Student Behavior Analysis Affecting Learning Achievement of Information Technology and Computer Science Students

Supaporn Bundasak

Abstract—The analysis of student behavior affecting the learning results with that data mining is research. To find behavioral factors that affects the student's academic performance. Learn more by the research team leading the reference information from collecting data from questionnaires of students of the Faculty of Science. To analyze the model of decision making to separate groups with similar behavior together and use grouping techniques (Clustering). Making decision to find out answers to behavioral factors that affect learning results. In small groups based on various behavioral factors such as Use of free time in everyday life Learning anxiety Concentrate on learning Motivation for learning and using K - Means Cluster and Random Tree to find a model that has Correctly Classified Instances when tested with 380 datasets of training data, with an average of 98% which is the average at a good level can Bring the model that has been Continued to develop the system.

Index Terms—Study behavior, learning results, data mining techniques, k-means, random tree.

I. INTRODUCTION

The learning behavior is concerned with finding that students had not passed the criteria of the University and some students had graduated more slowly than the specified curriculum and will be increase each year. One of the important things that affect the students' academic performance was the learning behavior of the students themselves. They had changed or were passion about wrong values or other factors. The important problems were affected the academic performance because it had many problems that negatively affect the students themselves and will negatively affect the advisor of that student as well, when having lower grades or being in a risk group causing students to fail Registration can be made in accordance with the plan specified by the university. Therefore, students must choose the course that was suitable for the study to adjust the academic results, not to result in poor academic performance .All problems may be caused by negligence of the students themselves or may be due to lack of experience Student learning planning and data mining was one way of analyzing large amounts of information Therefore, the researcher had the idea to use these academic observation to develop as a data warehouse and used to analyze the behavior of Kasetsart University Sriracha campus's student with data mining techniques to predict behavior that may affect the academic performance of Kasetsart University Sriracha Campus students. By using the student's academic data student behavior survey, which used a feature selection method combined with a data mining technique and then compared to get the best forecast in developing forecasting systems to be effective in forecasting and can be used to predict the behavior of students that affect their academic performance and the system can give advice to students.

II. RELATED LITERATURE

Due to the problem of poor academic performance which may result in problems, that cause students to have low grades and resulting in the removal of students from university before graduation, resulting in research. "Higher education results" by the study program: distance education (part-time) or High-school graduation GPA; Difference in years from the moment the student graduates high-school until he / she enrolls at university. 1000 students from "Nicolae Titulescu" University of Bucharest from the graduates' generations. "POOR RESULTS", the class "MEDIUM RESULTS" and the class "GOOD RESULTS". We classified students according to their GPA after the first year of study in three classes: POOR RESULTS - those students with GPA lower than 6; MEDIUM RESULTS– Those students with GPA between 6 and 8; "POOR RESULTS" was 86.6%, the class "POOR RESULTS" was 94.2% and the class "GOOD RESULTS" was 85.7%. The MSE obtained after training the network was 1.7%. The mean square error for the test data set was 1.91% [1].

Many people want to improve education. Many researchers argue that little difference since our efforts in the field of modern systems, which still operate in the past. The major question was: What should we changed in the current education system, and how do we know it will work? We argue that what we need a theory for predicting outcomes of such changes. Otherwise, we may be making changes that, well-intended, may not result in desired outcomes — possibly even making things worse, and not better. In this article, that provided an overview of the strategy based on questions. This was followed by how this theory can be applied to a fictional school system. Axiomatic Theories of Intentional Systems (ATIS) and examples of systems properties, 1) PESO: Predicting Education System Outcomes, and 2) APT & C: Analysis of Patterns in Time and Configuration. APT & C provided an alternative

measurement formula for investigating dynamic and structural relations in systems. Finally, this research has been discussed and developed [2].

A study of the grade prediction system found that the grade prediction system was focused on being developed with the focus on finding the accuracy of the predictions based on the existing grade data and give priority to speed in prediction and privacy for classification settings by using support vector machines (SVMs) in logistic regression that can predict precisely 76% [3].

III. MATERIALS AND METHODS

Principles and related predictions of learning behavior, that affect academic achievement study of students in computer science and information technology students, Faculty of Science, Sriracha, Kasetsart University Sriracha campus by studying and researching relevant information to bring data to develop reports, which can be categorized as follows.

