



THE 5<sup>TH</sup> INTERNATIONAL ANNUAL MEETING  
ON STEM EDUCATION 2022  
FACULTY OF MATHEMATICS AND NATURAL SCIENCE  
UNIVERSITAS NEGERI MALANG  
Jalan Semarang 5, Malang, Jawa Timur, 65145, Indonesia



**LETTER OF ACCEPTANCE**

**Dear Duemong, Fudailah; Kepan, Solaeh**

Thank you for your submission to The 5<sup>th</sup> International Annual Meeting on STEM Education (IAMSTEM) 2022, which will be held on **August 5-6<sup>th</sup>, 2022 in Malang, Indonesia.**

The committee has completed the evaluation and we are happy to announce that your article abstract:

ID : 37

Title : A Study of Factors Affecting the Achievement in the National test (NT) of Thailand for Students in Jurisdiction of the Basic Education Commission Yala Primary Educational Services Area Office, Area 1

is **ACCEPTED** for oral presentation. The presentations are available in two modes: offline and online. The offline presenters are those who come in person to the conference venue in Malang, Indonesia. The offline presenters are required to prepare a poster (together with a stand) 160 x 60 cm. The online presenters are required to prepare powerpoint presentation.

Do not forget to pay the early bird registration fee (*40 USD for each international presenter or Rp 450.000,- for the local presenter/article*) no later than June 14<sup>th</sup> 2022.

Bank Name: Bank Negara Indonesia 46-BNI 46

VA Number: 9888855513090001

Bank SWIFT Code: BNINIDJAMLG

Billing ID: FMIPA 1209

Name: STEM FMIPA

Description: International Annual Meeting on STEM Education

Please submit your payment receipt, full paper, as well as certificate of IC-STEM (if you have any) in CMT by following guidelines: [https://bit.um.ac.id/upload\\_payment\\_proof](https://bit.um.ac.id/upload_payment_proof) . Full paper should be submitted no later than July 20<sup>th</sup>, 2022 to our system: <http://2022.iamstem.um.ac.id> . You may also download the paper template at [http://2022.iamstem.um.ac.id/?page\\_id=1457](http://2022.iamstem.um.ac.id/?page_id=1457) .

\* Additional fee for publication will be incurred after the review process

For more information, you can also contact our committee through following numbers:

Erti Hamimi : +6281336075431

Isnanik Juni Fitriyah : +628563358825

Looking forward to seeing you at the conference. We will try to do our best to make this year's event fruitful and memorable, and we look forward to welcoming you at The 5<sup>th</sup> IAMSTEM 2022.

Sincerely



Dr.rer.nat. Safwatun Nida, M.Pd



# A Study of Factors Affecting the Achievement in the National test (NT) of Thailand for Students in Jurisdiction of the Basic Education Commission Yala Primary Educational Services Area Office, Area 1

Fudailah Duemong<sup>1, a)</sup> and Solaeh Kepan<sup>2, b)</sup>

<sup>1</sup>*Computer Education Program, Faculty of Technology Science and Agriculture,  
Yala Rajabhat University, Yala, Thailand*  
<sup>2</sup>*Information Technology Program, Faculty of Technology Science and Agriculture,  
Yala Rajabhat University, Yala, Thailand*

a) fudailah.d@yru.ac.th

b) solaeh.k@yru.ac.th

**Abstract.** The National Test (NT) is a basic exam given for 3<sup>rd</sup> grade students conducted by the Thai Government, through the Ministry of Education, the National Institute of Education of Educational Testing Service under the office of the Basic Education Commission. The aim of the NT exam is to measure the knowledge of primary school students to understand students' literacy, numeracy, and reasoning ability. For this research was studying the suitable variables with the students' data obtained from the NT from 110 schools by the jurisdiction of the Basic Education Commission Yala Primary Educational Services Area Office, Area 1 for 2,692 students in 22 attributes in academic year 2020 (Between year 2020 and 2021). In this research, Educational Data Mining techniques are used to find correlations and co-occurrences by using Association Rules. Apriori algorithm, which was used to study factors affecting achievement on the National Test. The research found that the most effective factors and get more impact on the achievement of the National Test were students' disability, the shortage of school uniforms, students' religion, and school commuting, respectively, in the best rules found.

