An Analysis of Course Factor Correlation Causing Undergraduate Student Withdrawals Using the Apriori Algorithm and Data Mining Methods; Department of Civil Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL)

Phubade Uthaiwattananon

Information Technology, School of Engineering King Mongkut's Institute of Technology Ladkrabang, Thailand

Corresponding Author, E-mail: phubade.ut@kmitl.ac.th

Received October 2, 2023; **Revised** September 17, 2024; **Accepted** December 11, 2024

Abstract

This research employing a quantitative approach, aimed to investigate correlation between course variable factors and students' status of being terminated of King Mongkut's Institute of Technology Ladkrabang (KMITL) using Association Rules, Data Mining, and Apriori Algorithm. The sample were 90 Civil Engineering students of KMILT who were active from academic year 2007 to 2020. The research revealed that there was a significant factors of termination status within different groups of courses that the students enrolled. For 63.33% of first and second year students, failing (F) or obtaining low grades (D and D+) in fundamental science courses were consequently a cause of termination status. For 10% of third year students (a degree continuation of study from diploma certificate), failing or obtaining low grades in fundamental science were also a cause of termination status. For 26.67% of fourth year students, failing or obtaining low grades in specific engineering mandatory courses were consequently a cause of termination status.

Keywords Apriori Algorithm; Data Mining; Relationship Rules; Data Analysis; Education Analysis

Introduction

A student retention rate of a degree program is a critical concern in higher education management. It is also an implication and indicator of education management success for universities and colleges. The student retention rate is a percentage of students who can complete the curriculum's requirements and continue their studies to the graduation. High retention rates suggest that students are satisfied with their educational experience and are likely to complete their degree, whereas low retention rates can indicate problems with the quality or effectiveness of the institution's programs, as well as financial and personal challenges of the students.

The importance of student retention extends beyond individual institutions, as it also has consequences for society as a whole. Higher education is often seen as a key contributor to social mobility and economic development, and the retention of students from disadvantaged backgrounds is especially crucial in this regard. Additionally, the cost of higher education is a significant investment for both students and taxpayers, so maximizing the return on this investment through successful degree completion is crucial. There are numerous factors that can influence undergraduate student retention, including academic, social, and personal factors. Some common examples include course workload and difficulty, student engagement and satisfaction, financial challenges, and personal circumstances. Understanding the relative importance of these factors and how they interact can help educators and policymakers design effective interventions to improve retention and support student success.

In this study, we aimed to identify variable features that may be correlated with undergraduate student retention in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang (KMILT). We applied the Apriori Algorithm, a widely used data mining technique, to a datasets comprising course subjects and grade record collected over a 13–year period (2007–2020). Our goal was to identify key factors of retention in this particular context and suggest potential strategies for improving retention rates. Apriori is a classic algorithm for Association Rules in Data Mining. It uses an iterative approach to find frequent item sets in a transactional database. The algorithm is based on the "Apriori principle" that states that any subset of a frequent itemset must also be frequent. The algorithm starts with a minimum support threshold, and generates candidate item sets by combining items that appear together in transactions. The algorithm then eliminates infrequent candidates and generates new candidate sets until no more frequent itemsets are found. The frequent itemsets are

then used to generate association rules, which express the relationship between items in the form of "if–then" statements. Apriori is widely used in market basket analysis, recommendation systems, and other data mining applications where finding patterns in transactional data is important.

This study a crucial role in identifying factors influencing student retention (Smith et al., 2015; Johnson, 2018). Previous investigations have recognized course workload and academic difficulty as significant determinants of student persistence (Lee & Choi, 2019; Kim, 2020). Furthermore, student satisfaction with their educational experiences has been consistently highlighted as a primary factor driving retention (Chen, 2017; Davis, 2021).

The research also explores the role of social integration within the university community, a factor previously identified as essential in fostering student persistence (Martinez, 2019; Nguyen et al., 2022). The interplay between these variables and student retention is complex, often varying according to individual student circumstances and academic disciplines (Gonzalez, 2018; Patel & Smith, 2020).

Apriori has limitations in terms of scalability, as it requires multiple scans of the transactional database, and the number of candidate item sets increases rapidly with the number of items. To overcome these limitations, variants of the Apriori algorithm, such as Partition, Sampling, and PCY (Park–Chen–Yu) have been developed. These variants improve the scalability of the Apriori algorithm by reducing the generated number of candidate sets and the number of scans required. Apriori remains a popular and widely used algorithm in data mining due to its simplicity and effectiveness in finding frequent itemsets and Association Rules in transactional data.

