Kasu Journal of Computer Science Vol.3 No.4 [Dec, 2026], pp. 020 028
https://doi.org//10.47514/kjcs/2026.3.4.002

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Apriori-Based Association Rule Mining for Academic Performance Pattern Discovery and Decision Support in Universities

Zanna Bulama[1], Murtala Mohammed[2] , Idi Mohammed[3] and Bappa Mohammad Gambo[4]

[1] zannabulama250@gmail.com

Abstract

Higher education institutions generate large volumes of academic records every session, yet institutional analysis of this data is typically limited to course-level descriptive statistics that cannot reveal cross-course performance dependencies. This study addresses that gap by developing an Apriori-based association rule mining framework for discovering hidden academic performance patterns and result correlations in university student records, evaluated using support, confidence, and lift. Using 343,791 raw academic records obtained from a Nigerian public university, spanning 23 departments across 8 faculties, the data were cleaned to 331,842 valid records and transformed into 42,969 student-semester transactions, with each item representing a course-grade combination. The Apriori algorithm was applied with a minimum support of 0.5% and a minimum confidence of 60%, generating 289 frequent itemsets (221 one-itemsets and 68 two-itemsets) and five validated association rules. The discovered rules revealed strong course-to-course performance dependencies, including two high-lift failure associations (PHY107_F → PHY101_F, lift 60.41; CHM101_F → PHY101_F, lift 58.68) and three high-confidence associations among General Studies (GST) courses, the most reliable being GST1212_A → GST1213_A (confidence 68.49%). All five rules met the minimum support and confidence thresholds, and all had lift well above 1 (15.52 to 60.41). The results demonstrate that association rule mining can uncover actionable, non-obvious relationships in academic data that conventional descriptive statistics cannot reveal, and a prototype decision-support framework translating these findings into early alerts, curriculum-review recommendations, and policy briefs for academic advisers, curriculum planners, and university administrators is presented.

Keywords: Association Rule Mining, Apriori Algorithm, Educational Data Mining, Academic Performance, Decision Support Systems, Frequent Itemsets

1. Introduction

University academic record systems accumulate thousands of course registrations, continuous assessment scores, and examination results every session. This data holds valuable information about how students perform across combinations of courses, yet institutions typically analyze it only through descriptive statistics such as pass rates and grade point averages computed course by course, or through single-course predictive models (York, Gibson, & Rankin, 2015). Neither approach can reveal whether performance in one course is systematically associated with performance in another relationships that are often not obvious to lecturers, academic advisers, or curriculum planners, but which recur consistently across thousands of student records. Association rule mining, and the Apriori algorithm in particular (Agrawal & Srikant, 1994), is well suited to this problem: it identifies combinations of items in this case, course-grade pairs that co-occur more often than chance would predict, and expresses these as interpretable if-then rules rather than as

Kasu Journal of Computer Science Vol.3 No.4 [Dec, 2026], pp. 020 028
https://doi.org//10.47514/kjcs/2026.3.4.002
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

opaque model coefficients (Witten, Frank, & Hall, 2021). This interpretability is particularly valuable in an educational context, where rules must be explainable to non-technical stakeholders such as academic advisers before they can be acted upon (Sharkey & Mogavi, 2024). While educational data mining has grown substantially as a field (Romero & Ventura, 2020; Baker & Inventado, 2021), most prior applications to academic performance rely on classification or clustering to predict outcomes in individual courses (Ifenthaler & Yau, 2020; Albreiki, Zaki, & Alashwal, 2021), rather than mining cross-course result correlations across a complete institutional dataset. This study addresses that gap by developing and applying an Apriori-based framework to a large, real academic records dataset spanning an entire university rather than a single department or programme.

