

Application of apriori algorithm to find relationships between courses based on student grades STMIK YMI Tegal

Muhamad Nur Hassan¹, Gunawan Gunawan², Zaenul Arif³
^{1,2,3}Informatics Engineering, STMIK YMI Tegal, Indonesia

ARTICLE INFO

Article history:

Received Mar 22, 2024
Revised Apr 4, 2024
Accepted Apr 4, 2024

Keywords:

Apriori Algorithm;
Curriculum Development;
Positive Relationships;
Quality of Learning;
Student.

ABSTRACT

This research explores the application of the Apriori algorithm to investigate the relationship between courses based on student grades at STMIK YMI Tegal. This research focuses on analyzing the relationship between courses to support curriculum development that is responsive and relevant to industry needs and improves the quality of learning. The main objective of this research is to identify and understand relationship patterns between various courses based on student analysis scores using the Apriori algorithm, an effective data mining methodology for uncovering association rules between items in large datasets. By using a quantitative approach and quasi-experimental design, this research succeeded in analyzing grade data from various semesters, identifying combinations of courses that often appear together with high grades, indicating a positive correlation between related courses. The results of the analysis reveal that several basic courses play a significant role in forming a strong foundation for advanced courses, highlighting the importance of a capable curriculum structure. Although the lift scores show a neutral relationship, these findings provide important initial insights for further understanding of interactions between courses. The implication for curriculum development is the need to emphasize the integration of courses that have positive relationships to support a coherent learning process and increase student success.

This is an open access article under the [CC BY-NC](#) license.



Corresponding Author:

Muhamad Nur Hassan,
Informatics Engineering,
STMIK YMI Tegal,
#1 Pendidikan Street, Tegal City, Central Java 52142, Indonesia
Email: dcaninfinity@gmail.com

1. INTRODUCTION

In the era of globalization and advances in information technology, the world of education is faced with the challenge of continuing to improve the quality of learning and the relevance of the material taught to the needs of the world of work. (Mohamed Hashim et al., 2022) (Al Husaeni et al., 2024). STMIK YMI Tegal as one of the higher education institutions in the field of information and computer technology, strives to continue to adapt its curriculum so that it can produce graduates who are competent and ready to compete. (Núñez-Canal et al., 2022) (Abelha et al., 2020). In this research, analyzing the relationship between courses is important to understand how each course contributes to graduate competency (Birdman et al., 2022) (De Villiers, 2021).

The problem faced is a lack of comprehensive understanding of the relationship between courses, especially based on student academic performance (Jia et al., 2022) (Chang et al., 2022). This is a problem because without in-depth understanding, curriculum development will not be effective in achieving the expected learning objectives (Kim et al., 2022). Apart from that, students also have difficulty choosing courses that can optimally support the development of their skills.

The application of the Apriori algorithm to find relationships between courses based on student grades at STMIK YMI Tegal has the potential to provide substantial benefits in various

aspects of education. First of all, this research is expected to provide in-depth insight into the dynamics between subjects and the relationship patterns that may exist between them. With a better understanding of how courses relate to each other, educational institutions can optimize their curricula, ensuring that the sequence and content of courses are more structured and relevant to students. Apart from that, it is also hoped that the results of this research can be used by lecturers and professors to design more effective and integrated learning strategies, which in turn can improve students' understanding and academic performance. By analyzing the relationships between courses, educational institutions can also identify areas that require special attention or improvement in their curriculum. Thus, it is hoped that this research can become a basis for the development of more efficient and results-oriented education, with a focus on improving students' learning experiences and preparing them to face future challenges.

This research was conducted to overcome this problem with the aim of finding relationships between courses based on student grades using the Apriori algorithm (Yan, 2024). This research is important because the results can be used to recommend more effective curriculum adjustments, facilitate students in choosing courses, and ultimately improve the quality of education and student satisfaction. (S. Lee, 2022).

To overcome this problem, we propose the application of the Apriori algorithm, which is an effective data mining method in finding relationships between items in a large dataset. (Sun, 2020). In this research, "items" can be interpreted as courses, and the dataset used is a collection of student grades (Mai et al., 2022) (Zhang et al., 2023). The Apriori algorithm was chosen because of its proven ability to efficiently find strong association rules between items (Telikani et al., 2020).