## INTRODUCTION

According to the statement about the present education by UNESCO in the Global Education Monitoring (GEM) report in 2020 [1], there were a variety of changes in society such as rapid technology advancements, climate changes, conflicts, impatience, hatred, social inequality, and the COVID-19 pandemic; therefore, these negatively impacted on more inequality and vulnerability in the society and education. It was found that one fifth of children and younger people around the world cannot access to education. It is very essential for those who play important roles in education in each country to acknowledge and understand the world's present changing situations and problems of their education systems in order to find out ways for learners to live in their flexible and equal societies as well as develop their education systems to suit real situations.

For education [2], [3], 21<sup>st</sup> Century Skills have been applied and focused on 3Rs (reading, writing, and arithmetic) and 4Cs (critical thinking, communication, creativity, and collaboration). Especially, primary school students are required to have more various skills such as communication and calculation to enhance the country's human capital and develop people at each age range qualitatively.

Nowadays for education in Thailand, manpower and potential of people at every age have been prepared and developed. According to the National Institute of Educational Testing Service under the office of the Basic Education

Commission, the National Test (NT) [1], [4], [5], which has been used since 1983, is aimed to investigate basic abilities of students' literacy, and numeracy to relevant to the Basic Education Core curriculum B.E. 2551 (A.D. 2008). In addition, this test is beneficial for teachers to know each student' potential and be able to improve learning lessons and teaching methods in the next semester. Plus, it is useful for the government to adjust the educational policies as well as for parents and students to improve their learning process at higher levels. The NT is a test for all 3<sup>rd</sup> grade students in Thailand under the Quality Assessment of Basic Education School projects. The test is used to measure two basic competency skills—Thai language and mathematics—in order to classify student quality at levels of the nation, regions, regional education offices, provincial education offices, education zone offices, and institutes with different contexts.

Regarding to the importance of the National Test, personal information and NT results of each 3<sup>rd</sup> grade student in academic year 2020 in Yala Primary Educational Services Area Office, Area 1, as well as schools' information were analyzed to find out correlation. By a technique of education data mining (EDM) and association rules with Apriori algorithm, frequent itemsets of the data were investigated to find out their correlation between students' characteristics and their NT results in order to become a guideline for educational quality development to suit area contexts.

## BACKGROUND AND RELATED WORK

### 1. Data Mining

Data mining [6]–[10] is considered an important part of discovering knowledge in a database. The purpose of data mining is to retrieve trends and rules of data to develop and apply to an organization or make decisions on operation improvements and methods. Outcomes of data mining are knowledge or new useful things used to support decision-making and understand types of data. The working process of data mining to retrieve knowledge as mentioned in Fig. 1.

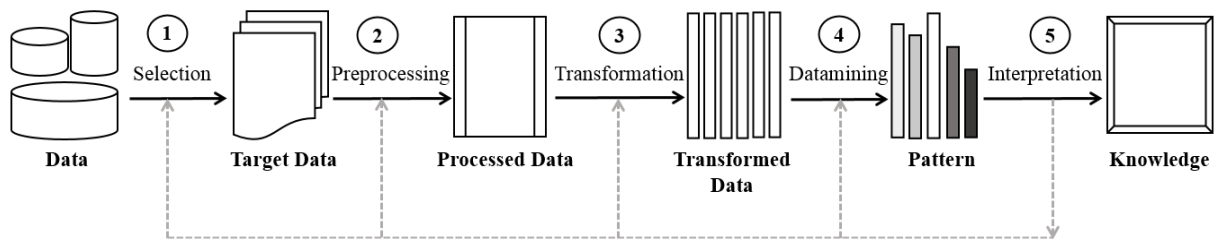


FIGURE 1. The process of discovering knowledge in a database

- Data selection is an indication of data sources and data retrieve for basic analysis before a step of data preprocessing. In fact, it is important to understand data in a database and their variables' meanings. Data's explanations, probable values, sources, and types should be clear in order to designate data target of outcomes or knowledge retrieved from a database. In this process of data search, data related to a target should specifically be retrieved to be inputs for the next step. Therefore, data selection varies depending on objectives of each job.
- Data preprocessing is one of the essential steps, focused on data investigation and data cleansing in order to have quality data. If data are not quality, outputs or outcomes will not be good. There are three methods of the data processing as follows.
  - Inspection of data scopes is a way to eliminate incorrect or irrelevant data. For example, salary data should be shown in numbers but not letters or symbols.
  - Elimination of missing or incomplete data can be done by correcting data with 'null' in gaps in order not to affect other data.
  - Elimination of noisy data, which means clearly different data from other data in the same database, is a method to investigate and eliminate data probably by statistical analysis methods for decision making.

