Prediction Students' Performance in Elective Subject Using Decision Tree Method

Suhainy binti Sulaiman[1]

[1](Information Technology and Communication Department, Politeknik Ungku Omar, Perak, Malaysia)

ABSTRACT *: In polytechnic system, a student must take the elective subjects at least 3 subjects to complete their study. The elective subjects were chosen based on their interest and first come first serve. The result obtained for elective subjects in final examination will affect their future. It is important to predict whether they pass or fail in final examination. Literature survey (LS) was used to obtain the information about students' profile and current approaches in predicting the students' performance using data mining. In this paper, the researcher uses data mining which is decision tree method to predict the students' performance in elective subject. The aim of this research is to evaluate the students result in choosing the correct elective subjects. This research is focused on the ICT students who select DBM3033 as an elective subject. Two phases involved which are prepossessing data and mining data. RapidMiner software is used in mining data process. Classification technique is applied for decision tree method. The research findings showed that students whose result weak in both SPM Mathematics and DBM1033 are predicted as fail in final examination for DBM3033.*

KEYWORDS *: Data Mining, Classfication, Decision Tree*

_____

I. INTRODUCTION

Students in polytechnic have to take the elective subjects to complete their diploma level. The subjects are depending on the programme structure. It is an optional subject and not a core subject. Basically, the students can choose it but it depends on a number of students in the class. Some students choose the incorrect subject and in consequence received unsatisfactory result in final examination or otherwise they do not finish their study.

Therefore, by realizing the important of choosing the correct elective subject, this research will highlight that the elective subject is important as the core subjects in having a good result. The selection of elective subjects based on first come first serve. Sometimes occuring the situation in which he or she do not interested in the subject but they had no choice. Some students choose the subject based on the lecturer. The normal senario in the elective class is the gap between the excellent and weak students. This situation creates difficulties to lecturers during their teacjhing and learning process. They have to spend more time to those students who have difficulties in the subject. This is the reason why the reasearch needs to be initiated to find out how the data mining can assist in solving the problem [1].

From the observation, there are some students failed to finish their study or having a good result due to making a wrong decision in choosing the elective subjects.   For ICT students, to complete their study, they have to take 3 elective subjects. Normally, they took those elective subjects while they were in fourth and fifth semester.

This study will involve a 3 years group of ICT students who are from June 2011 till December 2013. It will cover 6 semester group of students. It will focus on fourth and fifth semester students who take DBM3033 as an elective course. So, the analysis is conducted based on their mathematics 1st semester result to see the pattern and the correctively with the DBM3033 result.

Based on the observation, students who choose the right subject are getting a good result. To obtain a good result, the students should choose the suitable subject based on their pre knowledge and skills as well as their interest. The students who choose the correct subject basically will easy to understand the lecture and have not problems in attendance. Based on the findings, a further research to find the best solution to solve this problem is addressed.

There are three objectives of this research. The first is to identify the student performance according to their examination result. The second is to utilize the data mining technique for analyzing the student record. The third is to evaluate the students' result in choosing the correct elective subject.

According the previous researchers, data mining was used to predict the students' performance. Decision tree technique was applied to classify the students' data into students' categories [1]. Previous reseacher had also used the assosiation rules algorithm techniques to identify students' failures pattern [2]. The lecturers can use the pattern from the analyze result in making the decision to enhance the learning process. The result also helps to build the curriculum structure in order to improve students' academic performance and trim down failure rate.

The past researcher used the association rules algorithm in enhancing the quality and experiences of students' performance in higher education [3]. According to Jeevalatha [4] , data mining can be used in educational field to enhance the understanding of learning process to focus on finding, extracting and validating variables related to the student learning process

Knowledge Discovery in database (KDD) is very useful [5] in student centered learning. It concerns with developing new methods to discover knowledge from educational database in order to analyze student's trends and behaviors towards education. "Education Data Mining is one of field, that consider with developing methods in exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in." www.educationaldatamining.org.

To predict the values of variable, the classification can be used. Learning and classification are two processes those involve in the data classification process. In Learning, classification algorithm analyzes the training data. In estimating the accuracy of the classification rules, the test data are used [6]. Dorina Kabakchieva applied a decision tree as classification method in his research [7]. The objective of the research is to predict students based on their personal and pre-university performance using any patterns that available data. It presented the data mining initial result that was implemented at a Bulgarian university which aimed to reveal the high potential of data mining applications for university management.

