

# Development of Guidelines for the Academic Support of Students by Using the A priori Algorithm

Sumitra Nuanmeesri<sup>1</sup> and Wongkot Sriurai<sup>2</sup>

<sup>1</sup>Suan Sunandha Rajabhat University, Faculty of Science and Technology, Bangkok, Thailand; sumitra.nu@ssru.ac.th

<sup>2</sup>Ubon Ratchathani University, Ubon Ratchathani, Thailand; wongkot.s@ubu.ac.th

## Abstract

**Objective:** To examine the relationship of courses and tendency to attain weak study outcome to assist and protect students from school expulsion in higher education. **Statistical Analysis:** The data used for operating the Apriori algorithm in association rule was the grade points of the students from the Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, who each studied 23 courses in the general education category and the information technology category in total. This data had been recorded between 2011 and 2016 (3,200 records).

**Findings:** The accuracy of thirty association rules with test data of 1,200 student records was real study outcome of all students from academic year of 2017 to 2018. Data from these 2 academic years were not used to formulate the rules. According to evaluation of rules' efficiency defined by accuracy was 89%. **Application/Improvement:** The association rules to provide, assistance, approaches and mentorship to students with risk via online social media. These students would be able to adapt themselves and attained better study outcome.

**Keywords:** Academic Support, Apriori Algorithm, Educational, Prediction

## 1. Introduction

Education focuses on human development, especially learning management for those who are becoming adults, to be able to develop their society and their nation together. In regard to the aforementioned statement, there are three essential elements of higher education learning management which consists of individuals, knowledge, and the society<sup>1</sup>. In Thailand, education is provided by the Ministry of Education. The public sector has direct control over it and also provides an opportunity to the private sector to participate in primary education and higher education provision. Higher education is a significant educational level because it contributes to the students' future careers. The graduates should have qualities that entrepreneurs and society expect so every higher educational institution should work to enhance student performance. However, the road to graduation always

involves some obstacles. There are a number of low proficiency students who cannot graduate on time with some students dropping out of higher education. Therefore, most of the educational institutions are expected to help these low proficiency students by offering tutorial classes and academic advice.

According to "Factors Affecting the Dropping out of Rajamangala University of Technology Phra Nakhon Students" written<sup>2</sup> most of the factors affecting student drop out numbers were personal and family issues, the learning environment in the educational institution, and the teaching academics. Thus, the suggested solution is to develop a student support and advisory management system. From "The Use of Association Rule to Find the Factors Affecting the Risk of Low Proficiency Students" written<sup>3</sup> the finding shows that the factors affecting the low proficiency students were their approach to entry and their admission scores from different subjects. From

\*Author for correspondence

the aforementioned research, the researchers found that the association rule is a technique to find the relationships between each piece of data<sup>2</sup>. This research applied the association rule to examine the relationships between these courses that tend to cause low proficiency students<sup>4,5</sup>. The courses used for creating the association rule were the courses from the general education category and the information technology category. All of these courses were the courses under the Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University. The researchers had collected the data from the registration and assessment system as well as the education service division. The results found that the relationships between the aforementioned courses indicated that the courses had impacted the low proficiency students. This data is highly helpful in terms of implementing courses to better support students.

## 2. Theoretical Background

### 2.1 Association Rule

The association rule is one of the techniques for data mining. It analyzes the relationships between each piece of information embedded in a large database, finding frequent patterns, and predicts a phenomenon. An example of the association rule application is the market basket analysis; the association rule can find relationships between the customers and the products that they purchased. This purchase product data can support the market planning, for example, putting the frequently bought products on the same shelves for developing and implementing sales promotions.

