JUIT
ADMI Jurnal Ilmiah Teknik
OPEN ACCESS
JUIT Vol 4 No. 3 September 2025 | ISSN: 2828-6936 (Print), ISSN: 2828-6901 (online), Page. 109-117

COURSE SELECTION PATTERN ANALYSIS USING APRIORI ALGORITHM

Syifa Nurani Rahmayanti[1], Metty Mustikasari[2]*
[1]Information System Management, Syifanurani12@gmail.com, Universitas Gunadarma, Indonesia
[2]Information System Management, metty@staff.gunadarma.ac.id, Universitas Gunadarma, Indonesia

Article History
Received : September 16, 2025
Revised : September 30, 2025
Accepted : October 10, 2025
Avl. Online : October 13, 2025

Corresponding author*:
metty@staff.gunadarma.ac.id

Cite This Article:
Syifa Nurani Rahmayanti, & Mustikasari, M. (2025). Course Selection Pattern Analysis Using Apriori Algorithm. *Jurnal Ilmiah Teknik*, *4*(3), 109–117.

DOI:
https://doi.org/10.56127/juit.v4i3.2326

Abstract: This research discusses Association Rules as one of the data mining functions implemented using the Apriori Algorithm. The Institute for Computerization Development (LePKom) is a unit at Gunadarma University that organizes courses and workshops. The course and workshop participants are Gunadarma University students from the following undergraduate and diploma programs: Bachelor of Information Systems (S1-SI), Bachelor of Computer Systems (S1-SK), Diploma in Information Management (D3-MI), Diploma in Computer Engineering (D3-TK), and Bachelor of Informatics Engineering (S1-TI). New students must select three course topics from the six fundamental courses available at the Institute for Computerization Development (LePKom), and from these three choices, students will be assigned one course each semester. Association rules can be generated using the Apriori Algorithm to identify patterns of which course topics are most frequently selected together by new students each semester, as well as to optimize course infrastructure requirements.

Keywords: Data Mining, Association Rule, Apriori, FP Growth, Weka.

INTRODUCTION

Data mining is a decision support process where information patterns can be discovered in data. Data mining is data analysis using tools to find patterns and rules in datasets. Software is tasked with finding patterns by identifying rules and features in the data and is expected to be able to recognize these patterns in data with minimal input from users (Chee et al., 2019; Yang et al., 2024). Data mining is an algorithm that is used for the treatment of data to find hidden patterns from the data processed. The data processed then produces new knowledge from old data, which can be used in the determination of future decisions (Samboteng et al., 2022). Information generated from applying data mining techniques is used to explore and predict potential opportunities within an organization.

Data mining techniques have been widely employed in prediction. For example, developing a decision support model for determining the target multi-family housing complex for green remodeling using data mining techniques (Al-Refaie et al., 2023). Lv et al. proposed a complex data fusion and efficient learning algorithm (multi-graphics

processing unit (GPU)) to process the multi-dimensional and complex big data based on the compositive rough set model (Lv et al., 2022). Zhang et al. proposed a novel category-induced coarse-to-fine domain adaptation approach (C2FDA) for cross-domain object detection (H. Zhang et al., 2022). Zhang et al. analyzed the problems of existing container positioning methods and proposed a vision-based container position measuring system to provide precise parameters for container lifting operations (Y. Zhang et al., 2023). Mitici et al. employed dynamic predictive maintenance for multiple components using data-driven probabilistic remaining useful life prognostics illustrated by the case of turbofan engines (Mitici et al., 2023). Some early definitions of data mining included a focus on the automation process. Suo's research shows that data mining can clearly show behavioral pattern in the context of ship and efficiently uncover suspicious ship patterns (Suo et al., 2022).

