The Use of Multiple-Classifier Data Mining Technique for Study Guideline

Teerasak Sungsri

Abstract—**The purpose of this research is to recommend appropriate courses for students which use multiple-classifier data mining techniques to assist in the work. The process of creating a predictive model use Weka as a tool to build a model. Data classification techniques for finding answers and include results from the prediction of an individual's classifier techniques have majority votes. By segmenting the data used in the research into three groups based on the amount of data. The research found that recommended course, using multiple-classifier data mining techniques has worked with more accurate than using single data mining techniques. We have shown that the most accurate of 72%. In addition, also found the accuracy of the recommend is increased when the amount of data that is used to increase.**

Index Terms—**Data mining, classification, multiple classifier, education.**

I. INTRODUCTION

The Academy is an organization for using data mining technique since there is a lot of storage. Data can be analyzed and utilized in many areas, for example, student information teacher information registration information, etc. Each Academy needs to build a solid relationship between the long-lasting and sustainable institution with its own students by focusing on customer satisfaction. Confidence to the students before admission international students studying in Academy and graduates of the Academy. By applying the principles and concepts of CRM with the Academy called student relationship management [1]. Since the first studies in the Academy until graduating as a graduate. The problem is the lack of information to support the decision of the individual student and provide guidance on the impact on the institution and the learner is huge. The research is divided into several issues as follows: The first is the lack of proper guidance for students before admission to courses that suit their own ability to succeed in Academy. The second is the lack of proper guidelines for international students enrolled in each grade in order to plan properly. Finally, is the lack of guidelines for the proper education graduates to choose the right career. Issues arising from the impact on the learning of students make the grade slump. Students do not graduate on time on the course. Or graduation the impact on the choice of future career. Resulting in loss of time and a lot of money.

Manuscript received January 15, 2017; revised April 1, 2017.

Teerasak Sungsri is with Business Computer Program, Faculty of Management Science, Nakhon Ratchasima Rajabhat University, Thailand (e-mail: mrajaab@gmail.com).

Moreover, the institutions lack the proper education guidelines will result in a conviction. The relationship between the learner and the satisfaction with the institution itself. This research aims to create a predictive model for multiple classifiers to Education. The process of creating a predictive model using Weka as a tool to build a model. For the results obtained from the analysis to help support decisions about education. This will affect relationship satisfaction and confidence. Between the Academy and the student. This research focuses on the creation of a model for guidance courses appropriate for the class by using data classification. In addition, the researchers measured the performance of model. By comparison between the single-classifier data mining techniques with multiple-classifier data mining techniques. The population of this research is that students in the 2007-2010 academic year, Nakhon Ratchasima Rajabhat University.

II. RELATED WORK

The purpose of this research is to help solve the problem of choosing an appropriate course of study. By applying data mining techniques to help in the work. A study in the related research base.

Ref. [2] Such research has led the data mining technique to predict and analyze student's performance as well slow learners among them. The dataset of student academic records is tested and applied by Weka Tool on various classification algorithms such as Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree. The result show that among all data mining classifiers MultiLayer Perception performs best with 75% accuracy. [3] Such research has led the data mining technique todetect malware behavior based on the feature and behavior of each malware. This paper use Weka Tool and some classification algorithms such as NaiveBayse, BayseNet, IB1, J48, and regression algorithms. The result show that the regression method had best performance for classification of malware detection. [4] Such research has led the data mining technique to predict the performance of employees. This paper use the CRISP-DM data mining methodology to build the classification model and use the decision tree technique was the main data mining tool. [5] Such research has led the data mining technique to help guide the tertiary education of students in high school. The technical rules for relationships (Association Rule Discovery) in the search form relationships that affect education in higher education. This research used a sample divided by the proportion of teachers with test data into four types: 60:40, 65:35, 70:30 and 75:25, respectively, the findings have been average. Must include an 82.08 percent