In this study, the decision tree method is used. A decision tree is a flow-chart-like tree structure, where each internal node is denoted by rectangles, and leaf nodes are denoted by ovals. All internal nodes have two or more child nodes. All internal nodes contain splits, which test the value of an expression of the attributes. Arcs from an internal node to its children are labelled with distinct outcomes of the test. A class label associates with each leaf nodes [8]. Based on the accuracy, the decision tree is more accurate compared than Bayesian Networks [9].

II. METHODOLOGY

A. Research Methodology

This section elaborates the process of predicting the students' performance. During of data gathering, the literature survey (LS) was used to obtain information about students' profiles and current approaches in predicting the students' performance using data mining. For this research, student data was taken from iPUO systems. As a range of research, ICT students data from year 2011 to 2013 was taken and it is involved fourth and fifth semester students . The selected semesters are June 2011, December 2011, June 2012, December 2012, June 2013 and December 2013. Each session would take the students from final year that had fifty plus number of students who took DBM3033. Evidently, not all of the students would complete their study till the end of semester. With that facts, the researcher had to make sure the number of student must be same. From the final result spreadsheet only DBM3033 marks will be selected to be processed. During the phase of Preprocessing Data, the collected data would be copied into Ms. Excel file. As part of data preparation and preprocessing of the data set and to get better input data for data mining technique, it needs to do some preprocessing for the collected data before it can load into the data mining software. Figure 2.1 shows the proposed frame work for this research.
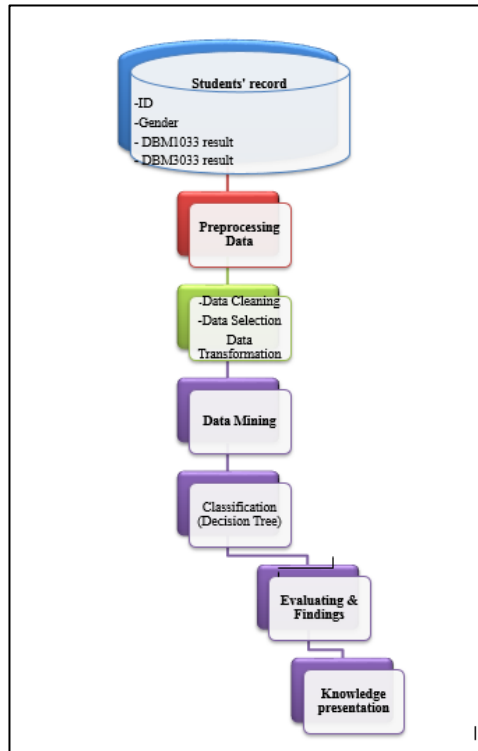
Figure 2.1 : Proposed frame work

B. Population of the research

In this paper, the focus area was department of Information Communication Technology (ICT). The sample consisted of 592 final year students. This study was focusing on DBM3033 subject as elective subject because in this subject it had final examination.

C. Research Tools

For preprocessing phase, the selected students data was cleaned and transformed before mining process using RapidMiner Software. For this research, RapidMiner version 5.3 as a software requirement will be used. The Rapid Miner needs to be installed with a Java Run Time Environment (JRE) version 1.5 or higher. RapidMiner is a commonly used data mining tool since it provides a friendly graphic user interface. The drag-and-drop feature of RapidMiner can extremely provide the facility to the engineers since the operation blocks are ready-made so that the complex codes are hidden behind [10].

D. Analyze Data

For this research, decision tree method was chosen to predict the student performance based on the selected elective subject. DBM1033 final result and Math SPM final result are the two factors that were considered in this research.

Figure 2.2 presents the strong rules that had been generated from the mining process which gained all of valuable results. All of the rules could be classified to predict the students' final result of elective subject (DBM3033) . This prediction was based on their SPM Math and DBM1033 result. For example , the student who had very good in SPM Math and DBM1033 result was predicted as excellent in her elective subject (DBM1033) final result. For those who weak in SPM Math and DBM1033 result, they were predicted as fail in final result for elective subject (DBM3033).