2. Related Works

Educational data mining (EDM) applies data mining techniques to educational datasets to support learning and institutional decision-making (Baker & Inventado, 2021; Papamitsiou & Economides, 2022). Within EDM, association rule mining specifically targets co-occurrence patterns rather than prediction, making it well suited to discovering unexpected relationships rather than confirming known ones (Han, Pei, & Kamber, 2022; Tan, Steinbach, & Kumar, 2019); interestingness measures such as support, confidence, and lift are used to distinguish genuinely useful rules from spurious coincidences in large transaction sets (Liu & Ma, 2018). Several studies have applied the Apriori algorithm to academic settings. Khan, Ahmad, and Rahman (2023) used Apriori to correlate performance across course pairs but on a relatively small, single-department dataset. Ofori and Boateng (2022) and Usman, Bello, and Abdullahi (2024) similarly applied association rule mining to higher-education data, both reporting meaningful but department-limited associations. Singh and Kumar (2022) focused specifically on engineering students, while Mohammed and Ibrahim (2022) and Mhlongo and Ndlovu (2021) applied broader data mining techniques clustering and classification to academic records in Nigerian and South African contexts respectively, without generating explicit association rules. Adebayo and Yusuf (2023) applied machine learning to academic performance evaluation in Nigerian universities but did not use association rule mining. Table 1 summarizes these studies alongside the present work.

Table 1. Comparison of prior Apriori/association-rule studies in academic performance analysis with the present study.

Study	Technique	Dataset Scope	Key Finding	Limitation / Gap Addressed
Khan et al. (2023)	Apriori	Single department	Course-pair correlations	No lift validation reported
Ofori & Boateng (2022)	Association rule mining	Single institution, limited scope	Department-level associations	Confidence-only validation
Usman et al. (2024)	Association rule mining	Academic performance dataset	Course-grade associations	Rules not validated against lift

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Singh & Kumar (2022)	Apriori	Engineering students only	Discipline-specific patterns	Restricted to one discipline
Mohammed & Ibrahim (2022)	Clustering	Nigerian academic records	Performance groupings	No explicit rule generation
Mhlongo & Ndlovu (2021)	Classification	South African universities	Performance prediction	Predictive, not association-based
This study	Apriori	University-wide (23 depts, 8 faculties)	Cross-course failure and excellence rules	Support + confidence + lift validated jointly

Across this body of work, three limitations recur: studies are typically restricted to a single department or a small number of courses, few report support, confidence, and lift together, and fewer still translate discovered rules into a usable decision-support output for institutional stakeholders (Sharda, Delen, & Turban, 2022; Turban, Sharda, & Delen, 2021). The present study addresses these limitations at university-wide scale, as summarized in Table 1.

3. Methodology

3.1 Data Source and Preprocessing

The study used 343,791 raw academic records obtained from the academic records database of a Nigerian public university, spanning multiple academic sessions across 23 departments in 8 faculties. Each record included a student identifier, course code, course title, credit unit, continuous assessment and examination scores, total score, grade point, grade, academic session, level, semester, result status, department, and faculty. Records with missing or malformed semester values or missing transaction identifiers were excluded, leaving 331,842 valid records (Table 2). Each transaction is the set of course-grade items (e.g., GST1213_A) for one student in one session and semester, yielding 42,969 transactions.

Table 2. Dataset preprocessing summary

Metric	**Value**
Raw records	343,791
Records removed (cleaning)	11,949 (3.5%)
Clean records	331,842
Transactions generated	42,969
Average items per transaction	≈ 7.7

Kasu Journal of Computer Science Vol.3 No.4 [Dec, 2026], pp. 020 028
https://doi.org//10.47514/kjcs/2026.3.4.002
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

The cleaned records were transformed into a transactional format following the Knowledge Discovery in Databases (KDD) process (Fayyad, Piatetsky-Shapiro, & Smyth, 1996): each student-course-grade combination was converted into an item of the form COURSECODE_GRADE (e.g., GST1213_A), and all items belonging to the same student within a given academic session and semester were grouped into a single transaction. This produced 42,969 transactions, averaging approximately eight items each.