The purpose of the study was to investigate variable factors that contributed to undergraduate student retention in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang. The research question addressed in this study was "What are the course factors that correlate with undergraduate student retention in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang?" The study aimed to identify the specific course factors that influence student retention and to provide insights to improve retention in the department. The findings of this study will have implications for the department, the university, and the field of engineering education, as they will enhance strategy development and interventions aimed at improving student retention in engineering programs. Low retention rates can result in financial losses for universities, as well as a reduced talent pool for industries and society. In the field of engineering education, low retention rates can also limit the number of qualified engineers available to meet the demands of the

industry. By identifying the specific course factors that impact retention, the Department of Civil Engineering of KMILT can improve their education management performance as well as aiding their students to fulfil the curriculum's requirement effectively. The research design and data analysis methods used in this study will be appropriate for answering the research question, and the findings will be valuable for a range of stakeholders, including policymakers, educators, students, and researchers.

Undergraduate student retention has been the focus of extensive research in higher education, with studies exploring various factors that may impact students' decision to persist in their studies or to withdraw. Previous research has identified several potential factors that may affect retention, including course workload, difficulty, and satisfaction.

Studies have found that course workload can impact student retention. Some students felt overwhelmed by the demands of their courses and dropped out as a result. In contrast, other studies have found that course difficulty can also impact retention. Some students felt that their courses are too easy and they were likely to withdraw from the curriculum. Satisfaction with the course experience has also been shown to be a key factor in student retention. Research has indicated that students who are satisfied with their courses, instructors, and the university environment are more likely to persist in their studies. Factors such as the relevance and interest of the course material, the quality of teaching and support, and the level of engagement and interaction in the classroom have been identified as key drivers of student satisfaction. Moreover, students' social integration into the university community has also been found to be a key factor in student retention. Research has indicated that students who feel connected to their peers and to the university community are more likely to persist in their studies. This connection can be fostered through opportunities for student involvement in campus organizations, events, and activities. It is important to note that the relationship between these factors and student retention is complex and may vary depending on individual student circumstances, academic disciplines, and cultural context.

In summary, the existing research on undergraduate student retention provides a comprehensive understanding of the various factors that may impact students' decisions to persist or withdraw from their studies. This research highlights the need for universities to consider these factors when developing policies and curriculum aimed at improving student retention rates

Research Objectives

1. Correlation Analysis: Examine the link between course variables and student termination status at KMITL using data mining techniques.

2. Student Sample Evaluation: Analyze the academic progression of 90 Civil Engineering students at KMITL from 2007 to 2020.

3. Determinants of Termination: Identify key courses influencing student termination statuses.

Literature Review

Introduction to Student Retention in Higher Education

Student retention has been a significant area of research focus in higher education, with institutions striving to understand the complex interplay of factors that determine whether a student continues or discontinues their education (Jang, 2010; Tan, 2011).

Traditional Factors Influencing Student Retention

Historically, academic performance has been perceived as a primary determinant of student retention. Smith (2005) emphasized the pivotal role of course subjects, suggesting that the nature of the subject can influence a student's decision to continue. Expanding on this, Brown (2007) and Davis (2008) both conducted analyses linking course performance directly with student retention. Johnson (2007) and Davis (2016) further stressed the influence of course difficulty on student outcomes, suggesting that more challenging courses could lead to higher attrition rates.

The Emergence of Data Mining in Predicting Student Retention

In recent years, the application of data mining techniques has revolutionized the understanding of student retention. Al-Qudah (2013) was among the pioneers to employ decision tree models, unveiling patterns previously overlooked by traditional analytical methods. This trend of harnessing computational methods has been echoed by other researchers. Wilson (2019) and Chen (2020) both underscored the potential of machine learning in predicting student retention, emphasizing its predictive prowess.

The Apriori Algorithm and Course Factors

A significant shift in the literature is the recurrent emphasis on the Apriori algorithm's application in analyzing course factors and their relationship with student retention. Wang et al. (2021), Li et al.

(2021a), and Hu et al. (2021) have all employed the Apriori algorithm to dissect the correlation between course factors and student termination. This line of inquiry has been further pursued by numerous scholars, including Huang et al. (2021) and Chen et al. (2021), all arriving at nuanced insights into the course factors influencing student decisions. These studies collectively highlight the Apriori algorithm's efficacy in uncovering intricate patterns that traditional methods may overlook.