A. Data Mining

Data mining or may be called knowledge searching in a database. The technique found the pattern images from a large amount of information automatically, organized as a process of pulling out knowledge from large data by using algorithms from statistics machine learning and pattern recognition or, in other words, data mining was a process that deals with a large amount of fact to find images, patterns, guidelines and relationships hidden in that data set based on recognition statistics machine learning and mathematics data mining according to the terms set by the Royal Institute means extraction or analysis Find the information you need from a large amount of information. Or in other words, data mining was a software suite that analyzes the data that was designed to support users' needs to find the information you need from a large amount of data [4].

B. Clustering

Clustering was a technique used to classify or divide a group (may be a person, animal or object, etc.) by grouping with the same or similar characteristics together. For groups that had similar or similar characteristics to another group, Cluster Analysis technique was divided into several categories by the technique chosen was group analysis Non-step (Nonhierarchical Cluster Analysis or sometimes called K - Means Cluster Analysis) was a technique to classify groups into small groups. Will be used when there was a large number of data group by having to specify the number of groups or the number of Cluster required by the 4-step analysis process. Step 1: Group the data into k groups, which can be divided in several ways as follows: - Randomly divided - divided by educators themselves. Step 2: Calculate the midpoint of each group, such as the central point, the group of groups that C is $\bar{x}_c$, Step 3: There were 2 types of considerations, which will be calculated in the first form. Calculate the distance from each unit to the midpoint of all groups and will consider moving the unit to the lowest distance group. Type 2 Calculates the square distance of each unit to the midpoint of the group that that unit was by

giving the Sum of Square Error (SSE) equal to the square distance of each unit to the center point as the equation 1.

$$SSE = \sum_{i=1}^{n} \quad (x_i - \bar{x}_{c(i)})(x_i - \bar{x}_{c(i)}) \tag{1}$$

Refers to a group of units i

SSE = the sum of the distance from each unit in the group to the midpoint of all groups, all groups with low SSE values, indicates that the units in that group are similar.

Step 4: Consider moving the group to use the moving criteria according to the calculated values in step 1, will move the unit i to the group that causes the distance from the unit i to the midpoint group that had the lowest value [2].

C. Decision Tree

The decision tree is one of the mining techniques. Information in the form of the best known classification method, often used information verification and build trees for prophecy for the structure of the decision tree will look like General tree structure by branching according to conditions or paths of branches and predicted data that will happen, Which will use the rules in the form "if already" to create a decision tree structure for the decision tree structure will consist of the internal node, Which was the node that represents feature that was used to divide the group of data with the root node at the top of the structure, Which was the node that had the most influence on the group classification branches was a link between the nodes used as conditions or the choice of action that comes from the results each of every Predictor or feature, Leaf Node was the node that shows the results of the condition. Or actions according to the conditions for creating the decision tree of each algorithm will have similar characteristics, starting with selecting the attributes that are most closely related to the class to be the top node of the tree (Root Node). After that, the branching attribute will continue to be removed. Until the data can be divided into Class is clear [5].

D. Random Tree

Random Tree technique was a technique used in classification as C4.5, with the principle of creating a tree from multiple random tree types on each node and choosing Processed without using Prune and because the number of trees increases rapidly, which was difficult to solve, creating all possible trees Components can create a random set of trees out of the various sets of trees. The Random Tree is randomly drawn at random from a set of possible trees. In this context, "random" means that each tree in the tree's set has equal opportunities. Of being an example the way to tell this is that the distribution of the tree is a "set" continuous random tree (CRT) is a random, real tree, T_e and has the shortest travel distance. The tree subset has a tree with n points which has several researchers a lot of random tree techniques have been used for classification [6], [7].

E. Weka

Weka program was a program used to analyze data using data mining techniques. Data mining technique was the process of analyzing data from large amounts of data to find relationships. Format and classification of data and Weka also collected many techniques for analyzing data

Techniques together Can also easily analyze these data Via the Weka GUI (Graphic User Interface) screen [4], [8].

IV. PROCESS OF WORK

Scope of data analysis, searching for prototype, student behavior forecasting. This study focused on finding techniques for data mining and created a model for student behavior analysis that affected the academic results of Kasetsart University Sriracha Campus students to find the most suitable algorithm for data using data mining techniques to analyze data to find out what behaviors affect the results studying. The data used are data obtained from conducting a survey, surveying Kasetsart University Sriracha, students in Faculty of Science. That had 190 Information Technology students and 190 computer science students. The preliminary analysis consists of the student profiles, the student's behavior and the results of learning that helps to design the fact to analyze. The forecast model was the most suitable method to analyze the student behavior affecting students' learning data from the data warehouse.