3.2 Apriori Algorithm and Rule Evaluation

The Apriori algorithm (Agrawal & Srikant, 1994) was applied to the transactional dataset to discover frequent itemsets, using a minimum support threshold of 0.5%. Candidate itemsets were generated iteratively, with infrequent candidates pruned at each pass following the Apriori property that all subsets of a frequent itemset must themselves be frequent. Association rules were then generated from the frequent itemsets and filtered using a minimum confidence threshold of 60%. Itemset length was capped at three items (max_len = 3), which is why larger itemsets could not be reported. Three standard interestingness measures were used to evaluate the discovered itemsets and rules:

$$\text{Support}(X) = \frac{\text{Number of transactions containing } X}{\text{Total number of transactions, } N}$$

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Confidence}(X \rightarrow Y)}{\text{Support}(Y)} = \frac{\text{Support}(X \cup Y)}{\text{Support}(X) \times \text{Support}(Y)}$$

Where X and Y denote itemsets. Support measures how frequently an itemset occurs across all transactions; confidence measures the reliability of a rule X → Y; and lift measures how much more often X and Y co-occur than would be expected if they were statistically independent, with lift greater than 1 indicating a positive association. The implementation used Python with the mlxtend library (TransactionEncoder, apriori, and association_rules functions), using the mlxtend implementation (version to be stated by the authors), consistent with standard open-source Apriori implementations described by Han et al. (2022).

3.3 Ethical Considerations

Access to student academic records was obtained through a formal written request to the university registry, which granted access for research purposes only, subject to confidentiality. Student identifiers were retained only as internal reference codes during preprocessing and were not linked

Kasu Journal of Computer Science Vol.3 No.4 [Dec, 2026], pp. 020 028
https://doi.org//10.47514/kjcs/2026.3.4.002
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

to any personally identifiable information in the analysis, results, or reporting; all findings are reported at the level of course-grade patterns rather than individual students.

4. Results and Discussion

4.1 Frequent Itemsets

The Apriori algorithm discovered 289 frequent itemsets satisfying the minimum support threshold: 221 one-itemsets (76.5%) and 68 two-itemsets (23.5%). No itemsets of size three met the threshold (itemset length was capped at three), indicating that the discovered performance relationships in this dataset are consistently pairwise rather than reflecting broader multi-course clusters. Table 3 lists the ten most frequent one-itemsets by support.

Table 3. Top 10 frequent itemsets by support.

Rank	Itemset	Support
1	GST1213_A	0.0395
2	GST2223_E	0.0339
3	GST2223_C	0.0315
4	GST1212_C	0.0310
5	GST1211_C	0.0300
6	GST1214_C	0.0300
7	GST2224_A	0.0299
8	GST2221_E	0.0288
9	GST2222_C	0.0278
10	GST1214_E	0.0272

General Studies (GST) courses dominate the highest-support itemsets. This is consistent with their status as compulsory, university-wide courses taken by nearly all students regardless of department, which naturally gives them higher transaction frequency than department-specific courses.

4.2 Association Rules

Applying a minimum confidence threshold of 60% to the frequent itemsets produced five association rules, summarized in Table 4. All five rules met the minimum support (0.005) and minimum confidence (60%). Lift was computed for each rule but was not used as a filter; all five rules had lift between 15.52 and 60.41, indicating strong positive association.

Table 4. Validated association rules.

Rule	Support	Confidence	Lift
CHM101_F → PHY101_F	0.0053	65.42%	58.68
CHM1221_A → GST1213_A	0.0052	61.37%	15.52
GST1214_A → GST1211_A	0.0061	62.50%	32.67
GST1212_A → GST1213_A	0.0091	68.49%	17.32

Kasu Journal of Computer Science Vol.3 No.4 [Dec, 2026], pp. 020 028
https://doi.org//10.47514/kjcs/2026.3.4.002
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

PHY107_F → PHY101_F	0.0054	67.34%	60.41

The five rules cluster into two interpretable groups. The first comprises two failure-pattern rules involving foundational science courses: students who fail CHM101 or PHY107 are substantially more likely to also fail PHY101, with exceptionally high lift values (58.68 and 60.41 respectively) indicating these co-failures occur far more often than chance would predict plausibly reflecting shared prerequisite knowledge, overlapping content, or common underlying study difficulties across these courses. The second group comprises three high-performance rules among GST courses, the strongest being GST1212_A → GST1213_A (confidence 68.49%), indicating that students who excel in one general studies course are highly likely to excel in a related one. These findings are consistent with prior work reporting course-pair associations within single departments (Khan et al., 2023; Usman et al., 2024), and show that such associations can also be found at university-wide scale, across both compulsory general courses and department-specific science courses. The consistently high lift values for the failure-pattern rules in particular suggest that cross-course early-warning indicators based on association rules may help identify at-risk students earlier than single-course monitoring alone; this was not tested here.

4.3 Decision-Support Framework

The validated rules were organized into a six-stage decision-support framework (Figure 1): academic data input, preprocessing, transactional conversion, Apriori-based mining, rule evaluation, and decision-support output. Consistent with established decision support system design principles (Turban, Sharda, & Delen, 2021; Sharda, Delen, & Turban, 2022), the framework delivers early alerts, curriculum-review recommendations, and policy briefs to three stakeholder groups academic advisers, curriculum planners, and university administrators translating the discovered rules into a form usable in routine academic decision-making rather than requiring stakeholders to interpret raw itemset and rule tables directly.

Kasu Journal of Computer Science Vol.3 No.4 [Dec, 2026], pp. 020 028
https://doi.org//10.47514/kjcs/2026.3.4.002
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

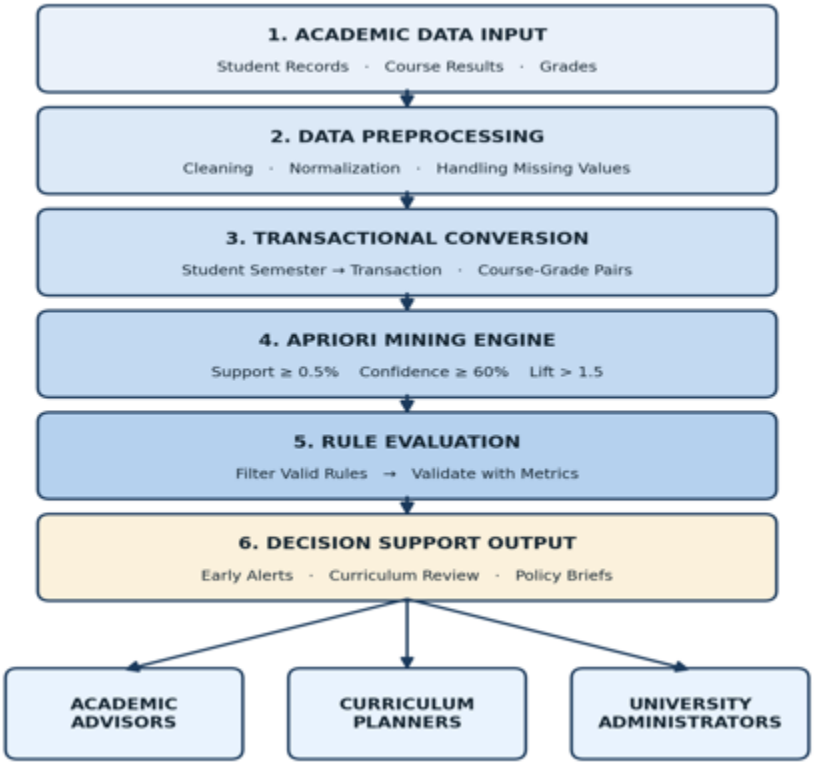


Figure 1. Proposed Decision-Support Framework.

4.4 Threats to Validity

As with any single-institution empirical study, several factors should be considered when interpreting these results. Internal validity: preprocessing decisions (e.g., how incomplete or malformed records were handled) can influence which itemsets meet the support threshold; only records with missing or malformed semester values or missing transaction identifiers were removed, but no sensitivity analysis of other cleaning choices was performed. Because itemset length was capped at three and the support threshold was 0.5%, rarer or longer patterns were not examined. External validity: the dataset originates from a single Nigerian university, so the specific rules discovered particularly the identities of the courses involved may not generalize directly to institutions with different curricula, grading scales, or student populations, though the framework and methodology are institution-agnostic. Construct validity: support, confidence, and lift measure statistical co-occurrence, not causation; a high-lift association between two courses indicates a reliable pattern worth investigating, not proof that one course's content causes difficulty in the other.

