Received: September 7, 2023 | Revised: December 12, 2023 | Accepted: December 14, 2023

JOURNAL OF APPLIED RESEARCH ON SCIENCE AND TECHNOLOGY
JARST
INSTITUTE OF RESEARCH AND DEVELOPMENT
RAJAMANGALA UNIVERSITY OF TECHNOLOGY THANYABURI
ISSN ONLINE: 2773-9473
https://ph01.tci-thaijo.org/index.php/
rmutt-journal

The development of association rules for student performance analysis using FP-Growth algorithm as a guideline for multidisciplinary learning

Wongkot Sriurai[1*] and Sumitra Nuanmeesri[2]

[1]Department of Mathematics Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani 34190, THAILAND

[2]Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok 10300, THAILAND

[*]Corresponding author: wongkot.s@ubu.ac.th

ABSTRACT

This study aims to develop association rules for student performance analysis using the FP-Growth algorithm. The data used for developing the association rules comprised 107 student reports. The reports, voluntarily provided by 107 junior high school students, consisted of student achievement results across 8 subject areas: Thai Language, Mathematics, Science, Social Studies, English Language, Computing Science, Visual Arts, and Home Economics. The data was applied to developing association rules using the FP-Growth algorithm towards WEKA, a machine learning software. The research team designed the process consisting of the following 5 stages: data collection, data preparation, model formulation, model evaluation, and model application. After achieving the association rules, the research team applied them to the prototype development of a student performance analysis system for promoting students' academic excellence. The system could be operated by Android mobile phones. According to the research results, the association rules developed by the algorithm provided a confidence level of 92%, and a rule of 7 rules will be generated. The findings indicated the correlations between the subject areas, which shared similar individual students' academic achievements (≥ 80 scores). The association rules could be applied to the multidisciplinary curriculum planning, which benefited students and promoted academic excellence. For example, by applying Rule, it could be assumed that students who earned 80 scores or higher in the English subject would likely earn identical scores from their Thai Language class. Therefore, they could effectively learn to integrate English and Thai languages. To illustrate, students may be asked to translate song lyrics from English to Thai, serve as tourist guides or translators, or even give welcome speeches to foreign guests.

Keywords: Association rules, FP-Growth, Multidisciplinary learning

INTRODUCTION

Education is a foundation of human beings' development process, which supports the maturation of their physics, emotions, intelligence, morality, and ethics. It also facilitates personal adaptation and a peaceful way of living. Thailand has paid attention to providing educational opportunities, human capital investment, and lifelong learning. This conforms to the National Education Act of B.E. 2542 (1999) and its second amendment, the National Education Act of B.E. 2545 (2002), [1]. Moral, ethical, and cultural intelligence allow individuals to live with each other happily. Therefore, today's educational management must focus on these qualities, especially during the COVID-19 pandemic.

Schools are educational institutions responsible for educational management. They are obliged to maintain academic standards in agreement with the demands of society as well as students' interests, talents, and capabilities. Each school attempts to enhance students' knowledge and skills for their future careers. In Thailand, opportunity expansion schools have provided 3-year education to countless junior high school students. A huge number of 6th graders who had graduated from different elementary schools entered these opportunity expansion schools. Some of them were students from other districts or provinces, creating diversities in talents and skills among the young learners. At the stage of lower secondary education, students can choose to study academic or vocational subjects based on their preferences. This helps the students to obtain academic and vocational knowledge appropriate for their age, market demand, and individual interests. Ultimately, these students will achieve their own personal ambitions and the society's common goals.

