VISUALIZING STUDENT PERFORMANCES
USING DATA MINING TECHNIQUES

SRI BANU A/P MUNISAMY

UNIVERSITI UTARA MALAYSIA 2007
VISUALIZING STUDENT PERFORMANCES
USING DATA MINING TECHNIQUES

A thesis submitted to the Graduate School in partial
Fulfillment of the requirement for the degree
Master of Science (Information Technology)
Universiti Utara Malaysia

By
Sri Banu A/P Munisamy
Sekolah Siswazah  
(Graduate School)  
Universiti Utara Malaysia  

PERAKUAN KERJA TESIS  
(Certification of Thesis Work)  

Kami yang bertandatangan di bawah, memperakukan bahawa  

We, the undersigned, certify that  

Sri Banu A/P Munisamy  
(nama penuh/full name)  

calon untuk ijazah  
candidate for the degree of Master Of Science (Information Technology)  
telah mengemukakan tesisnya yang bertajuk  
has presented his/her thesis with the following title:  

Visualizing Student Performances Using Data Mining Techniques  
(seperti yang tercatat di muka surat tajuk dan kulit tesis)  
as it appears on the title page and front cover of thesis  

dan tesis tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.  
The thesis is acceptable in form and content and that a satisfactory knowledge of the field is covered.  

AJK Tesis/Thesis Committee  

(i) Nama/Name: Dr. FAUZIAH AHMAD  
(Penyelia Utama/Principal Supervisor)  

Tandatangan  
Signature:  

(ii) Nama/Name: ____________________  

Tandatangan  
Signature: ____________________  

(iii) Nama/Name: ____________________  

Tandatangan  
Signature: ____________________  

Tarikh/Date: ____________________
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 purpose may be granted by my supervisor(s) or, in their absence by the Dean of the Graduate School. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall 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.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

Dean of Graduate School
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman.
ABSTRAK

Objektif kertas kerja ini adalah untuk menjalankan satu kajian untuk memaparkan prestasi pelajar di mana data yang digunakan adalah keputusan pelajar dan hasil maklumat yang diterima dipamerkan dalam bentuk grafik. Kajian ini memfokuskan kepada mengenalpasti faktor-faktor yang mempengaruhi prestasi pelajar dan juga menunjukkan keputusan tersebut dalam bentuk grafik untuk memudahkan pihak pengurusan. Pendekatan yang digunakan untuk mengkategorikan pembolehubah-pembolehubah yang berkaitan adalah analisis faktor. SPSS telah digunakan untuk membuat analisis ini. Keputusan analisis ini menunjukkan bahawa sikap pelajar dan latar belakang pensyarah adalah dua faktor utama yang mempengaruhi prestasi pelajar.
ABSTRACT

The goal of this qualitative paper is to conduct a study to visualize the student performances where the input data is the student's result and the output of the information retrieved is displayed in a graphical representation. The study focuses on identifying the factors that influence the student performances as well as visualizing the result for management's convenience. The approach that has been used to categorize the influencing variables is factor analysis. The analysis and visualization has been done using SPSS. Result of the analysis shows that student attitude and lecturer's background are the two factors most important factor known to influence students' performances.
ACKNOWLEDGEMENTS

This project would not have been possible without the support of many people. Many thanks to my supervisor Dr. Faudziah Ahmad, who provided valuable advice and guidelines, and read my numerous revisions and helped make sense of the confusion. Thanks to the lecturers of CADP; Ms. Naziffa, Mdm. Ainee, Ms. Usha Vellapan, Mdm. Julie and Mdm. Usha Devi for participating in the research experiment by providing valuable data during the interview. I would also like to take this opportunity to thank the students of CADP who participated in the survey. Finally, I would like to express my deepest gratitude to my parents who encouraged me to do my best, and my friends who endured this long process with me, always offering support and love.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERMISSION TO USE</td>
<td>I</td>
</tr>
<tr>
<td>ABSTRACT (BAHASA MALAYSIA)</td>
<td>II</td>
</tr>
<tr>
<td>ABSTRACT (ENGLISH)</td>
<td>III</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>IV</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>VII</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>VIII</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>IX</td>
</tr>
</tbody>
</table>

## CHAPTER 1: INTRODUCTION
1.1 Preface                                   | 1    |
1.2 Problem Statement                          | 3    |
1.3 Research Objectives                        | 4    |
1.4 Research Questions                         | 5    |
1.5 Scope of Study                             | 5    |
1.6 Research Outcomes                          | 6    |
1.7 Significance of Research                   | 6    |
1.8 Organizations of the Report                | 7    |
1.9 Summary                                    | 8    |

## CHAPTER 2: REVIEW OF LITERATURE
2.1 Background Study of CADP                  | 11   |

2.1.1 CADP Mission                             | 13   |
2.1.2 CADP Vision                              | 13   |

## CHAPTER 3: RESEARCH METHODOLOGY
3.1 Data Selection                             | 16   |
3.2 Data Preprocessing                         | 20   |
3.3 Data Mining and Visualization Technique    | 25   |

## CHAPTER 4: RESEARCH EXPERIMENT, RESULTS AND INTERPRETATION
4.1 Ranking the Subjects                       | 28   |
4.2 Identifying Factors                        | 31   |

