical because inadequately processed transaction data may produce biased and unrepresentative association patterns [15]. The preprocessing results indicate a reduction in the number of unique items and a more consistent transaction structure, ensuring that the dataset is suitable for association rule mining.

2.3 Implementation of Apriori and FP-Growth Algorithm

After data preparation, the Apriori and FP-Growth algorithms were applied to extract drug purchasing patterns in the form of association rules. Both algorithms were selected because they are widely used and relevant methods for transaction data analysis, particularly in the context of market basket analysis [3]. In this study, both algorithms were executed using identical parameter settings, namely a minimum support threshold of 0.01 and a minimum confidence threshold of 0.17. The use of equivalent parameters ensures a fair and consistent comparison between the two algorithms. The implementation focused on identifying combinations of drugs that frequently co-occur within customer transactions, without discussing the technical operational details of each algorithm.

2.4 Evaluation

The evaluation stage was conducted to compare the performance of the Apriori and FP-Growth algorithms on pharmacy transaction data. The comparison was based on the number of frequent itemsets generated, the number of association rules that satisfy the specified parameter thresholds, and the computational time required for analysis. This evaluation aligns with the primary objective of association rule mining, which is to identify frequently occurring item combinations and establish relationships among items based on their confidence levels [16]. These metrics are essential

because the strength of association rules is assessed using measures such as support and confidence, which represent itemset frequency and the reliability of inter-item relationships, respectively [17] [18]. The evaluation results were then used to assess the suitability of each algorithm in uncovering drug purchasing patterns and their processing efficiency within a pharmacy environment.

3. RESULTS AND DISCUSSION

This section presents the results of the analysis of drug sales transaction data from Apotek Gadi Lamba Condet based on the research methodology applied, including dataset characteristics, data preprocessing results, the implementation of the Apriori and FP-Growth algorithms, and the performance evaluation of both algorithms. The discussion focuses on interpreting the meaning of the identified drug purchasing patterns and their relevance within the pharmacy transaction context.

3.1 Datasets

The dataset used in this study consists of drug sales transaction data from Apotek Gadi Lamba Condet covering the period from January to June 2025. The dataset represents real-world customer purchasing activities and reflects actual drug consumption needs and purchasing behavior in a pharmacy setting. After the data collection and initial filtering stages, the dataset used for analysis comprised 7,038 valid transactions, each containing a set of drugs purchased together within a single transaction.

Table 1. Comparison of Dataset Scale Before and After Preprocessing

Description	Before Preprocessing	After Preprocessing
Number of Unique Transactions	8.269	7.038
Number of Unique Items	1.632	1.495

The dataset characteristics indicate that pharmacy transactions involve a high diversity of drug types with uneven frequency distributions. This condition reflects the typical nature of retail pharmacy transaction data, where a small number of drugs appear very frequently, while the majority of drugs occur less often. Such a dataset structure forms an important basis for applying association rule mining to identify meaningful drug purchasing patterns.

3.2 Data Preprocessing

The data preprocessing stage plays a critical role in ensuring data quality before applying the Apriori and FP-Growth algorithms. At this stage, raw transaction data originally stored in an item-per-row format were transformed into a transaction-based itemset format. This process aimed to eliminate redundancy, standardize drug name representations, and ensure that each transaction accurately represents a single, complete purchase event.

Table 2. Results of Data Transformation into Transaction Format

ID Transaksi	Items
202501000002	{PRIMOLUT-N 5 MG, MOLACORT 0,75 MG, ASPLETS THROMBO 100 MG}
202501000003	{FENOFIBRATE 100 MG, SULCOLON 500 MG, DEXAMETHASON 0,5 MG}
202501000004	{CLONIDIN 0,15 MG, METFORMIN 500 MG}
202501000005	{C. PROTAGENTA MINIDOSE, ACTIFED HIJAU EXP SYR}

The preprocessing results show a reduction in both the number of transactions and unique items, indicating that the data cleaning and transformation processes successfully improved data consistency. With the refined data structure, the dataset becomes more representative of actual drug purchasing behavior. This stage ensures that the patterns generated by the algorithms are not artifacts of data recording inconsistencies, but rather reflect genuine customer purchasing habits.

