Universal Journal of Educational Research 9(10): 1756-1764, 2021
DOI: 10.13189/ujer.2021.091004 http://www.hrpub.org

A Modeling of an Intelligent System for Learning Result Prediction to Reduce Drop-Out of Undergraduate Students

Budsaba Sungwanna*, Pallop Piriyasurawong

Division of Information and Communication Technology for Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Received July 12, 2021; Revised September 15, 2021; Accepted September 21, 2021

Cite This Paper in the following Citation Styles

(a): *[1] Budsaba Sungwanna, Pallop Piriyasurawong , "A Modeling of an Intelligent System for Learning Result Prediction to Reduce Drop-Out of Undergraduate Students," Universal Journal of Educational Research, Vol. 9, No. 10, pp. 1756 - 1764, 2021. DOI: 10.13189/ujer.2021.091004.*

(b): *Budsaba Sungwanna, Pallop Piriyasurawong (2021). A Modeling of an Intelligent System for Learning Result Prediction to Reduce Drop-Out of Undergraduate Students. Universal Journal of Educational Research, 9(10), 1756 - 1764. DOI: 10.13189/ujer.2021.091004.*

Abstract The objectives of this research were to; 1) analyze the factors of an intelligent system for learning result prediction to reduce drop-out of undergraduate students. 2) construct a modeling of an intelligent system for learning result prediction to reduce drop-out of undergraduate students. The samples were 141 undergraduate students who study English Education program in Academic year 2012-2014 at Kanchanaburi Rajabhat University by purposive sampling. The research results were as follows 1) the factors analysis was based on the attribute weight indexing technique using the Information Gain method. The learning results prediction factors had 14 factors, for example, mean of GPA from semester 1 to 5 and learning results about 9 subjects, 2) constructing a modeling of an intelligent system for learning result prediction to reduce drop-out of undergraduate students by measuring the quality with Cross-validation Test; 10-fold cross-validation and Naïve Bayes technique, the highest accuracy index is 84.33 percent, and followed by the creation of student's learning result prediction by using Decision Tree technique, 73.86 percent of accuracy index.

Keywords Intelligent System, Learning Result Prediction Factors, Drop-Out, Data Mining

1. Introduction

The university's academic administration faces a difficult task in analyzing the factors that influence student learning outcome prediction in order to lower the number of drop-out bachelor students. Rajabhat Kanchanaburi University's study regulations state that, the student's status is defined in associate's degree, bachelor's degree, and bachelor's degree (continuing program) education in 2008. The student status will be revoked if enrollment is not completed or if the grade point average falls below the stipulated levels. Furthermore, a student who has completed the entire curriculum and has a grade point average of at least 2.00 can be permitted to graduate [1]. According to the findings, students' academic performance is unrelated to their study plans, and they drop out due to a lower grade point average than the University's requirements. Because dropout is a serious problem at many colleges, it is necessary to examine the components in order to determine the causes or guidelines for resolving this issue. The use of classical association rule mining for student learning outcome forecasting [3], and the analysis of factors impacting undergraduate students choosing a major [4] are all based on the results of a research of analysis for student dropout in undergraduate using data mining technique. These are utilized to improve the quality of education and the efficiency with which educational

management is carried out.

To find the knowledge or relevance of the data which have not been used from the database of the students, the qualified data mining technique can be used to analyze the beneficial knowledge which helps improve the university quality and solve problems and factors that influence the drop-out of students thus leading to education planning and managing to reduce the drop-out of the students. The techniques used for data mining are as follows: 1) Association Rule [5]: shows the relevance of the situation or object happening concurrently, 2) Data classification [6], 3) Data clustering [7] and Virtual presentation.

From the data of student registration in the academic year 2012-2014, 3,913 students enrolled in different courses related to their majors. The summary of the student status analysis indicates that there are 1) 2,379 graduated students or 60.79 percent, 2) 243 ungraduated students or 6.21 percent, and 3) 1,291 drop-out students or 32.99 percent. As a result, the students have to drop out which is a significant issue for education management. To find the cause and solution to solve and prevent the issue occurs, the study of the data of student registration is useful by analyzing the past data to create the student performance prediction used for predicting, for example, student status, retirement opportunity of the student, graduation opportunity of the student, or study result of the student. Data mining and various techniques are used for studying past knowledge to analyze and create student performance predictions. The researcher is interested in analyzing factors and predicting the study result to resolve and reduce the number of drop-out students of Kanchanaburi Rajabhat University.