accuracy by increasing the proportion of teachers with more size. [6] Such research has led the data mining techniques to help in the search for the most appropriate department provides students with the technical classification. A decision tree (Decision Tree) to analyze registration data and personal information of students for careers with the Department of the Faculty of Engineering. The results are accurate more than 80 percent. [7] Such research has led the data mining techniques to help in the search for knowledge to guide students in the institution. The proposed algorithm Adaptive Candidate Apriori to find patterns in sequence learning courses. In the second semester, as well as in the classroom. Affecting GPA and improve the efficiency of the algorithm for the better. The findings can be adopted for recommendations of students in higher education as well. [8] Such research has led mining techniques to analyze data to help find patterns or features. To study the factors that affect the education of students in higher education institutions that are likely to graduate and graduate. Using decision tree (decision tree) to analyze trends in future studies. And learning the rules of the association (Association Rule Discovery) to analyze the relationship of the subject in the curriculum that the student is enrolled. The analysis of the decision tree model is accurate about 80 percent. [9] Such research has led the data mining techniques to help in career decisions in the future. The technical rules for relationships (Association Rule Discovery) to help in the search form. And the outlook for early career right after graduation. This research used data of a job to help in the analysis of the relationship. The segmentation of the trainers are three types of data trainer 60 percent accuracy percentage, 65.39, more trainers than 70 percent accuracy percentage of 74.72 and trainers 80 percent accuracy percentage 81.89 respectively. The results showed that the proportion of valid values increase proportionally increased.

III. DATA MINING

Data mining [10]-[12] is a technique that has become popular and is used in various fields. Widely used as the business finance field hospital insurance Telecom and network and science, such as the techniques used to search for new knowledge or relationship patterns hidden in large data warehouses. To be used for the organization of such behavior to create client packages. Or promotion the suit individual customer groups. The process of data mining. Data analysis with data mining techniques. There is a standard process used for data mining called. "Cross-Industry Standard Process for Data Mining" or abbreviated as "CRISP-DM" can break the process of data mining on all six stages as illustrated in Fig. 1.

Stage 1: Business Understanding is the initial step for data mining. The process begins understanding the problem. Or define the problem you want to use data mining techniques. To understand the results of the data analysis. Including planning work.

Stage 2: Data Understanding is a next step for collecting relevant data for data mining. The information must be reliable. The data volume must be large enough for work. The information provided must be appropriate the details for work.
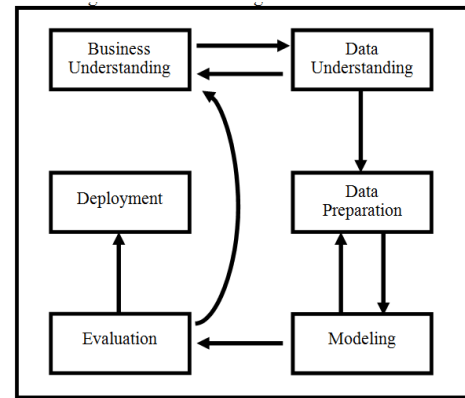

Fig. 1. The standard data mining: CRISP-DM.

Stage 3: Data Preparation is a process for the preparation of the information. To get the right information for each data mining techniques. And provide data integrity. Most accurate by this process, a process that takes the longest to get quality information. Affect the accuracy of the data analysis is divided into three phases: phase one is Data Selection phase two is Data Cleaning and phase three is Data Transformation.

Stage 4: Modeling is the process for creating a model. By using data mining techniques that is Classification Clustering and Association Rule.

Stage 5: Evaluation is the process of measuring the effectiveness or accuracy of the estimates underlying the forecast is generated. To analyze whether the model was built that is accurate. Suitable to use or not. And results to meet or meets the desired goals in the first step or not.

Stage 6: Deployment is the final step that be deployed.

IV. MULTIPLE CLASSIFIERS

The multiple classifiers [13] are being developed to increase the classification performance. It can be called ensemble classifiers or modular classifiers. Ensemble classifiers aim at obtaining highly accurate classifiers by combining less accurate one. They are proposed to improve the classification performance of a single classifier. The result is better performance than using a single classification. The method used to combine different techniques together to find ways as a result of the classification of Majority Voting Bagging and Boosting.
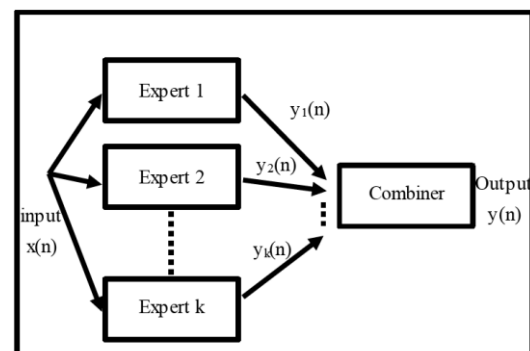

Fig. 2. The architecture of a classifier ensemble.