4.2.1 Factor Analysis Report                   | 32   |
4.2.1.1 Correlation Matrix                     | 33   |
4.2.1.2 KMO and Bartlett’s Test                | 35   |
4.2.1.3 Communalities                          | 36   |
4.2.1.4 Total Variance Explained               | 37   |
4.2.1.5 Scree Plot                             | 38   |
4.2.1.6 Component Matrix and Rotated Component Matrix | 40   |
4.2.1.7 Component Correlation Matrix           | 42   |
4.3 Visualizing Student Performances

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion 45
5.2 Recommendations and Future Work 47

REFERENCES 48

APPENDIX A: QUESTIONNAIRE 50

APPENDIX B: INTERVIEW QUESTIONS 53
LIST OF TABLES

Table 1: Subjects and the total number fo records 17
Table 2: Programming Language of Subjects 18
Table 3: Attributes 21
Table 4: Data Set for Variables 22
Table 5: Description of the Attributes 23
Table 6: Subject Ranking 30
Table 7: KMO Measure Level 35
LIST OF FIGURES

Figure 1: Steps in the research methodology 15
Figure 2: Sample of raw data 24
Figure 3: Sample data after categorization 25
Figure 4: Screenshot of SPSS software 26
Figure 5: Screenshot of Microsoft Excel 29
Figure 6: Correlation Matrix 34
Figure 7: KMO and Bartlett’s Test result 35
Figure 8: Communalities 36
Figure 9: Total Variance with Eigenvalues 38
Figure 10: Scree Plot 39
Figure 11: Component Matrix 40
Figure 12: Rotated Component Matrix 41
Figure 13: Component Correlation Matrix 42
Figure 14: Line Graph 43
Figure 15: Bar Chart 44
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAS</td>
<td>Academic Advisory System</td>
</tr>
<tr>
<td>ASP</td>
<td>Active Server Pages</td>
</tr>
<tr>
<td>BMU</td>
<td>Bunda Mulia University</td>
</tr>
<tr>
<td>CADP</td>
<td>Center of Affiliate and Diploma Programme</td>
</tr>
<tr>
<td>DBIS</td>
<td>Diploma in Business Information System</td>
</tr>
<tr>
<td>DIT</td>
<td>Diploma in Information Technology</td>
</tr>
<tr>
<td>DIC</td>
<td>Diploma in Internet Computing</td>
</tr>
<tr>
<td>GTTC</td>
<td>Ghana Telecom Training Centre</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
</tr>
<tr>
<td>INTI</td>
<td>International University College</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>KDD</td>
<td>Knowledge Discovery in Database</td>
</tr>
<tr>
<td>KDS</td>
<td>Kolej Disted Stamford</td>
</tr>
<tr>
<td>KMO</td>
<td>Kaiser-Meyer-Olkin</td>
</tr>
<tr>
<td>KPTMKL</td>
<td>International College Penang and Kolej Poly-Tech MARA Kuala Lumpur</td>
</tr>
<tr>
<td>MANOVA</td>
<td>Multivariate Analysis of Variance</td>
</tr>
<tr>
<td>MMU</td>
<td>Multimedia University</td>
</tr>
<tr>
<td>NEMISA</td>
<td>National Electronic Media Institute South Africa</td>
</tr>
<tr>
<td>PHP</td>
<td>Hypertext Preprocessor</td>
</tr>
<tr>
<td>SPSS</td>
<td>SPSS for Windows</td>
</tr>
<tr>
<td>TTC</td>
<td>Telekom Training College</td>
</tr>
<tr>
<td>WEKA</td>
<td>Waikato Environment for Knowledge Analysis</td>
</tr>
</tbody>
</table>
XMDV  
X11 Multi - Dimensional Visualization Tool
CHAPTER 1

INTRODUCTION

1.1 Preface

Data mining is the process of discovering hidden information from large databases. The data are analyzed to identify relationships and behaviors between the attributes. "Knowledge Discovery in Database (KDD) is the overall process of discovering knowledge from data." (Fayyad et al., 1996) Some of the steps involved in Knowledge Discovery in Database (KDD) are data cleaning, data preparation, data selection, data mining. Thus, data mining is one of the steps in Knowledge Discovery in Database (KDD) process where new patterns and trends are discovered from large amount of data stored in the databases by applying data analysis and discovery algorithms.

Several techniques are used to mine the data. Artificial neural networks, decision trees, genetic algorithms, statistical analysis, visualization, nearest neighbor method and rule induction are the most commonly used techniques. These methods are used by many organizations to reveal hidden information by exploring large amount of data. The result of mining the data is used to solve problems and to form
The contents of the thesis is for internal user only
REFERENCES

Basu, S. (1997) *DATA MINING*. Georgia State University Fall

Retrieved 21 February 2007 from

http://www.siggraph.org/education/materials/HyperVis/applicat/data_mining/
data_mining.html


performance on programming assignments to improve student understanding.

In *Proceedings of the International Conference on Engineering Education*


useful knowledge from volumes of data. *In Communications of the ACM* {*

*Data Mining and Knowledge Discovery in Databases*


discovery: An overview. *In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and

R. Uthurusamy, editors, Advances in Knowledge Discovery and Data Mining,*

pages 1–35. AAAI/MIT Press

48


Jing Luan (2004). Data mining applications in higher education, *SPSS Executive Report*

Keim, D.A., & Kriegel, H. (1996) Visualization techniques for mining large databases: A comparison. *IEEE Transactions on Knowledge and Data Engineering*