3.3 Implementation of Apriori and FP-Growth Algorithm

After the dataset was prepared, the Apriori and FP-Growth algorithms were applied using identical parameter settings, namely a minimum support value of 0.01 and a minimum confidence value of 0.17. The application of both algorithms resulted in ten association rules with identical support, confidence, and lift values. This similarity indicates that both algorithms are capable of uncovering the same drug purchasing patterns when applied to pharmacy transaction data with comparable characteristics and parameters.

Table 3. Association Rules Generated by the Apriori Algorithm

Antecedents (If Purchased)	Consequents (Then Purchased)	Support	Confidence	Lift
PARACETAMOL 500 MG	CETIRIZINE 10 MG TABLET	0,026286	0,170664	1,288771
CETIRIZINE 10 MG TABLET	PARACETAMOL 500 MG	0,026286	0,198498	1,288771
VOLTADEX 50 MG	PARACETAMOL 500 MG	0,010514	0,195251	1,267688
AMLODIPIN 10 MG	METFORMIN 500 MG	0,023586	0,174187	1,250945
METHYLPREDNISOLONE 4 MG	AMLODIPIN 5 MG	0,018187	0,185239	1,166110
METHYLPREDNISOLONE 4 MG	PARACETAMOL 500 MG	0,017334	0,176556	1,146309
MEFINAL 500 MG	PARACETAMOL 500 MG	0,011793	0,176221	1,144135
METFORMIN 500 MG	AMLODIPIN 5 MG	0,025149	0,180612	1,136985
ASAM MEFENAMAT 500 MG	PARACETAMOL 500 MG	0,010230	0,173077	1,123723
SIMVASTATIN 10 MG	PARACETAMOL 500 MG	0,013072	0,170370	1,106150

Table 4. Association Rules Generated by the FP-Growth Algorithm

Antecedents (If Purchased)	Consequents (Then Purchased)	Support	Confidence	Lift
PARACETAMOL 500 MG	CETIRIZINE 10 MG TABLET	0,026286	0,170664	1,288771
CETIRIZINE 10 MG TABLET	PARACETAMOL 500 MG	0,026286	0,198498	1,288771
VOLTADEX 50 MG	PARACETAMOL 500 MG	0,010514	0,195251	1,267688
AMLODIPIN 10 MG	METFORMIN 500 MG	0,023586	0,174187	1,250945
METHYLPREDNISOLONE 4 MG	AMLODIPIN 5 MG	0,018187	0,185239	1,166110
METHYLPREDNISOLONE 4 MG	PARACETAMOL 500 MG	0,017334	0,176556	1,146309
MEFINAL 500 MG	PARACETAMOL 500 MG	0,011793	0,176221	1,144135
METFORMIN 500 MG	AMLODIPIN 5 MG	0,025149	0,180612	1,136985
ASAM MEFENAMAT 500 MG	PARACETAMOL 500 MG	0,010230	0,173077	1,123723
SIMVASTATIN 10 MG	PARACETAMOL 500 MG	0,013072	0,170370	1,106150

The results reveal that Paracetamol 500 MG is the most dominant drug in the identified association rules, appearing frequently in combination with various other drugs such as Cetirizine 10 MG Tablet, Voltadex 50 MG, Asam Mefenamat 500 MG, Mefinal 500 MG, and Simvastatin 10 MG. The central role of Paracetamol 500 MG in the association rules indicates its function as a complementary drug that is often purchased alongside other medications. This pattern reflects common pharmacy purchasing practices, where analgesic and antipyretic drugs are frequently added to primary therapies.

In addition, the association between Metformin and Amlodipin highlights a purchasing pattern related to chronic disease management. This pattern suggests that customers undergoing long-term treatment tend to purchase multiple medications in a planned and recurring manner. Consequently, the generated association rules not only represent drug co-occurrence frequencies, but also capture underlying therapeutic needs and customer purchasing behavior.

