PREDICTION OF STUDENTS ACADEMIC ACHIEVEMENT IN UniMAP USING DATA MINING APPROACH

by

FATIHAH BINTI AZIZ
(1232110765)

A thesis submitted in fulfilment of the requirements for the degree of Master of Science in Engineering Mathematics

Institute of Engineering Mathematics
UNIVERSITI MALAYSIA PERLIS

2016

ACKNOWLEDGMENT

Alhamdulillah, with the blessing of Allah S.W.T., The Most Gracious and the Most Merciful, this thesis was completed as a requirement in attaining the master of science in mathematical engineering.

Special thanks to both of my parents, Mr Aziz bin Abu and Mrs Rokieh binti Ishak for their love and support that they bestowed upon me. A million years from now you both know that I will love you still. Also a million thanks to my sisters, Fatin binti Aziz and Farah Adibah binti Aziz who has faith in me and keeps supporting me.

I am indebted to JMP Pro 11 and Microsoft Visual Basic Express 2010 team that had developed such a useful and practical computer programming that helps me in building my decison tree model and simple prediction system.

A million thanks to my supervisor, Assoc. Prof. Dr. Abdul Wahab bin Jusoh for helping and guiding me through the completion of the project. Also to the other co-supervisor, Mr. Syafarudy Abu for his guide and concern during the completion of this project.

I wish to acknowledge the help and assistance received from my colleagues, lecturers and friends for their help in various ways that they had given me for the whole years. May Allah bless all of you.

©This item is protected by original copyright

TABLE OF CONTENTS

	PAGE
THESIS DECLARATION	i
ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS	xii
LIST OF SYMBOL	xiii
ABSTRAK	xv
ABSTRACT	xvi
CHAPTER 1 INTRODUCTION	
1.1 Overview of the study	1
1.2 Problem statements	3
1.3 Objectives of the study	4
1.4 Research Scope	5
1.5 Significance of study	5
1.6 Limitation of study	6
1.7 Organization of thesis	7

©This item is protected by original copyright

CHAPTER 2 LITERATURE REVIEW

2.1	Introduction	10
2.2	An Overview of Knowledge Discovery	11
	2.2.1 Data Mining process	13
	2.2.2 Data Mining in various sectors	14
2.3	Data Mining in Prediction and Classification	15
2.4	Educational Data Mining (EDM)	18
2.5	Educational Data Mining in Malaysia	19
2.6	Factors that influence students' academic achievement	21
2.7	Classification Model	25
2.8	Studies on the prediction of students' academic achievement using C4.5 algorithms	26

CHAPTER 3 METHODOLOGY

3.1	Introduction	31
3.2	Framework for the research methodology	32
3.3	Universiti of Malaysia Perlis	33
	3.3.1 Collecting Data from UniMAP	34
3.4	Preprocessing of the collected data	37
	3.4.1 Selection of data	37
	3.4.2 Transformation of data	37
3.5	Statistical Analysis using Pearson's Chi-Square test	38
3.6	Validating	40
3.7	Decision Tree Model	41
	3.7.1 C4.5 algorithm	43
	3.7.2 Information and Entropy	45