At present, association rules have been applied to discover the rules in large data sets. For example, Dharmaraajan and Dorairangaswamy investigated the

DOI: 10.60101/jarst.2023.253807

correlation patterns of a weblog's data using FP-Growth and Apriori algorithms to classify user behavior when accessing the weblog's data. They compared the effectiveness of the two algorithms in data correlation analysis with the Rapid Miner program. Their research findings showed that both algorithms could analyze the weblog's usage patterns as well as characteristics of user behavior. The data could be used to improve the web design or offer services to the users more productively. After comparing the effectiveness of both algorithms in terms of data size and time spent on data processing, FP-Growth spent less time on data processing and provided higher productivity than Apriori [2]. Gashaw and Liu assessed the performance of popular data mining algorithms, including Eclat, Apriori, and FP-Growth. They employed three sets of weblog data to analyze association rules by comparing the effectiveness of the algorithms in two aspects: 1) the amount of time spent on data processing and 2) memory usage. They found that Apriori performed relatively better when the support value was high regarding the amount of time spent on big data processing. However, if the support value decreased, all three algorithms spent more time processing data. Only FP-Growth could create a new dataset within the expected duration, while Eclat required the most significant amount of time in all cases. Regarding memory usage, FP-Growth demanded the smallest number of memory units compared to other algorithms when the support value was high. Eclat and Apriori required more memory units when the support value was low. Meanwhile, FP-Growth constantly uses the same number of memory units [3]. Faza et al. investigated the association rules of graduate student data gained from an Indonesian university. They applied the FP-growth algorithm by which the data used for constructing association rules consisted of types of schools, types of admission, grade point average (GPA), and duration of study. Their research results demonstrated that attributes correlating with graduation were public high schools outside Medan City, regular admission, GPA between 3.00-3.49, and study duration. According to the findings, the research team could identify which schools were suitable for encouraging students to enroll at this university and graduate within the regular timeframe [4]. Jongkasikit applied a data mining technique to explore factors affecting higher education enrollment at the Faculty of Industrial Technology, Lampang Rajabhat University. She accumulated data from 334 students from Years 1-3 at the faculty and then reduced the dimensionality of the data by employing the Evolutionary Selection technique. The dimensionally reduced data was accordingly used to find association rules for applying FP-Growth. Their findings disclosed 36 factors affecting higher education enrollment of these students. 35 association rules were constructed based on the discovered factors [5]. According to the previously mentioned studies, each research applied the association rules to find correlations between different item sets. Overall, it can be assumed that association rules can be practically used for investigating hidden data correlations.

Thailand has constantly modified its educational management patterns and techniques to equip young learners with knowledge and skills in Thai language, mathematics, science, social studies, English language, computing science, visual arts, and home economics. These subjects are compulsory for every student. Thus, the research team aims to explore students' aptitudes by analyzing the student's past academic performance through association rule mining [6]. The association rules can signify the relationships between subject areas where students share similar achievements. Once the association rules have been formed, they are applied to the multidisciplinary curriculum planning. The newly developed curriculum is expected to benefit the students and promote their academic excellence to the greatest extent. This research aims to develop association rules for student performance analysis by using the FP-Growth algorithm and to develop the information system by using the FP-Growth algorithm for designing a multidisciplinary curriculum.

The paper is organized as follows. The next section describes the research methodology. Subsequently, we report on the association rules for student performance analysis created by the FP-Growth algorithm via the WEKA. We conclude by summarizing our contributions.

MATERIALS AND METHODS

Research methodology

After reviewing relevant theories and previous studies, the research team designed the process consisting of the following 5 stages: data collection, data preparation, model formulation (using FP-Growth Algorithm), model evaluation, and model application (See Figure 1).

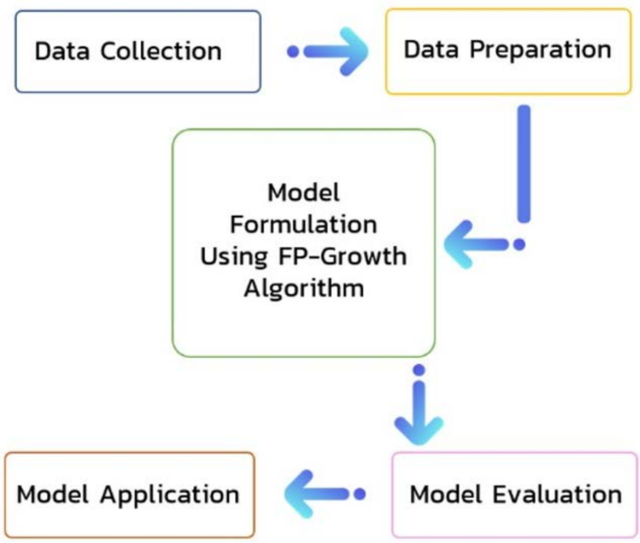


Figure 1 Research process.