2. Research Objectives

1. To analyze the factors of an intelligent system for learning result prediction to reduce drop-out of undergraduate students.
2. To construct a modeling of an intelligent system for learning result prediction to reduce drop-out of undergraduate students.

3. Theories and Literature Review

Factors analysis or properties selection influencing the prediction of study learning results to reduce the drop-out of students by reducing the data can be divided into 3 approaches: 1) Filter Approach, 2) Wrapper Approach, and 3) Embed Approach. The Filter Approach is used to arrange the importance index attribution from the calculation of weighted or relevance value index attribution, for example, Chi-square, or Information Gain. The approach used for reducing the data in this research is Filter Approach: Information Gain.

Decision Tree [8] is the technique providing the result as a tree structure. Within the tree structure includes nodes. Each node has the testing property. To create the decision tree, the attribute with the most relevance to the class is selected as the root node. The following attributes are selected. To find the relevance of the attributes, the Information Gain is used in the calculation as follows [6,8]:

IG (parent, child) = entropy(parent) – [p(c1) × entropy(c1) + p(c2) × entropy(c2) + …] when entropy(c1) = -p(c1) log p(c1) and p(c1) is the probability value of c1.

Naïve Bayes is the technique of calculating the probability by predicting from the past situation happened. The equation is as follows [9]:

$$P(B|A) = \frac{P(A|B)xP(B)}{P(A)}$$

P(B|A) is the probability of situation B when situation A happens priorly,

P(A|B) is the probability of situation A when situation B happens priorly,

P(A) is the probability of the occurrence of situation A,

and P(B) is the probability of the occurrence of situation B.

Three principles [8] used for dividing the data quality testing are 1) Self-Consistency Test, the model-creating data and model-testing data are the same, 2) Split Test, to divide the data by randomizing the data into 2 parts, such as 70:30 or 80:20 percent when the first part is used for creating the model and the second part is used for testing the model quality, 3) Cross-validation Test, to divide the data into many parts, such as 10-fold cross-validation when each part has the equal data amount and a part of the data is used for testing the quality of the model, and all the divided data amount are used for testing. To test the model quality of this research, the 10-fold cross-validation approach is used.

The techniques used for testing the quality of the model to classify the data type (classification) [8] are 1) Precision, the probability value used for predicting, 2) Recall, the correct amount of the prediction used for testing the accuracy of the model, and 3) Accuracy, the correct amount of every class. This is the model-accuracy testing considered from every class.

Literature Review

Paphorn [2] indicated the objectives of analysis for student dropout in undergraduate using data mining technique are 1) to analyze the factors related to the drop-out of bachelor students, 2) to synthesize the model used for predicting the drop-out of bachelor students, and 3) to compare the effectiveness of data classification of the model by using the Rule Induction, K-Nearest Neighbor, Decision Tree, and Naïve Bayes techniques. The data were collected from 14 students of Rajamangala University of Technology Isan, the academic year 2014-2018. The result indicated that 12 factors were related to the drop-out of

students. After developing the model tested its quality by using 10-fold cross-validation and accuracy testing, the model created by using the Rule Induction technique had the highest quality with 94.70 of the accuracy. The most related 5 factors were grade point average, academic year, old school, major, and career of the parents.

From the factor analysis with data mining technique in higher educational student drop out by Nontawat [10], the data were analyzed to create the stimulated model classifying data by using the decision tree and algorithm J48 to test the model by using the cross-testing technique, and the data were divided into percent. The data were collected from 3,604 students of the Faculty of Science, Buriram Rajabhat University, the academic year 2013-2016. The result indicated that 11 factors related to high-school education and during university education influenced the drop-out of students. The correct was 96.73. The precision was 96.6. The recall was 96.7. And the f–measure was 96.5.

From Solaet [11]: the analysis of factors influencing the dismissal of students using students and graduates' data during the academic year 2012–2015. There were 97 records from the Computer Science program and 202 records from the Information Technology program that consisted of 26 features. Data mining techniques like the Decision tree technique, Back Propagation Neural Network (BP-NN), and Support Vector Machine (SVM) were employed to propose forecasting models. The models were tested and compared using a 10-fold Cross-Validation. The results revealed that 3 factors influencing the dismissal of the students were Physics grade points, Platform Technology grade points, and the first-semester grade point average of the second academic year while the Data Structure grade points were the only factor influencing the dismissal of the Computer Science program students.