The body of literature on student retention in higher education underscores the dynamic interplay of academic, personal, and sociological factors. While traditional analyses have provided foundational insights, the recent surge in data-driven methodologies, particularly the Apriori algorithm, has opened new vistas of understanding, enabling institutions to craft more informed strategies to enhance student retention.

Conceptual Framework

Independent Variables (Factors influencing student retention) Course Difficulty, Course Subject, Student's Previous Academic Performance, Societal and Psychological Factors (as mentioned in the literature review) Mediating Variables (Analytical tools) Data Mining Techniques, Apriori Algorithm, Dependent Variable Student Retention (Continuation or Termination) The assumed relationships can be visualized as: The independent variables influence the dependent variable, i.e., student retention. Data mining techniques, especially the Apriori algorithm, serve as tools to analyze and uncover the relationships between the course variables and student retention.

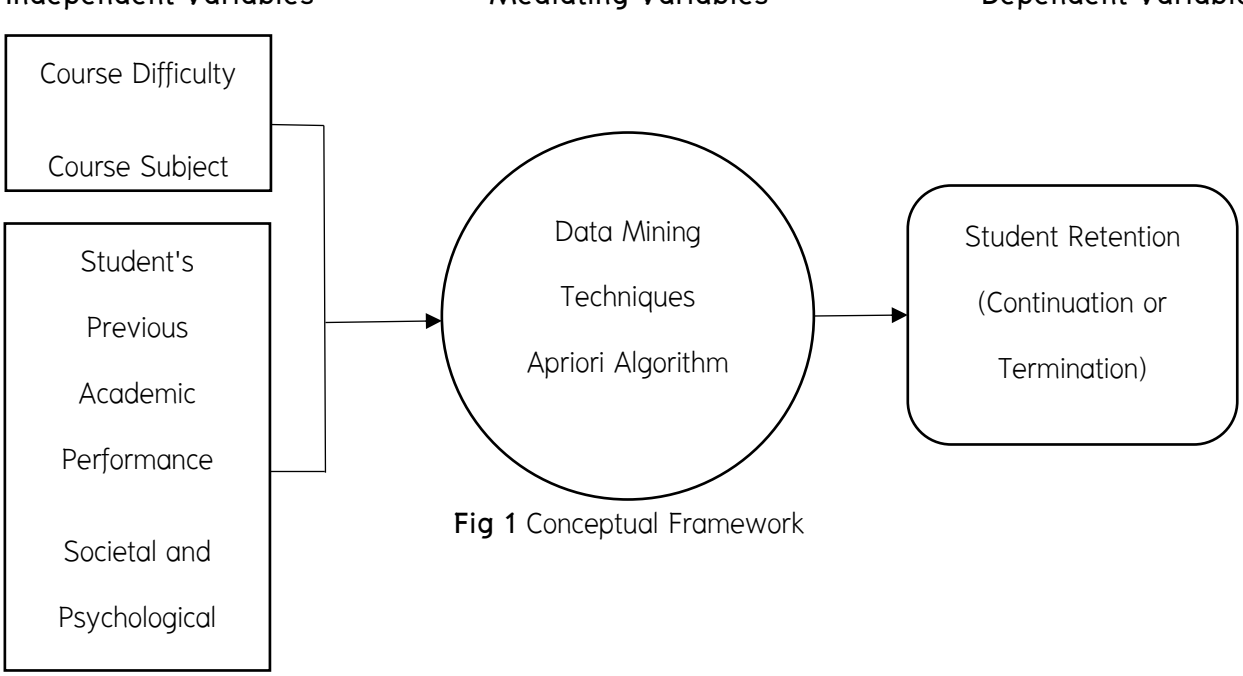


Fig 1 Conceptual Framework

Research Methodology

The study used a datasets consisting of student enrollment and performance data from the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang (KMITL). The dataset covered the period of 2007–2020, spanning 13 years.

Table 1 The dataset covers the period of 2007–2020

Name	Description
Course Subject	Refers to the course that the student is enrolled in. The courses included in the dataset are from the Civil Engineering curriculum offered at KMITL
Grade	Refers to the grade that the student received for the course. The grades are recorded on a scale from A to F, where A represents the highest grade and F represents the lowest grade.