One form of pattern that data mining can produce is association rules. Association rules can be used to find relationships or cause-and-effect connections. Association rules can be generated using the Apriori Algorithm. WEKA is a tool for machine learning algorithms and data preprocessing. With this tool, dataset preprocessing can be performed, input can be provided to learning schemes, and the resulting classifiers can be analyzed. WEKA's working areas include standard methods for data mining: regression, classification, clustering, association rule mining, and attribute selection.

In higher education institutions, data can be obtained from historical records, so data will continuously accumulate, for example, student data. The new student admission process at a university produces abundant data from these new students. The Institute for Computerization Development (LePKom) is a unit at Gunadarma University that organizes courses and workshops that equip students with practical computer-based information and communication technology materials needed in the workforce.

These courses prioritize student activeness and independence in understanding, learning, and exploring their chosen course materials. Course implementation techniques include pretests, text/print-based learning, video-based learning, practical activities (creating programs, running simulations, etc.), and posttests. Course and workshop participants are Gunadarma University students from Bachelor of Information Systems (S1-SI), Diploma in Information Management (D3-MI), and Bachelor of Informatics Engineering (S1-TI) programs.

In one semester, students can choose one module from the available course modules, where the materials provided become more specific as the semester advances. New students must select two or three course options from six fundamental courses available at the Institute for Computerization Development (LePKom), and from those two or three choices, students will be assigned one course each semester. Students are required to choose two or three courses as backups in case their first or second course choice does not fit their available schedule. The fundamental materials that must be chosen by new students include DBMS, ERP, Desktop Programming, Web Programming, Networking, and Server OS. In their final level, students are required to complete a project or take national or international certification exams.

With the data obtained, it is expected that patterns of course materials frequently selected together in the first semester by new students in each cohort can be identified to optimize course infrastructure requirements. The course data is processed using data mining techniques, employing Association Rules with the Apriori algorithm. Additionally, the FP Growth algorithm is also used as a comparison with the Apriori algorithm (Idris et al., 2022).

LITERATURE REVIEW

The Apriori algorithm is easy to implement and very simple, used to extract all frequent items in a database. This algorithm is very effective in finding the relationship pattern of one or more itemsets in a large data set so that it is effective in calculating data and finding patterns of combinations (Fey, 2023). This algorithm allows multiple database scans to find frequent itemsets by using k-itemsets to generate k+1 itemsets. Each k-itemset must be greater than or equal to the minimum support threshold as frequency. Otherwise, these candidate itemsets are excluded. In the first iteration, the algorithm scans the database to find the frequency of items in the database. The frequency of 1-itemsets is used to find items in 2-itemsets, which in turn are used to find 3-itemsets, and so on, until no more k-itemsets are found. If an itemset is not frequent, its large subsets are also not frequent.
Association analysis is defined as a process to discover all association rules that meet the minimum support requirements and minimum confidence requirements.

The Institute for Computerization Development (LePKom), specifically the virtual-based LePKom Mandiri, was established at Gunadarma University in 2013 as a subdivision

of LePKom Gunadarma University. Virtual-based LePKom Mandiri provides education and training activities for students beyond their regular lectures and practicum sessions.