This research contributes to filling the existing gap by providing new insights into the relationships between courses based on student grade analysis, which has not previously been widely explored. By using data mining methods, especially the Apriori algorithm, this research proposes innovations in curriculum analysis and education program development (Huang, n.d.).

The state of the art of this research refers to the use of the Apriori algorithm in various fields to find association rules, but its application in an analysis of relationships between courses based on student grades is one of the innovative efforts proposed by this research (Qi et al., 2024). It is hoped that, through this research, significant patterns of relationships between courses can be found based on student grades, which can later be used as a basis for recommending more relevant and effective curriculum adjustments (Chen et al., 2020). Apart from that, it is also hoped that this research can contribute to improving the quality of learning and student satisfaction at STMIK YMI Tegal.

2. RESEARCH METHOD

Research Design

This research uses a quantitative approach with a quasi-experimental design to investigate the relationship between courses based on student grades at STMIK YMI Tegal (Culver & Bowman, 2020). The quasi-experimental approach was chosen because it allows the effect of the intervention to be observed without the need to randomize the research subjects, which in this case are students' courses and grades (Gopalan et al., 2020). This design allows research to be conducted under more controlled conditions compared to observational studies, but still provides flexibility in data collection and analysis (K. J. Lee et al., 2021).

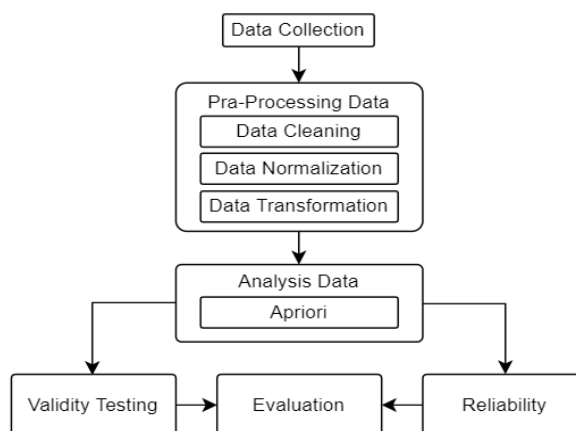


Figure 1. Research flow

In figure 1. describes the research flow that begins with "Data Collection". After the data is collected, the next process is "Data Pre-Processing", which includes three sub-stages, namely "Data Cleaning", "Data Normalization", and "Data Transformation". This process aims to prepare the data so that it is ready for analysis. Next, the data that has been processed goes to the "Data Analysis" stage using the "Apriori" method, which is an analysis technique for finding relationships between items in a large dataset. After analysis, the next stage is "Evaluation", where the results of the analysis are evaluated based on "Validity Testing" to ensure the data represents what was researched and "Reliability" to ensure the consistency of the results (Xu et al., 2021). This flow describes the method process in research data to ensure the quality and accuracy of research results.

Data collection

The data used in this research is a collection of student scores from various courses at STMIK YMI Tegal. Data collection was carried out by accessing the institution's academic database to obtain student grades for the last few semesters. The data collected includes student identity, course name, and grades obtained. To ensure confidentiality and privacy, all data used was anonymized prior to analysis.

Data Pre-processing

Before analysis, data pre-processing is needed to ensure the quality of the data to be analyzed. This step includes cleaning the data from missing or incomplete values, normalizing the value scores to ensure consistency, and transforming the data into a format suitable for analysis using the Apriori algorithm. This process is important to increase the accuracy of analysis results.

Research procedure

This research was carried out through several stages. First, data pre-processing was carried out as previously explained. Once the data is ready, the Apriori algorithm is applied to find relationships between subjects. This algorithm looks for combinations of courses that often appear together in a collection of high student grades, indicating a positive correlation between these courses.

Data analysis

Data analysis was carried out using the Apriori algorithm to identify association rules between subjects. The results of this analysis are then analyzed further to determine significant relationship patterns. In addition, statistical analysis was also carried out to test the strength and significance of the relationships between subjects found. The basic formula in the Apriori Algorithm relies on three main concepts, namely support, confidence, and lift. Support measures how often an itemset appears in a transaction dataset. In measuring course grades, support shows the proportion of transactions (students) who get a certain combination of grades.

$$\text{Support}(x) = \frac{\text{Number of transactions that contain } X}{\text{Total transaction amount}} \quad (1)$$

Confidence measures how often item Y appears in transactions containing X. In this analysis, it shows how often course B grades appear when course A grades already exist.