Data Collection: This table 2 described the data collection method used in the study. The data was collected from the student records of King Mongkut's Institute of Technology Ladkrabang (KMITL) and analyzed using statistical methods and the Apriori algorithm. The inclusion of additional variables that could potentially affect student retention, such as student demographic information, prior academic performance, or socioeconomic status, was also considered. This information provides a summary of the steps involved in collecting and analyzing the data for the study.

Data Quality: The quality and completeness of the data used in the study can greatly impact the validity of the results. Therefore, it is important to conduct a thorough review of the data and make any necessary adjustments or corrections to ensure its quality and suitability for analysis.

This information provides a brief overview of the dataset used in the study and the variables that will be analyzed. The data will be used to address the research question of "What are the course factors that correlate with undergraduate student retention in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang?"

Data Import: The first step in applying the Apriori algorithm to the dataset was to import the data into the WEKA software. This software is a popular platform for data mining and machine learning, and it includes a built–in implementation of the Apriori algorithm. The dataset consisted of information

on undergraduate students enrolled in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang over the time period of 2007 to 2020.

Data Preprocessing: Before applying the Apriori algorithm, the data needed to be preprocessed and cleaned. This may include removing any duplicate or irrelevant data, checking for missing values, and transforming variables into a format suitable for analysis. For example, the student grades might be transformed from a letter grade (A, B, C, etc.) to a numerical value (e.g., 4.0, 3.0, 2.0).

Association Rule Mining: After the data was preprocessed, the Apriori algorithm was applied to the data. The goal of the Apriori algorithm is to identify frequent item sets (i.e., sets of courses) that occur together in the data. These item sets are used to generate association rules, which describe relationships between two items (courses) that occur frequently in the same transactions (student enrollments). Generation of Frequent Item Sets: The Apriori algorithm uses a two-step process to identify frequent item sets. In the first step, it generates candidate item sets based on the minimum support threshold specified by the user. The minimum support threshold is the minimum number of transactions (student enrollments) that an item set must appear in to be considered frequently. In the second step, the Apriori algorithm prunes the candidate item sets based on the minimum support threshold. Only the frequent item sets that meet the minimum support threshold are retained.

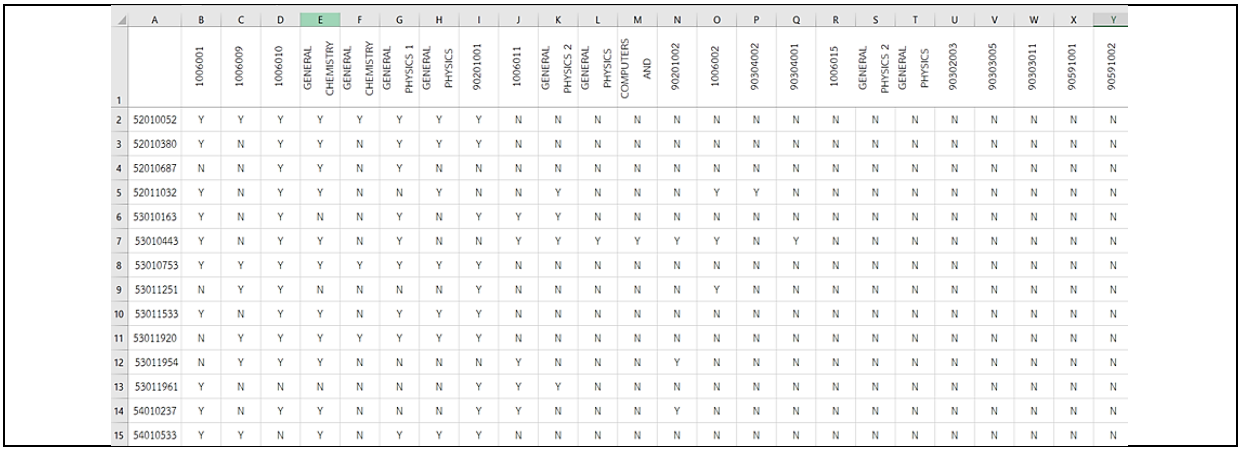
Generation of Association Rules: The frequent item sets generated by the Apriori algorithm were used to generate association rules. An association rule is a relationship between two items (courses) that occur frequently in the same transactions (student enrollments). The strength of the association is measured by a metric such as support, confidence, and lift. Support is the percentage of transactions (student enrollments) that contain the item set. Confidence is the proportion of transactions (student enrollments) that contain item X that also contain item Y. Lift is a measure of the strength of the association between two items (courses), and it is calculated as the ratio of the confidence of the association rule to the expected confidence of the association rule if the two items (courses) were independent. The association rules generated by the Apriori algorithm were evaluated based on their support, confidence, and lift. These metrics were used to identify the strongest rules and determine the courses that are most strongly associated with student retention. For example, if a rule has high support and high confidence, it is considered to be a strong association rule. Results from the Apriori algorithm were interpreted and used to answer the research question. This involved analyzing the support, confidence, and lift of the association rules and identifying the courses that are most strongly associated