The general forms of the association rule are “ $X \rightarrow Y$ . X is the condition; Y is the result.” and “If.... Then...” in which the variable X and Y represent the sets or groups of products (so-called the ‘item set’ in which each item represents a product) and the arrow symbol represents the relationship between the product sets. For example,  $X \rightarrow Y$  (s, c) means when customers buy the product X, they also buy the product Y. ‘S’ symbol represents the support value; it is used for measuring the purchase frequency of X and Y products. This can be calculated by the following equation<sup>4</sup>:

$$\text{Support } (X \rightarrow Y) = \frac{\text{Number of transactions consisting of } X \text{ and } Y \text{ at the same time}}{\text{Number of all transactions}}$$

‘C’ symbol represents the confidence value. This value refers to the accuracy or reliability of this rule. This can be calculated by the number of frequencies of item X and Y’s appearances shown in all transactions. X can be calculated by the following equation<sup>4</sup>:

$$\text{Confidence } (X \rightarrow Y) = \frac{\text{Number of transactions consisting of } X \text{ and } Y \text{ at the same time}}{\text{Number of } X \text{ transactions}}$$

After identifying the support and confidence values to measure the frequency and reliability of the rule, there are two following steps to examine the relationships according to the association rule as described below:

1. Find all of the frequent item sets; however, the frequent item sets should not have a support value lower than the minimum support level, and
2. Generate the association rule from the frequent item sets; however, this rule should not have a confidence value lower than the minimum confidence level.

There are a number of algorithms available for the association rule, but the most popular algorithms are Apriori<sup>6,7</sup> and FP-growth<sup>7</sup>. In this research, the researchers decided to use the Apriori algorithm to create the association rule.

### 2.2 Apriori Algorithm

The Apriori is a popular algorithm used for finding the information based on the association rule<sup>8</sup>. Firstly, it has to explore the relationships in the macro level then counts the transactions by identifying items and examining item sets that appear at each stage. It starts from the item sets that have only one member. If any itemset has a support value lower than the minimum support level, that itemset will be cut out and it will not be used at the next stage. The operation of this algorithm will continuously repeat until it covers every stage, or there is no other itemset left for the next stage. To count the number of transactions, Apriori will examine the transactions only once at each stage. It will examine which item sets that the transactions contain. The prominent point of this algorithm is the speed of finding the frequent item sets as it will not consider the item sets that have a low frequency when compared with the standard. The operation of Apriori algorithm includes the following stages:

- First: generate the candidate item sets starting from the candidate item sets that have a size equal to k-item sets in which  $k = \{1, 2, 3, \dots, n\}$ .

The candidate item sets, found at the k level, came from the join between large item sets at the k-1 level. For example, the itemset {ABC} is resulted from the itemset {AB} joined with the item set {AC}, and

- Second: find the large item sets at k level by firstly reading the data in the database to count the support value of each candidate itemset at k level. Next, check which candidate item sets have a support value more than or equal to the minimum support level, since those candidate item sets have the potential to become large item sets at the k level. Then, generate the candidate item sets to find the large item sets at the k+1 level. This operation will repeat until it cannot generate the candidate item sets from the k+1 level anymore.

### 3. Materials and Methods

The researchers divided the research methodology into three main stages consisting of: 1. research and data collection, 2. information preparation, and 3. the association rule operation (Figure 1).

#### 3.1 Research and Data Collection

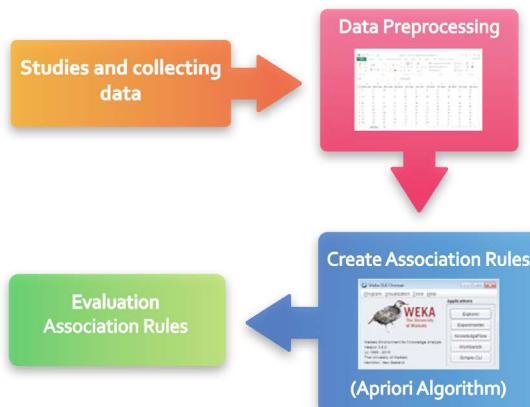
The researchers had studied the data of the students in the Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University. The information used in the application of the association rule were the grade points from the courses under the general education category and the information technology category 23 courses in total (3,200 records). These were the students' grade points accumulated between 2011 and 2016 in the Department of Information

**Table 1.** The data of the courses for operating the association rule

| No. | Courses   | Assessment  |
|-----|---|---|
| 1   | Aesthetic Appreciation (GEH1101)                        | - The assessment takes on the form of letter grades: A, A-, B+, B, B-, C+, C, C-, D+, D, D-, and F. |
| 2   | Thai Society in Global Context (GEH1102)                |   |
| 3   | Thai for Academic Purpose (GEL2201)                     |   |
| 4   | Truth of Life (GEH2202)                                 |   |
| 5   | Computer Programming Development (INT1104)              |   |
| 6   | Data Structure and Algorithm (INT1107)                  |   |
| 7   | Computing Platform Technology (INT1402)                 |   |
| 8   | Computer Operating System (INT1403)                     |   |
| 9   | Object Orientation System Analysis and Design (INT1503) |   |
| 10  | Object Orientation Programming Development (INT2104)    |   |
| 11  | Database System Design (INT2202)                        |   |
| 12  | Network Technology and Internet (INT2302)               |   |
| 13  | Management and Design Network System (INT2303)          |   |
| 14  | Information Security (INT2304)                          |   |
| 15  | Data Communication and Networks (INT2305)               |   |
| 16  | Web System and Technology (INT2401)                     |   |
| 17  | Mobile Application Development (INT3103)                |   |
| 18  | Applied Computer Graphics (INT3107)                     |   |
| 19  | Multimedia Technology (INT3110)                         |   |
| 20  | English for Research Information Technology (INT3513)   |   |
| 21  | Electronic Commerce (INT3514)                           |   |
| 22  | Fundamental of Mathematics (MAT1201)                    |   |
| 23  | Statistics for Research (STA2101)                       |   |

Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University. The researchers collected this data from the registration and assessment system and the education service division. The data is decisively in detail as follows:

1. There were four courses under the general education category which are:
  - Aesthetic Appreciation,
  - Thai Society in the Global Context,
  - Thai for Academic Purpose, and
  - Truth of Life.
2. There were nineteen courses under the information technology category which are:
  - Computer Programming Development,
  - Data Structure and Algorithm,
  - Computing Platform Technology,
  - Computer Operating System,
  - Object Orientation System Analysis and Design,
  - Object Orientation Programming Development,
  - Database System Design,
  - Network Technology and Internet,
  - Management and Design Network System,
  - Information Security,
  - Data Communication and Networks,
  - Web System and Technology,
  - Mobile Application Development,
  - Applied Computer Graphics,
  - Multimedia Technology,
  - English for research Information Technology,
  - Electronic Commerce,
  - Fundamentals of Mathematics, and
  - Statistics for Research.



**Figure 1.** The Model of development of guidelines for the academic support of students by using the Apriori algorithm.

### 3.2 Information Preparation

According to the data collected from Step 1, the researchers had transformed this information into a suitable application (Table 1).

Based on the data for operating the association rule as shown in Table 1, the researchers used the Microsoft Excel program to collate data. This file was saved in .CSV in order to generate the association rule with Weka 3.8, software for data mining (Figure 2).

|    | A       | B       | C       | D       | E       | F       | G       | H       | I       | J       |
|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1  | GEH1101 | GEH1102 | GEH2202 | GEL2201 | INT1104 | INT1107 | INT1402 | INT1403 | INT1503 | INT2104 |
| 2  | C-      | D+      | C-      | C-      | C-      | D-      | C-      | B       | C-      | C*      |
| 3  | F       |         |         |         | F       | F       | F       |         |         | F       |
| 4  | C+      | D       | C       | C-      | B-      | D+      | D+      | D       | B       | D+      |
| 5  | F       |         |         |         | F       | F       | F       |         |         | F       |
| 6  | B       | C       | C       | C       | C       | D+      | C-      | C       | A       | C+      |
| 7  | C-      | C       | C+      | C-      | B-      | D+      | C       | B-      | C+      | D+      |
| 8  | C       | B-      | C       | C       | D       | D-      | C+      | B-      | C+      | D       |
| 9  | C-      | C+      | D+      | C       | C       | D-      | C       | C       | B       | C+      |
| 10 | C+      | C+      | D       | C       | C+      | D       | C-      | C       | C+      | C+      |
| 11 | D+      | C+      | B       | D+      | C+      | C-      | C-      | C-      | B       | D       |
| 12 | D+      | D-      | B       | D       | C-      | D       | D+      | C       | C       | C-      |
| 13 | C       | D+      | D+      | D       | C-      | C-      | C       | B-      | B-      | D       |

**Figure 2.** Examples of information used for operating the association rule.

### 3.3 The Association Rule Operation

The researchers had analyzed the data to find determine the relationships between each piece of information. The data was analyzed by using Apriori algorithm with Weka 3.8 to operate the association rule<sup>4</sup>. While running the program, the lower Bound Min Support value was set to 0.1 min Metric was set to 0.9.

According to the results from the association rule operation using Weka 3.8, the researchers had found thirty rules in total with a confidence value of 100%. In this study, ten rules were selected to illustrate the details (Table 2).

## 4. Results

This research verified the accuracy of association rules with data of 1200 student records. This data was real study outcome of all students from academic year of 2017 to 2018. Data from these 2 academic years were not used to formulate the rules. Efficiency of the rules was evaluated by using their accuracy show in equation<sup>4</sup>.

**Table 2.** The results from the association rule operation and the meaning of rules

| Rules Resulted from Weak 3.8   | Meanings  |
|--|---|
| 1. INT1107=D ==> Status=t_weak <conf:(1)>                                | If students get a 'D' in their Data Structure and Algorithm course, they tend to have a lower than the standard proficiency.  |
| 2. INT2305=D ==> Status=t_weak <conf:(1)>                                | If students get a 'D' in their Data Communication and Networks course, they tend to have a lower than standard proficiency.   |
| 3. INT1403=F ==> Status=t_weak <conf:(1)>                                | If students get an 'F' in their Computer Operating System course, they tend to have a lower than standard proficiency.  |
| 4. INT2202=F ==> Status=t_weak <conf:(1)>                                | If students get an 'F' in their Design for Database System course, they tend to have a lower than standard proficiency.   |
| 5. MAT1201=D ==> Status=t_weak <conf:(1)>                                | If students get a 'D' in their Fundamental of Mathematics course, they tend to have a lower than standard proficiency.  |
| 6. INT1104=F and INT1107=F ==> Status=t_weak <conf:(1)>                  | If students get an 'F' in their Computer Programming Development and Data Structure and Algorithm courses, they tend to have a lower than standard proficiency.     |
| 7. INT1107=F and INT2202=F ==> Status=t_weak <conf:(1)>                  | If students get an 'F' in their Data Structure and Algorithm and Fundamental of Mathematics courses, they tend to have a lower than standard proficiency            |
| 8. INT1403=F and INT2202=F ==> Status=t_weak <conf:(1)>                  | If students get an 'F' in their Computer Operating System and Fundamental of Mathematics courses, they tend to have a lower than standard proficiency.              |
| 9. INT1107=F and INT1403=F ==> Status=t_weak <conf:(1)>                  | If students get an 'F' in their Data Structure and Algorithm and Computer Operating System courses, they tend to have a lower than standard proficiency.            |
| 10. INT1104=F and INT1107=F and INT2202=F ==> Status = t_weak <conf:(1)> | If the students get an 'F' in their Computer Programming Development and Data Structure and Algorithm courses, they tend to have a lower than standard proficiency. |

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

TP = Higher grades in a given course tend to have a higher overall grade.

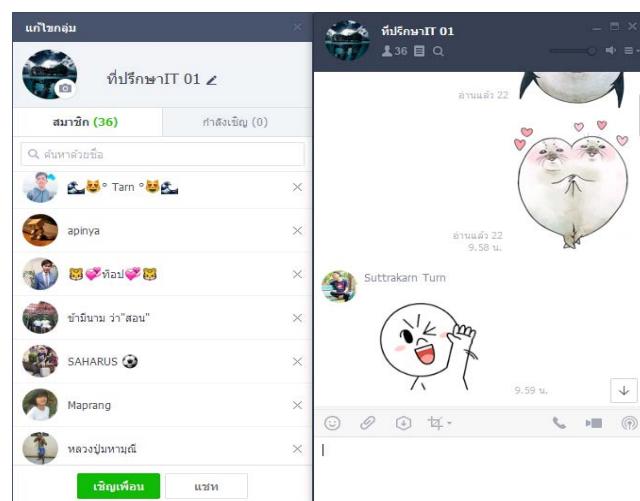
FP = Higher grades in a given course tend to have a lower overall grade.