DOI: 10.60101/jarst.2023.253807

1. Data collection

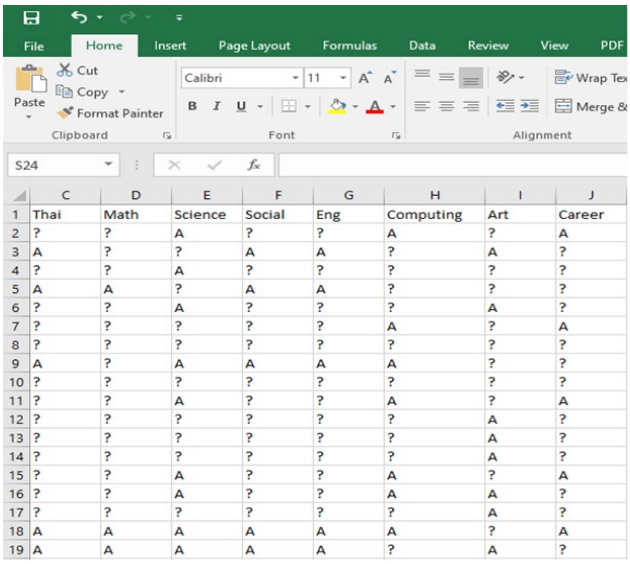
The data used for developing the model consisted of school reports voluntarily provided by 107 students from a junior high school. Ethics approval for the conduct of this study was obtained. The data comprised student grades of 80% or higher across 8 subject areas, including Thai language, mathematics, science, social studies, English language, computing science, visual arts, and home economics. These subjects were represented by the variables shown in Table 1.

Table 1 Examples of variables indicating each subject.

Variables	Descriptions
Thai	Thai Language
Math	Mathematics
Science	Science
Social	Social Studies
Eng	English Language
Computing	Computing Science
Art	Visual Arts
Career	Home Economics

2. Data preparation

Following data collection, the data cleaning process was operated. At this stage, the incomplete datasets were eliminated. In this case, the student grades lower than 80% were excluded. Once the data cleaning process had been completed, the data was converted into a .csv file, as illustrated in Figure 2. Subsequently, the converted data was applied to the model development process.



	C	D	E	F	G	H	I	J
1	Thai	Math	Science	Social	Eng	Computing	Art	Career
2	?	?	A	?	?	A	?	A
3	A	?	?	A	A	?	A	?
4	?	?	A	?	?	?	?	?
5	A	A	?	A	A	?	?	?
6	?	?	A	?	?	?	A	?
7	?	?	?	?	?	A	?	A
8	?	?	?	?	?	?	?	?
9	A	?	A	A	A	A	?	?
10	?	?	?	?	?	?	?	?
11	?	?	A	?	?	A	?	A
12	?	?	?	?	?	?	A	?
13	?	?	?	?	?	?	A	?
14	?	?	?	?	?	?	A	?
15	?	?	A	?	?	A	?	A
16	?	?	A	?	?	A	A	?
17	?	?	?	?	?	?	A	?
18	A	A	A	A	A	A	?	A
19	A	A	A	A	A	?	A	?

Figure 2 Examples of data used for model development.

Figure 2 displays examples of data used for model development. The developed model was expected to investigate the association rules for student academic performance analysis effectively. 'A' represents 'Grade 4' (scores ≥ 80%) in each subject.