- Data Transformation which has been cleansed is transformed to be ready data for analysis. In the step of data transformation, correction is needed to facilitate working on data search and retrieve correct data as targeted.
- Data mining includes various data mining techniques which are important to consider existing data and outputs. After a technique is selected, an algorithm should be selected for data search in a database. It is necessary to consider types of inputs, processing duration, and existing resources to select a suitable algorithm without problems on processing for analysis. For instance, if knowledge in a form of association rules, a technique called 'association' and Apriori algorithm should be selected for data processing
- Data Interpretation is outputs or outcomes after the process of data mining may not easily be understandable, so they should be interpreted for easy understanding and use in this step. After data are interpreted, the data will be shown in tables, graphs, or figures. If outputs are errored or irrelevant to purposes, the data have to be reprocessed in the aforementioned steps (1-4) in order to get the most correct and useful data. When data are ready in a required form, the data and knowledge can be useful for decision-making, organization development, or new product creation, and so on.

## 2. Educational Data Mining

Educational Data Mining [3], [7], [9], [11], [12] is an integration of each field of research to understand learner's learning behavior and learning results and analyze learning styles. It is considered a way to give real knowledge of education system, which is very useful for administrators to make decisions, promote learning and teaching processes, and develop learner-based strategies. The goals of educational data mining contain the following three components.

- Prediction of learner's behavior is conducted by identifying each learner's characters such as knowledge, perception, motivation, and attitudes. A model of learners can also be set with the mentioned characteristics, and learners can be classified into each type.
- Learning and teaching materials can be designed to make learners enjoy their classes. In addition, teachers can more easily set concepts of teaching for students.
- Efficient learning and teaching can be promoted to reflect learning outcomes in the long term by setting short-term goals which can be brought to be teaching methods and learner-centered learning to positively impact on each learner's behavior.

## 3. Data Mining Techniques

There are various data mining techniques [8], [11]–[15] used to analyze data and process the data to get data types, data models or data prediction as follows.

- **Statistics and Visualization:** There are a lot of statistical techniques used to analyze data, especially numerical data. Data mining for big-sized data can be presented in visualization which is interesting and effective to make understanding of data types, data distributions, or data outliers.
- **Classification** is a form of data analysis through classifiers to classify related data categories or classes.
- **Clustering** is a process to group data with high resemblance into one group or separate data into each group. This technique is useful to distribute data and eliminate data outliers.
- **Prediction** is a technique which is similar to classification, but it is different because data are separated for classification depending on behavior or value prediction in the future. The past data are models to predict or explain what will happen in the future. Regression is one of statistical techniques usually used to create an equation from the existing data to predict the future data.
- **Neural Network** is a way to create a computer to duplicate human brain's working process so that the computer can think and learn similarly to human's neural network in order to memorize uncertain patterns, and predict or classify jobs into each category or class.
- **Association** is one technique of data mining to find out correlation of data in data groups. Each data characteristic can identify another. There are three groups of association as follows:
  - association discovery such as market basket analysis to plan for promotions or develop products to meet customer needs,

- sequential pattern discovery such as connection of customer's purchase, focusing on purchasing behavior in the long term, for example, a customer having bought a computer and tending to buy another computer later,
- Similar time sequence discovery which is used to find out connection between two sets of data at the related period of time such as similar time of selling higher amount of alcohol and snacks.

## 4. Association Rules

Association rule [10], [16], [17] was firstly developed by researchers at IBM (International Business Machines Corporation), the United States. Its objective is to investigate interesting correlation hidden in a data set. The clear example of these rules' applications was that an association rule for customers' purchasing in one cart at a supermarket (Market Basket Analysis) in order to understand behavior of customers. For instance, when a customer buys milk, he tends to buy bread as well. The most important step of the association rule is to find out frequent itemsets. In fact, it often takes the longest period of time to finish this step. Therefore, each algorithm should be selected to suit types of data in a database in order to reduce period of time and memory capacity.

Generally, the association rule is divided into two steps as shown Fig. 2. First, a step of searching for frequent itemsets is done by searching items in a database containing each relevant set of items. A support value of these items should be more than or equal to the minimum support value determined by users to consider being frequent itemsets of data. Second, a step of finding out an association rule of retrieved frequent itemsets. An association rule can be accepted with a confidence value which is more than or equal to the minimum confidence value determined by users. Mining frequent itemsets is considered an important step for an association rule. Frequent itemsets can be searched to create a great number of association rules. This step takes a lot of time and memory capacity; therefore, users need to select the most suitable algorithm for each set of data in each database, resources of existing computers, or outputs.