©This item is protected by original copyright

3.7.3 Information Gain 46

3.7.4 Gain Ratio 46

3.7.5 Stopping Criteria 47

3.8 Knowledge Representation 48

3.9 Model Evaluation 49

3.9.1 Confusion Matrix 50

3.9.1.1 Sensitivity, Specificity and Precision 51

3.9.1.2 Receiver Operating Characteristic Curve (ROC) 53

CHAPTER 4 ANALYSIS OF DATA

4.1 Introduction 56

4.2 Data Distribution of students in UniMAP 57

4.3 Preprocessing of the collected data 62

4.3.1 Selection of Data 62

4.3.2 Transformation of Data 64

4.4 Statistical Analysis using Pearson's Chi-Square test 68

4.5 Development of the decision tree model using the students' data from UniMAP 73

4.6 Model Evaluation 86

4.6.1Confusion Matrix 86

4.6.2 Receiver Operating Characteristic Curve (ROC) 89

4.6.2.1 Plotting the ROC Curve 89

CHAPTER 5 RESULTS AND DISCUSSION

5.1 Introduction 96

5.2 Statistical Analysis 97

5.2.1 Overall result of the Pearson's Chi square Test 97

©This item is protected by original copyright

5.3 Contingency Table between the Factor Variables and Response Variable 99

5.3.1 Contingency Table between Gender and Category of Graduate CGPA 99

5.3.2 Contingency table between class of age and category of graduate CGPA 100

5.3.3 Contingency table between Between Entry Qualification and Category of Graduate CGPA 102

5.3.4 Contingency Table between Class of Entry CGPA and Category of Graduate CGPA 104

5.3.5 Contingency Table between Classes of CGPA in Semester 1And Category of Graduate CGPA 105

5.3.6 Contingency Table between Class of CGPA in Semester 2 and Category of Graduate CGPA 106

5.3.7 Contingency Table between Schools and Category of Graduate CGPA 107

5.4 Prediction Result by Decision Tree Model 109

5.5 Evaluation of the Decision Tree 115

5.5.1 Confusion Matrix 115

5.4.2 Receiver Operating Characteristic Curves (ROC) 117

CHAPTER 6 CONCLUSION

6.1 Introduction 119

6.2 Conclusion 119

6.3 Future Research 122

REFERENCES 123

©This item is protected by original copyright

APPENDIX 132

LIST OF PUBLICATIONS 141

©This item is protected by original copyright

LIST OF FIGURE

NO.		PAGE
2.1	Knowledge Discovery Process	13
3.1	The framework for this research methodology	33
3.2	The summary of data types	35
3.3	The steps for chi-square test for independence	40
3.4	The decision tree represents the students who would be graduated as excellent (blue) or average (red)	42
3.5	The output of decision tree represents the student who would be graduated as Excellent (blue) or Average (red).	48
3.6	The example of ROC curve for graduated CGPA	54
3.7	The steps for developing the ROC curve	55
4.1	The variable Semester 1 (X3) is the root node for decision tree from 25 samples of students in UniMAP (1st split)	82
4.2	The second leaf node splits on the variable Entry Qualification (2nd split)	84
4.3	Final result of the decision tree model using 25 samples of students' data taken randomly from UniMAP	84
4.4	The ROC curve	94
5.1	ROC curve for the validation data set	118

© This item is protected by original copyright

LIST OF TABLES

NO.		PAGE
3.1	The raw data collection from Unimap's students	36
3.2	Data before the transformation process	38
3.3	Data after the transformation process	38
3.4	Decision Tree Algorithms	43
3.5	Confusion matrix for two classes	51
3.6	General interpretation of area under the curve for ROC	54
4.1	Distributions of class of age	57
4.2	Distributions of gender	58
4.3	Distributions of entry qualifications	58
4.4	Distributions of class of entry CGPA	59
4.5	Distributions of Schools	59
4.6	Distributions of class of semester 1	60
4.7	Distributions of class of semester 2	61
4.8	Distributions of class of graduate CGPA	61
4.9	Raw data of students in UniMAP	63
4.10	Raw data of students in UniMAP	64
4.11	The selected variables for the UniMAP's students	66
4.12	The transformation data of the UniMAP's students	67
4.13	The Category of graduate CGPA by Gender	69
4.14	The observed and expected value of category of graduate CGPA by gender	70
4.15	25 samples of data from student of UniMAP	74
4.16	Categories of Graduated CGPA	75

©This item is protected by original copyright

4.17	Gender by Category of Graduate CGPA	76
4.18	Class of Age by Category of Graduated Students	76
4.19	Entry Qualification by Category of Graduate CGPA	76
4.20	Class of Entry CGPA by Category of Graduate CGPA	76
4.21	Result of Semester 1 by Category of Graduate CGPA	76
4.22	Results of Semester 2 by Category of Graduate CGPA	77
4.23	Schools by Category of Graduate CGPA	77
4.24	The value of entropy for the factor variables which are relatives to Category of Graduate CGPA	78
4.25	The value of Information Gain for the factor variables which are relatives to Categories of Graduate CGPA	79
4.26	The value of Information Gain for the variables which are relatives to Categories of Graduate CGPA	80
4.27	The value of Gain Ratio of the variables which are relatives to Category of Graduate CGPA	81
4.28	The calculations of 2nd upper leaf node	83
4.29	Confusion matrix show the result from the decision tree	87
4.30	The samples of data for building the ROC curve	91
4.31	The sorted data according to the probability in decreasing order	92
4.32	The sorting data according to the probability of excellent students	93
4.33	The confusion matrix for cutoff value 0.9243	93
4.34	The cut value for TP, TN, FP, FN and FPR	9494
5.1	The result of the Chi square test for independence of all the variables	98
5.2	Contingency Table Between Gender And Category Of Graduate CGPA	99
5.3	Contingency Table Between Class Of Age And Category Of Graduate	100

