DETERMINING THE EFFECT OF CURRICULUM AND FACILITIES ON ACADEMIC ACHIEVEMENT USING DATA MINING APPROACH

A thesis submitted to the Faculty of Information Technology in partial fulfillment of the requirement for the degree Master of Science (Intelligent System) Universiti Utara Malaysia

By
SURENAH SEDEK

©Surenah Sedek, 2008. All Rights Reserved.
Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

SURENAH SEDEK

calon untuk ijazah
(candidate for the degree of) MSc. [Intelligent System]

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

DETERMINING THE EFFECTS OF CURRICULUM AND FACILITIES
ON ACADEMIC ACHIEVEMENT USING DATA MINING

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan
dar: meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory
knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): ASSOC. PROF. FADZILAH SIRAJ

Tandatangan
(Signature) : 

Nama Penyelia Kedua
(Name of 2nd Supervisor): MISS NOORAINI YUSOFF

Tandatangan
(Signature) : 

Tarikh (Date) : 17/06/08
PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor, in her absence, by the Dean of the Faculty of Information Technology. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain should not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Request for permission to copy or to make use of material in this thesis, in whole or in part should be addressed to:

Dean of the Faculty of Information Technology
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman
ABSTRAK
(BAHASA MELAYU)

mengukur item kurikulum dan latar belakang dengan pencapaian akademik responden. Secara keseluruhannya, model rangkaian neural memperoleh peratusan lebih daripada 80 % manakala statistik hanya menghasilkan kurang daripada 50%.
ABSTRACT (ENGLISH)

Education domain offers many interest and challenge in data mining applications that potentially identified as a tool to help both educators and students to improve the quality of education system. Data Mining applies modern statistical and computational technologies to the problem of finding useful hidden patterns within large databases. Thus, this study applied data mining technique to identify the hidden information that affects the academic achievement among respondents. The respondents for this study are consists of all public university students which complete their study within year 2007. The questionnaire that has been used in this study was adopted from Kajian Pengesanan Graduan, Kementerian Pengajian Tinggi and it focuses on curriculum and facilities that have been provided by universities. The aims of this study is to determine whether the curriculum and facilities that provided by university has an effect on students academic achievement. 55,315 respondents data were used for descriptive task while 39,801 data for predictive task. Both data mining approaches, namely the descriptive and predictive have been utilized to perform the analysis prior to build the model. For descriptive purposes, frequency, cross tabulation and correlation coefficients were computed to check whether significant correlation exists. For predictive modeling, logistic regression and neural network were used. Statistical Pakages for Sosial Science (SPSS) was used for regression technique and Statistical Analytical Software (SAS) for Neural Network modeling. Then, the online questionnaire was integrated with Neural Network model to predict future student academic achievement. The findings in this study suggest neural network is the best model compared to logistic regression to measure the effect of curriculum and facilities on academic achievement. The highest accuracy from neural network is 89.47%, when demographics and curriculum become the contributing variables to academic achievement. Most of the neural network model accuracy is over than 80% while logistic regression accuracy is below than 50 %.
ACKNOWLEDGEMENT

In the name of Allah, Most Gracious, Most Merciful. Peace upon the prophet, Muhammad S.A.W, a foremost praise and thankful to Allah for His blessing, giving me the strength in completing this research.

On the other hand, my special thanks to Universiti Utara Malaysia (UUM) for facilities and resource that has been provided to all students. With all the facilities that have been provided, it helped me to make my research completed and succeeded.

I would like to extend my thanks and gratitude to my supervisor Associate Professor Fadzilah Siraj and Miss Nooraini Yusoff for the guidance, patience, encouragements, advice and flourish of knowledge during completing this research.

My warm appreciation dedicates to the Ministry of Higher Education who has gives their cooperation during this research.

Last but not least, a lasting heartfelt to my family and all my friends for all of the love and support in completing this course, as well as to my siblings.
# TABLE OF CONTENTS

## DESCRIPTIONS

- PERMISSION TO USE                                  i
- ABSTRAK (BAHASA MELAYU)                            ii
- ABSTRACT (ENGLISH)                                  iv
- ACKNOWLEDGEMENTS                                    v
- TABLE OF CONTENT                                    vi
- LIST OF FIGURES                                      viii
- LIST OF TABLES                                       x

## CHAPTER 1 : INTRODUCTION

1.1 Background                                       1
1.2 Problem Statement                                 5
1.3 Objectives                                        6
1.4 Scope of the study                                7
1.5 Significance of the study                         7
1.6 Organization of the report                        8

## CHAPTER 2 : LITERATURE REVIEWS

2.1 Academic Achievement                              9
2.2 Attribute for Measuring Academic Achievement      10
2.3 Data Mining                                       12
2.4 Data Mining Technique                             15
2.5 Conclusion                                        20

## CHAPTER 3 : METHODOLOGY

3.1 Overview of Methodology                           22
3.2 Phase 1 - Survey                                  23
3.3 Phase 2 – CRISP- Data Mining                      25
3.4 Phase 3 – Information System Research Design     34
3.5 Conclusion                                        37