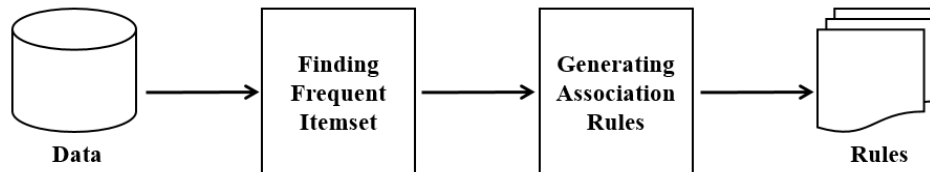


FIGURE 2. The data mining process of discovering the association rules

## 5. Apriori Algorithm

Apriori algorithm [10] which was firstly proposed by Apriori in 1993 is popularly used to create association rules. It is also considered to be an original algorithm to mine frequent itemsets. The working process of the Apriori algorithm is shown below.

- Each item appearing in a database are firstly read to count frequency rates.
- Frequency values of each item are inspected to calculate support values which are more than or equal to the minimum support value to be considered as frequent itemsets with a size of 1 item,  $L_1$ : Frequent 1-itemset.
- Then all of  $L_1$  in the second step is combined to create candidate Itemsets with a size of 2 items,  $C_2$ : Candidate 2-itemset
- Each item in the database is read again and count the frequency rates of  $C_2$ . Then  $C_2$  having fewer support values than the minimum support value is eliminated.  $C_2$  having support values which are more than or equal to the minimum support value will become  $L_2$ .
- The third and fourth steps are repeated until  $C_k$  cannot be created from  $L_{k-1}$  when k means a size of itemsets. The candidate itemsets will be finished creating and the process is ended, resulting in frequent itemsets.

## 6. Related Research Works

Predicting students' performance [7], [9], [16], [18] is one of the most interesting and challenging issues in educational data mining (EDM) for researchers due to more data in educational systems. Students' performance can be predicted, together with identifying students' risk of academic failures. Their performance was also useful for instructors to take some actions to help students as early as possible, provide with support in selecting courses, or evaluate learners as to optimize learning or design appropriate future study plans for students. Furthermore, patterns of education management have been changed from traditional education (Offline Education) to be online education. It has resulted in changes of various dimensions such as teachers' less important roles, flexibility of study locations, convenience of selecting study time, and cheaper expenses of online study than offline study. For example, Xiaodong et al. [19] applied Apriori algorithm to seek for correlation between courses and factors leading to high or low grades of each course. Association rules were employed to 34 courses and 100 students. The outcomes of this data mining became references for improvements of each course's teaching quality. Abdallah et al. [20] also studied whether a factors of convenience, to access the Internet via smart phones and wireless devices for e-learning, and students' participation affected students' performance or not. His research results of search for association rules by Apriori algorithm showed that both factors had positively correlation to better performance. In addition, it was found that the factors most frequent affecting the students' performance were grades and their previous learning achievements, their e-learning activities, their population data, and their social information. It was also indicated in the research that the most frequently used data mining techniques to predict and classify students' factors were decision trees, Naïve Bayes classifiers, and artificial neural networks.

## RESEARCH METHODOLOGY

The main aim of this study is to investigate factors affecting achievements in the National Test (NT) that affect the students' performance, as well as, the most common education data mining techniques applied to study correlation between learners' characteristic and each student's quality evaluation. This research is based on the mining process, which consist of 4 states as shown in Fig. 3.