©This item is protected by original copyright

CGPA

5.4 Contingency table between class of age and entry sessions 100

5.5 Contingency Table between Entry Qualification and Category of Graduate CGPA 102

5.6 Contingency Table between Class of Entry CGPA and Category of Graduate CGPA 104

5.7 Contingency Table between Classes of CGPA in Semester 1And Category of Graduate CGPA 105

5.8 Contingency Table between Class of CGPA in Semester 2 and Category of Graduate CGPA 106

5.9 Contingency Table between Schools and Category of Graduate CGPA 107

5.10 Contingency Table between Schools and Entry Sessions 108

5.11 The result of the prediction from decision tree model 11

5.12 The result of the confusion matrix for the validation data set 115

5.13 Results of sensitivity, specificity, precision, accuracy and error rate of the decision model 115

©This item is protected by original copyright

LIST OF ABBREVIATIONS

AIRS2	Artificial Immune Recognition System 2
ANN	Artificial Neural Network
APU	Asian Pacitic University
C4.5	Classifier Version 4.5
CART	Classification and Regression Tree
CGPA	Cumulative Grade Point Average
EDM	Educational Data Mining
GPA	Grade Point Average
ID3	Iterative Dichotomiser 3
IPTA	Institute of Higher Learning
JMP	Visual Statistical Discovery Software
KDD	Knowledge Discovery in Database
MAIRS	Modified Artificial Immune Recognition System 2
UPNM	Universiti Pertahanan Nasional Malaysia
NKRA	National Key Result Area
ROC	Receiver Operating Characteristic Curve
STPM	Malaysian High School Certificate
SVM	Suport Vector Mechine
UniMAP	Universiti of Malaysia Perlis
UPSI	Universiti Pendidikan Sultan Idris
WCIT	World Congress on Computing and Information Technology

©This item is protected by original copyright

LIST OF SYMBOLS

c	total column of the table
r	total row of the table
H_0	null hypothesis
H_A	alternative hypothesis
α	alpha
χ^2	chi square
A_n	Element of variable ID
B_n	Element of variable Age
C_n	Element of variable Gender
df	Degree of Freedom
D_n	Element of variable Nationality
E_n	Element of variable Sessions
FN	False Negative
F_n	Element of variable Entry Qualification
FP	False Positive
FPR	False Postive Rate
G_n	Element of variable Entry CGPA
H_n	Element of variable Programmed
J_n	Element of variable Semester 1
K_n	Element of variable Semester 2
L_n	Element of variable Semester 3
M_n	Element of variable Semester 4

©This item is protected by original copyright

N_n	Element of variable Semester 5
P	Information Gain for variables
P_best	Highest Information Gain
P_n	Element of variable Semester 6
Q_n	Element of variable Semester 7
R_n	Element of variable Semester 8
S_n	Element of variable Graduated CGPA
TN	True Negative
T_n	Element of variable Status
TP	True Positive
TPR	True Posotive Rate
X1	Class of age
X2	Gender
X3	Entry Qualification
X4	Result of Semester 1
X5	Result of Semester 2
X6	Entry CGPA
X7	Faculty

©This item is protected by original copyright

Ramalan Pencapaian Akademik Pelajar di UniMAP Menggunakan Pendekatan Perlombongan Data