This research was a descriptive research for the behavior of students affecting the academic results of the Information Technology students and Computer Science students. The tools used in the research were divided into two parts. Part one: Personal information, including student history information Student behavior Student results. Part two: Questionnaire for factors affecting learning was a questionnaire developed from various researches. There were five factors as follows: 1) Factors for students' leisure time. 2) Learning factors. 3) Achievement motivation factors. 4) Factors of learning anxiety and 5) Attention factors in studying.

The interpretation criteria were as follows: Score 1 was equal to never. Score 2 was equal to less. Score 3 was equal to mare. Score 4 was equal to true. Score 5 was the most realistic. Data collection the researcher collects data. The students found the purpose and asked for cooperation in answering the questionnaire and in this research found that received a complete return questionnaire 190 information technology students and 190 computer science students and clustering 3 groups.

V. THE PROPOSED METHOD

Analyzing data using cluster analysis (Clustering) was a technique used to classify or divide groups (may be people, animals or things, etc.) by grouping with the same or similar characteristics together. For groups that had similar or similar characteristics to another group of techniques this research used the k-mean algorithm n dividing the group of students with similar characteristics, 190 information technology students can be divided into 3 groups, and the computer science students can be divided into 3 groups, but each group can find answers to behaviors that affect academic results. Subjects were grouped as follows, was the general subject group, the major subjects of the branch and the free subjects by using RandomTree techniques, algorithms were a method for categorizing, checking data

and build trees for prophecy for the structure of the decision tree will look liked General tree structure by branching according to conditions or paths of branches and predictable data that will occur. Which will use the rules in the form "If (condition) already (result)" (If then Rule) to build the decision tree structure [9].
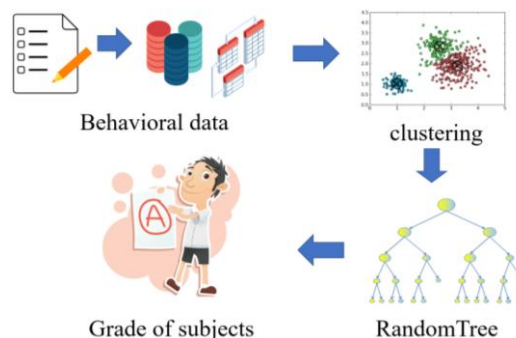

Fig. 1. The process of research.

VI. RESULTS

From the study of behaviors that affect students' learning results bachelor's degree of Information Technology and Computer Science, Kasetsart University Sriracha Campus. The results can be summarized as follows.

A. General Characteristics of Respondents

Most sample students were female. The third-year student in the highest level, followed by the second year and fourth year students. The students had a cumulative grade point average of 2.00-2.99, followed by a cumulative grade point average of 3.00-4.00 and 1.00- 1.99 with the least amount, most students live in non-university dormitories and not dormitories in the university and not working part-time than students who work outside of school hours.
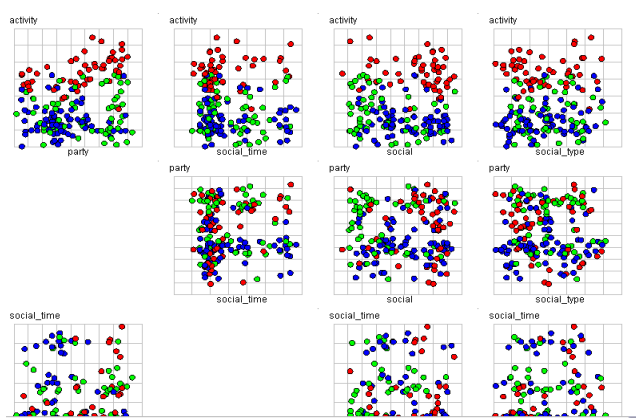

Fig. 2. The results of the clustering for information technology students.