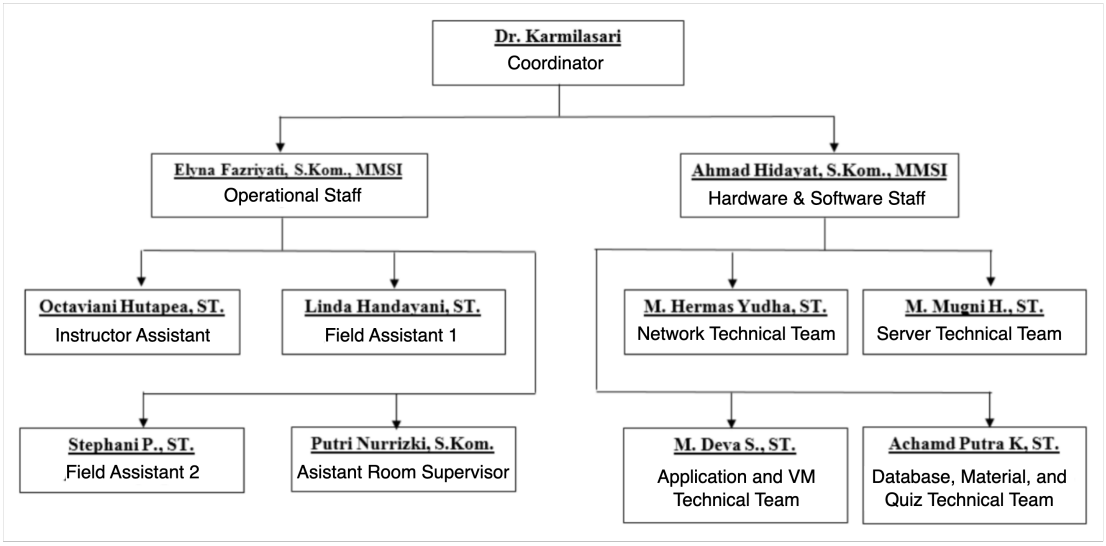

Figure 1. Organizational structure of virtaul-based LePKom Mandiri

Organizational structure is a component that demonstrates work relationships, authority, and responsibilities from superiors to subordinates, creating an orderly work system. With an organizational structure in place, activities are expected to run smoothly as intended.

RESEARCH METHOD

The stages involved in research consist of Problem Formulation, Data Collection, Data Processing with Data Mining, Final Analysis, and Conclusion.

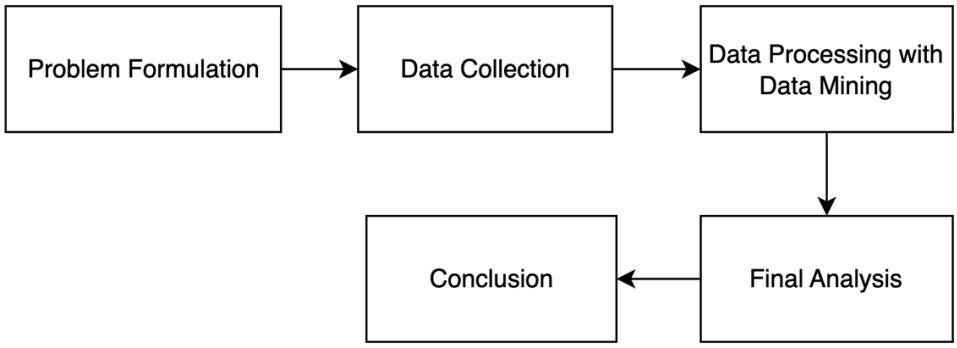

Figure 2. Research Method

Apriori data analysis consist of Data Selection, Pre-processing, Data Transformation, Data Mining with the following steps preparing .csv data, removing unused attribute,

setting parameters, and analysis results. The parameters set include: LowerBoundMinSupport (0.05 or 5%), MetricType (Confidence), minMetric (0.5 or 50%), and numRules (maximum 20 rules). The basic methodology of association analysis is divided into two stages:

1. High-Frequency Pattern Analysis

   This stage searches for item combinations that meet the minimum support value requirements in the database. Support is a measure that indicates the level of dominance of an item or itemset from all transactions. The formula for calculating the support value is:

   $$Support\ A = \frac{Number\ of\ Transaction\ consist\ of\ A}{Total\ Transaction} \quad (1)$$

   While the support value for 2-items uses the formula:

   $$Support(A, B) = \frac{\Sigma\ Number\ of\ Transaction\ consist\ of\ A\ and\ B}{\Sigma\ Transaction} \quad (2)$$