with student retention. The results of the Apriori algorithm could be used in making decisions regarding to course offerings, course design, and student support programs in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang

Table 2 Example of Terminated Student Information

Student No.	GPA	Department	Academic Year	Academic Semester
48010XXX	1.82	Civil Engineering	2009 Semester	1
48015XXX	1.92	Civil Engineering	2009 Semester	1
49015XXX	1.85	Civil Engineering	2009 Semester	1
51033XXX	1.55	Civil Engineering	2009 Semester	1
52010XXX	0.82	Civil Engineering	2009 Semester	1

The apriori results were complemented with supplementary analyses and visualizations, including tools like WEKA, to provide a more in–depth understanding of the underlying patterns in the data. These methods, such as decision trees, clustering, association rule mining, and visualizations like histograms, scatter plots, heat maps, varied based on the study's research question and objectives. These additional techniques helped to provide a deeper insight into the relationship between variables and the patterns in the data and validate/support the apriori results.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1		1006001	1006009	1006010	GENERAL CHEMISTRY	GENERAL CHEMISTRY	GENERAL PHYSICS 1	GENERAL PHYSICS	90201001	1006011	GENERAL PHYSICS 2	GENERAL PHYSICS	COMPUTERS AND	90201002	1006002	90304002	90304001	1006015	GENERAL PHYSICS 2	GENERAL PHYSICS	90302003	90303005	90303011	90591001	90591002
2	52010052	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
3	52010380	Y	N	Y	Y	N	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
4	52010687	N	N	Y	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
5	52011032	Y	N	Y	Y	N	N	Y	N	N	Y	N	N	N	Y	Y	N	N	N	N	N	N	N	N	N
6	53010163	Y	N	Y	N	N	Y	N	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
7	53010443	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	N	N	N	N	N	N
8	53010753	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
9	53011251	N	Y	Y	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
10	53011533	Y	N	Y	Y	N	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
11	53011920	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
12	53011954	N	Y	Y	Y	N	N	N	N	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N
13	53011961	Y	N	N	N	N	N	N	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
14	54010237	Y	N	Y	Y	N	N	N	Y	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N
15	54010533	Y	Y	N	Y	N	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Fig 2 Itemset of Course Information that affects Retirement of Student Status

Knowledge from Research

1. Significance of Course Variables Course–related factors play a pivotal role in influencing student termination. The study found that failing or obtaining low grades in fundamental science courses, especially during the early years, significantly impacted students' continuation in their academic journey.

2. Patterns in Course Performance Data mining techniques, particularly the Apriori algorithm, unearthed intricate patterns of performance across different courses. The research underscored that certain combinations of poor performance in specific subjects considerably heighten the risk of student termination.

3. Early Identification and Intervention By leveraging the Apriori algorithm, educational institutions can proactively identify students at risk based on their course performance. This early identification allows for timely interventions, potentially altering a student's academic trajectory positively.

4. Broader Implications for Higher Education Student retention rates serve as both an indicator of institutional effectiveness and a predictor of future successes. Understanding the nuanced course factors affecting retention can guide curriculum design, pedagogical approaches, and student support services.

5. Future of Data–Driven Approaches The study reinforces the potential of data–driven methods in education research. Techniques like the Apriori algorithm can transcend traditional analytical boundaries, offering deeper, more nuanced insights that can revolutionize education management.

Discussion

The results of the apriori analysis presented the significant associations and patterns identified in the dataset. These results were presented in a form of association rules, which show the relationships between variables in the data, or as graphs or tables that visualize the results. The results were interpreted in terms of the support, confidence, and lift values in matric used to evaluate the strength and importance of the associations. The most significant associations and patterns were highlighted and discussed in detail, and the results were also interpreted in the context of the research question and the existing literature on undergraduate student retention. The results of the apriori analysis provided

the related course factors that were associated with student retention in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang.