## CHAPTER 4 : RESULT AND FINDINGS

4.1 The Sampling Data Set                             38
# LIST OF FIGURES

<p>| Figure 1 | The information required for the study | 5 |
| Figure 2 | Data Mining Technique and Tasks | 15 |
| Figure 3.1 | Flows for this study | 22 |
| Figure 3.2 | Example of questionnaire in section Demographics | 23 |
| Figure 3.3 | Example of questionnaire in section Curriculum | 24 |
| Figure 3.4 | Phase of the CRISP-DM Reference Model | 25 |
| Figure 3.5 | Example of questionnaire in part curriculum | 27 |
| Figure 3.6 | Example of questionnaire in part facilities | 27 |
| Figure 3.7 | Sample of raw data | 39 |
| Figure 3.8 | The interface of SAS | 32 |
| Figure 3.9 | Example of data partition | 32 |
| Figure 3.10 | Selection of logistic regression | 33 |
| Figure 3.11 | Selection of correlation Bivariate | 33 |
| Figure 3.12 | Information System Research Design | 34 |
| Figure 3.13 | Feed Forward Algorithm | 36 |
| Figure 4.1 | Graph respondent by IPTA | 40 |
| Figure 4.2 | Graph respondent by Gender | 40 |
| Figure 4.3 | Respondent by Academic Qualification | 41 |
| Figure 4.4 | Respondent by Field of Study | 41 |
| Figure 4.5 | Respondent by CGPA | 42 |
| Figure 4.6 | Respondent by Mode of Study | 42 |
| Figure 4.7 | Respondent by Sponsorship | 43 |
| Figure 4.8 | Respondent by family income | 43 |
| Figure 4.9 | Number of respondent with level of satisfying for Curriculum | 45 |
| Figure 4.10 | Distribution of respondents who are extremely satisfied with the curriculum. | 46 |
| Figure 4.11 | Number of Respondent by CGPA | 53 |
| Figure 4.12 | Academic Performance by Gender | 54 |
| Figure 4.13 | Academic Performance by Age | 55 |
| Figure 4.14 | Academic Performance by Academic Qualification | 55 |
| Figure 4.15 | Academic Performance by Field of Study | 56 |
| Figure 4.16 | Academic Performance by Mode of Study | 56 |
| Figure 4.17 | Academic Performance by Family Income | 57 |
| Figure 4.18 | The comparison accuracy between NN and Logistic Regression | 61 |
| Figure 4.19 | Interface of Prediction System | 62 |</p>
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.20</td>
<td>User key-in the level of satisfied.</td>
<td>63</td>
</tr>
<tr>
<td>4.21</td>
<td>The output that user key-in.</td>
<td>63</td>
</tr>
<tr>
<td>4.22</td>
<td>Result of Prediction System.</td>
<td>64</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table 3.1</th>
<th>Range of CGPA</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.2</td>
<td>List of Variable for Demographic</td>
<td>39</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>List of Variable for Curriculum Section</td>
<td>30</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>List of Variable for Facilities Section</td>
<td>30</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>The related attribute that involved in this study</td>
<td>39</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Number of respondent with level of satisfying for Curriculum</td>
<td>44</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Distribution of respondents who are extremely satisfied with the curriculum.</td>
<td>46</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Distribution of respondents who are not at all satisfied with the curriculum.</td>
<td>47</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Number of respondent with extremely satisfied level on curriculum items by CGPA</td>
<td>48</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Number of respondent with not at all satisfied level on curriculum items by CGPA</td>
<td>49</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>Mode of Study by level of satisfaction with Curriculum</td>
<td>50</td>
</tr>
<tr>
<td>Table 4.8</td>
<td>Number of respondent with level satisfying for Facilities</td>
<td>51</td>
</tr>
<tr>
<td>Table 4.9</td>
<td>The significant Value for Each Attribute</td>
<td>58</td>
</tr>
<tr>
<td>Table 4.10</td>
<td>Academic achievement for NN Model</td>
<td>59</td>
</tr>
<tr>
<td>Table 4.11</td>
<td>Academic Achievement for Logistic Regression Model</td>
<td>61</td>
</tr>
<tr>
<td>Table 4.12</td>
<td>Comparative accuracy between Neural Network and Logistic Regression</td>
<td>61</td>
</tr>
</tbody>
</table>
CHAPTER ONE

INTRODUCTION

This chapter presents the overview of the study that involves the curriculum and facilities are effect the academic achievement is applying the data mining technique. In addition, this chapter also includes the problem statement, objectives, scope, and significance of study and thesis organization.

1.1 BACKGROUND

Data mining is defined as the extraction from large amount of data. It is a powerful new technology with great potential to help companies focus on the most important information in their data warehouse (Berson, 2000). According to Tsantis and Castellani (2001), data mining applies modern statistical and computational technologies to the problem of finding useful patterns hidden within large database. In addition, data mining uses a combination of an explicit knowledge base, sophisticated analytical skills, and domain knowledge to uncover hidden trends and
The contents of the thesis is for internal user only
References


Akey, T. M. (2006) School Context, Student Attitudes and Behaviours, and Academic Achievement, An Exploratory Analysis, MDRC.

Averett, S. L., & McLennan, M. C. (2002). Exploring the effect of class size on student achievement: What have we learned over the past two decades?, Department of Economic and Business Lafayette College Easton.


Choi, Y (2006) Academic Achievement and Problem Behaviours among Asian
Pacific Islander American Adolescents. School of Social Service Administration, University of Chicago.


Hussain. T. (2007) Student Achievement In Saudi Arabia: The Importance Of Teacher Factors. *Faculty of the Graduate School of Arts & Sciences at Georgetown University.*


Luan, J. (2001). Data Mining As Driven by Knowledge Management in Higher Education. *SPSS Public Conference, UCSP.*


Predicting Student Performance: An Application of Data Mining Method with an Educational Web-Based System., Frontiers in Education Conference.


Romanowski, C. J. (2001) A Data Mining for Design and Manufacturing:


