

# Association Rule Mining with Permutation for Estimating Students Performance and Its Smart Education System



Nongnuch Ketui<sup>1\*</sup>, Warawut Wisomka<sup>2</sup>, Kanitha Homjun<sup>3</sup>

<sup>1,2,3</sup> Computer Science, Faculty of Science and Agricultural Technology,  
Rajamangala University of Technology Lanna Nan, Thailand  
{nongnuchketui<sup>1</sup>, kanithahomjun<sup>3</sup>}@rmutl.ac.th, i-y<sup>2</sup>@hotmail.com

Received 15 December 2018; Revised 15 February 2019; Accepted 15 February 2019

**Abstract.** Nowadays, Education Data Mining is the business domain to help the quality of institute. This paper proposed Association Rule Mining (ARM) to discover the interesting patterns between various academic achievements in dataset. To examine the experiment, we used 17,875 academic achievements within 483 students. Experimental result shown that 248 rules are return at confidence 0.2 and support 0.7 while the performance of the rules with a new set equals to 76.00% After getting the rules, we implement a smart education system as the web application for guiding the students registration.

**Keywords:** apriori algorithm, association rule, education mining, permutation, smart education

## 1 Introduction

Data mining technology can apply in the education for improving learning such as the associating a knowledge gaps, the predicting the student achievement, and the discovering the suitable major. Educational data mining (EDM) deals with the application of data mining tools and techniques to inspect the data at educational institutions for deriving knowledge [1]. Association rules mining is one of the data mining technique which the interesting patterns are extracted. Association rules are considered both of a minimum support and a minimum confidence. Association rule generation consists of two steps: First, the given a frequent itemset is applied with a minimum support. Second, their frequent itemset rules are indicated by using a minimum confidence. Finally, using a lift value is related each of item in the association rules. Association rules will focus on rules of the type: if A then B where A and B can be particular items, values, words, etc. The relationship between items is measured by the support, confidence, and lift. All association rules in the association rule mining are considered with the condition of minimum support and minimum confidence. Two steps of association rules mining were proposed by Agarwal [2]; (1) finding the frequent itemsets as the join step and (2) comparing all the rules with the threshold values to prune with both of minimum support and confidence. The famous algorithms are Apriori and FP-Growth which are used to discovery the frequent itemsets. After getting the association rules, we apply the mathematical method called permutation to select the left-hand side subject with the the support, confidence, and lift. The permutation can order or arrange a set of things. Here, the association rules will duplicate subjects in the left-hand side. The association rules will be used to implement a smart education system for advisors and theirs students. This paper uses association rules to extract the student performance pattern with Apriori algorithm which will be helps the education institution to analyze the student performance and to improve it in order to provide good placement for the students. While the permuting method is used to select the good association rule and organize the output showing on the web-based application called the smart education application. The purpose of this work, we introduce data mining process in Section 3 and then describe the association rule mining and Apriori algorithm for discovering the rule patterns in Section 3.1 and 3.2. The permutation method is described in Section 4. Method and procedure are shown in Section 5. Experimental results and

---

\* Corresponding Author

discussions are described in Section 6. A smart education system and GUI are given in Section 7. Finally, conclusion and future work are represented in Section 8.

## 2 Related Work

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data. Educational Data Mining (EDM) is still in a part of interested mining. Many data mining approaches are used to study in EDM such as the classification algorithm, the association rule discovery, or the clustering. For the classification algorithm, many researchers reviewed that the decision tree algorithm is quite to use for predicting on student's academic performance [3]. They implemented various decision tree algorithms ID3, J48/C4.5, random tree, multilayer perception, rule based and random forest with Weka tool. To evaluate the performance percentage split method or cross validation method is used. The results could improve the student's academic achievement. In the research [4], they studied the the capabilities of information mining within the context of the next Education Institute and tries to get new specific information by applying data processing techniques to academic knowledge of Technological academic Institute of Athens. They used the decision tree algorithm to analyze the student's questionnaires within the categories inside the analysis method of every department of the Institute. Other hands, Ahmed and his team [5] focused on the classification models to predict the final grade of students. They selected the decision tree (ID3) method for their experiment. Including to the research [6] of Kaur and his team, they studied the classification and prediction based data mining algorithms to predict slow learners in education sector. They applied on various classification algorithms such as Multilayer Perception, Naive Bayes, SMO, J48 and REPTree using WEKA potential tool. The latest research is created by the new framework, called cloud-service decision tree classification for education platform [7]. This research studied the ideological education model of university students, provided the multi-objective weight self-adaptation form, and introduced the cloud-service decision tree classification algorithm. The measures are used to verify the performance of algorithms. Another the data mining approach, the association rule discovery is the one choice to study the relation between items. Data mining is used to apply in various research works. One famous technique for extracting the pattern discovery is called the association rule [8-9]. According to researchers [10], the association rules were extracted from the relationships among the large datasets. They implemented on the marketing, finance, medical and many database. The association rules approach discovered all rules by determining the minimum support and minimum confidence [11]. Agrawal and her team [2] firstly implemented the Apriori algorithm to mine single-dimensional Boolean association rules from transactional databases. Then, this algorithm could describe the rule generation to be easily and understand [12]. They used the Apriori algorithm to generate the frequent item sets and designed the model for economic forecasting. In [13-14], theirs researches were applied to discover association rules and then the discovered association rules were sorted according to the lift value. The minimum support and minimum confidence were used to consider the relation rules. In the education data mining, the recommender system technique was used for predicting the performance of the students. This technique mainly focused on reducing the information overload and act as information filters [15]. In case study of [16] discussed various EDM techniques to improve students' performance. This dataset consisted of 3360 record and 18 attribute and evaluated with various techniques such as association rule mining, classification, outlier detection, and clustering to identify the various factors that affects various performance. Many researchers tried to build advising systems to facilitate the process of registration for students. Two processes were implemented firstly, the student registered theirs course following the advisor suggestion [17]. Second, the interactive between the students provided his/her preferences and the advisor builds his/her recommendations based on these preferences [18]. This type was more common and usually provided better outcomes compared to the first one. In [19] developed a web-based advising system for undergraduate students. This system took as input students' preferences and builds its recommendations upon these preferences. The students input their preferences and then got their recommendations from advisors via a web browser. In another application to build a smart advising system, [20] developed and designed a prototype for a rule-based expert advising system with an object oriented database. Finally, the system used association rule mining to find associations between courses that have been registered by students in many previous semesters [21].

### 3 Data Mining

Data Mining (DM) or Knowledge Discovery in Database (KDD) is a process to find patterns and relation from large data sets with the various methods such as machine learning, statistics, and database systems [11]. There are three processes of data mining as the described following:

(1) Preprocessing: prepare the suitable dataset and will arrange in a useful format. A noise data should be eliminated from raw dataset. The remaining dataset which having a relation, will be used for creating a model.

(2) Modeling: use the selected data for creating a model. A problem or business understanding is guided to choose the appropriate technique. Then, the model should be evaluated with the new cases.

(3) Postprocessing: deploy the model to the real situation as the application.

#### 3.1 Association Rule Mining

Association rule mining is a rule-based machine learning method for discovering interesting relations between variables in large databases. The measurements are used to identify the strong rules which discovered in the large dataset [12]. In order to select interesting rules from the possible rules set, constraints on various measures of significance and interest are used. The support, confidence, and lift values are the famous conditions are set minimum thresholds. Let  $L$  be an itemset,  $L \rightarrow R$  an association rule and  $T$  a set of transactions of a given database

(1) Support measures the frequency of itemset occurs in the dataset.

$$\text{sup}(L) = \frac{\text{freq}(L)}{\text{freq}(T)}. \quad (1)$$

$$\text{sup}(R) = \frac{\text{freq}(R)}{\text{freq}(T)}. \quad (2)$$

Here,  $\text{freq}(L)$  is the frequency of item  $L$  (left-hand side) which occurring in all transaction  $\text{freq}(T)$  and  $\text{freq}(R)$  is the frequency of item  $R$  (right-hand side) which occurring in all transaction  $\text{freq}(T)$ .

(2) Confidence indicates the rule has been often found to be true.

$$\text{conf}(L \rightarrow R) = \frac{\text{sup}(L, R)}{\text{sup}(L)}. \quad (3)$$

Here,  $\text{sup}(L, R)$  is the frequency of itemset  $L$  and  $R$  which occurring together.

(3) Lift is the ratio of the observed support to that expected if  $L$  and  $R$  were independent.

$$\text{lift}(L \rightarrow R) = \frac{\text{sup}(L, R)}{\text{sup}(L) \times \text{sup}(R)}. \quad (4)$$

Here,  $\text{sup}(L)$  is the radio of the frequency of item  $L$  (left-hand side) and all transaction  $\text{freq}(T)$  and  $\text{sup}(R)$  is the radio of the frequency of item  $R$  (right-hand side) and all transaction  $\text{freq}(T)$ .

The association rules can be generated by (1) to find the frequent of all itemsets (frequent itemsets) and (2) discovery the relation rule from the frequent itemsets considering the support, confidence, and lift values.

#### 3.2 Apriori Algorithm

Apriori [2] is an algorithm for considering frequent itemsets mining over transactional datasets. It proceeds by identifying the frequent itemsets in database and extending them to larger itemsets until those item sets appear sufficiently often in the database. The Apriori algorithm use k-itemsets for finding (k + 1)-itemsets candidate  $C_k$  under the minimum support condition  $L_k$  and then iterating from 1-itemsets through transactional database until the frequent itemsets cannot be found. Then, the Apriori algorithm can be used to determine the frequent itemsets to be association rules (see the process in Fig. 1).

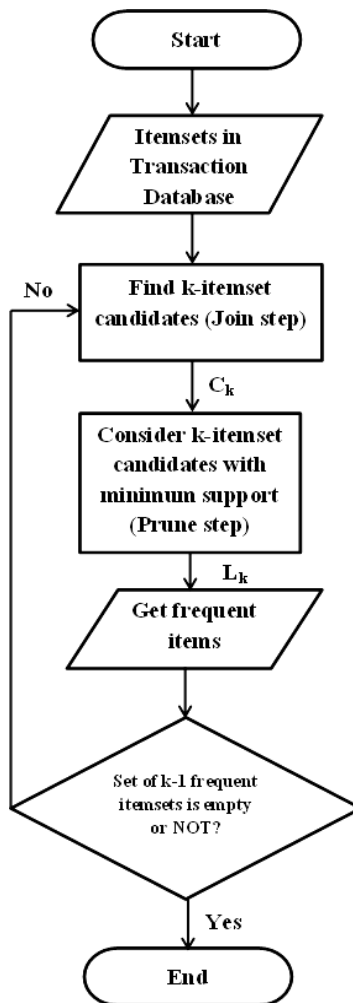


Fig. 1. The process of apriori algorithm

#### 4 Permutation Method

A permutation method is a part of mathematical technique which arranging the members of a set into some order.

This process is called permuting. For example, there are six permutations of the set {A, B, and C} or three members. We can calculate the number of permutation set with  $n!$ . See the output below;

- (1)  $A \rightarrow B \rightarrow C = ABC$
- (2)  $A \rightarrow C \rightarrow B = ACB$
- (3)  $B \rightarrow A \rightarrow C = BAC$
- (4)  $B \rightarrow C \rightarrow A = BCA$
- (5)  $C \rightarrow A \rightarrow B = CAB$
- (6)  $C \rightarrow B \rightarrow A = CBA$

Here,  $n$  is the number of members. If we need to rearrange each of member in one by one and  $r$  is represented the number of arrangement, the other calculation equals to  $\frac{n!}{(n-r)!}$ . If  $r$  equals to 2, we can

see the permuting output as

- (1)  $A \rightarrow B = AB$
- (2)  $A \rightarrow C = AC$
- (3)  $B \rightarrow A = BA$
- (4)  $B \rightarrow C = BC$
- (5)  $C \rightarrow A = CA$
- (6)  $C \rightarrow B = CB$

## 5 Method and Procedure

The method in this experiment, we focus on the data mining process such as preprocessing, modeling, and postprocessing which are described below:

### 5.1 Preprocessing

**Data cleaning.** 17,875 records within 483 students examine the performance of academic achievement; we only focus on the students of Faculty of Science and Agricultural Technology (see in Table 1).

**Table 1.** A number of dataset divided by department

| Dept.Id | Dept.Name                        | Student (person) |
|---------|----------------------------------|------------------|
| 11      | B.Sc.Animal Science              | 100              |
| 21      | B.Sc.Plant                       | 88               |
| 23      | Dip.Landscape Technology         | 12               |
| 31      | B.Sc.Fisheries                   | 54               |
| 61      | B.Sc.Science and Food Technology | 64               |
| 90      | B.Sc.Computer Technology         | 140              |
| 92      | B.Sc.Computer Science            | 25               |
| Total   |                                  | 483              |

**Data selection.** 8 attributes such as year, semester, subject id, subject name, department id, department name, grade, and credit are used. However, the experimental dataset only selects the records which having the active status (“Y”). Grade “I” is eliminated while the missing values are set to “N/A”. Table 2 is shown an example of selection dataset.

**Table 2.** A Example of Transformed Dataset

| Stu.No | Year/Sem. | Subject Id | Subject Name      | Dept.Id | Grade | Credit | Transformed Grade |
|--------|-----------|------------|-------------------|---------|-------|--------|-------------------|
| 1      | 1/2009    | 01210001   | Writing Report    | 11      | B     | 3      | High              |
| 1      | 1/2009    | 01320101   | English I         | 11      | A     | 3      | High              |
| 2      | 1/2009    | 01610001   | Physical Exercise | 21      | A     | 1      | High              |
| 2      | 2/2009    | 03222102   | Animal Production | 21      | D+    | 3      | Low               |
| 2      | 2/2009    | 13011131   | Calculus I        | 21      | D     | 3      | Low               |
| 3      | 2/2009    | 13020101   | General Chemistry | 31      | C     | 3      | Middle            |

**Data transformation.** A to F grade level are changed into three group i.e., “High” equals A, B+, or B level, “Middle” displays C+ or C, and “Low” represents D+, D, or F (see in Table 2). Then, we prepare the dataset in a excel file and convert it to .arff format for Weka environment [14]. The transformed data is shown in the last column of Table 3.

**Table 3.** A example of dataset for analyzing in weka enviornment

| Stu.No | Course Id |          |          |          |          |          |          |          |
|--------|-----------|----------|----------|----------|----------|----------|----------|----------|
|        | 01210001  | 01320101 | 01610001 | 03222102 | 13011131 | 13020101 | 13041101 | 20000101 |
| 1      | H         | H        | H        | L        | L        | L        | M        | H        |
| 2      | M         | M        | H        | L        | M        | L        | M        | M        |
| 3      | H         | L        | H        | M        | L        | L        | M        | H        |
| 4      | M         | L        | M        | L        | L        | L        | L        | M        |
| 5      | M         | M        | H        | L        | M        | L        | M        | M        |
| ...    | ...       | ...      | ...      | ...      | ...      | ...      | ...      | ...      |

### 5.2 Modeling

The association rule model is generated by the student achievement of subjects in each student (see in Table 3) and we can describe one example as following:

$$[01210001 = H] + [01320101 = H] \rightarrow [01610001 = H]$$

The above relation, if the student studies the subject ‘01210001’ which getting the grade in high level and learns the subject “01320101” with high level (“A, B+, or B”), then he/she is possible to achieve grade A, B+ or B in the subject “01610001”. This relation of student academic achievement is up to the student performance. We need to find the association rule with the minimum support and confidence of all dataset. The output of association rule discovery is shown in Fig. 2. The association rules from this experiment, 520 rules are returned so we need to remove some rules with conditions as following:

- (1) Set the minimum support and minimum confidence to consider the frequency itemset is quite high and the confidence is too suitable to students.
- (2) Eliminate the ordered subject which is occurred in the curriculum plan and select the left-hand side of subject that students have studied and gotten theirs grade already.
- (3) Use the permutation to reduce the number of association rules for displaying and finding the related subject (described in the next subsection).

|                   |               |               |               |               |               |               |             |             |              |              |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|-------------|-------------|--------------|--------------|
| 1. 1.3022001E7=H  | 2.2031102E7=H | 2.2034202E7=H | 151 ==>       | 2.0000101E7=H | 147           | <conf:(0.97)> | lift:(2.19) | lev:(0.17)  | [79]         | conv:(16.76) |
| 2. 1.3022001E7=H  | 2.2034202E7=H | 176 ==>       | 2.0000101E7=H | 170           | <conf:(0.97)> | lift:(2.17)   | lev:(0.19)  | [91]        | conv:(13.95) |              |
| 3. 2.2031102E7=H  | 2.2034202E7=H | 161 ==>       | 2.0000101E7=H | 154           | <conf:(0.96)> | lift:(2.15)   | lev:(0.17)  | [82]        | conv:(11.17) |              |
| 4. 2.0000101E7=H  | 2.2031102E7=H | 2.2034202E7=H | 154 ==>       | 1.3022001E7=H | 147           | <conf:(0.95)> | lift:(1.87) | lev:(0.14)  | [68]         | conv:(9.45)  |
| 5. 2.0000101E7=H  | 2.2034202E7=H | 179 ==>       | 1.3022001E7=H | 170           | <conf:(0.95)> | lift:(1.86)   | lev:(0.16)  | [78]        | conv:(8.78)  |              |
| 6. 2.0000101E7=H  | 2.2031102E7=H | 175 ==>       | 1.3022001E7=H | 165           | <conf:(0.94)> | lift:(1.85)   | lev:(0.16)  | [75]        | conv:(7.81)  |              |
| 7. 2.2031102E7=H  | 2.2034202E7=H | 161 ==>       | 1.3022001E7=H | 151           | <conf:(0.94)> | lift:(1.84)   | lev:(0.14)  | [69]        | conv:(7.18)  |              |
| 8. 2.0000101E7=H  | 215 ==>       | 1.3022001E7=H | 198           | <conf:(0.92)> | lift:(1.81)   | lev:(0.18)    | [88]        | conv:(5.86) |              |              |
| 9. 2.2031102E7=H  | 2.2034202E7=H | 161 ==>       | 1.3022001E7=H | 2.0000101E7=H | 147           | <conf:(0.91)> | lift:(2.23) | lev:(0.17)  | [81]         | conv:(6.33)  |
| 10. 2.2034202E7=H | 199 ==>       | 2.0000101E7=H | 179           | <conf:(0.9)>  | lift:(2.02)   | lev:(0.19)    | [90]        | conv:(5.26) |              |              |
| 11. 1.3022001E7=H | 2.0000101E7=H | 2.2031102E7=H | 165 ==>       | 2.2034202E7=H | 147           | <conf:(0.89)> | lift:(2.16) | lev:(0.16)  | [79]         | conv:(5.11)  |
| 12. 2.2034202E7=H | 199 ==>       | 1.3022001E7=H | 176           | <conf:(0.88)> | lift:(1.74)   | lev:(0.15)    | [74]        | conv:(4.07) |              |              |
| 13. 2.0000101E7=H | 2.2031102E7=H | 175 ==>       | 2.2034202E7=H | 154           | <conf:(0.88)> | lift:(2.14)   | lev:(0.17)  | [81]        | conv:(4.68)  |              |
| 14. 1.3022001E7=H | 2.0000101E7=H | 2.2034202E7=H | 170 ==>       | 2.2031102E7=H | 147           | <conf:(0.86)> | lift:(1.57) | lev:(0.11)  | [53]         | conv:(3.18)  |
| 15. 2.0000101E7=H | 2.2034202E7=H | 179 ==>       | 2.2031102E7=H | 154           | <conf:(0.86)> | lift:(1.56)   | lev:(0.11)  | [55]        | conv:(3.09)  |              |
| 16. 1.3022001E7=H | 2.0000101E7=H | 198 ==>       | 2.2034202E7=H | 170           | <conf:(0.86)> | lift:(2.08)   | lev:(0.18)  | [88]        | conv:(4.01)  |              |
| 17. 1.3022001E7=H | 2.2034202E7=H | 176 ==>       | 2.2031102E7=H | 151           | <conf:(0.86)> | lift:(1.56)   | lev:(0.11)  | [54]        | conv:(3.04)  |              |
| 18. 2.2034202E7=H | 199 ==>       | 1.3022001E7=H | 170           | <conf:(0.85)> | lift:(2.08)   | lev:(0.18)    | [88]        | conv:(3.91) |              |              |
| 19. 2.0000101E7=H | 2.2031102E7=H | 175 ==>       | 1.3022001E7=H | 2.2034202E7=H | 147           | <conf:(0.84)> | lift:(2.31) | lev:(0.17)  | [83]         | conv:(3.84)  |
| 20. 1.3022001E7=H | 2.2034202E7=H | 176 ==>       | 2.0000101E7=H | 2.2031102E7=H | 147           | <conf:(0.84)> | lift:(2.31) | lev:(0.17)  | [83]         | conv:(3.74)  |
| 21. 1.3022001E7=H | 2.2031102E7=H | 198 ==>       | 2.0000101E7=H | 165           | <conf:(0.83)> | lift:(1.87)   | lev:(0.16)  | [76]        | conv:(3.23)  |              |
| 22. 1.3022001E7=H | 2.0000101E7=H | 198 ==>       | 2.2031102E7=H | 165           | <conf:(0.83)> | lift:(1.51)   | lev:(0.12)  | [55]        | conv:(2.62)  |              |
| 23. 2.0000101E7=H | 215 ==>       | 2.2034202E7=H | 179           | <conf:(0.83)> | lift:(2.02)   | lev:(0.19)    | [90]        | conv:(3.42) |              |              |
| 24. 2.0000101E7=H | 2.2034202E7=H | 179 ==>       | 1.3022001E7=H | 2.2031102E7=H | 147           | <conf:(0.82)> | lift:(2)    | lev:(0.15)  | [73]         | conv:(3.2)   |
| 25. 2.0000101E7=H | 215 ==>       | 2.2031102E7=H | 175           | <conf:(0.81)> | lift:(1.48)   | lev:(0.12)    | [56]        | conv:(2.36) |              |              |
| 26. 2.2034202E7=H | 199 ==>       | 2.2031102E7=H | 161           | <conf:(0.81)> | lift:(1.47)   | lev:(0.11)    | [51]        | conv:(2.29) |              |              |
| 27. 1.3022001E7=H | 246 ==>       | 2.0000101E7=H | 198           | <conf:(0.8)>  | lift:(1.81)   | lev:(0.18)    | [88]        | conv:(2.79) |              |              |
| 28. 1.3022001E7=H | 246 ==>       | 2.2031102E7=H | 198           | <conf:(0.8)>  | lift:(1.46)   | lev:(0.13)    | [62]        | conv:(2.26) |              |              |

Fig. 2. The relative rules from association rule model

### 5.3 Postprocessing

In this process, the model is used to predict the student academic achievement by considering from the registered subjects and their grade level together with the association rules. The left-hand side rule is previously the registered subject and its grade while the right-hand side rule is the predicted academic achievement of each subject.

For example in the experiment, if we need to predict the academic achievement of student who will register the subject “20000101”, the association rules can displayed based on the previous dataset as following:

- (1) [13062016 = H] → [20000101 = H] confidence = 1.00 and lift = 2.25
- (2) [13062016 = H] + [13022001 = H] → [20000101 = H] confidence = 1.00 and lift = 2.43
- (3) [13031101 = L] + [22034202 = H] → [20000101 = H] confidence = 0.99 and lift = 2.23

From the above result, all rules can predict the academic achievement of subject “20000101” in the various results. So that, the permutation method is used to consider all subject in the left-hand side (n = 4).

So that, n = {13062016 : H; 13022001 : H; 13031101 : L; 22034202 : H}.

We apply the mathematics as the permutation method to select the appropriate rules. First, this method iterates from the selected subject r = 1 to n. Then, the calculation with formula  $\frac{n!}{(n-r)!}$  (see in Section 4).

In this case, the experiment shows that the maximum subject in the left-hand side equals to 4 so that the iteration of process is set to p(4; 1); p(4; 2); p(4; 3); p(4; 4). The permuting calculation, the number of candidate rules of p(4; 1); p(4; 2); p(4; 3); p(4; 4) equals to 4, 12, 24, and 24 respectively. Then, the process iterates all candidate rules in the list of association rule to get their confidence and lift values.

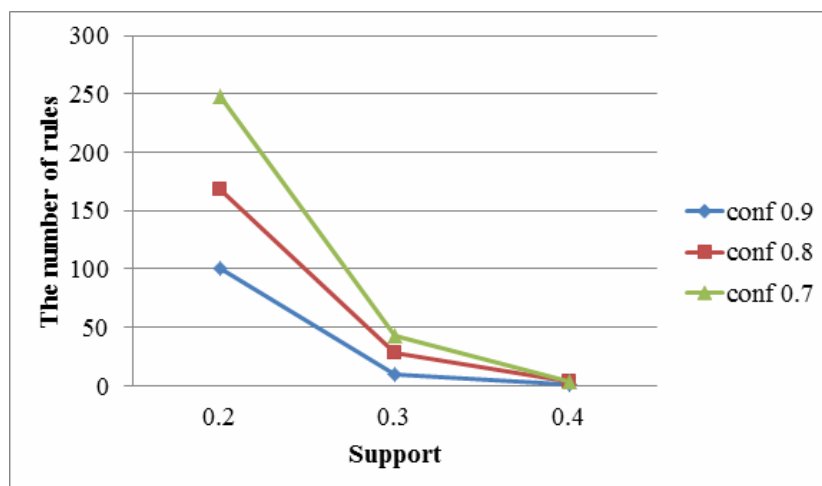
Second, we find the maximum confidence values in each candidate rules. If the maximum confidence is the same, the lift value is used to consider. After calculating the permutation method, the most related rule is the second rule ([13062016 = H] + [13022001 = H] → [20000101 = H]).

## 6 Result and Discussion

In the data mining analysis, the range of confidence and support is between 0.1 and 1.0. To consider the confidence, support and the number of rules in Table 4, we found that the minimum support at 0.2, 0.3, 0.4 and the minimum confidence at 0.7 could appear the association rules. In the opposite of the minimum support at 0.5, the association rule is not shown. Another, the association rules are the lowest at the minimum confidence at 0.7, 0.8, and 0.9 and the minimum support. Fig. 3 shows the comparison of the number of association rules in each confidence and support value. The line graph displays that the number of rules will decrease when the minimum support values increase. To evaluation the model, we tested with thirty students from B.Sc. Animal Science, B.Sc. Fisheries, and B.Sc. Computer Technology (10 persons per department). The result shows that the average performance of model gets 0.75 (from B.Sc. Animal Science at 0.76, B.Sc. Fisheries at 0.72, and B.Sc. Computer Technology 0.76). However, this model is the supporter or guideline to advisors and students for planning the registration in the future. If the output is predicted in the group D+, D, or F, the students should concentrate their class. Otherwise, if the predicted grade appears in high level (A, B+, or B), the students may neglect those subjects. The main objective of our research can help the performance of student to be increase.

**Table 4.** The association rule result

| Confident | Support | Rules |
|-----------|---------|-------|
| 0.7       | 0.2     | 248   |
| 0.7       | 0.3     | 43    |
| 0.7       | 0.4     | 4     |
| 0.8       | 0.2     | 168   |
| 0.8       | 0.3     | 28    |
| 0.8       | 0.4     | 3     |
| 0.9       | 0.2     | 101   |
| 0.9       | 0.3     | 9     |
| 0.9       | 0.4     | 1     |



**Fig. 3.** Comparison of the number of rules in each confidence

## 7 A Smart Education System and GUI

In this work, we have developed the smart education application<sup>1</sup> with the concept of data mining with the student academic achievement, selecting the appropriate rules with the famous mathematics as the permutation, and predicting the performance of academic achievement with the association rule discovery. For our implementation, the following software and hardware requirement are used. The system as web-based application was constructed by PHP (Personal Home Page) programming. A MySQL database was used for storing original exams and the statistic results. We edited all programs with Notepad++ and designed the database to support the association rules. The maximum of left-hand side equals to four whereas we set the attribute name L1, L2, L3, and L4 to store the registered subject and create attribute grade L1, grade L2, grade L3, and grade L4 to save the academic performance (High, Middle, and Low). The attribute R1 and grade R1 are used to store the right-hand side of rules. Each of the association rules indicates the confidence and lift values. The system can run as the web-based application. In the system of smart education application, the student academic achievement can be gathered from register data. Fig. 4 shows the user interface (UI) snapshots when the system allows users (lecturers or students) to select subjects and the academic achievement. The number in the Figure 4 describe the detailed as

- No.1: Select the subject that registered in the previous semester.
- No.2: Add the subject if you need to register more than one subject with its performance.
- No.3: Click button to predict the student academic achievement.
- No.4: Click button to clear the entry subject and put the new subject.
- No.5: Display the predicted result and its confidence and lift in the percentage format.



Fig. 4. The main smart education system

## 8 Conclusion and Future Work

In our work, we applied association rules to extract the student performance pattern with the Apriori algorithm for analyzing the student performance and improve it in order to providing a good achievement for the students. To use 17,875 records of student registration in Faculty of Science and Agricultural Technology, RMUTL Nan. The permutation method as the mathematical approach is used to reduce the association rules model. The experimental results showed that 520 association rules are generated and

<sup>1</sup> <http://sat.nan.rmutl.ac.th/arm/index.php>



applied to implement the smart education system as the web-based application. In the future works, we will evaluate with other data mining technique for improving the student performance.

## Acknowledgements

We would like to thank registration staffs at Education Division of University for preparing the raw dataset. We give special thanks to Computer Center of University for managing the database server in order to clapping my hand to lecturers and students at Science department, RMUTL Nan for all kind helps.

## References

- [1] M. Al-Razgan, A.S. Al-Khalifa, H.S. Al-Khalifa, Educational data mining: a systematic review of the published literature 2006- 2013, in: Proc. 2013 the 1st International Conference on Advanced Data and Information Engineering, 2013.
- [2] R.-H. Agrawal, R. Srikant, Fast algorithms for mining association rules. in: Proc. 1994 the 20th International Conference on Very Large Data Bases, VLDB, 1994.
- [3] K. Hardeep, A literature review from 2011 TO 2014 on student's academic performance prediction and analysis using decision tree algorithm, Journal of Global Research in Computer Science 9(5)(2018) 10-15.
- [4] NV-K. Rao, M. Rakesh, S. Cherukurada, J. Thirupathi, B. Madhuravani, A review on data mining approach used in education data mining using decision tree algorithm, Journal of Advanced Research in Dynamical and Control Systems 7(Special Issue)(2018) 1735-1738.
- [5] A.-B. El Din Ahmed, I.S. Elaraby, Data mining: a prediction for student's performance using classification method, World Journal of Computer Application and Technology 2(2)(2014) 43-47.
- [6] P. Kaur, S. Manpreet, S. J. Gurpreet, Classification and prediction based data mining algorithms to predict slow learners in education sector, Procedia Computer Science 57(2015) 500-508.
- [7] C. Wang, J. Wang, Cloud-service decision tree classification for education platform, Cognitive Systems Research 52(2018) 234-239.
- [8] J. Hipp, U. Guntzer, N. Gholamreza, Algorithm for association rule mining: a general survey and comparison, ACM SIGKDD Explorations Newsletter 2(1)(2000) 58-64.
- [9] D. Magdalene, D. Angeline, Association rule generation for student performance analysis using apriori algorithm, The SIJ Transactions on Computer Science Engineering and Its Applications 1(1)(2013) 12-16.
- [10] U.-M. Fayyad, P.-S. Gregory, Eds., Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, Cambridge, 1996.
- [11] B. Liu, W. Hsu, Y. Ma, Integrating classification and association rule mining, in: Proc. the Fourth International Conference on Knowledge Discovery and Data Mining, 1998.
- [12] Y. Ma, B. Liu, C. Wong, P.-S. Yu, S.M. Lee, Targeting the right students using data mining, in: Proc. 2000 International Conference on Knowledge discovery and Data Mining, Boston, 2000.
- [13] A.M. El-Halees, Mining students data to analyze learning behavior: a case study. <[https://www.researchgate.net/publication/228571634\\_Mining\\_Students\\_Data\\_to\\_Analyze\\_Learning\\_Behavior\\_A\\_Case\\_Study](https://www.researchgate.net/publication/228571634_Mining_Students_Data_to_Analyze_Learning_Behavior_A_Case_Study)>, 2009.
- [14] N. Kittiphattanabawon, T. Theeramunkong, Relation discovery from Thai news articles using association rule mining, in: H. Chen, C.C. Yang, M. Chau, S.H. Li (Eds.), Intelligence and Security Informatics, PAISI 2009, Lecture Notes in Computer

- Science, Vol. 5477 Springer, Berlin, 2009, pp. 118-129.
- [15] N. Thai-Nghe, Factorization techniques for predicting student performance, *Educational Recommender Systems and Technologies: Practices and Challenges* (2011) 129-153.
- [16] M.-M. Abu Tair, A.-M. El-Halees, Mining educational data to improve students' performance: a case study, *International Journal of Information and Communication Technology Research* 2(2)(2012) 140-146.
- [17] B.-B. Crookston, A developmental view of academic advising as teaching, *Journal of College Student Personnel* 13(1)(1972) 12-17.
- [18] C.-M. Chando, Predicting advising style preference from student characteristics, [dissertation] Memphis: University of Memphis, 1997.
- [19] S. Hsu, O. Marques, M. Ilyas, X. Ding, Web-based undergraduate academic advising system, in: *Proc. International Conference on Engineering Education*, 2002.
- [20] M. Al Ahmar, A prototype student advising expert system supported with an object-oriented database. <<https://thesai.org/Publications/ViewPaper?Volume=1&Issue=3&Code=SpecialIssue8SerialNo=16>>, 2011.
- [21] R. Shatnawi, Q. Althebyan, B. Ghalib, M. Al-Maolegi, Building a smart academic advising system using association rule mining. <<http://arxiv.org/abs/1407.1807>>, 2014.