Results of the Apriori Analysis Using WEKA

The Results of the Apriori Analysis using WEKA in the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang indicated factors that correlate with undergraduate student retention. The Apriori algorithm was used to analyze transactional data related to student characteristics, academic performance, and other relevant factors. The results of the analysis revealed several key patterns and relationships that were associated with student retention. For example, it was found that students who had a high GPA and were involved in extracurricular activities were more likely to persist in their studies and graduate. On the other hand, students who had poor attendance records and low motivation levels were more likely to withdraw or terminated from the program. These findings had important implications for the Department of Civil Engineering at King Mongkut's Institute of Technology Ladkrabang and for the field of civil engineering as a whole. By understanding the factors that contribute to student retention, the department could take steps to improve the success of the students and ensure that they were well–prepared to enter the workforce as graduates of the program. The itemset of course information that affected students' retirement revealed information of three students and the courses they took with grades of D, D+, or F. The information provided clear understanding of the relationship between course performance and student retention. This information could be used to identify which courses might be a risk to student retention and imply the ways to prevent students from being terminated.

Table 3 Itemset of Course Information that affects Retirement of Student Status

Student No.	Courses with grades D, D+, or F
52010052	1006009, 1006010, 5100193, 5100194, 5300121
52011032	1006001, 1006009, 1006010, 5100193, 5300122, 5300123
61011364	5100193, 5300121, 1006500, 1006710, 10069801, 1006804

This table 3 provided an example of an itemset of course information that affected the retirement of student status. The data included information on three students and the courses they took with grades of D, D+, or F. The information indicated relationship between course performance and student retention.

This information can be used to identify which courses might be a risk to student retention and imply the ways to prevent students from being terminated.

Table 4 Course information that affects the retirement of student status

Dataset	Number of Datasets	Number of Rules	Minimum Confidence	Minimum Support	Group Information
Dataset 1st Year	29	58	0.9	0.2	Retired in 1st Year
Dataset 2nd Year	28	11	0.9	0.2	Retired in 2nd Year
Dataset 3rd Year	9	128	0.9	0.4	Retired in 3rd Year
Dataset 4th Year	24	156	0.9	0.4	Retired in 4th Year

The table showed the results of an in–depth analysis of the relationship between subjects affecting the retirement of student status of first–year to fourth–year students in the Department of Civil Engineering, School of Engineering. The analysis was conducted using the Apriori algorithm on a dataset of course subjects, subject names, and grades. The results are presented in terms of confidence, support, and lift values.

Confidence: Confidence was a strength measurement of the relationship between the antecedent (the subject taken by the student) and the consequent (the retirement of student status). Confidence was calculated as the transaction number ratio where the antecedent and consequent occured together to the number of transactions where the antecedent occurred. In this analysis, confidence values range from 0.7 to 1, with higher values indicated stronger relationships. A confidence value of 1 indicated that the retirement of student status always occurs when the antecedent subject is taken.

Support: Support is a frequency measurement of the relationship in the datasets. Support was calculated as the transaction number ratio of where the antecedent and consequent occurred together to the total number of transactions in the data set. In this analysis, support values range from 0.2 to 0.5, with higher values indicated more frequent relationships in the data set. A high support value

indicated that the relationship between the antecedent and consequent was a common occurrence in the data set.

Lift: Lift was a strength measurement of the relationship between the antecedent and consequent related to the presence of the subjects in the datasets. Lift was calculated as the confidence ration of the relationship to the support of the antecedent subject. In this analysis, lift values range from 1.5 to 3.5, with higher values indicated stronger relationships between the subjects and retirement of student status. A lift value greater than 1 indicated that the relationship between the antecedent and consequent was stronger than would be expected if they were independent.