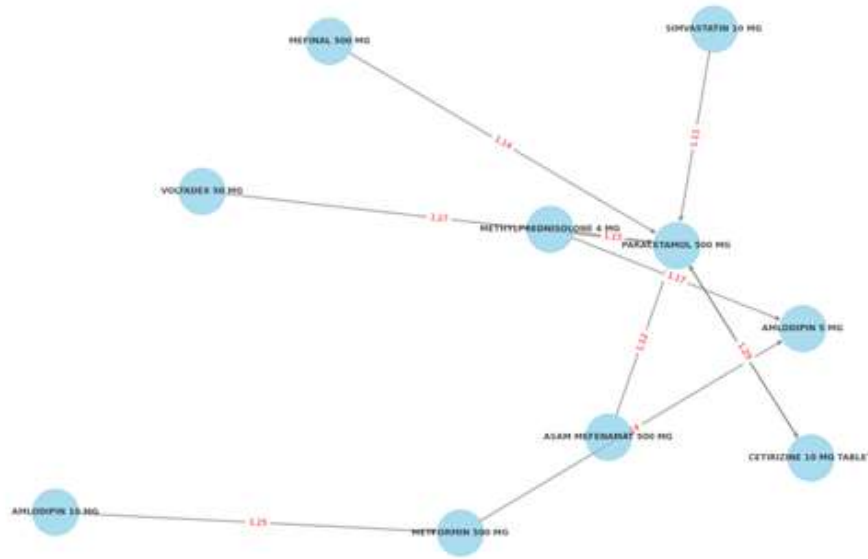


Figure 2. Association Rule Network Graph Generated by the Apriori Algorithm

Figure 2 illustrates the association rule network generated by the Apriori algorithm based on pharmacy transaction data. In this network, nodes represent drug items, while edges indicate association rules that satisfy the predefined minimum support and confidence thresholds. The numerical values displayed along the edges correspond to lift values, which indicate the strength of associations between drug pairs. The network shows that Paracetamol 500 MG acts as a central node, reflecting its frequent co-occurrence with several other drugs such as Cetrizine 10 MG Tablet, Asam Mefenamat 500 MG, Mefinal 500 MG, Simvastatin 10 MG, and Voltadex 50 MG. This structure indicates that Paracetamol 500 MG is commonly purchased alongside other medications in pharmacy transactions. In addition, the association between Metformin 500 MG and Amlodipin reflects a purchasing pattern related to chronic disease therapy.