From Watanyuta[12]: The study of associated factors of the resignation decision and the efficiency comparison of various predictive models of employee resignation: A case study of an insurance company. The analyzed data set was the data of resigned employees and employees who were still working from the year 2013-2017. The data of 1,000 items with 11 attributes were analyzed. Moreover, five predictive models for the resignation of employees using 1) decision tree, 2) support vector machine, 3) neural network, 4) Naïve Bayesian, and 5) nearest-neighbor techniques had been conducted. The efficiency comparison among those five models was performed based on the 5-fold cross-validation technique. It was found that the highest accuracy was at 91.03%, while the support vector machine technique's predictive model yielded an accuracy of 90.93%. Neural Network technique predictive model yielded accuracy at 90.75%. Naïve Bayes and K-Nearest Neighbors models yielded 89.60% and 82.10% of accuracy, respectively.

Since the study uses many analyses of factors and learning result prediction by using qualified data mining techniques according to the technique and structures of each data and cannot state the best data analysis technique, researcher analyzes the factors of prediction to reduce individually with the measurement of weighted average of the factors affecting the prediction, and compare the techniques used for creating the prediction model of Naïve Bayes and Decision Tree to find the best prediction model used for predicting the intelligent study to reduce the drop out of the student.

4. Research Methodology

4.1. Population and Sample

1) Population is 2,379 graduated students of Kanchanaburi Rajabhat University, in the academic year 2012-2014.
2) Sample is 141 graduated students of Kanchanaburi Rajabhat University, in the academic year 2012-2014, majoring in English.

4.2. Research Instrument

1) Study primary information, such as dismissal document of students, student data from the office of academic promotion and registration, Kanchanaburi Rajabhat University, data from the academic sources, academic titles, and academic titles about factors of study result prediction, factors of drop-out of high-education student prediction, and predictive model creation, etc.
2) Analyze the data of the parents, grade point average, student status, drop-out status of the student from the database of the student from the office of academic promotion and registration, Kanchanaburi Rajabhat University.
3) Analyze the factors influencing the study learning result prediction to reduce the drop-out of students from the database, the academic year 2012-2014.

4.3. Research Process

This research presents the data mining techniques by using the Rapid Miner program. CRISP-DM [13] includes 6 steps connecting as follows:

Source: CRISP-DM 1.0 step-by-step data mining guide. by Chapman. P., et al. (2000). SPSS inc.

Figure 1. The process of CRISP-DM

1. Business Understanding: to analyze the data mining to study the factors or causes of drop-out of bachelor students by using 1) Decision Tree, and 2) Naïve Bayes. The data were collected and analyzed from 2,379 bachelor students of Kanchanaburi Rajabhat University, the academic year 2012-2014.
2. Data Understanding: to collect the data from the student database with the different course enrollments according to students' majors. The background data of students, parents, grade point average, and student status of 2,379 students, the academic year 2012-2014, were studied. As the program structures of each major were different, the factors of studies result of students of each major must be analyzed individually. The sample of 141 graduated students in the academic years 2012-2014 were selected to analyze the factors influencing the study result prediction. 34 most necessary variables: 33 independent variables and 1 dependent variable, were selected to classify the relevance of the data.
3. Data Preparation: to convert the raw data to the appropriate form and analyze them in the following step. The data were converted into the same scale or filled on the lacking data.
    1) Data Selection: to select the specific data related to this research. 34 attributes: sex, background education, father status, mother status, grade point average per academic term, grades of 26 courses, and grade point average of graduation, were selected.

Table 1. Detail of attributes

Attribute	Detail
sex	0=male 1=female
Old_ed	11=grade 12 1=Vocational Certificate
Fat_status	0= deceased 1=alive 9=unknown
Mor_status	0= deceased 1=alive 9=unknown
Grade_T1	VH = 3.50-4.00 H = 3.00-3.49 M = 2.50-2.99 L = 2.00-2.49 VL = 0.00-1.99
Grade_T2	
Grade_T3	
Grade_T4	
Grade_T5	
Studies results of 24 courses	4.0 = A 3.5 = B+ 3.0 = B 2.5 = C+ 2.0 = C 1.5 = D+ 1.0 = D 0.0 = F
GPA	vLow = 2.00-2.24 Low = 2.25-2.80 Middle = 2.81-3.19 Good = 3.20-3.53 High = 3.54-4.00

2) Data Cleaning adjusted and corrected values, deleted or fixed error values or missing values.
3) Data Transformation was used for converting adjusted values into a ready-to-use format which was appropriated with the chosen algorithm. From the students' data, it is found that the values were in form of nominal and numeric. It should be adjusted into the same format.

