



Application of Data Mining Using the Apriori Algorithm in Analyzing Subject Selection Patterns of Tutoring Students

Muhammad Rizky Ferdiansyah¹, Naufal Renanda², Afriza Akhid Khoiruddin³, Arya Subastian⁴, Muhammad Arifin^{5*}

^{1,2,3,4,5} Faculty of Engineering University Muria Kudus

202353089@std.umk.ac.id¹, 202353071@std.umk.ac.id², 202353084@std.umk.ac.id³, 202353066@std.umk.ac.id⁴, arifin.m@umk.ac.id^{5*}

Abstract

This study examines the application of data mining using the Apriori algorithm to analyze subject selection patterns among tutoring students in Kudus, Central Java. With the increasing number of students attending tutoring, understanding subject selection patterns is crucial to improve the effectiveness of educational services. The Apriori algorithm, a popular association rule mining technique, is used to identify relationships between frequently selected subjects. The research dataset consists of student subject selection transaction data, including information such as student name, student ID number, tutoring branch, and selected subjects. The analysis process included data preprocessing, data transformation into transaction format using Transaction Encoder, application of the Apriori algorithm with a minimum support of 0.05, and formation of association rules with a minimum confidence of 0.3. The results show frequent itemsets indicating the most popular subjects and association rules that describe students tendencies in selecting subject combinations. These findings can be utilized by tutoring managers to design more effective learning packages, optimize the allocation of teaching resources, and provide subject recommendations tailored to student needs. This research contributes to the development of educational data mining in the context of tutoring institutions in Indonesia.

Keywords: Apriori Algorithm; Association Rule Mining; Data Mining; Educational Data Mining; Subject Selection; Tutoring

1. Introduction

The development of information technology has brought significant changes in various sectors, including education. One notable impact is the growing volume of data generated from learning activities, both in formal schools and non-formal educational institutions such as tutoring centers. If properly managed and analyzed, this data can provide valuable insights for improving the quality of educational services [1].

Tutoring centers constitute one of the important components of the education system in Indonesia. Students attend these programs to gain a deeper understanding of specific subjects, prepare for examinations, and improve their academic achievement. Each student has diverse needs and preferences in determining the subjects they choose. Therefore, understanding subject selection patterns can help tutoring managers design more effective and efficient learning programs [2].

Data mining is the process of extracting interesting knowledge or patterns from large datasets. One popular data mining technique is association rule mining, which aims to discover relationships or associations among items in a dataset [3]. The Apriori algorithm is one of the widely used methods in association rule mining to find frequent itemsets based on support and confidence values [4].

Research by Hao et al. (2023) demonstrated that the Apriori algorithm can be used for cognitive intervention analysis in college students' sports health [5]. Research by Zhou (2025) proved that improvements to the Apriori algorithm can increase efficiency in educational decision-making systems, particularly in analyzing student behavior and predicting academic risk [6]. This study aims to implement the Apriori algorithm in analyzing subject selection patterns by students at tutoring institutions in Kudus, Central Java.

Through the identification of subject combinations that are frequently selected simultaneously, this research is expected to reveal patterns that can be utilized in the development of learning packages more relevant to student needs, optimize the distribution of teaching staff, and provide more accurate subject selection recommendations.

2. Research Methodology

2.1. Type of research

This study employs a quantitative approach with a descriptive method. The objective is to identify subject selection patterns of tutoring students through the application of data mining techniques, specifically using the Apriori algorithm.

2.2. Data source

The data used in this study is secondary data obtained from a tutoring center information system in Kudus, Central Java. The data includes information on students and the subjects they selected, with the following attributes: createdBy (name of the admin who entered the data), modifiedAt (date and time of the last modification), cabang (tutoring branch), siswa (student name), nis (student identification number), and mata_pelajaran (subjects selected by the student, which may contain more than one subject separated by commas).

2.3. Research stages

2.3.1. Data collection

Data were collected from the tutoring center information system that records transaction data on student subject selections. The data is stored in Excel format (.xlsx) and was subsequently loaded using the Pandas library in Python.