3. Model formulation using FP-Growth algorithm

The model representing the correlations between subject areas that shared similar individual student achievements (scores ≥ 80%) was formulated at this stage. It was constructed using the FP-Growth algorithm via WEKA, a machine learning software. The following parameters were identified to create association rules: lowerBoundMinSupport = 0.1, minMetric = 0.9, and numRulesToFind = 10.

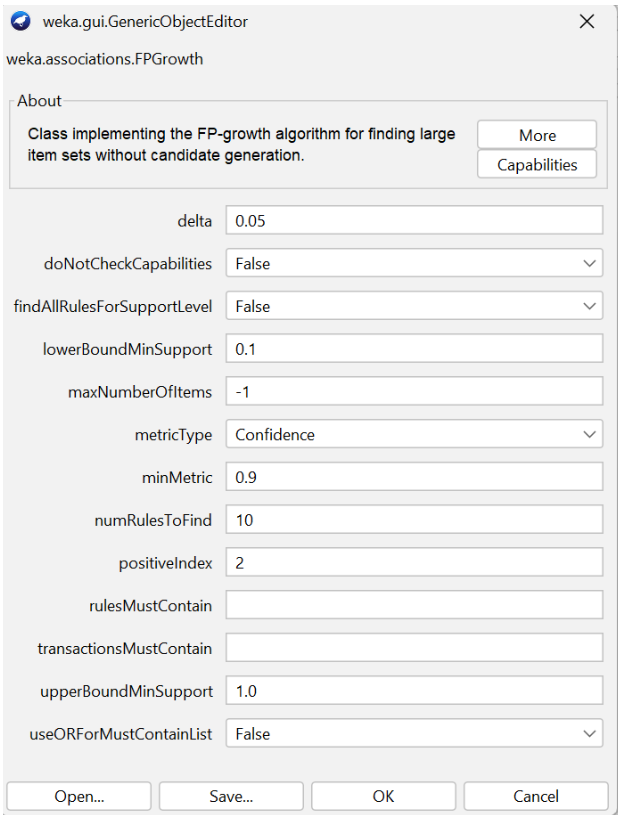


Figure 3 Association rules parameters in WEKA.

The FP-Growth algorithm searches for frequently co-occurring itemsets. FP-Growth algorithm was used to find the interesting rules from education data. Its pattern growth consists of the following steps [7, 8].:

1. First data scanning: The frequency of each dataset in the database is counted. Then, the itemsets with a minimum support value higher than or equal to the identified target are ordered based on the frequency from high to low to formulate the Header table.

2. Second data scanning: A FP-Tree is created by scanning each data list in the database. The item sets disappearing in the Header table are excluded. The remaining datasets are re-ordered in the Header table. These datasets are used to create node trees added to the FP tree. The nodes with the same itemsets are connected and then added to the Header table.

3. Conditional pattern base formulation: Each dataset's FP-Tree conditions are created to find frequently co-occurring itemsets. A conditional pattern base is a sub-database comprising sets of frequently co-occurring itemsets in each path. Each itemset is set with the

DOI: 10.60101/jarst.2023.253807

same frequency as the itemset currently determined by the FP-tree. Afterwards, the FP-tree is formulated in the conditional pattern base, so-called a 'conditional FP-tree'. It is derived from the total sum of frequencies of each item set from all paths. Only the itemsets with the acceptable minimum support value are selected for creating frequently co-occurring itemsets further.

4. Finalization: The frequently co-occurring itemsets are searched by constructing the conditional pattern base and the conditional FP-tree of each dataset towards the 'divide and conquer' technique.

4. Model evaluation

After obtaining the association rules, the reliability of the outputs was evaluated. The evaluation considered the support, confidence, and lift values of the association rules before applying them to curriculum planning for the development of academic excellence. If the consistency index is less than 1, it means there is an inconsistent correlation. On the other hand, if the consistency index is equal to or greater than 1, it means there is a consistent correlation. The values of support are calculated with Formula 1. The confidence calculation of the association rule A=>B is shown in the following formula 2 [7, 8].

$$support(A => B) = support(AUB)/P(AB) \quad (1)$$

$$confidence(A => B) = support(AUB)/support(A) \quad (2)$$

5. Model application

Once the association rules of students' academic achievements (at a grade of 80% or above) had been obtained through the model developed by the FP-Growth algorithm, they were applied to developing an information system for designing a multidisciplinary curriculum. The curriculum was aimed to promote students' academic excellence based on their preferences and talents.

RESULTS AND DISCUSSIONS

The association rules of students' academic achievements were discovered by constructing the aforementioned model towards the use of FP-Growth via WEKA. According to the investigation of association rules across the subject areas, the association rules simultaneously emerged and passed the determined minimum support value. Even if their order was rearranged, similar association rules were subsequently obtained.

These association rules were found to have a confidence level of 92% or higher, meaning they were statistically reliable. As a result, the acquired association rules could be applied to curriculum planning for promoting students' academic excellence. Figure 4 illustrates the association rules for student performance analysis created by the FP-Growth algorithm.

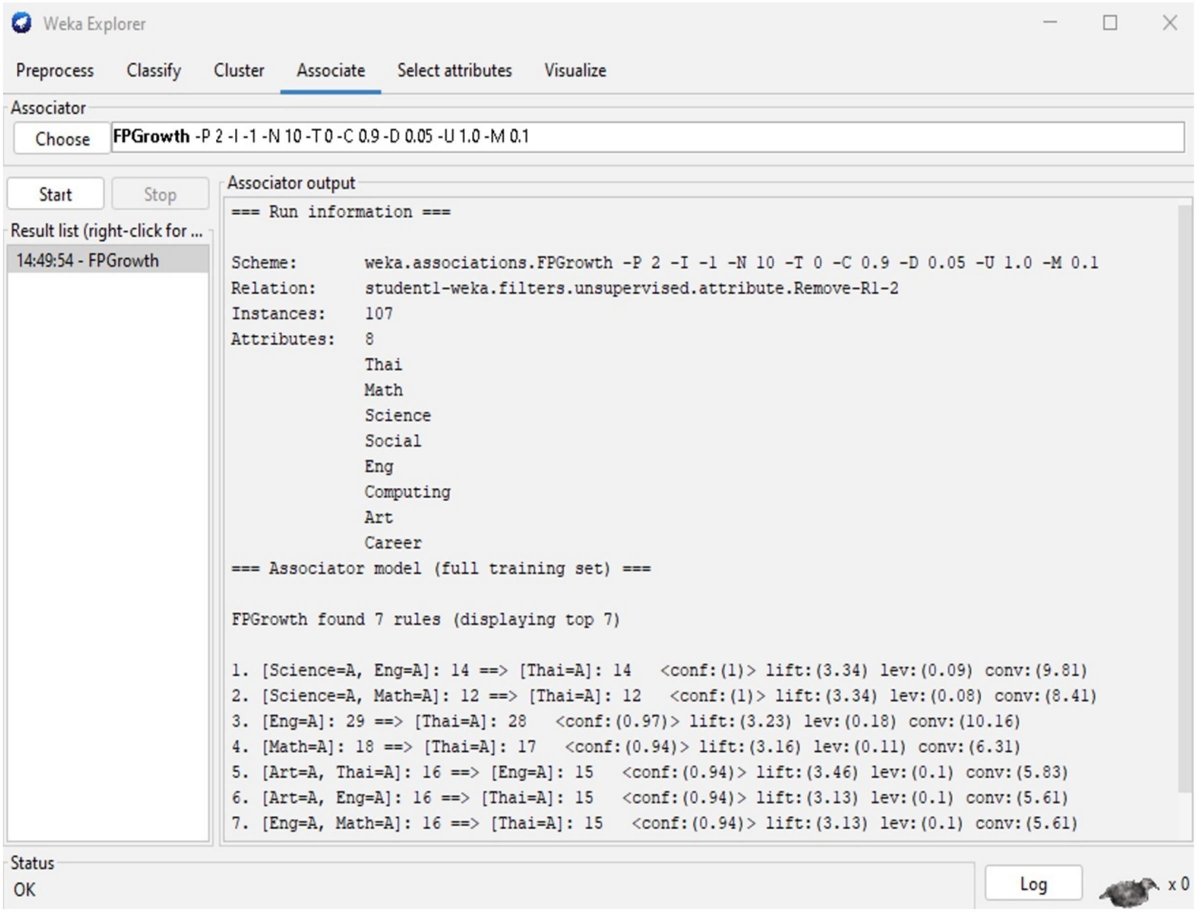


Figure 4 Association rules gained from the analysis of student's academic achievements.

DOI: 10.60101/jarst.2023.253807

According to Figure 4, there were 7 association rules of students' academic achievements as follows.

Rule 1: Students who received A grades in Science and English Language were more likely to receive an A grade in the Thai Language subject (at a confidence level of 100%).

Rule 2: Students who received A grades in Science and Mathematics were more likely to receive an A grade in the Thai Language subject (at a confidence level of 100%).

Rule 3: Students who received an A in English were more likely to receive an A in the Thai language subject (at a confidence level of 97%).

Rule 4: Students who received an A in Mathematics were more likely to receive an A in the Thai Language subject (at a confidence level of 94%).

Rule 5: Students who received A grades in Visual Arts and the Thai Language were more likely to receive an A grade in the English Language subject (at a confidence level of 94%).

Rule 6: Students who received A grades in Visual Arts and English Language were more likely to receive an A grade from the Thai Language subject (at a confidence level of 94%).

Rule 7: Students who received A grades in English Language and Mathematics were more likely to receive an A grade in the Thai Language subject (at a confidence level of 94%).

As the association mentioned above rules demonstrated, multidisciplinary curriculum planning could be achieved. For instance, by applying Rule 3, it could be assumed that students who earned 80 or higher scores in the English subject were likely to earn the same scores from their Thai language class. Therefore, they could effectively learn to integrate English and Thai languages. To illustrate, students may be asked to translate song lyrics from English to Thai, serve as tourist guides or translators, or even give welcome speeches to foreign guests. Another example is the application of Rule 4 (Mathematics scores ≥ 80% = Thai scores ≥ 80%). Interactively, students may be asked to summarize and evaluate media reliability while learning about statistical data collection (the combination of Chapter 4, Statistics in the Mathematics subject and Module 1, Creative Communication in the Thai language subject).

After achieving the association rules, the research team applied them to the prototype development of a student performance analysis system for promoting students' academic excellence. The system could be operated by Android mobile phones, as shown in Figure 5.