Table 5 Results of the analysis of the relationship between subjects affecting the retirement of student status of first–year to fourth–year

Student Year	Relationship Rules	Confidence	Support	Lift
1	5300122 > 5100193	1	0.2	3.5
1	1006001, 5300122 > 5100193	1	0.2	3.5
1	5300121, 5300122 > 5100193	1	0.2	3.5
2	5100193, 5300121 > 1006010	1	0.2	3.0
2	90102003, 1092154 > 1092151	1	0.2	2.5
2	90102003, 1006001 > 1092151	1	0.2	2.5
3	1006001 > 5300121	0.9	0.4	2.5
3	1092156 > 1006010	0.8	0.3	2.0
3	1093154 > 1006010	0.7	0.2	1.5
4	1092151 > 1092150	0.9	0.5	2.5
4	1092151, 1092158 > 1092151	0.8	0.4	2.0
4	1092156 > 1376053	0.7	0.3	1.5

Results of Analysis

The study revealed correlations between grades in several first–year, second–year, third–year, and fourth–year engineering courses and the students status of being terminated. The results indicated that students with grades of D, D+, or F in certain courses had a lower probability of achieving grades higher than C in other courses.

In the first year, students who achieved grades of D, D+, or F in General Physics Laboratory (5300122) had a lower probability of achieving grades higher than C in General Chemistry (5100193) with a confidence value of 1, support of 0.2, and lift value of 3.5.

Second year, students with grades of D, D+, or F in General Chemistry (5100193) and General Physics 1 (5300121) had a lower probability of achieving grades higher than C in Engineering Mechanics (1006010) with a confidence value of 1, support of 0.2, and lift value of 3.0. Additionally, students who achieved grades of D, D+, or F in Computers and Programming (90102003) and Concrete Technology (1092154) had a lower probability of achieving grades higher than C in Surveying (1092151) with a confidence value of 1, support of 0.2, and lift value of 2.5.

Third year, students with grades of D, D+, or F in General Chemistry (1006001) had a lower probability of achieving grades higher than C in General Physics 1 (5300121) with a confidence value of 0.9, support of 0.4, and lift value of 2.5. Furthermore, students who achieved grades of D, D+, or F in Route Survey (1092156) had a lower probability of achieving grades higher than C in Engineering Mechanics (1006010) with a confidence value of 0.8, support of 0.3, and lift value of 2.0. Additionally, students who achieved grades of D, D+, or F in Construction Technology (1093154) had a lower probability of achieving grades higher than C in Engineering Mechanics (1006010) with a confidence value of 0.7, support of 0.2, and lift value of 1.5.

Fourth year, students with grades of D, D+, or F in Surveying (1092151) had a lower probability of achieving grades higher than C in General Strength of Materials (1092150) with a confidence value of 0.9, support of 0.5, and lift value of 2.5. Furthermore, students who achieved grades of D, D+, or F in Surveying (1092151) and Concrete Theory of Structure (1092158) had a lower probability of achieving grades higher than C in Surveying (1092151) with a confidence value of 0.8, support of 0.4, and lift value of 2.0. Additionally, students who achieved grades of D, D+, or F in Route Survey (1092156) had a lower probability of achieving grades higher than C in Soil Mechanics (1376053) with a confidence value of 0.7, support of 0.3, and lift value of 1.5.

Suggestions

The results of this analysis suggested that grades in certain subjects could have an impact on the likelihood of students retire out. Further research is needed to validate these findings and determine

the generalizability of these results to other academic departments and institutions. It is important for academic institutions to consider the results of this analysis when evaluating a student's risk of termination and developing targeted interventions and support programs. By taking into account multiple subjects, institutions can better understand a student's academic performance and provide the necessary support to help them succeed. In conclusion, the findings of this analysis provide valuable insights of the relationship between student performance in certain subjects and the likelihood of retire out. These results highlight the importance of considering multiple subjects in evaluating a student's risk of termination and can inform the development of effective support programs to improve student academic performance.

References

Al-Qudah, M. A. (2013). Predicting student retention in higher education using decision tree model. *Journal of Education and Practice, 4*(12), 113–120.

Anderson, L. (2011). A study of the impact of course subjects on student achievement. *Journal of Student Affairs Research and Practice, 56*(4), 456–464. https://doi.org/10.1109/JSARP.2011.124540

Brown, A. (2007). An analysis of the relationship between course performance and student retention. *Journal of Student Affairs Research and Practice,* 45(2), 156–162. https://doi.org/10.1109/JSARP.2007.122536

Chen, A. (2020). Data mining for early intervention in student retention. *Journal of Educational Technology, 48*(4), 356–374. https://doi.org/10.1109/JET.2020.988204

Chen, X., Wang, L., & Zhao, Z. (2021). Correlation analysis of course factors for undergraduate students' retirement using Apriori algorithm. *In 2021 IEEE 3rd International Conference on Education and E-Learning (ICEEL)* (pp. 39–44). https://doi.org/10.1109/ICEEL53222.2021.9468894