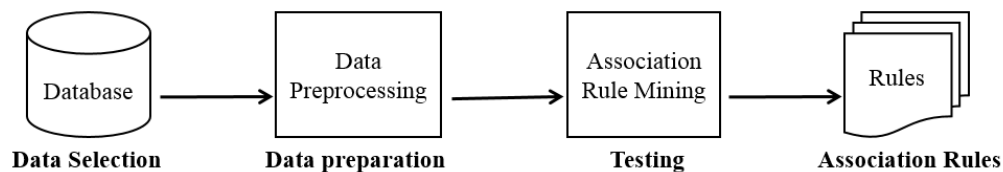


FIGURE 3. The mining process for discovering association rules

1. **State of data selection:** The data used in the research were the data of the National Test results and each 3rd grade student's profile in Yala Primary Educational Services Area Office, Area 1 in the academic year 2020. The reason why the data of the academic year 2020 were selected was that there were the higher number of schools taking part in this test than those in the academic year 2021 possibly due to the COVID-19 pandemic. Factors used in the research were genders, ages, religions, a number of siblings, marital status of parents, father's income, mother's income, parents' incomes, shortage of opportunities, shortage of uniforms, shortage of stationery objects, shortage of lunch meals, disability, distance from a house to a school, means of transportation to school, and GPA scores.
2. **State of data preparation:** After the data were retrieved from the database, the next step was data preparation by transforming the data to suit data processing. The data of NT results and each student's profile were required to in a form of usable value for processing.
3. **State of testing:** Testing is a step to use prepared data to test by a data mining technique thorough Apriori algorithm to find out association rules. Document files for testing were in a form of ".rtff" (ARFF format). A package program used in the research was WEKA (Waikato Environment for Knowledge Analysis) which was developed from JAVA language by Waikato University in New Zealand in 1997. It was created and

focused on machine learning and data mining. Association rules of NT results and each student's personal information were mined by considering their factors in order to be a guideline on quality and efficient teaching and learning development. Apriori algorithm was used to determine the minimum frequency rate for itemsets by considering factors gained from the association rules. Then frequent itemsets were created from itemsets. The support values of frequent itemsets must be greater than or equal to the minimum support value. This step needed to be repeated until frequent itemsets could not be found.

4. **State of Association Rules:** Association rules were gained from factors in the previous step. An association rule was in a form of "IF – THEN". Each rule contained two parts: the left part of 'IF' and the right part of 'THEN'. The left part might include more than one true condition to make a condition in the right part become true. For example, "IF X THEN Y" could be symbolized instead of "X -> Y". When there is X, there will be Y. This rule was considered a model to seek for itemsets to be a guideline on quality and efficient teaching and learning development. An association rule could be precise by measuring a support value which was probability of two data sets. Confidence values were used to search for probability that when there is X, there will be Y. In addition, lift values were used to measure rule importance or correlation happening in each situation resulting in another situation.

## RESULTS OF EXPERIMENTAL

In this research, NT results (Thai language and mathematics) of 2,692 third grade students in the academic year 2020 (2020-2021) in Yala Primary Educational Services Area Office, Area 1 were analyzed. There were also 22 lists of the related data gathered in the research: mathematic scores, Thai language scores, total scores of these two subjects, genders, ages, religions, a number of siblings, Parent status, father's income, mother's income, patron parents, patron's incomes, shortage of opportunities, shortage of uniforms, shortage of stationery objects, shortage of textbooks, shortage of lunch meals, disability, distance from a house to a school, means of transportation to school, and GPA scores. For research methodology, a data mining technique in terms of association rules and Apriori algorithm was applied in order to study correlation between students' characteristics and NT results of each student. Furthermore, the research results could be applied for development of quality teaching and learning to suit each area context. The results shown in Table 1 refer to the experimental results of this research we focus on the Apriori algorithm for association rule mining with WEKA.

**TABLE 1.** Represents testing examples of association rules

No.	Rules	Confidence	Lift	Conviction
1	If the students had not been disabled, they would not have lacked their uniforms	1	1	0.90
2	If the students had been Muslims, they would not have lacked their uniforms.	1	1	0.85
3	If they students had gone to school by public bus, they would not have lacked their uniforms.	1	1	0.77
4	If the students had been Muslims, and they had not been disabled, they would not have lacked their uniforms.	1	1	0.75
5	If the students had not been disabled, and they had gone to school by public bus, they would not have lacked their uniforms.	1	1	0.69
6	If the students had been Muslims, and their mothers' incomes had been less than 3,000 Baht, they would not have lacked their uniforms.	1	1	0.67
7	If the students had been Muslims, and they had lacked their stationery tools, they would not have lacked their uniforms.	1	1	0.66
8	If the students had lacked their stationery tools, and they had not been disabled, they would not have lacked their uniforms.	1	1	0.65
9	If the students had lacked their stationery tools, they would not have lacked their uniforms.	1	1	0.37



**TABLE 1.** Represents testing examples of association rules

No.	Rules	Confidence	Lift	Conviction
10	If the students had lacked their textbooks, they would not have lacked their uniforms.	1	1	0.36