CONDITION	GENDER	SPM MATH	DBM1033	FINAL
IF	MALE	VERY GOOD	VERY GOOD	EXCELLENT
	MALE	GOOD	GOOD	HONORS
	MALE	FAIR	GOOD	PASS
	MALE	WEAK	WEAK	FAIL
	MALE	VERY GOOD	GOOD	HONORS
	MALE	GOOD	VERY GOOD	HONORS
	MALE	GOOD	FAIR	PASS
	MALE	WEAK	FAIR	PASS
	MALE	FAIR	WEAK	PASS
	FEMALE	VERY GOOD	VERY GOOD	EXCELLENT
	FEMALE	GOOD	GOOD	HONORS
	FEMALE	FAIR	GOOD	PASS
	FEMALE	WEAK	WEAK	FAIL
	FEMALE	VERY GOOD	GOOD	HONORS
	FEMALE	GOOD	VERY GOOD	HONORS
	FEMALE	GOOD	FAIR	PASS
	FEMALE	WEAK	FAIR	PASS
	FEMALE	FAIR	WEAK	PASS

Figure 2.1 : Prediction rule Model

E. Implementation

The experiment used classification technique. In classification, the decision tree algorithm would be applied.

III. RESULT AND DISCUSSION

For this research, all experiments are conducted using RapidMiner Software. It split into two data set to validate it. The larger set data is used as training and the smaller data set used as testing. The accuracy for both data set is not a big difference. By referring to Figure 3.1, if there are no differences between the real final result and prediction final result, it is the best accurate. The mean accuracy for training data set is 76.19% whereas for testing is 77.5%. Referring to the experiments result, there is a small number of students whose result did not predict correctly. he miss accurate values are also measured. There are four items which are missed accurate value; 2 increased grade, 1 increased grade, 2 decreased grade and 1 decreased grade. The research findings result show that two factors those influence the prediction final result are DBM1033 and SPM Math result. Students whose result are very good in DBM1033 and SPM Mathematics will score excellent in the final result for DBM3033 whereas who is weak result in both will fail in the final. So, students whose result did not good enough or not strong in one or both factor subjects can choose another elective subject that offered. Therefore, the students who want to select DBM3033 as their elective subject must be strong in DBM1033 and SPM Mathematics in order to score in the final for DBM3033.

ROW NO	GENDER	FINAL	CLASS FINAL	PREDICTION (FINAL)	CLASS FINAL	BA102	SPM MATH	DIFFERENCES
1	MALE	PASS	2	PASS	2	GOOD	GOOD	0
2	FEMALE	EXCELLENT	4	EXCELLENT	4	VERY GOOD	GOOD	0
3	MALE	PASS	2	PASS	2	FAIR	FAIR	0
4	MALE	FAIL	1	FAIL	1	WEAK	WEAK	0
5	MALE	FAIL	1	FAIL	1	WEAK	WEAK	0
6	MALE	HONORS	3	PASS	2	FAIR	FAIR	1
7	FEMALE	PASS	2	PASS	2	FAIR	FAIR	0
8	MALE	PASS	2	EXCELLENT	4	VERY GOOD	VERY GOOD	-2
9	MALE	HONORS	3	EXCELLENT	4	VERY GOOD	VERY GOOD	-1
10	FEMALE	EXCELLENT	4	PASS	2	GOOD	VERY GOOD	2
11	FEMALE	PASS	2	PASS	2	GOOD	GOOD	0
12	MALE	FAIL	1	PASS	2	FAIR	FAIR	1
13	FEMALE	PASS	2	PASS	2	FAIR	FAIR	0
14	FEMALE	PASS	2	PASS	2	GOOD	GOOD	0
15	FEMALE	PASS	2	PASS	2	GOOD	GOOD	0
16	FEMALE	EXCELLENT	4	EXCELLENT	4	VERY GOOD	VERY GOOD	0
17	FEMALE	EXCELLENT	4	EXCELLENT	4	VERY GOOD	VERY GOOD	0
18	FEMALE	EXCELLENT	4	EXCELLENT	4	VERY GOOD	VERY GOOD	0
19	FEMALE	PASS	2	PASS	2	GOOD	FAIR	0
20	MALE	FAIL	1	FAIL	1	WEAK	WEAK	0

Figure 3.1 Classification on Grade for Training Data Set

NO	EXPERIMENT DATA	NUMBER OF STUDENTS	ACCURACY	PERCENTAGE THE STUDENTS WHOSE RESULT PREDICT NOT CORRECT			
				INCREASE		DECREASE	
				2 GRADE	1 GRADE	2 GRADE	1 GRADE
1	JUNE 2014	92	77.17%	0.00%	14.13%	0.00%	8.70%
2	DECEMBER 2013	128	59.38%	3.13%	11.71%	10.94%	14.84%
3	JUNE 2013	98	77.55%	0.00%	3.06%	0.00%	19.39%
4	DECEMBER 2012	171	69.24%	0.00%	2.92%	0.50%	27.40%
5	JUNE 2012	42	76.19%	0.00%	4.76%	0.00%	19.05%
6	DECEMBER 2011	61	67.12%	0.00%	24.95%	0.00%	8.20%