ABSTRAK

Kajian ini adalah mengenai ramalan pencapaian akademik pelajar-pelajar di Universiti Malaysia Perlis (UniMAP) dengan menggunakan teknik perlombongan data. Objektif utama kajian ini adalah untuk meramal pencapaian akademik pelajar-pelajar yang akan berjaya, yang akan meneruskan pengajian mereka di peringkat yang lebih tinggi serta pelajar yang memerlukan bantuan dalam pembelajaran mereka. Selain itu, objektif khusus kajian ini adalah untuk mengenal pasti faktor-faktor yang mempengaruhi pencapaian akademik pelajar, untuk membangunkan model pokok keputusan bagi membuat ramalan pencapaian akademik para pelajar dan juga untuk menilai prestasi model pokok keputusan yang dibina. Metodologi penyelidikan ini bermula dengan pengumpulan pelajar data dari UniMAP dari tahun 2002 sehingga 2012. Kemudian, diteruskan dengan pelaksanaan model pokok keputusan dengan menggunakan algoritma C4.5 untuk membangunkan model bagi meramalkan pencapaian akademik para pelajar di UniMAP. Algoritma C4.5 adalah pengganti ID3 algoritma yang telah dicipta oleh JR Quilan pada tahun 1986. Di dalam kajian ini, perisian JMP Pro 11 telah digunakan dalam mengembangkan model pokok keputusan tersebut. Dari hasil kajian ini, terdapat enam faktor yang mempengaruhi pencapaian akademik pelajar di UniMAP. Faktor-faktor variabel tersebut adalah jantina, umur, kelayakan kemasukan, kemasukan PNGK, fakulti serta keputusan dari semester 1 dan 2. Selain itu, dari hasil pokok keputusan, perwakilan pengetahuan akan diterjemahkan ke dalam peraturan keputusan menggunakan peraturan IF-THEN. Daripada model ramalan tersebut, 25 peraturan keputusan telah diterjemahkan kepada pengetahuan yang berguna. Kemudian, 25 peraturan keputusan tersebut telah diimplikasikan ke dalam sistem ramalan mudah akademik bagi pelajar-pelajar di UniMAP. Dengan terhasilnya sistem ramalan ini, ia dapat memudahkan serta dapat membantu pelajar-pelajar baru untuk meramalkan hasil akhir mereka pada semester sebelumnya. Daripada kajian ini, didapati bahawa pelajar yang berumur diantara 24 hingga 30 tahun telah menamatkan pengajian dengan keputusan yang cemerlang berbanding dengan pelajar kategori umur yang lain. Manakala, pelajar perempuan dari universiti ini telah menamatkan pengajian dengan keputusan yang lebih cemerlang daripada pelajar lelaki. Dan akhir sekali , daripada kajian ini didapati bahawa pelajar yang masuk universiti dengan kelayakan Matrikulasi bergraduasi dengan keputusan yang lebih baik daripada pelajar Diploma dan STPM.

© This item is protected by original copyright

Prediction of Students Academic Achievement in UniMAP Using Data Mining Approach

ABSTRACT

This research is about the prediction of students' academic achievement in Universiti Malaysia Perlis (UniMAP) by using the data mining approach. The main objective of this research is to predict the academic achievement of the students in UniMAP who will be succeed, will be continued their study in higher level and also students whose need help in their study. Besides that, the specific objectives of this research are to identify the factors which influence the academic achievement of the students, to develop the decision tree model for prediction of the academic achievement of the students and lastly, to evaluate the performance of the decision tree model. Our research methodology started with the collection of students' data from UniMAP since 2002 until 2012. Then, continued with the implementation of the decision tree model by using the C4.5 algorithm to develop the model for predicting the students' academic achievement in UniMAP. C4.5 algorithm is a successor of ID3 algorithm which had been invented by J.R Quilan in 1986. In this research, JMP Pro 11 software has been used in developed the decision tree model. In addition, the finding of this research stated that there were five factor variables had influenced the academic achievement of the students in UniMAP. The factor variables were gender, age, entry qualification, entry CGPA and results from semester 1 and 2. The response variable was the result of graduate students. Moreover, from the result of the decision tree, the knowledge representation has been interpreted into decision rules using IF-THEN rules. From the predictive model, there were 25 decision rules which had interpreted. Later, the rules had been implemented into a simple academic prediction system for the students in UniMAP. This simple prediction system could help new students to predict their final result in the earlier semester. Some of the significant finding from this research, found that students who in between 24 to 30 years old had graduated with excellent result than younger and older adult learners. Female students from this university had graduated with more excellent result than the male students. Lastly, the finding of this research found that students who entry the university with the Matriculation qualification graduated with more excellent result than Diploma and STPM.

© This item is protected by original copyright

CHAPTER 1

INTRODUCTION

1.1 Overview of the study

Malaysia is recognized as the most competitive nation among 142 economies in the world. Malaysia has leveled up its ranking as a result to become a developed country by the year 2020. Malaysia makes the initiative to induce the country from a middle income economy to a high income nation. Furthermore, Malaysia has not only made the initiatives to increase in the economic sector, but also in the other sectors, especially the educational sector which now has been improved to the world standard. The Global Competitiveness Report 2010-2011 released by the World Economic Forum reported that Malaysia has been positioned in 23rd rank among 139 economies based on the quality of the educational system.