TN = Lower grades in a given course tend to have a lower overall grade.

FN = Lower grades in a given course tend to have a higher overall grade.

According to evaluation of rules' efficiency defined by accuracy was 89% when program applied 30 rules in the curriculum meeting and revision. Lecturers, advisors, and students paid attention to students who had study records that complied these 30 rules. These rules also were guidelines to treat and help other students who had risk to make weak study outcome. Students would be notified and be able to adapt themselves on time. They would be protected from school expulsion due to study outcome is

lower than school criteria. There would be activities or approaches to monitor and advise them.

**Figure 3.** online social media consulting group.



**Figure 4.** Advisor use online social media consulting group to consult the students.

1. Lecturers paid attention to students who had risk and provided activity and supplement for them,
2. Program acknowledged risk and provided, assistance for them on time before they were removed or banned from school,
3. Advisor provided guidance and solution for students who had risk via online social media. These students with risk had to study extra time with supplement classes (Figure 3 and 4), and
4. Students with risk, they would realize this risk that might affect their study. They would be able to adapt their study behavior on time.

## 5. Conclusions

This research indicates that the application of the association rule may assist researchers in finding the relationship between the courses that tend to produce low proficiency students. Researchers collected the grade point data of

the students enrolled in the Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, between 2011 and 2016. This data consisted of the grade points obtained in the courses under the general education category and the information technology category, of which there are 23 courses in total. This information was analyzed by Weka 3.8 to find determine which courses tend to produce low proficiency students the most. There were thirty association rules that resulted from the association rule operation. These rules can contribute to guidelines for assisting students enrolled in Information Technology courses. Tuition classes or a special academic advisor may be provided for low proficiency students. Result from applying association rules to provide, assistance, approaches and mentorship to students via online social media. Finding relationship by using weak study outcome of students was helpful to monitor students with risk. These students would be able to adapt themselves and attained better study outcome. Finally, they would graduate as stipulated in the curriculum book. Better study outcome and graduation were expected by lectures, students, program and uni-versity. Application of association rules to find out relationship of courses and tendency to attain weak study outcome was a good approach to assist and protect students from school expulsion.

## 6. Acknowledgements

We would like to express our gratitude to the Faculty of Science and Technology and Institute for Research and Development, Suan Sunandha Rajabhat University and Faculty of Science, Ubon Ratchathani University, who offered us the opportunities to conduct this research.

## 7. References

1. Ranumas M. Effective teaching and learning in higher education, Journal of Southern Technology. 2011; 9(2):169–76.
2. Nawarat K, Varoontanya K. Factors Affecting The Drop Out of Rajamangala University of Technology Phra Nakhon Students in 2014 Academic Year. Research Funded by Rajamangala University of Technology Phra Nakhon; 2015. p. 1–97 .
3. Nuttid S, Anthika S. The use of association rules to find the factors effected to the risk of low proficiency students Case study: Prince of Songkla University, Journal of Management Sciences. 2011; 28(1):65–79.

4. Data Mining. Date accessed: 17/09/2018. [https://en.wikipedia.org/wiki/Data\\_mining](https://en.wikipedia.org/wiki/Data_mining).
5. Larose DT. Discovering knowledge in data: an introduction to data mining. John Wiley and Sons; 2005. PMCid: PMC1157106.
6. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc.; 1994. p. 487–99.
7. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data; 2000. p. 1–12. <https://doi.org/10.1145/342009.335372>.
8. Goswami DN, Chaturvedi A, Raghuvanshi CS. An algorithm for frequent pattern mining based on apriori, International Journal on Computer Science and Engineering. 2010; 2(4):942–47.