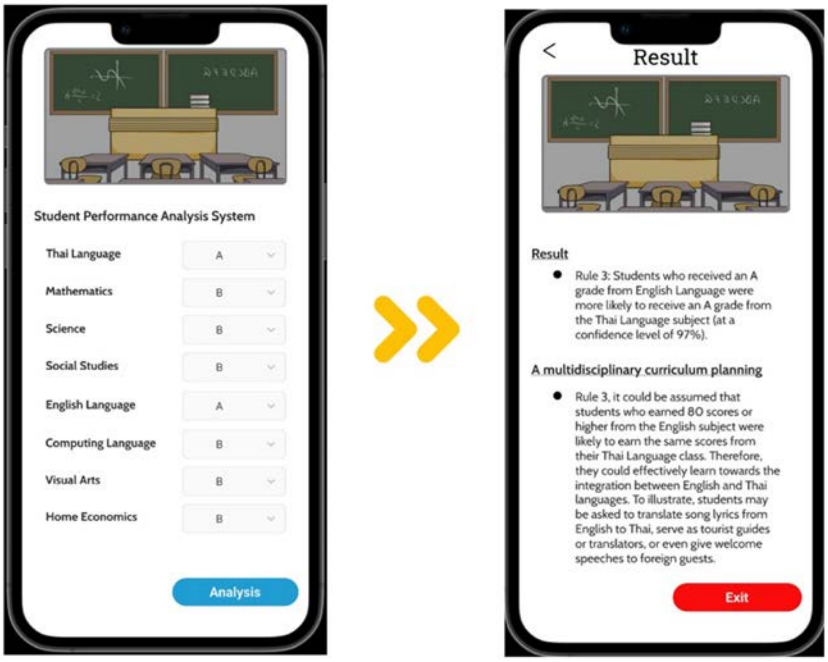


Figure 5 Example of student performance analysis system.

CONCLUSION

Educational management based on individual talents and preferences can contribute to students' higher education and job opportunities. Particularly for students at opportunity expansion schools with high talent diversity, instructors should analyze and determine the best ways to teach each student effectively and appropriately. This study formulated the association rules of students' academic achievements using the FP-Growth algorithm via the WEKA data mining software. The research findings revealed that the association rules developed by FP-Growth could achieve a confidence level of 92% and higher. This implies that these association rules could be applied to multidisciplinary curriculum planning aimed at

DOI: 10.60101/jarst.2023.253807

academic excellence. For instance, students who earned 80 or higher in the English subject were likely to earn identical scores in their Thai Language class. Therefore, they could effectively learn to integrate English and Thai. The findings indicated the correlations between the subject areas, which shared similar individual students' academic achievements (≥ 80 scores). Therefore, teachers can design teaching and learning using shared activities between subjects. Our research results are consistent with a study conducted by Gashaw and Liu [3] as well as research carried out by Jongkasikit [5] and Ashika et al [6]. These previous studies applied the association rules to the examination of data correlations. They also discovered that the FP-Growth algorithm could explain the data correlations.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Mathematics Statistics and Computer, Faculty of Science, Ubon Ratchathani University, and Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, for supporting and giving work opportunities.

REFERRENCES

1. Ministry of Education. National Education Act of B.E. 2542 and the Second and Third Amendments (B.E. 2545 and B.E. 2553), Relevant Laws, and Compulsory Education Act B.E. 2545. Bangkok: Office of the National Education Commission; 2007.
2. Dharmaraajan K, Dorairangaswamy MA. Analysis of FP-Growth and Apriori Algorithms on Pattern Discovery from Weblog Data. 2016 IEEE International Conference on Advances in Computer Applications (ICACA); 2016 Oct 24; Coimbatore, India: IEEE; 2017.
3. Gashaw Y, Liu F. Performance Evaluation of Frequent Pattern Mining Algorithms using Web Log Data for Web Usage Mining. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); 2017 Oct 14-16; Shanghai, China: IEEE; 2018.
4. Faza S, Rahmat RF, Nababan EB, Arisandi D, Effendi S. The association rules search of Indonesian university graduate's data using FP-growth algorithm. 10th International Conference Numerical Analysis in Engineering, Banda Aceh, Indonesia; 2017 Aug 24-25; Banda Aceh, Indonesia: IOP Conference Series: Materials Science and Engineering; 2018.
5. Jongkasikit NN. An Application of FP-Growth Algorithm to Find Factors in Choosing to Study in the Faculty of Industrial Technology Lampang Rajabhat University. Industrial Technology Lampang Rajabhat University Journal. 2018;11(2):29-39.
6. Ashika KJ, Jagruthi MN, Nikhil N, Pooja S. Educational data analytics using association rule mining for student job prediction. International Research Journal of Engineering and Technology (IRJET). 2020;7(6): 3979-84.
7. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems). 3rd ed. Morgan Kaufmann; 2011.
8. Sinsomboonthong S. An efficiency comparison in prediction of Khao Dok Mali 105 paddy rice classification with data mining techniques. Thai Science and Technology Journal (TSTJ). 2020;28(3): 394-402.

DOI: 10.60101/jarst.2023.253807