One of the initiatives that have been taken by government is through the implementation of National Key Result Area (NKRA) in the educational sector for improving the students' performance. Through this, every citizen has the chance to receive a good education and will not be left behind to pursue their studies at the highest level. Therefore, higher institutions need to take the responsibility by producing high quality students in academic and also co-curriculum. However, the academic performance of students in Malaysia can be affected by internal and external factors. According to Hasan et. al, (2010) some of the factors that influence students' academic performance is the socio demographic; for example age, gender and marital status.

©This item is protected by original copyright

Academic factors such as students' admission qualification and program taken at the university are also included as the causes that can influence student's academic achievement (VanRooyen et. al, 2006). Thus, factors that can influence the performance of the students will be discussed in chapter 2.

Over the past decades, the development of information technology in the educational sector has collected a large of database collection. Some of the researchers have made an interest in manipulating this precious data for advance decision making. Extraction of beneficial information from the huge database is called knowledge discovery in database. Knowledge discovery in database involves these steps: data selected from the data warehouse, preprocessing and selection of the target data, transformation of preprocessing data, data mining and data interpretation into understandable knowledge.

Data mining process is a part of the process in the knowledge discovery. The general meaning of data mining is the process of extracting hidden patterns from the large databases using specific method to produce the useful information for advance decision making. The use of data mining has become a very popular tool to manipulate databases in forecasting, classifying, clustering and making correlation between attributes (Robert, Hilderman, & Howard, 1999) and (Han & Kamber, 2006).

©This item is protected by original copyright

1.2 Problem Statements

The development of the education sector is in conjunction with the growth of technological advances. Many institutions of higher education have their own system for storing the information of students. However, an abundance of data is not used and exploited. The institutions have a lot of data, but still poor of information. In the case of UniMAP, the data mining approach can be done by the university's management, so that the students' database could benefit them. In addition, the result from discovering knowledge can be used to predict the future decision making.

Yet, there is still no documentation available on the prediction of UniMAP students' academic achievement using data mining in degree level. Therefore, UniMAP can acquire useful knowledge to improve the decision making in their management process especially in academic achievement.

In addition, students' academic achievement is important in evaluating the university quality and helping students to find better jobs after graduating. Suja et al., (2012) said that the students' academic achievement is measured by the cumulative grade point average (CGPA). The grade points for each of these grades are between 4.0 and 1.0 which is A until D, while the F grade gives a grade point of 0. The students' academic performance could be affected by socio demographic factors such as age, gender and academic factors; entrance qualification and courses taken in university (Pitt at. el, 2012). Therefore, an appropriate methodology needs to be proposed for UniMAP, to analyze and identify the factors that can influence students' academic achievements. Hence, the application of a suitable method can help the university management in predicting the students: who will be successful, who will need assistance and who will be likely to further their studies.

©This item is protected by original copyright

However, the prediction of students' academic achievement might become complex when there are factors which influence the academic achievement need to be considered. Therefore, the university needs to develop a mathematical model to predict the students' academic achievement. Through this methodology, the university can identify the factors that may influence students' academic achievement and a make a prediction towards the similar objective more precisely and accurately.

1.3 Objectives of the Study

The main objective of this study is to develop the classification model for predicting the students' academic achievement; those who need assistance to graduate, students who will excel and students who will likely to pursue their study in higher education through data mining approach. The specific objectives are:

i. To develop a decision tree model for early phase prediction of students' academic achievement.
ii. To identify the factors that influences the students' academic achievement.
iii. To validate the performance of decision tree models in predicting the students' academic achievement.

©This item is protected by original copyright

1.4 Research Scope

This study focuses on the students' academic performance in Universiti Malaysia Perlis. Factors that affect the students' academic performance have been investigated, such as age, gender, entry qualification, schools, entry cumulative grade point average value (CGPA), result for semester 1 and semester 2 and result of graduated CGPA.

1.5 Significance of the Study

This study could assist the university to predict the factors that affect students' academic achievement. By using the information from this prediction, the UniMAP management may supervise the student's academic achievement based on the predictive factors that have been identified. This scenario could help them to make the decision making for assisting the students who had bad achievement in their studies before expelling them from the institution. The academic advisor also may refer to the finding of this research while providing an advice to the problematic students.