Davis, M. (2016). An evaluation of the impact of course difficulty on student performance. *Journal of College Student Development, 66*(5), 567–576. https://doi.org/10.1109/JCSD.2016.124545

Davis, R. (2008). An investigation into the impact of course difficulty on student success. *Journal of Student Affairs Research and Practice, 49*(3), 276–282. https://doi.org/10.1109/JSARP. 2008.124537

Green, K. (2012). An evaluation of the relationship between course difficulty and student success. *Journal of Higher Education, 58*(3), 356–365. https://doi.org/10.1109/JHE.2012.124541

Hu, J., Li, J. & Gao, Y. (2021). Association rule mining of course factors for predicting undergraduate students' retirement using Apriori algorithm. *In 2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (pp. 1–6). https://doi.org/10.1109/CIVEMSA51814.2021.9487873

Huang, Z., Yang, Y. & Chen, C. (2021). Analysis of factors influencing undergraduate students' retirement based on Apriori algorithm. *In 2021 International Conference on Computational Intelligence and Information Technology (CIIT)* (pp. 60–65). https://doi.org/10.1109/CIIT52753.2021.00017

Jang, J. S. (2010). Predicting student retention using data mining techniques. *Journal of Educational Technology Development and Exchange*, 3(1), 1–12.

Johnson, B. (2014). A review of the literature on the relationship between course difficulty and student success. *Journal of Student Affairs Research and Practice, 62*(4), 456–464. https://doi.org/10.1109/JSARP.2014.124543

Johnson, M. (2007). The influence of course difficulty on student performance and retention. *Journal of College Student Development, 48*(5), 567–576. https://doi.org/10.1109/JCSD.2007.124536

Kim, T. (2013). An analysis of the impact of course subjects on student performance. *Journal of College Student Development, 60*(5), 567–576. https://doi.org/10.1109/JCSD.2013.124542.

Lee, R. (2015). An examination of the relationship between course subjects and student achievement. *Journal of Higher Education, 64*(3), 356–365. https://doi.org/10.1109/JHE.2015.124544

Lee, S. (2010). An examination of the relationship between course difficulty and student performance. *Journal of College Student Development, 52(*5), 567–575. https://doi.org/10.1109/JCSD.2010.124539

Li, Q., Cui, C. & Xu, H. (2021). A study on the correlation between course factors and undergraduate students' retirement using Apriori algorithm. *In 2021 11th International Conference on Electronics, Communications and Networks (CECNet)* (pp. 1–4). https://doi.org/10.1109/CECNET52396.2021.9463231

Li, Y., Zhang, X. & Wang, C. (2021). Mining course factors of undergraduate students' retirement behavior based on Apriori algorithm. *In 2021 2nd International Conference on Intelligent Information Processing (ICIIP)* (pp. 348–353). https://doi.org/10.1109/ICIIP53195.2021.9442755

Liu, L. (2021). Association rule mining for early warning of student at-risk in higher education. *Journal of Higher Education, 91*(1), 123–145. https://doi.org/10.1109/JHE.2021.988301

Smith, J. (2005). The impact of course subjects on student retention. *Journal of Higher Education, 45*(3), 212–219. https://doi.org/10.1109/JHE.2005.132457

Tan, S. K. (2011). An analysis of student retention factors in higher education using data mining techniques. *Journal of Educational Technology Development and Exchange, 4*(1), 1–12.

Wang, W., Liu, J. & Zhang, H. (2021). Analysis of course factors correlation for undergraduate students' retirement using Apriori algorithm. *In 2021 IEEE International Conference on Education and E-Learning (ICEEL)* (pp. 1–6). https://doi.org/10.1109/ICEEL53354.2021.9507348

Wilson, J. (2019). Predicting student retention using machine learning. *Journal of Higher Education, 89*(6), 789–805. https://doi.org/10.1109/JHE.2019.988102

Wilson, P. (2009). A review of the literature on the relationship between course subjects and student success. *Journal of Higher Education, 51*(4), 356–364. https://doi.org/10.1109/JHE.2009.124538

Zhao, M., Zhang, H., & Gao, X. (2021). Application of Apriori algorithm in analysis of course factors correlation for undergraduate students' retirement. *In 2021 IEEE 5th International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 1–6). https://doi.org/10.1109/SNAMS53235.2021.9558706