3. Results and Discussion

3.1. Data description

The dataset used in this study consists of subject selection transaction data by students at a tutoring institution in Kudus, Central Java. The data covers various subjects, including Indonesian Language, English, Mathematics, Biology, Geography, Physics, Chemistry, Economics, and History. Each record represents one student who selected one or more subjects.

3.2. Data processing results

The preprocessing process successfully separated subjects contained in a single entry into separate lists. For example, for student CHATARINA AURELLIA who selected 'Mathematics, Biology', the data was transformed into the list ['Mathematics', 'Biology']. This process is important to ensure that each subject can be identified as a separate item in the Apriori algorithm.

3.3. Data transformation results

Data transformation using TransactionEncoder produced a dataframe in one-hot encoding format. Each subject is represented as a column with boolean values, where True indicates that the student selected that subject, while False indicates otherwise. This format allows the Apriori algorithm to work more efficiently in identifying frequent itemsets.

	Bahasa Indonesia	Bahasa Inggris	Biologi	Ekonomi	Fisika	Geografi	Kimia	Matematika
0	True	False	False	False	False	False	False	False
1	False	True	False	False	False	False	False	False
2	False	False	True	False	False	False	False	True
3	False	False	False	False	False	True	False	False
4	False	True	False	False	False	False	False	False

Fig. 1: Data transformation table

3.4. Frequent itemsets

Applying the Apriori algorithm with a minimum support of 0.05 produced frequent itemsets showing subjects or combinations of subjects frequently selected by students. Frequent itemsets are divided into several categories.

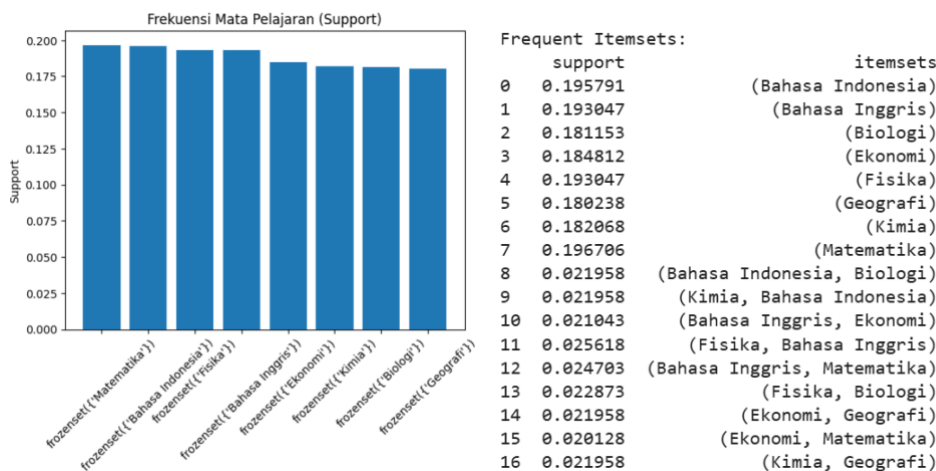


Fig. 2: Frequent itemset data

3.4.1. Single item frequent itemsets

Single item frequent itemsets represent individual subjects most frequently selected by students. These findings provide an overview of student interest levels in each subject at the tutoring center. Visualization in the form of bar charts is used to display the support value for each subject, reflecting the proportion of students who selected that subject.

3.4.2. Multi-item frequent itemsets

Multi-item frequent itemsets show combinations of subjects frequently selected together. These patterns are important for understanding relationships between subjects and can be used to design learning packages that match student needs.

3.5. Association rules

Association rules are formed from frequent itemsets with a minimum confidence of 0.3. Each rule takes the form: if a student selects subject X, then the student tends to also select subject Y. Rules are evaluated based on three main metrics: (1) Support, which measures how often that subject combination appears in the dataset; (2) Confidence, which measures how often the rule is true (conditional probability); and (3) Lift, which measures the strength of the relationship compared to if both subjects were selected independently. A scatter plot displaying the relationship between support and confidence provides visualization that facilitates the identification of the most meaningful rules.

3.6. Interpretation of results

3.6.1. Subject selection patterns

Frequent itemsets reveal the subjects most in demand by students. These patterns can be used to optimize the allocation of teaching staff based on the most popular subjects, prepare better learning materials for popular subjects, and identify subjects with low demand for evaluating marketing strategies or program development.