A. Majority Voting

The simplest method to combine classifiers is majority

voting. [13] The binary outputs of the k individual classifiers are pooled together. Then, the class which receives the largest number of votes is selected as the final classification decision. In general, the final classification decision that reaches the majority of (k+1)/2 votes is taken. Fig. 2 shows the general architecture of a classifier ensemble.

V. METHODOLOGY

Our methodology is categorized into four phases, which include: i) data collections, ii) data preparation, iii) modeling and iv) testing. Fig. 3 presents the phases of our methodology with input and output in each phase. The process beginning with gathering information related to the research. Then, access to data preparation phase. In data preparation, data will be ready with ARFF format to load in Weka software and in modeling step, Modeling step is a step for creating a model to be used for test data by using three techniques for modeling. Finally, Testing is a step for the information to find the answer or want to suggest a proper course. By the answers from each model is calculated by a majority voting and recommend appropriate courses out only one answer.
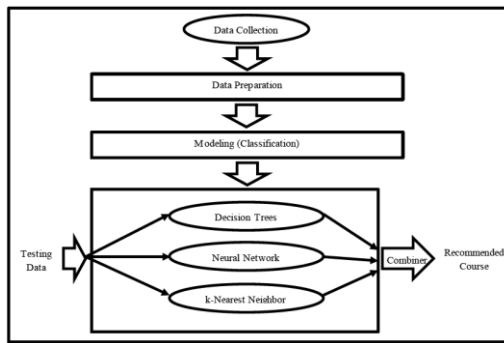

Fig. 3. Research methodology steps.

A. Data Collection

The data set in this research. As the data set in order to create a predictive model for guiding the right course for the class. We used data of new students enrolled in the 2007-2010 academic year. Such as personal information grades and course study.

B. Data Preparation

A process for the preparation of the information. To get the right information for each data mining techniques which provide data integrity and most accurate. We put each group of data through the selection process. Screening information and transform data to get the information that is most suitable for deployment in the process of creating a model.

C. Modeling

The data mining is used to analyze the data for modeling. The technique used in this work include Decision Tree K-nearest neighbors and Neural Network.

D. Testing

A step in bringing information to find the answers to recommend the right course. By the results of each model through the combiner and use majority voting technique to apply the answers only to recommend the right course out.

VI. EXPERIMENTAL RESULT

The purpose of this research is to recommend appropriate courses for students which use multiple-classifier data mining techniques to assist in the work. The data used in the research is a random number of new students from 4000 people. Table I illustrates a random number of data used in the research.

The sample data were collected to be used as Weka software for data analysis. And the data is taken into the Weka program to modeling and each model has to be tested in both the single and multiple classifiers. The single classifier techniques include decision trees K-Nearest Neighbors and Neural Network which is the results of each technique to show only one answer. The multiple classifier techniques will be multiple answers and were combined using majority voting for only one answer. Fig. 4 shows a comparison of performance between the multiple-classifiers with the single classifier. By Series1 is the data set contains 700 data. Series2 is the data set contains 2100 data and Series3 is the data set contains 3500 data.

TABLE I: A RANDOM NUMBER OF DATA USED IN THE RESEARCH

Courses	Data Set1	Data Set2	Data Set3
Bachelor of Science	100	300	500
Bachelor of Business Administration	100	300	500
Bachelor of Education	100	300	500
Bachelor of Arts	100	300	500
Bachelor of Laws	100	300	500
Bachelor of Communication Arts	100	300	500
Bachelor of Industrial Technology	100	300	500

Fig. 4 shows the percentage of accuracy that would suggest a proper course. It can be seen that the decision tree technique has accurate minimum percentage and multiple classifiers has accurate maximum percentage. In addition, the percentage of correct predictions is increased by the amount of data used in the research. To see that the most percentage accurate is multiple classifiers technique. The accuracy was 72%.