4. Modeling was a synthesis procedure for analyzing intelligent prediction method for grade estimation to prevent students drop-out from the students' data of Kanchanaburi Rajabhat University from 2012 A.Y. to 2015 A.Y. by using filter approach method, analyzed attributes weight index by using information gain by making modeling from obtained factors by using mining techniques consisted of 1) Decision Tree analysis and 2) 10-fold cross validation for Naïve Bayes classifier.
5. Modeling Evaluation was a tool used for data classification efficiency measurement included with 1 (Accuracy 2) Precision 3) Recall [8].

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

By TP was the amount of correct data which is extracted.

FP was the amount of error data which is extracted.

TN was the amount of correct data which is not extracted.

FN was the amount of error data which is not extracted.

The result of comparing intelligent prediction method for grade estimation efficiency from selected attributes by using Information Gain with 10-fold Cross Validation by using Decision tree analysis and Naïve Bayes classifier found that intelligent prediction method modeling which used 33 factors would be less accurate than intelligent prediction method modeling that reduced related factors. From the selection, the significant factors for intelligent prediction method for grade estimation, in 6 factors of Decision tree analysis gave an accuracy value at 73.86 percent and in 14 factors of Naïve Bayes classifier 14 factors gave an accuracy value at 84.33 percent.

Table 2. The result of comparing accuracy efficiency of modeling

Algorithm	Accuracy
Decision Tree	**73.86**
Naïve Bayes	**84.33**

6. Development was a procedure for using the most accurate intelligent prediction method for grade estimation to predict students' grades in next semester for choosing the most appropriate study plan for students and to predict students' grades when they graduated for making a solution or preventing student drop-out.

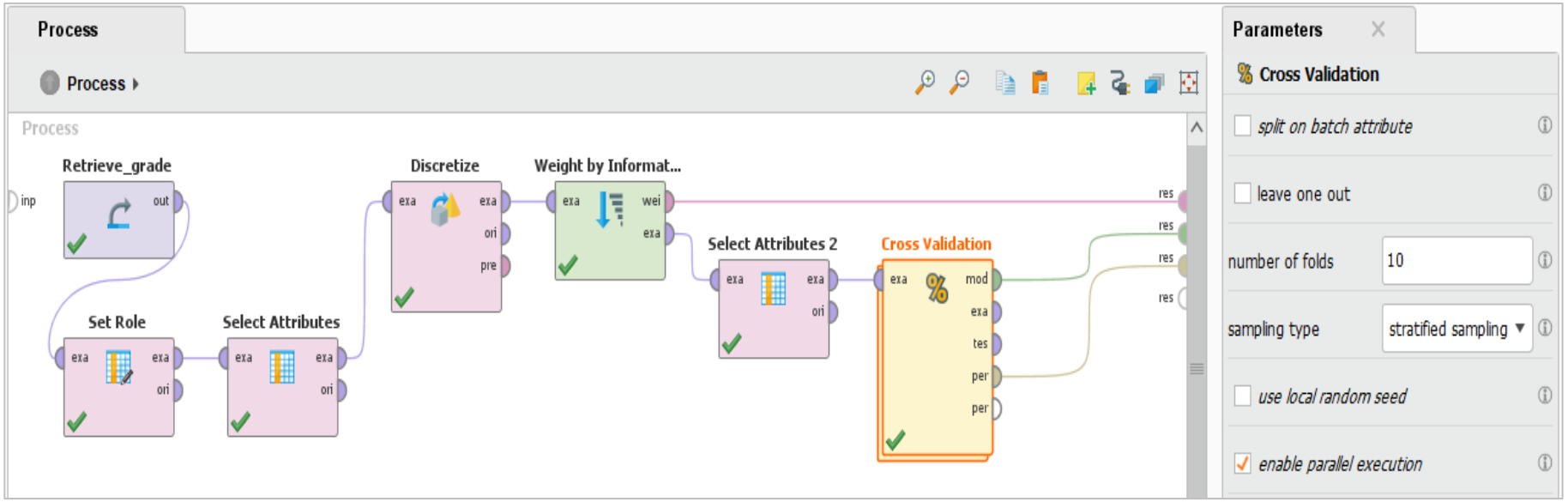

Figure 2. 10-fold Cross Validation Efficiency Measurement

accuracy: 84.33% +/- 9.45% (micro average: 84.40%)

	true Middle	true Good	true High	true Low	class precision
pred. Middle	40	4	0	4	83.33%
pred. Good	1	28	5	0	82.35%
pred. High	0	4	33	0	89.19%
pred. Low	4	0	0	18	81.82%
class recall	88.89%	77.78%	86.84%	81.82%	

Figure 3. The efficiency of prediction modeling by using Naïve Bayes classifier

accuracy: 73.86% +/- 10.75% (micro average: 73.76%)

	true Middle	true Good	true High	true Low	class precision
pred. Middle	34	10	0	8	65.38%
pred. Good	7	23	5	0	65.71%
pred. High	0	3	33	0	91.67%
pred. Low	4	0	0	14	77.78%
class recall	75.56%	63.89%	86.84%	63.64%	

Figure 4. The efficiency of prediction modeling by using Decision tree analysis

5. Result

The result of analyzing intelligent prediction method for grade estimation to prevent the drop-out of Bachelor degree students by calculating weight index using Information Gain from intelligent prediction method modeling by using Naïve Bayes classifier found that 14 factors which affected intelligent prediction method for grade estimation were GPA semester 2, GPA semester 3, GPA semester 4, GPA semester 1, subject code 202507 grade, GPA semester 1, subject code 202506 grade, subject code 105131 grade, subject code 202101 grade, subject code 901101 grade, subject code 202211 grade, subject code 202207 grade, subject code 101501 grade and subject code 903301 grade. The details were shown in Table 3

The synthesis result of intelligent prediction method for grade estimation to prevent students drop-out by efficiency measuring using 10-fold cross validation found that Naïve Bayes classifier had the highest accuracy index at 84.33 percent and Decision Tree analysis had the accuracy index at73 . 86percent, respectively.

Table 3. The result of analyzing the weight ratio of the intelligent prediction method for grade estimation factors

Row No.	Attribute	Weight
1	grade_T2	0.928247
2	grade_T3	0.917103
3	grade_T4	0.810689
4	grade_T1	0.775998
5	202507	0.729367
6	grade_T5	0.717749
7	202506	0.68555
8	105131	0.655477
9	202101	0.648303
10	901101	0.624847
11	202211	0.596433
12	202207	0.572321
13	101501	0.524012
14	903301	0.502386

6. Discussion

The purpose of this research was for analyzing the intelligent prediction method for grade estimation to prevent students' drop-out. The result of analyzing intelligent prediction method for grade estimation factors by using Rapid Miner from 33 factors. When the attributes weight index was analyzed by using Information Gain found that every factor is significant in the intelligent prediction method for grade estimation. But when the factors were decreased by one to the lowest to obtain the highest prediction accuracy value, it is found that in Naïve Bayes classifier, the factors for intelligent prediction method for grade estimation accuracy had 14 factors included GPA semester 2, GPA semester 3, GPA semester 4, GPA semester 1, subject code 202507 grade, GPA semester 1, subject code 202506 grade, subject code 105131 grade, subject code 202101 grade, subject code 901101 grade, subject code 202211 grade, subject code 202207 grade, subject code 101501 grade and subject code 903301 grade. In Decision Tree Analysis, there were 6 factors intelligent prediction method for grade estimation accuracy included GPA semester 2, GPA semester 3, GPA semester 4, GPA semester 1, subject code 202507 grades and GPA semester 1. About making intelligent prediction method modeling and comparing intelligent prediction method for grade estimation efficiency to prevent students drop-out by choosing attributes through Information Gain together with data mining method, Naïve Bayes classifier gave accuracy value at 84.33 percent and Decision tree analysis gave accuracy value at 73.86, respectively. This study corresponded to the study of Solaet[11] that analyzed students drop-out factors by using data mining method from the data of studying students and graduated students between 2012 A.Y to 2015 A.Y. from 97 records of students studied in computer science curriculum and 202 records of students studied in information technology curriculum included with 26 factors by using Decision tree analysis, reverse artificial neural network and support vector machine, making intelligent prediction method modeling and comparison modeling efficiency by using 10-fold Cross Validation. The result stated that background information was not a factor for predicting drop-out of students who studied in these 2 curriculums. The factors that affected to students drop-out of students who studied in information technology were 3 factors consisted of basic physics subject grade, platform technology subject grade and GPA of sophomore students in semester 1 and the factors that affected students drop-out of students who studied in computer science was 1 factor which was Data structure subject grade. Therefore, to prevent the amount of students drop-out, students' grade management is significant. Students' grade should be in the university criteria and requirements which is not lower than 2.0

7. Conclusions

The research process is to develop educational quality rely on various formats to obtain the best result and must consider many factors. In this study, the researcher presents a research method in two steps as follows: 1. Analyze the factors of the intelligent prediction method for grade estimation to prevent students' drop-out and 2 and make the model of intelligent prediction method for grade estimation to prevent students' drop-out. From the factors analysis result and making model by using data mining technique found that in Naïve Bayes classifier has the highest accuracy index value at 84.33 and has highest related factors to 14 factors included GPA semester 2, GPA semester 3, GPA semester 4, GPA semester 1, subject code 202507 grade, GPA semester 1, subject code 202506 grade, subject code 105131 grade, subject code 202101 grade, subject code 901101 grade, subject code 202211 grade, subject code 202207 grade, subject code 101501 grade and subject code 903301 grade. The result of the study is desirable and can apply to develop an intelligent prediction method for grade estimation to prevent students' drop-out in the future.

Acknowledgements

I would like to thank the Graduate College of King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand for providing financial support for the research fund.

REFERENCES

[1] Kanchanaburi Rajabhat University. "The Regulations of Kanchanaburi Rajabhat University on Diploma, Bachelor', and Bachelor's Degree(Continuing program) B.E. 2008," Kanchanaburi Rajabhat University, July 26., 2008.

[2] Laopilai P., Sanrach C., "Analysis for Student Dropout in Undergraduate Using Data Mining Technique," The Science Journal of Phetchaburi Rajabhat University, vol. 16, no. 2, pp. 61-71., 2019.

[3] Ruxpakawong P., Ruxpakawong U., "The Use of Traditional and Fuzzy Association Rule Mining for Student Learning Outcome Forecasting," KKU Science Journal, vol. 43, no. 3, pp. 542-551, 2015.

[4] Thepprasit R., Sanrach C., "The analysis of factors affecting choosing a major of undergrate students of the faculty of education by using data mining technique," Journal of Graduate Studies Valaya Alongkorn Rajabhat University, vol. 14, no. 1, pp. 134-146, 2020.

[5] Srikant, R., Vu, Q., Agrawal, R., "Mining association rules with item constraints," Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, August 14, 1997, pp. 67-73.

[6] Han j., Kamber M., Pei j., "Classification: Basic Concepts," in Data Mining Concepts and Techniques, 3rd ed. USA: Morgan Haufmann, 2012, pp. 327-380.

[7] Pacharawongsakda E., "Clustering," in An introduction to data mining techniques, 2nd ed, ASIA digital press, Thailand, 2014, pp. 27-49.

[8] Pacharawongsakda E., "Classification," in An introduction to data mining techniques, 2nd ed, ASIA digital press, Thailand, 2014, pp. 76-82.

[9] Vilailuck S., Jaroenpuntaruk V., Wichadakul D., "Utilizing Data Mining Techniques to Forecast Student Academic Achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center," Veridian E-Journal Science Technol Silpakorn University, vol. 2, no. 2, pp. 1-17, 2015.

[10] Taweechat N., Phengprajon O., Yathongchai W., Yathongchai C., "Factor Analysis with Data Mining Technique in Higher Educational Student Drop Out," The 5th National Conference on Technology and Innovation Management of Rajabhat Maha Sarakham University, Maha Sarakham, Thailand, March 5., 2019, pp.1-7.

[11] Kepan S., Leelapatarapun P., Yokkhun A., "Analysis of Factors Influencing the Dismissal of Students Using Data Mining Techniques Case Study: Computer Science Program and Information Technology Program of Yala Rajabhat University," Veridian E-Journal of Science and Technology Silpakorn University, vol. 5, no. 4, pp. 96-110., 2018.

[12] Neelaphatrakun W., Beokaimook C., "The Study of Associated Factors of the Resignation Decision and the Efficiency Comparison of Various Predictive Models of Employee Resignation: A Case Study of an Insurance Company," Association of Private Higher Education Institutions of Thailand (APHEIT), vol. 8, no. 1., pp. 46-63., 2019.

[13] Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., Wirth R., "The CRISP-DM reference model," in CRISP-DM 1.0 step-by-step data mining guide, USA: SPSS., 2000, pp. 11-15.