3.6.2. Relationships between subjects

Association rules reveal interesting relationships between subjects. For example, students who choose science subjects such as Mathematics and Physics tend to also select Chemistry. This information can be used to design learning packages that combine subjects frequently selected together, provide additional subject recommendations to students based on subjects already selected, and arrange class schedules so that frequently co-selected subjects do not conflict.

3.6.3. Implications for learning strategy

The findings of this study have practical implications for developing learning strategies in tutoring centers. Personalized learning: the system can provide personalized subject recommendations based on the selection patterns of students with similar characteristics. Resource optimization: management can allocate teachers and classrooms more efficiently based on projected subject demand. Package development: tutoring centers can develop more attractive learning packages based on frequently co-selected subject combinations.

4. Conclusion

The Apriori algorithm was successfully applied to analyze subject selection patterns of tutoring students using a minimum support of 0.05 and a minimum confidence of 0.3. The discovered frequent itemsets indicate individual subjects and subject combinations most frequently selected by students, providing insight into student learning preferences. The generated association rules reveal meaningful relationships between subjects that can be used to provide subject recommendations and design more effective learning packages. The analysis results

can be utilized by tutoring managers to optimize resource allocation, develop learning strategies, and improve student satisfaction through more personalized learning program design. The application of educational data mining techniques through the Apriori algorithm has proven effective in extracting valuable knowledge from subject selection transaction data at tutoring institutions.

5. Recommendations

5.1. For tutoring practitioners

1. Implement an association rule-based subject recommendation system to assist students in determining subject choices that suit their needs and preferences.
2. Develop bundled learning packages based on frequent itemsets to improve learning efficiency while increasing student satisfaction.
3. Utilize the analysis results for better resource planning, including teacher allocation and class scheduling.

5.2. For future research

1. Explore other algorithms in association rule mining, such as FP-Growth or Eclat, to compare performance and results with the Apriori algorithm.
2. Consider adding other variables, such as student academic achievement, level of education, or field of study, to produce more comprehensive analysis.
3. Implement an improved Apriori algorithm by applying optimization techniques, such as hash-based methods or parallel computing, to increase computational efficiency for large datasets.
4. Develop a hybrid recommendation system that integrates association rule mining with collaborative filtering or content-based filtering to improve accuracy and relevance of recommendation results.
5. Adopt a longitudinal study approach to analyze the dynamics of changes in subject selection patterns over time and identify the factors that influence them.

References

- [1.] S. K. Mohapatra and D. Sahoo, Application of big data analytics in education: A review, *Education and Information Technologies*, vol. 27, pp. 1025-1045, 2022.
- [2.] R. Pratama, A. Nugroho, and S. Wibowo, Analysis of student learning preferences using data mining techniques, *International Journal of Emerging Technologies in Learning*, vol. 18, no. 3, pp. 45-56, 2023.
- [3.] M. Al-Hagery and H. Al-Assaf, A survey of data mining techniques in educational systems, *IEEE Access*, vol. 9, pp. 116389-116404, 2021.
- [4.] Z. Tang, Z. Jiang, Y. Li, H. Yuan, J. Han, and C. Chen, Association analysis of online learning behavior based on Apriori algorithm, *Frontiers in Computing and Intelligent Systems*, vol. 9, no. 2, pp. 18-22, 2024.
- [5.] H. Hao et al., Application of Apriori algorithm in cognitive intervention of college students sports health, *Journal of Healthcare Engineering*, 2023.
- [6.] Zhou et al., Improvements to the Apriori algorithm and its application in educational decision-making systems, in *Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security*, ACM, 2025.
- [7.] A. Rahman and S. Das, Data mining for students trends analysis using Apriori algorithm, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 6, no. 4, 2017.
- [8.] An effective learning management system for revealing student performance attributes, *arXiv preprint arXiv:2403.13822*, 2024.
- [9.] Apriori algorithm based prediction of students mental health risks in the context of artificial intelligence, *Frontiers in Public Health*, 2025.
- [10.] Y. Zhou, Improved Apriori algorithm for educational decision support systems, *Education and Information Technologies*, 2025.