Figure 3.2 Combination All of Experiments Result

	STUDENTS WHOSE RESULT PREDICT CORRECT	PERCENTAGE (%)	STUDENTS WHOSE RESULT DID NOT PREDICT CORRECT	PERCENTAGE (%)
TESTING DATA	93	77.5	27	22.5
TRAINING DATA	421	71.11	171	28.89

Figure 3.3 Summarizing Result

IV. CONCLUSION

In this research, decision tree is proposed to predict the students' performance in selecting the correct elective subjects. The decision tree is an approached method for classification. Some attributes are collected from the students' database record to predict their final result in DBM3033. From the experiments, the average of training data accuracy is 71.11% and the accuracy for the testing data is 77.50%. Therefore, it looks like the accuracy is still in good range.

Based on the findings, it can evaluate the students' result in choosing the correct elective subject. Therefore, it can improve their performance in that subject. From the experiments, the data mining model that has been used is a good enough to predict students' DBM3033 result. It is supported by the range value 2 happened in a very small percentage of students. There might be some factors affected their performance such as family contribution, attendance, peers influence, awareness, study group and the others.

Prediction of the performance students in selecting the correct elective subject helps the lecturers to prepare a better way in their teaching and learning process   It also like a warning system to detect the potential students such as the students who weak in the subject will get more attention from the lecturers. Moreover, the academic advisory can play the role to advise them to drop or register other subjects that suitable with their knowledge or skill to prevent failure in the examination. All the approaches can improve the percentage number of students who score flying colors in the final examination and directly, it helps the department to improve the targeted CGPA 3.5 and above.

For future study, more factors such as attendance, CGPA and GPA should be considered to get more patterns   of   the   result

REFERENCES

[1]   F. Ahmad, N. H. Ismail, and A. A. Aziz, "The Prediction of Students ' Academic Performance Using Classification Data Mining Techniques," *Appl. Math. Sci.*, vol. 9, no. 129, pp. 6415–6426, 2015.

[2]   A. A. Aziz, N. U. R. H. Ismail, and F. Ahmad, "MINING STUDENTS ' ACADEMIC PERFORMANCE," *J. Theor. Appl. Inf. Technol.*, vol. 53, no. 3, 2013.

[3]   Anwar, M.A. and Ahmed, N. 2011. Knowledge Mining in Supervised and Unsupervised Assessment Data of Students ' Performance. *2011 2nd International Conference on Networking and Information Technology IPCSIT vol.17 (2011) © (2011) IACSIT Press, Singapore Knowledge* 17(vol 17), pp. 29–36.

[4]   Jeevalatha, T. 2014. Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms. *International Journal of Computer Applications (0975-8887* 108(15), pp. 27–31.

[5]   R. B. Bhise, S. S. Thorat, and A. K. Supekar, "' Importance of Data Mining in Higher Education System ,'" *IOSR J. Humanit. Soc. Sci.*, vol. 6, no. 6, pp. 18–21, 2013.

[6]   Bhardwaj, B.K. 2011. Data Mining : A prediction for performance improvement using classification. *Internationa Journa of computer Science and Information Security* 9(4), p. 5.

[7]   Kabakchieva, D. 2013. Predicting Student Performance by Using Data Mining Methods for Classification. *Cybernetics and Information Technologies* 13(1), pp. 61–72. Available at: http://www.degruyter.com/view/j/cait.2013.13.issue-1/cait-2013-0006/cait-2013-0006.xml [Accessed: 8 June 2014].

[8]   Badr, A. et al. 2014. Data Mining : A prediction for Student ' s Performance Using Classification Method. *World Journal of Computer Application and Technology* 2(2), pp. 43–47.

[9]   Kabra, R.R. and Bichkar, R.. 2011. Performance Prediction of Engineering Students using Decision Trees. *International Journal of Computer Application (0975-8887)* 36(11), pp. 8–12.

[10]   Kitcharoen, N. et al. 2013. RapidMiner framework for manufacturing data analysis on the cloud. *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 149–154. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6567336.