2. Association Rule Formation

   After all high-frequency patterns are found, association rules that meet the minimum confidence requirements are then identified by calculating the confidence of associative rule A → B. Confidence is a measure that indicates the conditional relationship between two items. The formula for calculating the confidence value is:

   $$Confidence = P(B|A) = \frac{\Sigma\ Number\ of\ Transaction\ consist\ of\ A\ and\ B}{\Sigma\ Transaction} \quad (3)$$

FP-Growth is an alternative algorithm that is both efficient and effective for discovering frequent itemsets in large datasets. While FP-Growth demonstrates superior speed in producing results compared to the Apriori algorithm, it falls short in generating high confidence values. The advantages of FP-Growth include its ability to recognize objects nonlinearly, facilitate the mapping of inputs to outputs without understanding the underlying process, exhibit robust parallel processing capabilities, and demonstrate error tolerance (Harianto et al., 2023).

FP-Growth algorithm, which is a data mining technique based on FP-Tree, can discover a set of complete frequency patterns. FP-Tree is an extended prefix-tree structure

to store important and quantitative information related to frequency patterns, avoiding the shortcomings of the Apriori-based approach (Jang et al., 2021).

RESULT AND DISCUSSION

Graphical representation of the most taken courses and course combinations frequently selected together by the 2015, 2016, and 2017 cohorts.

Tabel 1. Total Course Selection

Cohorts	Courses						Student Count
	DBMS	ERP	Desktop	Web	Networking	Server OS	
2017	946	401	641	1042	826	512	1467
2016	1078	219	859	1273	605	385	1594
2015	989	365	868	796	710	340	1515

The table above presents the combined total of courses taken by students from the Informatics Engineering (TI), Information Systems (SI), and Information Management (MI) programs.

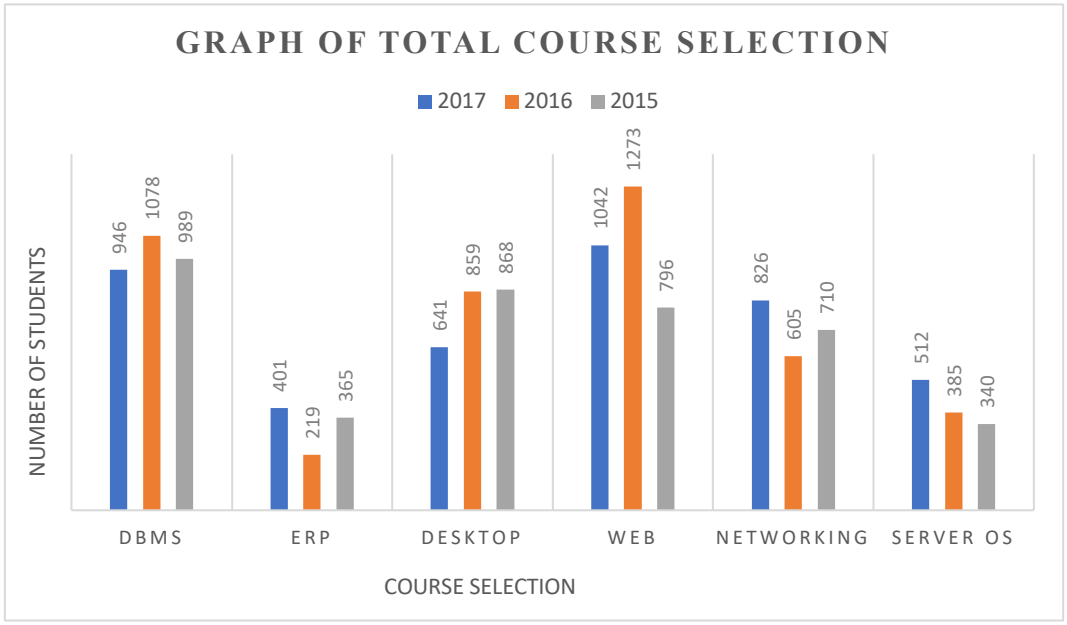

Figure 3. Graph of Total Course Selection

The graph illustrates the most frequently selected courses in each cohort. The 2015 cohort showed the highest interest in the DBMS course, while the 2016 and 2017 cohorts both favored the WEB course, with the 2017 cohort having 1,042 students selecting it out of a total enrollment of 1,467 students.