B. Factors that Affect Academic Performance

It was found that the following factors affecting students' behavior affecting learning group1 (cluster0) in information technology Factors in learning concentration and learning factors, etc., and behavior affecting the results of general subjects of information technology students were accurate 100 percent accuracy behavior affecting learning in the main subjects of the branch of information technology students had 97.83 percent accuracy and behaviors that affect learning in the free subjects of information technology

students were correct 98.91 percent accuracy, Most of the behaviors found that students spend their free time lying down, resting and playing YouTube. Followed by spending time traveling with friends or family and the next order was reading books, the least behavior was playing games, watching movies, listening to music, affecting student behavior affecting information technology students. Group2 (cluster1) found that the behavior affecting learning in general subjects of information technology students was 100 percent accuracy correct and behavior affecting learning in the core subjects of the branch of information technology students, with 100 percent accuracy correct and behaviors that affect learning in the liberated subjects of information technology students was 100 percent accuracy correct the behavior that was found to affect the behavior affecting students. Group2 of Information Technology Found that most behaviors that affect learning were Leisure time behavior and achievement motivation behavior, etc. Factors affecting student behavior affecting learning of information technology students Group3(cluster2) It was found that the behavior affecting learning in general subjects of information technology students was 100 percent accuracy correct and the behavior affecting learning in the main subjects of information technology students was 100 percent accuracy correct and behaviors that affect learning in the free subjects of information technology students were correct 100 percent accuracy correct. Group3 of Information Technology students found the behavior affecting students that most behaviors that affect learning were attention to learning behavior and leisure time behavior etc.

Factors affecting students' learning results in group1(cluster0) in Computer Science students showed that the behavior affecting learning in general subjects of computer science students was 100 percent correct. Behavior affecting learning in the subject group. The principles of computer science students were 100 percent accuracy and behavior affecting learning in the liberalized subjects of students in the field of science. The computer was accurate 100 percent.

Behaviors that affect the learning results of students in group1, most of which were factors of leisure time Factors of learning anxiety and learning factors, Factors affecting student learning results in group2 (cluster1) in Computer Science students showed that the behavior affecting learning in general subjects of computer science students was 100 percent correct. The behavior affecting learning in the subject group the core of the computer science students was correct 100 percent accuracy correct and behaviors that

affect learning in the liberal subjects of computer science students were correct. 94.44 percent accuracy correct behavior affecting students' learning results in groups that most of them were learning factors and achievement motivation.

TABLE I: THE RESULTS OF THE PREDICTIONS FOR INFORMATION TECHNOLOGY STUDENTS

IT 190	General subjects Correctly (%)	Mandatory subject Correctly (%)	Elective subject correctly (%)	All Subject correctly (%)
Treej48	82.11	84.21	90.53	87.37
Naivebays	74.74	72.10	72.11	72.105
Bayesnet	68.42	74.74	72.63	73.685
Neural network	95.78	95.23	95.79	95.51
IBK	98.42	97.89	97.37	97.63
Randomtree	98.42	97.89	97.64	97.765
Randomforest	98.42	97.89	97.37	97.63
SMO	97.89	96.84	97.37	97.105
Logistic	98.42	97.89	97.37	97.63

TABLE II: THE RESULTS OF THE PREDICTIONS FOR COMPUTER SCIENCE STUDENTS

CS 190	General subjects Correctly (%)	Mandatory subject Correctly (%)	Elective subject correctly (%)	All Subject correctly (%)
Treej48	81.05	63.15	59.47	67.89
Naivebays	51.79	52.63	68.95	57.79
Bayesnet	74.21	54.21	72.11	66.84
Neural network	75.79	71.05	75.26	74.03
IBK	81.05	77.67	80.53	79.75
Randomtree	81.05	65.79	100.00	82.28
Randomforest	66.31	78.42	80.53	75.09
SMO	76.84	76.32	69.47	74.21
Logistic	65.79	70.53	72.63	69.65

TABLE III: THE PREDICTIONS FOR COMPUTER SCIENCE STUDENTS USE CLUSTERING METHOD AND RANDOMTREE METHOD

CS Student	num	General subjects	Mandatory subject	Elective subject	All Subject
cluster0	58	100	100	100	100
cluster1	54	100	100	94.44	98.15
cluster2	59	100	100	100	100