$$\text{Confidence}(X \Rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)} \quad (2)$$

Lift measures how often X and Y occur together compared to what would be expected if they were statistically independent. A lift greater than 1 indicates that Y is likely to appear more often if X already exists than expected; a lift of less than 1 indicates otherwise.

$$\text{Lift}(X \Rightarrow Y) = \frac{\text{Confidence}(X \Rightarrow Y)}{\text{Support}(Y)} \quad (3)$$

Validity and Reliability Testing

To ensure the validity and reliability of the research results, several tests were carried out. Data validity testing is carried out to ensure that the Apriori algorithm is successful in identifying relationships that actually exist between courses. The reliability of the analysis was tested by repeating the research process on different data samples to ensure consistency of results.

Evaluation

The evaluation was carried out to assess the effectiveness and usefulness of research findings in developing curriculum development at STMIK YMI Tegal. This involves discussions with relevant stakeholders, including lecturers and study program managers, to understand the practical implications of the relationships between courses found and how this information can be used to improve the curriculum.

3. RESULTS AND DISCUSSIONS

This research succeeded in applying the Apriori algorithm to find relationships between courses based on student grades at STMIK YMI Tegal. From data processing involving dozens of student grade entries from the third semester, the Apriori algorithm identified several combinations of courses that frequently appeared together with high grades, indicating a positive correlation between these courses (Li et al., 2022). This indicates that student success in one course tends to be linked to success in other courses identified in the association rules.

Table 1. Student grade data

NIM	Web Programming 1	Data Analysis and Visualization	Indonesian	Object Oriented Programming 1
22166001	A	AB	A	AB
22166002	A	AB	A	AB
22166003	A	AB	A	AB
22166006	AB	AB	A	AB
22166007	A	AB	A	A
22166008	A	AB	A	AB
22166009	A	AB	A	AB
22166010	A	AB	A	A
⋮	⋮	⋮	⋮	⋮
22205070	EF	E	A	DE
22205073	B	B	BC	AB

Table 2. Student grade data

NIM	Database 2	Computer Architecture and Organization	Linear Algebra
22166001	B	A	B
22166002	A	A	A
22166003	B	A	B

NIM	Database 2	Computer Architecture and Organization	Linear Algebra
22166006	B	A	B
22166007	B	A	AB
22166008	A	A	AB
22166009	AB	A	B
22166010	A	A	A
⋮	⋮	⋮	⋮
22205070	D	D	AB
22205073	B	A	DE

Further analysis revealed that several basic courses have a significant role in forming a strong foundation of knowledge for advanced courses, confirming the importance of placement and selection of courses in the curriculum. For example, basic programming and algorithm courses have been shown to have a strong relationship with success in advanced software development courses. This suggests that a strong understanding of basic concepts facilitates better understanding of more complex courses.

Validity and reliability testing shows that the analysis results have a high level of confidence, with the Apriori algorithm successfully identifying consistent relationships between subjects in various data samples. This confirms the effectiveness of the Apriori algorithm in analyzing educational datasets to reveal valuable relationship patterns.

Table 3. Analysis results

	antecedents	consequents	support	confidence	lift
0	(B)	(A)	0.6	0.6	1.0
1	(A)	(B)	0.6	1.0	1.0
2	(C)	(B)	0.4	1.0	1.0

Based on the calculations in table 3, it was obtained from data processing carried out by researchers using Python tools. covers several important metrics, including support, confidence, and lift, for various combinations of subjects referred to as antecedents (causes) and consequents (effects). First, support describes how often a combination of two particular subjects appears in the dataset. For example, in the first row, the combination of subjects B and A has a support of 0.6, which means that this combination appears in the data in 60% of the total transactions. Second, confidence measures how often the combination of two subjects appears together relative to the occurrence of consequent subjects. For example, in the second row, the confidence for the combination of subjects A and B is 1.0, which means that every time subject A appears, subject B also appears, with a confidence level of 100%. Third, lift is a measure of how strong the relationship between two subjects is compared to their occurrence independently. A lift value equal to 1 indicates that there is no relationship between antecedents and consequents. A lift value greater than 1 indicates that the combination has a stronger relationship than would be expected at random. In this table, the lift value for all combinations is 1.0, which indicates that there is no significant increase or decrease in the probability of the combination occurring.

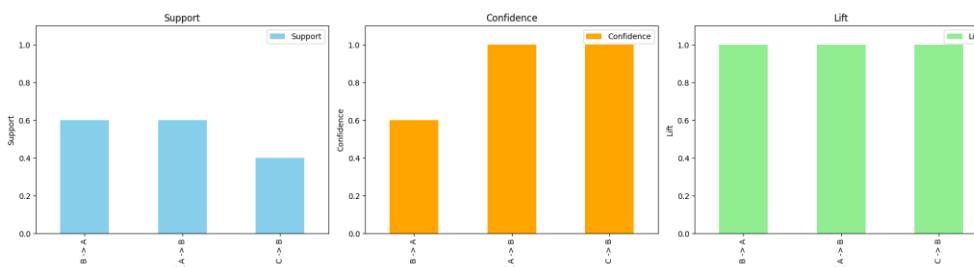


Figure 2. Diagram of analysis results

A priori analysis that has been carried out on the student grades dataset reveals some interesting relationships between different course grades, using support, confidence and lift metrics to assess the strength and relevance of these relationships. The results of the analysis show three important association rules, each of which provides insight into the patterns that emerge in student grade data.

In the first rule, the antecedent is the value 'B', and the consequent is the value 'A', with support of 60% and confidence of 60%. This shows that the value pair 'B' and 'A' is quite common, occurring in 60% of all transactions. However, a lift value of 1.0 indicates that this relationship is neutral, meaning that the presence of a 'B' value does not significantly increase the likelihood of the presence of an 'A' value, and this correlation may be nothing more than chance.

The second rule highlights the relationship between the values 'A' and 'B', where the value 'A' as an antecedent is strongly correlated with the presence of the value 'B' as a consequent, proven with 100% confidence. However, the lift also having a value of 1.0 confirms again that this relationship is not special or unique, indicating that although the correlation looks strong, it does not indicate the existence of a statistically significant dependency between the two values.

The third rule shows that the value 'C' (antecedent) is related to the value 'B' (consequent), with support of 40% and confidence of 100%. Although the combination of 'C' and 'B' occurs less frequently than the previous two rules, the presence of the value 'C' always includes the value 'B'. Once again, a lift of 1.0 suggests that this relationship, while consistent, may be nothing more than a chance pattern that does not indicate a meaningful association.

From the results of this research, several strategic recommendations for curriculum development at STMIK YMI Tegal can be identified. Firstly, there needs to be more emphasis on basic subjects to ensure students have a strong foundation. Second, adjustments to the curriculum structure can be made to integrate courses identified as having a positive correlation, thus facilitating more coherent and systematic learning.

In conclusion, this research provides valuable insight into the relationships between courses at STMIK YMI Tegal, offering a strong basis for curriculum improvements oriented towards improving student learning outcomes and learning effectiveness.

4. CONCLUSION

Research conducted on the application of the Apriori algorithm to analyze relationships between courses based on student grades at STMIK YMI Tegal has produced significant findings that show certain patterns of relationships between various courses. The analysis successfully identified association rules that provide insight into possible dependencies between student success in different courses. Although the lift values suggest that these relationships may be coincidental, these results nonetheless provide useful initial insight into further understanding of the structure and effectiveness of existing curricula.

Based on the results of this study, suggestions for future research are to investigate more deeply the relationships that have been identified, perhaps incorporating a qualitative approach to understand the values and factors that influence these relationships. Additionally, future research could examine the implementation of curriculum changes based on the results of the analysis to see the impact on student achievement. Finally, course adjustments and integration based on the results of the Apriori analysis can be considered to increase cohesion and synergy in the curriculum, with the ultimate goal of increasing overall student success.

This research provides an important contribution in understanding the relationship between courses based on student grades using the Apriori algorithm at STMIK YMI Tegal. However, there are several limitations that need to be considered. First, this study only considered data from one specific educational institution, so the generalizability of the findings to other contexts may be limited. In addition, the analysis is based only on student grades without considering other factors that may influence academic results, such as learning styles or personal factors. For future research, it is recommended to involve more educational institutions to expand the scope and validity of the findings. Additionally, integrating additional factors, such as learning styles or psychosocial factors, may yield a more comprehensive understanding of the relationship between coursework and student achievement.

REFERENCES

- Abelha, M., Fernandes, S., Mesquita, D., Seabra, F., & Ferreira-Oliveira, A. T. (2020). Graduate employability and competence development in higher education—A systematic literature review using PRISMA. *Sustainability*, 12(15), 5900.
- Al Husaeni, D. F., Al Husaeni, D. N., Nandiyanto, A. B. D., Rokhman, M., Chalim, S., Chano, J., Al Obaidi, A. S. M., & Roestamy, M. (2024). How technology can change educational research? definition, factors for improving quality of education and computational bibliometric analysis. *ASEAN Journal of Science*

- and *Engineering*, 4(2), 127–166.
- Birdman, J., Wiek, A., & Lang, D. J. (2022). Developing key competencies in sustainability through project-based learning in graduate sustainability programs. *International Journal of Sustainability in Higher Education*, 23(5), 1139–1157.
- Chang, C., Hwang, G., & Gau, M. (2022). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology*, 53(1), 171–188.
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, 8, 75264–75278.
- Culver, K. C., & Bowman, N. A. (2020). Is what glitters really gold? A quasi-experimental study of first-year seminars and college student success. *Research in Higher Education*, 61(2), 167–196.
- De Villiers, R. (2021). Seven principles to ensure future-ready accounting graduates—a model for future research and practice. *Meditari Accountancy Research*, 29(6), 1354–1380.
- Gopalan, M., Rosinger, K., & Ahn, J. Bin. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, 44(1), 218–243.
- Huang, X. (n.d.). Applying Statistical Analysis Methods to Study the Key Factors of Innovative Curriculum Systems in Teacher Education. *Applied Mathematics and Nonlinear Sciences*, 9(1).
- Jia, C., Hew, K. F., Bai, S., & Huang, W. (2022). Adaptation of a conventional flipped course to an online flipped format during the Covid-19 pandemic: Student learning performance and engagement. *Journal of Research on Technology in Education*, 54(2), 281–301.
- Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies*, 27(5), 6069–6104.
- Lee, K. J., Tilling, K. M., Cornish, R. P., Little, R. J. A., Bell, M. L., Goetghebeur, E., Hogan, J. W., & Carpenter, J. R. (2021). Framework for the treatment and reporting of missing data in observational studies: the treatment and reporting of missing data in observational studies framework. *Journal of Clinical Epidemiology*, 134, 79–88.
- Lee, S. (2022). Factors affecting the quality of online learning in a task-based college course. *Foreign Language Annals*, 55(1), 116–134.
- Li, S., You, M., Li, D., & Liu, J. (2022). Identifying coal mine safety production risk factors by employing text mining and Bayesian network techniques. *Process Safety and Environmental Protection*, 162, 1067–1081.
- Mai, T. T., Bezbradica, M., & Crane, M. (2022). Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data. *Future Generation Computer Systems*, 127, 42–55.
- Mohamed Hashim, M. A., Tlemsani, I., & Matthews, R. (2022). Higher education strategy in digital transformation. *Education and Information Technologies*, 27(3), 3171–3195.
- Núñez-Canal, M., de Obesso, M. de las M., & Pérez-Rivero, C. A. (2022). New challenges in higher education: A study of the digital competence of educators in Covid times. *Technological Forecasting and Social Change*, 174, 121270.
- Qi, C., Liang, H., Zuo, S., & Li, R. (2024). Comparing competency-oriented student activities between expert and novice teachers in China: Insights from an epistemic network analysis (ENA). *Education and Information Technologies*, 1–28.
- Sun, L. (2020). An improved apriori algorithm based on support weight matrix for data mining in transaction database. *Journal of Ambient Intelligence and Humanized Computing*, 11(2), 495–501.
- Telikani, A., Gandomi, A. H., & Shahbahrami, A. (2020). A survey of evolutionary computation for association rule mining. *Information Sciences*, 524, 318–352.
- Xu, Y., Yan, C., Shi, J., Lu, Z., Niu, X., Jiang, Y., & Zhu, F. (2021). An anomaly detection and dynamic energy performance evaluation method for HVAC systems based on data mining. *Sustainable Energy Technologies and Assessments*, 44, 101092.
- Yan, Y. (2024). Analysis and Research of Psychological Crisis Behavior Model Based on Improved Apriori Algorithm. *International Journal of Human-Computer Interaction*, 1–13.
- Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2023). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education*, 33(2), 290–324.