The DBMS course provides students with knowledge of Relational Database Concepts and familiarizes them with DBMS software such as MySQL, SQL Server, and Oracle. Students learn to understand database connections using XAMPP with MySQL, as well as concepts related to DDL (Data Definition Language) and DML (Data Manipulation Language).

Tabel 2. Speed Performance Test on WEKA

Data count	Apriori	FP Growth
1,000	0.33 second	0.34 second
5,000	0.35 second	0.37 second
10,000	0.66 second	0.71 second

The experimental results of processing speed tests in the Weka application indicate that the Apriori algorithm demonstrates faster processing times than the FP-Growth method when applied to datasets of 1,000, 5,000, and 10,000 records.

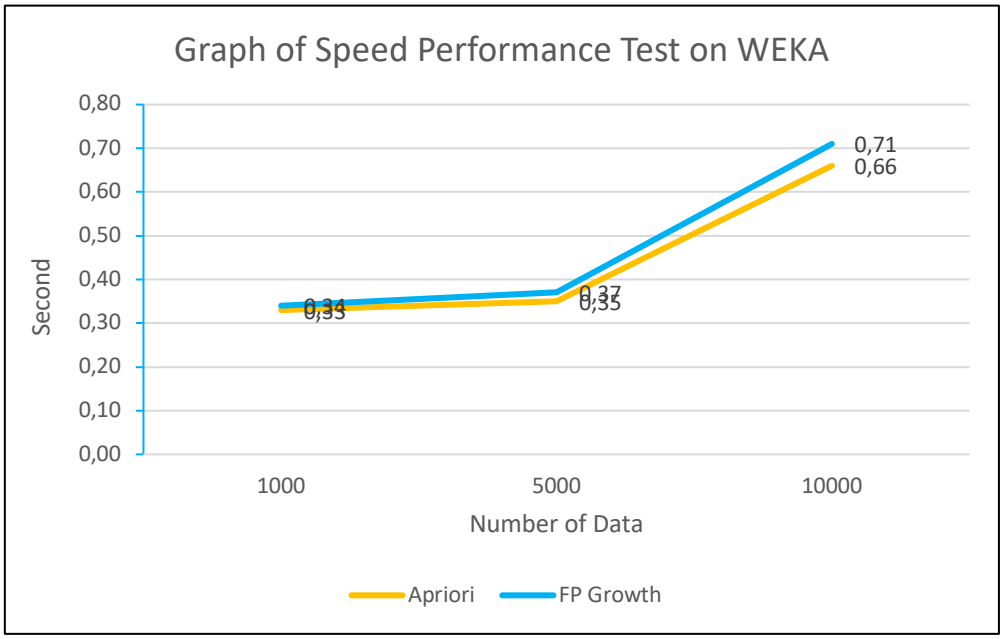

Figure 4. Graph of Speed Performance Test on WEKA

This experiment was carried out to evaluate the performance of the Apriori and FP-Growth algorithms. The FP-Growth analysis results were used as a benchmark for comparison with the Apriori analysis, enabling the identification of differences between the two algorithms. The graph presents the experimental results of both algorithms in Weka, where the processing time is expressed in seconds. The processing speed varies according to the volume of data and the number of attributes involved.

CONCLUSION

Based on the research conducted on the Virtual LePKom Mandiri of Gunadarma University using the Apriori Algorithm method, several conclusions can be drawn.

The Apriori Algorithm method can be used to address issues related to predicting students' interests in course selection, which exhibit varying patterns and tendencies. This is demonstrated by the differences in the rules generated in the final process.

