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. FAUDZIAH AHMAD Tandatangan
(Penyelia Utama/Principal Supervisor) *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

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAS	Academic Advisory System
ASP	Active Server Pages
BMU	Bunda Mulia University
CADP	Center of Affiliate and Diploma Programme
DBIS	Diploma in Bussiness Information System
DIT	Diploma in Information Technology
DIC	Diploma in Internet Computing
GTTC	Ghana Telecom Traning Centre
HTML	Hypertext Markup Language
INTI	International University College
IT	Information Technology
KDD	Knowledge Discovery in Database
KDS	Kolej Disted Stamford
KMO	Kaiser-Meyer-Olkin
KPTMKL	International College Penang and Kolej Poly-Tech MARA Kuala Lumpur
MANOVA	Multivariate Analysis of Variance
MMU	Multimedia University
NEMISA	National Electronic Media Institute South Africa
PHP	Hypertext Preprocessor
SPSS	SPSS for Windows
TTC	Telekom Training College
WEKA	Waikito Environment for Knowledge Analaysis

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

Cooper, D.R., & Schindler, P.S. (2006) *Business Research Methods*. New York: McGraw- Hill/Irwin.

Edwards,S.H., Pérez-Quiñones,M.A., Phillips,M., & RajKumar,J.(2006) Graphing performance on programming assignments to improve student understanding. In *Proceedings of the International Conference on Engineering Education*

Fayyad,U., Piatetsky-Shapiro,G., & Smyth,P.(1996) The KDD process for extracting useful knowledge from volumes of data. *In Communications of the ACM { Data Mining and Knowledge Discovery in Databases*

Fayyad,U., Piatetsky-Shapiro,G., & Smyth,P.(1996) From data mining to knowledge 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*

Fayyad,U. & Uthurusamy,R. (1999). Data mining and knowledge discovery in databases: *Introduction to the special issue.Communications of the ACM,* 39(11)

George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A simple Guide and Reference*. Pearson Education.

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*

Mazza,R., & Dimitrova,V. (2004) Visualising Student Tracking Data to Support Instructors in Web-Based Distance Education. *13th International World Wide Web Conference (WWW 2004) - Alternate Educational Track. 2004 May 17-22. pp.154-161.*

Merceron,A., & Yacef, K. (2005). Educational Data Mining: a Case Study

Minaei-Bidgoli,B., Kashy,D.A., Kortemeyer,G., & Punch,W.F. (2003). Predicting student performance: an application of data mining methods with the educational web-based system. *LON-CAPA. Proceedings of ASEE/IEEE Frontiers in Education Conference*