From the lecturers' point of view, this study helps them to improve their teaching technique for each respective group of students. Thus, more guidance can be provided to the students who are weak in their academic performance. For example, using this predictive model the students who have problems can be detected at the early phase. Therefore, the lecturers can help those students by providing a better lecture or more training to improve their graduate CGPA.

Moreover, students' academic achievement is important in evaluating the university reputation and thus become a main factor for the graduate students to get better jobs. The main focus of UniMAP is to produce graduates with high potential and

©This item is protected by original copyright

better quality in the future, in conjunction with the government recommendation in the National Keys Results (NKRs). Therefore, it is the responsibility of the university to ensure the graduates get excellent results. By fully utilizing the results of this study, the university can improve the services provided to students and manage the students well.

1.6 Limitation of the Study

There are some obstacles occurring during conducting this study:

i. Difficulties to obtain a complete data due to some relevant sensitive factors such as the races and religion which could be the factors that influence the students' academic performance.

ii. This study only focuses on the external factors which influence the students' academic performance. Some internal factors, for instance students' psychology, stress and motivation could also be the factors which affect the academic performance of the students is not being studied.

iii. The predictive model only focuses on the early phase of the semester, which is semester 1 and 2. The first year might not be too difficult for students to score in their CGPA as there are still no extreme subjects included. Therefore, there would be some bias in the prediction results. Hence, for the better result of the prediction model, the result of mid semester could be used as the predictive factors.

©This item is protected by original copyright

1.7 Organization of the thesis

This thesis has been organized into six chapters. The first chapter would be the introduction of the thesis. This chapter consists with seven subtopics. The first subtopic was an overview of the study, this subtopic would summarize about the background of the research. Then, continued with the problem statement subtopic. This subtopic would focus on the problem statement of this research. From the problem statement of the research, we continued the thesis report with the objectives of the research. This subtopic would be summarized about the main objective and also specific objectives of the research. In this chapter 1, we also focused on the research scope which was the students of UniMAP. Then the significance of research would be focused in next subtopic. Then, second last subtopic would be the limitation of the research.

The chapter 2 of this thesis would be the literature review. This chapter consists of eight subtopics. The first subtopic would be the introduction of the chapter 2. Then, this chapter would focus on the overview of the knowledge discovery, data mining process and also the using of data mining in several sectors. In this chapter, the summarization of data mining in prediction and classification would be also the part of the literature review. In this subtopic, the previous research that related to this research would be presented. Then, the subtopic of the chapter would be continued with the educational data mining in the large scope and also focused on the previous research of the educational data mining, especially in Malaysia. Moreover, the subtopic of factors that influence the students' academic achievement would be also included in this chapter. Finally, the last subtopic would summarize the studies on the prediction of students' academic achievement using the C4.5 algorithms.

©This item is protected by original copyright

The third chapter of this thesis was the methodology of the research. In this chapter, there were five subtopics that would be presented. The first subtopic was the introduction of this chapter, and then continues with the framework for the research methodology. The third subtopic was the collection of the students' data from the UniMAP. In this subtopic, there would be separated into two other subtopics which were the background of the UniMAP and also the process of collection of students' data. Furthermore, this chapter would be consisting of the subtopic for the preprocessing of the collected data. Finally, the last subtopic for this chapter would be the statistical analysis using the chi square test for the independence. This chapter has been summarized in the first stage of the methodology in this research.

The chapter 4 would be the analysis of data and this chapter would focus on the calculation involved in this research. This chapter also summarized about the development of the decision tree model using the sample data from the data collection of students' in UniMAP. The development of the decision tree model for this research had established using the JMP Pro 11. In this chapter, the demonstration of steps started with a collection of data, preprocessed of data, statistical analysis until the development of predictive model would be shown. Then, this chapter finally would focus on the evaluation of the performance of the decision tree model using the confusion matrix and ROC curve.

The fifth chapter was the result and discussion. This chapter consists of four subtopics, and the first subtopic was the result of the statistical analysis. This subtopic separated into two other subtopics which were the result of chi square test and also result of the associations between the factor variables and the response variable. Then, the result of the decision tree model would also be represented in this chapter. Based on the leaf report of the decision tree model, the useful knowledge had been extracted and

©This item is protected by original copyright