## CONCLUSION AND FUTURE WORK

Under the principles of data mining, association rules by Apriori algorithm were applied via Weka tool to find out maximal frequent itemsets. The outputs of association rules from the dataset were searched for seven times. At the seventh time, the minimum support was determined at 0.65, and the minimum confidence was designated at 0.90 which was the most suitable confidence value related to the dataset of the research. The findings showed the best 10 rules as follows in Table 1. The outcomes of association in Weka by Apriori algorithm under the techniques of datamining revealed that all the best rules were in the same direction or related to the student's uniforms. However, the research was only conducted by Apriori algorithm. Other research of comparison should be done by other algorithms such as FP-Growth algorithm, classification techniques in data mining, Decision Tree, and J48 algorithm. In addition to Weka, other tools for association rules could be used such as RapidMiner and Tanagra to analyze and seek for the best rules in order to be a guideline on educational quality development in terms of educational budgets, educational policy planning, and learning applications to highly support teachers and learner.

## REFERENCES

1. UNESCO, "Global Education Monitoring Report 2020 - Inclusion and education - All means all," Paris, Jun. 2020.
2. V. Srinivasan and H. Murthy, "Improving reading and comprehension in K-12: Evidence from a large-scale AI technology intervention in India," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100019, 2021.
3. W. Xiao, P. Ji, and J. Hu, "A survey on educational data mining methods used for predicting students' performance," *Eng. Reports*, vol. 4, no. 5, May 2022.
4. National Institute of Educational Testing Service, "Thailand - Ordinary National Educational Test," *National Institute of Educational Testing Service - Public organization*.
5. The National Institute of Educational Testing Service (Public Organization), "National Testing and Assessment," *The National Institute of Educational Testing Service (Public Organization)*.
6. S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, Sep. 2012.
7. G. Feng, M. Fan, and Y. Chen, "Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining," *IEEE Access*, vol. 10, pp. 19558–19571, 2022.
8. U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Mag.*, vol. 17, no. 3, pp. 37–54, 1996.
9. P. Bachhal, S. Ahuja, and S. Gargrish, "Educational Data Mining: A Review," *J. Phys. Conf. Ser.*, vol. 1950, no. 1, p. 012022, Aug. 2021.
10. R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Rec.*, vol. 22, no. 2, pp. 207–216, Jun. 1993.
11. C. Romero and S. Ventura, "Data mining in education," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 3, no. 1, pp. 12–27, Jan. 2013.
12. F. Ouyang and P. Jiao, "Artificial intelligence in education: The three paradigms," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100020, 2021.
13. K. K. Nivethithaa and S. Vijayalakshmi, "Survey on Data Mining Techniques, Process and Algorithms," *J. Phys. Conf. Ser.*, vol. 1947, no. 1, p. 012052, Jun. 2021.
14. S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Mining in Educational Data: Review and Future Directions,"

2020, pp. 92–102.

15. V. RAMESH, P.PARKAVI, and K. Ramar, “Predicting Student Performance: A Statistical and Data Mining Approach,” *Int. J. Comput. Appl.*, vol. 63, pp. 975–8887, Jul. 2013.
16. T. Wang, B. Xiao, and W. Ma, “Student Behavior Data Analysis Based on Association Rule Mining,” *Int. J. Comput. Intell. Syst.*, vol. 15, no. 1, p. 32, Dec. 2022.
17. F. Duemong, L. Preechaveerakul, and S. Vanichayobon, “FIAST: A Novel Algorithm for Mining Frequent Itemsets,” *2009 Int. Conf. Futur. Comput. Commun.*, pp. 140–144, 2009.
18. A. Abu Saa, M. Al-Emran, and K. Shaalan, “Factors Affecting Students’ Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques,” *Technol. Knowl. Learn.*, vol. 24, no. 4, pp. 567–598, Dec. 2019.
19. W. Liu and P. Zhou, “Research on the Relationship between Entrepreneurial Intention of College Students and Their Characteristics,” in *Proceedings of the 2nd International Seminar on Education Research and Social Science (ISERSS 2019)*, 2019, pp. 1–3.
20. A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, “Relationship Between Student Engagement and Performance in E-Learning Environment Using Association Rules,” in *2018 IEEE World Engineering Education Conference (EDUNINE)*, Mar. 2018, pp. 1–6.