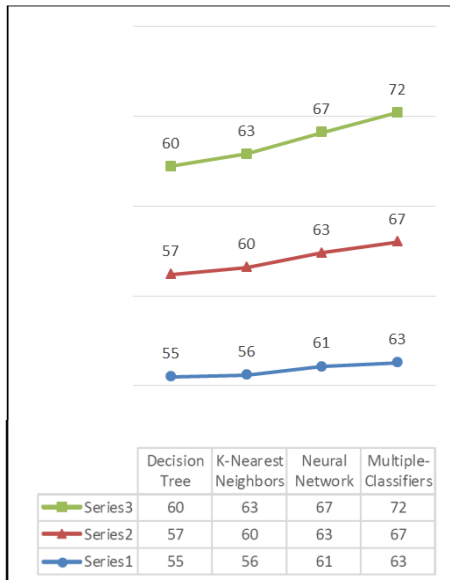|	Decision Tree	K-Nearest Neighbors	Neural Network	Multiple-Classifiers
Series3	60	63	67	72
Series2	57	60	63	67
Series1	55	56	61	63

Fig. 4. A comparison of performance between the multiple-classifiers with the single classifier.

VII. CONCLUSION

This research aims to improve the accuracy of the recommended courses. By adopting a multiple-classifiers and majority voting to help in the work. The work found that the results of using the multiple-classifiers accuracy than a single-classifiers (such as decision tree k-nearest neighbors and neural network).We have shown that the most accurate of 72% from 3500 data. In the other word, the operating result shows that the use of a single classification in predicting the percentage is less accurate than using multiple techniques to classify. Due to the use of a single classification. Each classification may be predictors wrong but when brought together, chances are forecast to increase. In the future, we will improve the accuracy to better than this. By other techniques involving multiple-classifiers to help in the work (such as Bagging Boosting) and we can use the various courses and increase the number of attributes that are relevant to this work in the future. In addition, this research can be used as a guide for choosing the right course for the future.

ACKNOWLEDGMENTS

REFERENCES

[1] Sripatum University. Student relationship management: SRM. [Online]. Available: http://aster.spu.ac.th/file/user/39/39/upload/Radio/SRM.pdf
[2] P. Kaur, M. Singh, and G. S. Josan, "Classification and prediction based bata mining algorithms to predict slow learners in education sector," in *Proc. 3rd International Conference on Recent Trends in Computing*, 2015.
[3] M. Norouzi, A. Souri, and M. S. Zamini, "A data mining classification approach for behavioral malware detection," *Journal of Computer Networks and Communications*, vol. 2016, Article ID 8069672.
[4] Q. A. Al-Radaideh and E. Al Nagi, "Using data mining techniques to build a classification model for predicting employees performance," *International Journal of Advanced Computer Science and Applications,* vol. 3, 2012.
[5] N. Jareankieatbovon, "Applied data mining technique for guiding to higher education," Department of Computing Graduate School, Silpakorn University. Thailand, 2006.
[6] C. Songsiri, T. Rakthanmanon, and K. Waiyamai, "Applying a data mining technique to help students in selecting their majors," Engineering Symposium of Kasetsart University, Thailand, 2001.
[7] K. Haruehansapong, "Extracting knowledge from student database using data mining techniques: A case study of Walailak University," Management of Information Technology Graduate School, Walailak University, Thailand, 2006
[8] P. Yamansabideen, "Student relationship management in Hegher education institutions with application of data mining," Information Technology Graduate School, Sripatum University, Thailand, 2006.
[9] B. Phengsuan, "Applied data mining techniques to develop a decision support system for graduate production of bachelor's degrees," Department of Computing Graduate School, Silpakorn University, Thailand, 2005.
[10] E. Pacharawongsakda, "An introduction to data mining techniques," *Data Cube*, Thailand, 2014.
[11] J. Han and M. Kamber, *Data Mining Concepts and Techniques,* Second Edition, University of Illinois at Urbana-Champaign, 2006.
[12] T. Theeramunkong, *Introduction to Concepts and Techniques in Data Mining and Application to Text Mining*, Second Edition, Thammasat University, Thailand, 2012.
[13] C. F. Tsai and C. Hung, "Automatically annotating image with keywords: A review of image annotation systems," *Recent Patents on Computer Science*, 2008.

Teerasak Sungsri is a lecturer in Business Computer Program Faculty of Management Science at Nakhon Ratchasima Rajabhat University in Thailand. He received his MSc. in computer science from Khon Kaen University, Thailand. His research interests include data mining education data mining and text mining. In addition, he was also interested in the programming as a C#.NET, ASP. Net, PHP, R etc.