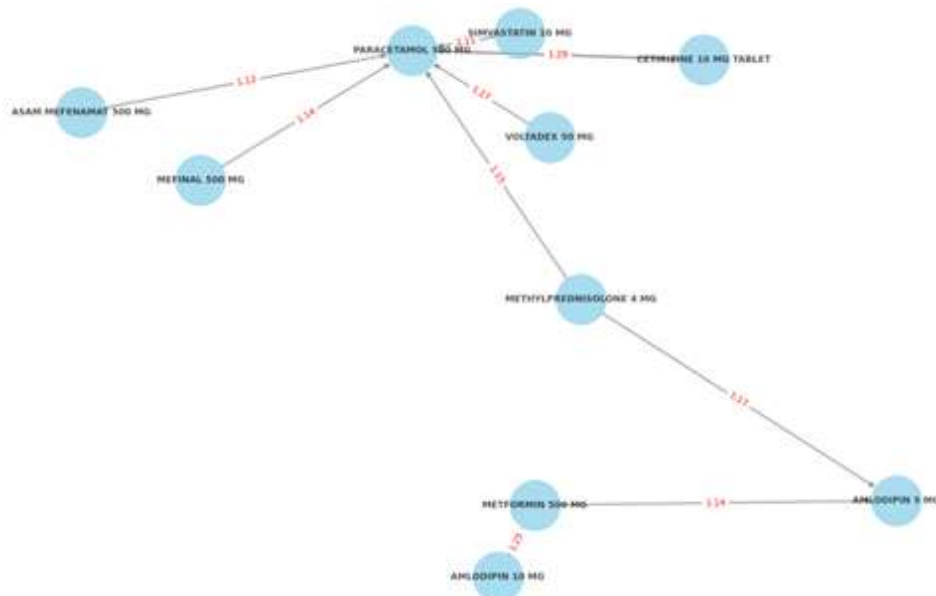


Figure 3. Association Rule Network Graph Generated by the FP-Growth Algorithm

Figure 3 presents the association rule network generated by the FP-Growth algorithm using the same transaction dataset and parameter settings as Figure 2. Each node represents a drug item and each edge represents an association rule, with lift values displayed to indicate relationship strength. The resulting network structure closely resembles that produced by the Apriori algorithm, with Paracetamol 500 MG again appearing as a central node connected to multiple other drugs. This similarity indicates that FP-Growth identifies the same dominant inter-drug purchasing patterns under identical dataset characteristics and parameter settings.

3.4 Evaluasi

The performance evaluation was conducted to compare the effectiveness of the Apriori and FP-Growth algorithms in processing pharmacy transaction data. The evaluation focused on the number of frequent itemsets generated, the number of association rules produced, and the computational time required.

Table 5. Performance Comparison of Apriori and FP-Growth Algorithms

Metric	Apriori	FP-Growth
Number of Frequent Itemsets	105	105
Number of Generated Rules	10	10
Computation Time (seconds)		1,065

The evaluation results show that both algorithms generated the same number of frequent itemsets and association rules, confirming that the quality of the extracted drug purchasing patterns is equivalent. The primary difference lies in computational time, where the Apriori algorithm demonstrated faster processing than FP-Growth for this medium-scale pharmacy transaction dataset.

These findings indicate that the selection of an association rule mining algorithm in the context of pharmacy transaction analysis is more strongly influenced by dataset characteristics than by differences in the algorithms' ability to generate meaningful purchasing patterns.

4. CONCLUSION

This study demonstrates that pharmacy sales transaction data can be effectively processed to uncover meaningful drug purchasing patterns based on real transaction records. Through systematic data cleaning and transformation, a total of 7,038 valid transactions were obtained, representing sets of drugs purchased together within single purchase events and enabling a systematic analysis of inter-drug purchasing relationships. The results indicate that the Apriori and FP-Growth algorithms generate identical association rules when applied to pharmacy transaction data using a minimum support threshold of 0.01 and a minimum confidence threshold of 0.17, as reflected in the identical drug combinations as well as their corresponding support, confidence, and lift values. All identified association rules exhibit lift values greater than one, indicating that the discovered relationships are meaningful and represent actual drug purchasing behavior. The main finding of this study suggests that, for medium-scale pharmacy transaction datasets, the primary difference between the Apriori and FP-Growth algorithms lies in computational efficiency rather than in the quality of the extracted purchasing patterns; therefore, the selection of an association rule mining algorithm in pharmacy transaction analysis should be driven primarily by dataset characteristics rather than algorithmic complexity.