The Apriori Algorithm can assist in meeting the course infrastructure needs of the Virtual LePKom Mandiri at Gunadarma University. The courses most frequently taken together across cohorts are Web Programming and Database Management Systems (DBMS).

Compared to the FP-Growth algorithm, the results of the Apriori Algorithm analysis using WEKA show identical support and confidence values. The main difference between the two methods lies in the processing time, where the Apriori Algorithm demonstrates faster execution than FP-Growth.

REFERENCES

Al-Refaie, A., Abu Hamdieh, B., & Lepkova, N. (2023). Prediction of Maintenance Activities Using Generalized Sequential Pattern and Association Rules in Data Mining. *Buildings*, *13*(4). https://doi.org/10.3390/buildings13040946

Chee, C. H., Jaafar, J., Aziz, I. A., Hasan, M. H., & Yeoh, W. (2019). Algorithms for frequent itemset mining: a literature review. In *Artificial Intelligence Review* (Vol. 52, Issue 4). https://doi.org/10.1007/s10462-018-9629-z

Fey, F. P. (2023). Application of the Apriori Algorithm to Purchase Patterns. *The Indonesian Journal of Computer Science*, *12*(2). https://doi.org/10.33022/ijcs.v12i2.3105

Harianto, H., Sudianto, A., & Bahtiar, H. (2023). Analisa Data Transaksi Penjualan Barang Menggunakan Algoritma Apriori Dan FP-Growth. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, *6*(2). https://doi.org/10.47970/siskom-kb.v6i2.371

Idris, A. I., Sampetoding, E. A. M., Ardhana, V. Y. P., Maritsa, I., Sakri, A., Ruslan, H., & Manapa, E. S. (2022). 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)*, *6*(2). https://doi.org/10.30865/ijics.v6i2.4535

Jang, H. J., Yang, Y., Park, J. S., & Kim, B. (2021). Fp-growth algorithm for discovering region-based association rule in the iot environment. *Electronics (Switzerland)*, *10*(24). https://doi.org/10.3390/electronics10243091

Lv, Z., Chen, D., & Lv, H. (2022). Smart City Construction and Management by Digital Twins and BIM Big Data in COVID-19 Scenario. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*(2). https://doi.org/10.1145/3529395

Mitici, M., de Pater, I., Barros, A., & Zeng, Z. (2023). Dynamic predictive maintenance for multiple components using data-driven probabilistic RUL prognostics: The case

of turbofan engines. *Reliability Engineering & System Safety*, *234*, 109199. https://doi.org/10.1016/J.RESS.2023.109199

Samboteng, L., Rulinawaty, Kasmad, M. R., Basit, M., & Rahim, R. (2022). Market Basket Analysis Of Administrative Patterns Data Of Consumer Purchases Using Data Mining Technology. *Journal of Applied Engineering Science*, *20*(2). https://doi.org/10.5937/jaes0-32019

Suo, Y., Ji, Y., Zhang, Z., Chen, J., & Claramunt, C. (2022). A Formal and Visual Data-Mining Model for Complex Ship Behaviors and Patterns. *Sensors*, *22*(14). https://doi.org/10.3390/s22145281

Yang, J., Bian, X., Qi, Y., Wang, X., Yang, Z., & Liu, J. (2024). A spatial-temporal data mining method for the extraction of vessel traffic patterns using AIS data. *Ocean Engineering*, *293*. https://doi.org/10.1016/j.oceaneng.2023.116454

Zhang, H., Luo, G., Li, J., & Wang, F. Y. (2022). C2FDA: Coarse-to-Fine Domain Adaptation for Traffic Object Detection. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 12633–12647. https://doi.org/10.1109/TITS.2021.3115823

Zhang, Y., Huang, Y., Zhang, Z., Postolache, O., & Mi, C. (2023). A vision-based container position measuring system for ARMG. *Measurement and Control*, *56*(3–4), 596–605. https://doi.org/10.1177/00202940221110932