TABLE IV: THE RESULTS OF THE PREDICTIONS FOR INFORMATION TECHNOLOGY STUDENTS USE CLUSTERING METHOD AND RANDOMTREE METHOD

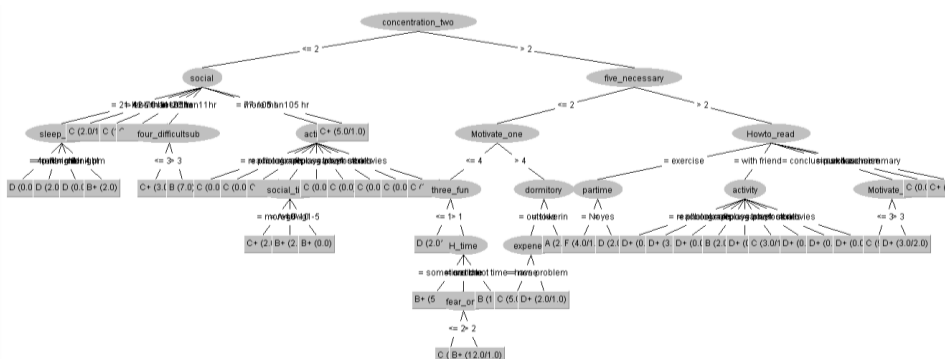| IT Student	num	General subjects	Mandatory subject	Elective subject	All Subject
cluster0	92	100	97.83	98.91	98.91
cluster1	39	100	100	100	100
cluster2	59	100	100	100	100

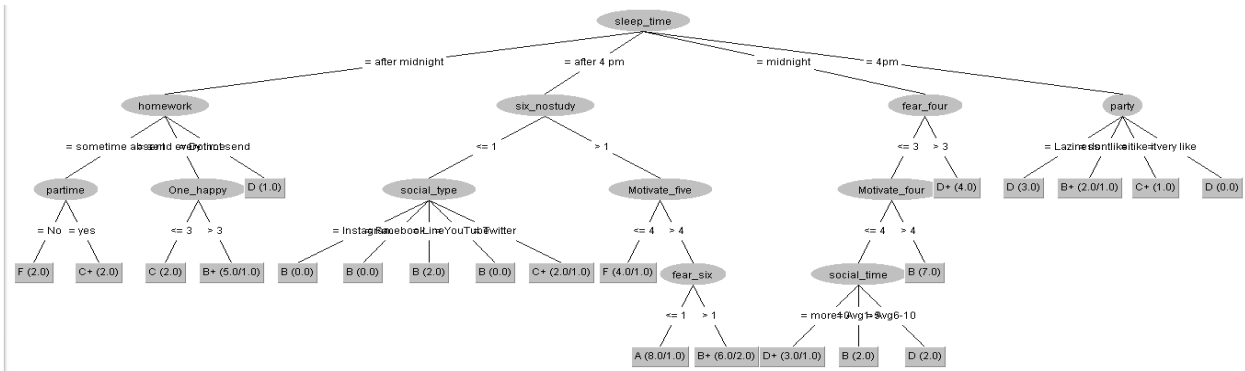Fig. 3. The results of the clustering for information technology students.

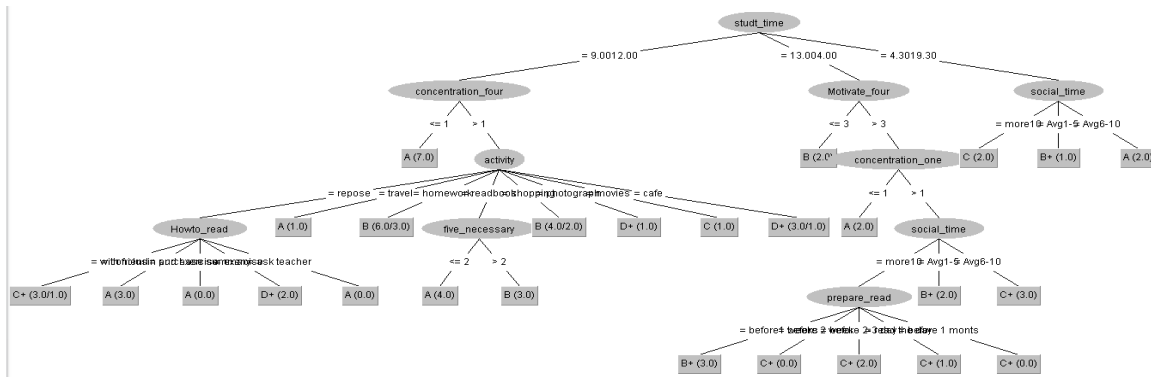Fig. 4. Student behavior model affecting computer science study Group1 in mandatory subject.

Fig. 5. Student behavior model affecting computer science learning Group1 in elective subjects.