REFERENCES

- [1] M. R. D. Bahari and R. Kurniawan, "Analysis of Drug Sales Patterns in the Belawan Naval Hospital Pharmacy Using Apriori Algorithm," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 4, pp. 1759–1768, Oct. 2024, doi: 10.47709/cnahpc.v6i4.4805.
- [2] K. D. Widodo and A. P. Lemantara, "Analysis and Implementation of the Apriori Algorithm for Strategies to Increase Sales at Sakinah Mart," *JUITA: Jurnal Informatika.*, vol. 11, no. 2, p. 203, 2023, doi: 10.30595/juita.v11i2.17341.
- [3] R. Amelia, D. Darmansyah, and A. M. Rismadin, "Perbandingan Algoritma Apriori dan Fp-Growth dalam Pengaplikasian Market Basket Analysis untuk Strategi Bisnis Retail," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 1, Jun. 2024, doi: 10.47065/bits.v6i1.5388.
- [4] 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," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 5, no. 2, p. 291, Dec. 2022, doi: 10.37600/tekinkom.v5i2.692.
- [5] D. Dwiputra, A. Mulyo Widodo, H. Akbar, and G. Firmansyah, "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*, vol. 2, no. 8, pp. 1229–1248, Aug. 2023, doi: 10.58344/jws.v2i8.403.
- [6] M. M. Lawal and O. T. Matthew, “FP-Growth Algorithm: Mining Association Rules without Candidate Sets Generation,” *Kasu Journal of Computer Science*, vol. 1, no. 2, pp. 392–411, Jun. 2024, doi: 10.47514/kjcs/2024.1.2.0016.
- [7] B. Mohanty, M. Tripathy, and S. Champati, “Performance Analysis of Association Rule Mining Algorithms: Evidence from the Retailing Industry,” *Journal of Engineering Science and Technology Review*, vol. 16, no. 5, pp. 108–122, 2023, doi: 10.25103/jestr.165.14.
- [8] 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.
- [9] A. A. C. Pabendon and H. D. Purnomo, “Penerapan Algoritma Apriori dan FP-Growth Untuk Market Basket Analisis Pada Data Transaksi NonPromo,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, p. 975, Jul. 2023, doi: 10.30865/mib.v7i3.6153.
- [10] A. I. Idris *et al.*, “Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores,” *The IJICS (International Journal of Informatics and Computer Science)*, vol. 6, no. 2, p. 107, Jul. 2022, doi: 10.30865/ijics.v6i2.4535.
- [11] P. Misra, “Purchase intention toward E-pharmacy: the consumption value perspective,” *Int J Pharm Healthc Mark*, vol. 19, no. 2, pp. 181–208, Apr. 2025, doi: 10.1108/IJPHM-12-2023-0107.
- [12] J. Nugraha and C. Yustia Purnamawati, “Application of the Association Method Using FP-Growth Algorithm to Find Pattern Of Medicine Purchasing Transactions at Pharmacy,” *Journal of Applied Research In Computer Science and Information Systems*, vol. 1, no. 2, pp. 48–53, Nov. 2023, doi: 10.61098/jarcis.v1i2.50.
- [13] B. Anwar, A. Ambiyar, and F. Fadhilah, “Application of the FP-Growth Method to Determine Drug Sales Patterns,” *Sinkron*, 2023, doi: 10.33395/sinkron.v8i1.12004.
- [14] D. Agushinta R. and M. M. Putri, “ASSOCIATION RULE ANALYSIS OF FP-GROWTH ALGORITHM ON DRUG PURCHASE PATTERNS,” *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 27, no. 3, pp. 196–212, 2022, doi: 10.35760/tr.2022.v27i3.4626.
- [15] Nurhachita and E. S. Negara, “A Comparison Between Naïve Bayes and The K-Means Clustering Algorithm for The Application of Data Mining on The Admission of New Students,” *Jurnal Intelektualita: Keislaman, Sosial, dan Sains*, vol. 17, pp. 43–48, Mar. 2020, doi: 10.5120/2237-2860.
- [16] S. Irianto Nadeak and Y. Ali, “Analysis of Data Mining Associations on Drug Sales at Pharmacies with APRIORI Techniques,” *International Journal of Information System & Technology Akreditasi*, vol. 5, no. 1, pp. 38–44, 2021.
- [17] F. Bao, L. Mao, Y. Zhu, C. Xiao, and C. Xu, “An Improved Evaluation Methodology for Mining Association Rules,” *Axioms*, vol. 11, no. 1, Jan. 2022, doi: 10.3390/axioms11010017.
- [18] K. Y. Huh and I. Song, “Analyzing collaborations in clinical trials in Korea using association rule mining,” *Transl Clin Pharmacol*, vol. 32, no. 4, pp. 177–186, Dec. 2024, doi: 10.12793/tep.2024.32.e17.