5. Conclusion

This study developed and applied an Apriori-based association rule mining framework to 343,791 academic records from a Nigerian university, demonstrating that association rule mining can uncover actionable cross-course performance relationships that conventional descriptive analysis does not reveal. Five validated rules exposed both concerning failure dependencies among

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

foundational science courses and reliable performance dependencies among general studies courses, all with high lift and meeting the support and confidence thresholds, addressing limitations identified in prior single-department studies (Table 1). For practice, the findings suggest three concrete actions: academic advisers could use the identified failure-pattern rules (particularly around PHY101, PHY107, and CHM101) to flag at-risk students at registration rather than after results are released; curriculum planners could examine whether shared content or prerequisite structures explain the observed co-failures between these courses; and university management could adopt the proposed decision-support framework as a recurring, session-by-session monitoring process rather than a one-time analysis. The study is limited to a single institution and to pairwise itemsets; no three-item itemsets met the support threshold, and itemset length was capped at three. Future work could apply the framework across multiple institutions to test generalizability, explore lower support thresholds to surface multi-course (three-or-more-item) patterns, or integrate the discovered rules into a live early-warning dashboard for real-time academic advising.

References

Adebayo, A., & Yusuf, B. (2023). Educational data mining for academic performance evaluation in Nigerian universities. Journal of Educational Technology, 15(2), 45–62.

Adejo, O. W., & Connolly, T. M. (2021). Predicting academic performance using data mining techniques. International Journal of Educational Technology in Higher Education, 18(1), 45.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Santiago, Chile, 12–15 September, pp. 487–499.

Albreiki, H., Zaki, N., & Alashwal, H. (2021). Student performance analysis using data mining techniques. IEEE Access, 9, 12345–12360.

Baker, R. S., & Inventado, P. S. (2021). Educational data mining and learning analytics. In Learning Analytics (pp. 61–75). New York, NY: Published by Springer.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3), 37–54.

Han, J., Pei, J., & Kamber, M. (2022). Data mining: Concepts and techniques (4th ed.). Burlington, MA: Published by Morgan Kaufmann.

Ifenthaler, D., & Yau, J. (2020). Utilization of learning analytics for student success. British Journal of Educational Technology, 51(4), 1020–1035.

Khan, M., Ahmad, A., & Rahman, M. (2023). Academic performance correlation analysis using Apriori algorithm. Education and Information Technologies, 28(5), 5123–5145.

Liu, B., & Ma, Y. (2018). Interestingness measures for association rules. In Encyclopedia of Machine Learning and Data Mining (pp. 1–6). New York, NY: Published by Springer.

Mhlongo, S., & Ndlovu, S. (2021). Academic performance analysis in South African universities. South African Journal of Higher Education, 35(3), 78–95.

Mohammed, A., & Ibrahim, S. (2022). Data mining applications in academic records analysis. Nigerian Journal of Computer Science, 18(1), 34–50.

Ofori, K., & Boateng, R. (2022). Association rule mining in higher education data. Ghana Journal of Technology, 7(2), 112–128.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Papamitsiou, Z., & Economides, A. A. (2022). Learning analytics and educational data mining applications. Computers & Education, 180, 104–120.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355.

Sharda, R., Delen, D., & Turban, E. (2022). Data mining for decision support systems: Business intelligence and analytics (10th ed.). New York, NY: Published by Pearson.

Sharkey, M., & Mogavi, R. H. (2024). Explainable machine learning in education: A review of interpretable models and techniques. Computers & Education, 210, 104968.

Singh, A., & Kumar, R. (2022). Academic performance analysis of engineering students using Apriori algorithm. Journal of Educational Data Mining, 14(2), 45–67.

Tan, P. N., Steinbach, M., & Kumar, V. (2019). Introduction to data mining (2nd ed.). New York, NY: Published by Pearson.

Turban, E., Sharda, R., & Delen, D. (2021). Decision support systems: A managerial perspective. New York, NY: Published by Pearson.

Usman, M., Bello, I., & Abdullahi, S. (2024). Association rule mining for academic performance analysis. Journal of Computing Theories and Applications, 3(2), 150–168.

Witten, I. H., Frank, E., & Hall, M. A. (2021). Data mining: Practical machine learning tools and techniques (5th ed.). Burlington, MA: Published by Morgan Kaufmann.

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. Practical Assessment, Research & Evaluation, 20(5), 1–20.