Factors affecting students' learning results in Computer Science students Group3 (cluster2) found that the behavior affecting learning in general subjects of computer science students was accurate. 100 percent accuracy correct behavior affecting learning in the subject group. The principles of computer science students were 100 percent accuracy correct and behaviors that affect learning in the liberalized subjects of computer science students were correct 100 percent accuracy correct behavior affecting students' academic performance in the group that most of the factors were learning concentration and learning factors [10]

VII. CONCLUSION

Searching for data mining techniques to create behavioral analysis models of students at Kasetsart University Sriracha Campus, Information Technology students and Computer Science students. Using the clustering of students before classifying the answers makes the results more accurate, which, in the grouping process, can do more if there was enough student behavior data and found that Ramdomtree was a good way to analyze the answers in the set of data for students in both fields and the more interesting methods such as Ramdomforest. Logistic which had similar results. Due to Ramdomtree can be converted into decision-making rules and see factors that affect the overall path of the analysis of the results of all group predictions were satisfactory. The accuracy was 98.15-100 percent accuracy correct. The results of this research can be used as a guideline for analyzing behavior affecting future academic performance of university students and was also a great help for advisors to advise students or may be an aid to parents of students in solving various problems that occur with their children and students can predict behavior that will affect the results of the study faster.

The predictions of the Decision tree are limited by basic answers to each attribute. The same training factor in the training data sets makes it limited to remembering. Studies have shown that models from the random forest have fewer errors and predictable results. For situations that do not fit too much, it can be seen that from the rules of trees, decide on predictions. Grade from the study behavior of students in each group, student behavior and analyze the class groups in the three big groups and not the same, primary factors different. This research also has a development plan to use information on past academic performance of students in computer science in both branches to help analyze, make decisions, and anticipate courses. This research is also done to find ways to improve learning behavior in order to promote and create behavior that affects good academic performance of that student in the future.

CONFLICT OF INTEREST

This research was no conflict of interest. Since it was a study of behavior that respondents were willing to give facts about the behavior of each person that affects learning in various courses and was also a development to find ways to improve behavior that affects students.

AUTHOR CONTRIBUTIONS

The researcher (Supaporn) conducted the research by collecting data to Information Technology and Computer Science students from Faculty of Science at Sriracha, Kasetsart University Sriracha Campus. This research was funded for research to develop an analysis of educational results by studying learning behavior by categorizing learning and using the information received to analyze and develop applications. Summarize and write articles

REFERENCES

[1] B. Oancea, R. Dragoescu, and S. Ciucu, "Predicting students' results in higher education using neural networks," in *Proc. International Conference on Applied Information and Communication Technologies*, Jelgava, Latvia, April 2013, pp. 25-26.

[2] A. M. Fahim, A. M. Salem, F. A. Torkey, and M. A. Ramadan, "An efficient enhanced k-means clustering algorithm," *Journal of Zhejiang University Science*, May 11, 2006.

[3] Y. Meier, J. Xu, O. Atan, and M. V. D. Schaar, *Predicting Grades*, Mar. 2016.

[4] L. H. Escobar, N. H. Gress and T. I. Matis, "Bayesian based approach learning for outcome prediction of soccer matches," *ICCS Camera Ready Version,* 2018.

[5] Z. Elouedi, K. Mellouli, and P. Smets, "Classification with belief decision trees," *Lecture Notes in Computer Science*, January 2000.

[6] A Comparative study on decision tree and random forest using R tool, *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, issue 1, pp. 196-199, January 2015.

[7] S. Kalmegh, "Analysis of WEKA data mining algorithm REPTree, simple cart and RandomTree for classification of Indian news," *International Journal of Innovative Science, Engineering & Technology*, vol. 2, issue 2, February 2015.

[8] Weka. (December 3, 2016). [Online]. Available: http://www.cs.waikato.ac.nz/ml/index.html

[9] M. Pojon, "Using machine learning to predict student performance," M. Sc. Thesis, p. 35, June 2017.

[10] K. Wisaeng, "A comparison of different classification techniques for bank direct marketing," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, issue4, pp. 116-119, September 2013.

Supaporn Bundasak was born in Chonburi, Thailand in November 1974. She received her BSc. degree in computer science from Burapha University, Thailand in 1997 and received her MSc. degree in computer science from Kasetsart University, Thailand in 2004. She works as a lecturer at the Computer Science and Information Technology Department of Kasersart University Sriracha campus. Her current research interests include recommendation systems, machine learning, data mining, deep learning, big data analytics and data science.