ANALYZING PRIMARY STUDENT DATA USING DATA MINING

CHONG SZE WEI

UNIVERSITI UTARA MALAYSIA
2009

ANALYZING PRIMARY STUDENT DATA USING DATA MINING

A dissertation submitted to the Division of Applied Sciences, College of Arts and Sciences in partial fulfillment of the requirements for the degree Master of Science (Information Technology), Universiti Utara Malaysia.

By

Chong Sze Wei

PERMISSION TO USE

In presenting this dissertation 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 dissertation 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 College of Arts and Sciences. It is understood that any copying or publication or use of this dissertation 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 dissertation.

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

Division of Applied Sciences

College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman.

ABSTRACT (ENGLISH)

Nowadays, academic achievement has become the most important evidence for establishing the value of Malaysia's education boundary. In this study, the primary students' examination data is collected on the previous examination mark yet sake to be analyzed for their future study plan. The selection of using data mining approaches was based on the capability of data mining as a grateful tool for academic analysis purposes. Focused on educational boundary, data mining approaches can be used for the process of uncovering hidden information and patterns that can help school community forecast the students' academic achievement. Therefore, the other relevant data such as student performance information and family income also engaged in this study. The overall relevant raw datasets is used for preprocessed and analyzed using statistical method. In addition, the result from the statistical manner analysis point out the considerable contribution of these attributes to the academic achievement plan.

ABSTRACT (MALAY)

Saban hari ini, pencapaian dalam bidang akademik sudah menjadi suatu tanda yang kukuh untuk menunjukkan nilai dalam batasan pendidikan Malaysia. Merujuk kepada penyelidikan ini, data peperiksaan yang lepas bagi pelajar pada peringkat sekolah rendah akan dikumpulkan untuk dianalisis sebagai rujukan pelan pelajaran mereka pada masa hadapan. Pemilihan menggunakan teknik data mining adalah ikutan kepada kebolehan Data Mining sebagai satu alatan yang canggih untuk kegunaan bagi analisis tentang bidang akademik. Dengan menumpu kepada batasan pendidikan, teknik data mining boleh digunakan untuk menunjukkan informasi yang tersembunyi dan untuk menunjukkan pola-pola atau corak yang boleh membantu komuniti sekolah menerokai. pencapaian pelajar dalam bidang akademik. Justeru, data-data yang berkaitan seperti data kemahiran pelajar dan data latar belakang keluarga juga akan tergolong dalam penyelidikan ini. Secara keseluruhan, data mentah akan digunakan untuk pemprosessan dan kaedah Statistik akan digunakan untuk analisis. Tambahan pula, keputusan yang telah diperolehi akan menunjukkan sumbangan yang boleh ditimbangi tentang hal-hal yang berkaitan dengan pelan kecemerlangan dalam bidang akademik.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my supervisor, Associate Professor Fadzilah Siraj, who willing to surpervise and lead me patiently with plot a routing and charitably sahring her abundant source of knowledge in this dissertation. Indeed, I really grateful for my supervisor because without her assistance numerous of beneficial comments, this study would have never been possible.

Therefore, I wish to send my warmest gratitude to my beloved parents, Mr. Chong Chan Mui and Mrs. Yoo Bee Eng for all their love and support by furnishing me a stunning position in their heart. In addition, thankful to my siblings which includes Chong Sze Hui, Chong Sien Joo, and Chong Sze Yin respectively for their help and encouragement.

Last but not least, a million thanks to all those who had lent a helping hand in permitting me to become visible of this project.

Chong Sze wei

Division of Applied Sciences

College of Arts and Sciences

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF ABBREVIATIONS

ASP	Active Server Page
BICC	Business Intelligence Competency Centre
BM1	Bahasa Melayu 1
BM2	Bahasa Melayu 2
DM	Data Mining
DMS	Data Mining System
ENG	English
GEP	Gifted Education Programme
ITS	Intelligent Tutoring System
KDD	Knowledge Discovery in Database
MATH	Mathematic
MOE	Ministry of Education
PMR	Penilaian Menengah Rendah
RM	Ringgit Malaysia
SMM	Sistem Maklumat Murid
SPM	Sijil Pelajaran Malaysia
SRKBG	Sekolah Rendah Kebangsaan Grik
STPM	Sijil Tinggi Pelajaran Malaysia
UPSR	Ujian Penilaian Sekolah Rendah
USA	United State

CHAPTER 1

INTRODUCTION

This chapter discusses the background of the study that consists of several sub-parts about subject of the study. These include overview on data mining techniques used for education area, a short description on the problem statements, research objectives, research scope and significance of this study. Lastly, this chapter presents the study organization by describing the structure of this report.

1.1 Background of Study

Data mining (DM) can be defined as the processes of extracting interesting information such as non-trivial, implicit, potentially valuable or previously unknown information from a huge amount of data storage area such as data warehouse, relational database and so on (Chen *et al*. 1996). It is a very valuable way of analyzing a huge amount of data, especially when humans are not capable to analyze such datasets manually. Also DM is well known as one of the core processes of Knowledge Discovery in Database (KDD).

DM also is known as one of the key features of many homeland safety initiatives. Regularly used as a means for detecting fraud, product retailing and others, data

The contents of the thesis is for internal user only

REFERENCES

Anjewierden, A., Koll"offel, B. & Hulshof, C. (2007). Using data mining methods for automated chat analysis to understand and support inquiry learning processes. *Proceeding of Towards Educational Data Mining*. Enschede, The Netherlands. pp. 27-36.

Arnold, A., Beck, J. & Scheines, R. (2004). An Inductive Approach. *Proceeding of Feature Discovery in the Context of Educational Data Mining.* Pittsburgh, PA 15213, USA.

Beikzadeh, M. R. & Delavari, N. (2004). A new analysis model for data mining processes in higher educational systems. *Proceeding of 5th international conference ITHET*. MMU, Cyberjaya, Malaysia. pp. 5-8.

Beal, C. R. & Cohen, P. R. (2006). Temporal Data Mining for Educational Applications.

Bravo, J., Vialardi, C. & Ortigosa, A. (2007). A problem-oriented method for supporting AEH authors through data mining. *Proceeding of International Workshop on Applying Data Mining in e-Learning (ADML'07).* Madrid, Spain. pp. 53-62.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (1999). CRISP-DM 1.0. *Proceeding of Step-by-step Data Mining Guide*. pp. 9-10.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (1999). CRISP-DM 1.0. *Proceeding of Step-by-step Data Mining Guide*. pp. 9-10.

Chen, M.-S., Jan, J. & Yu, P.S. (1996). Data mining: An overview from a database perspective. *Proceeding of IEEE Transaction on Knowledge and Data Engineering*, 8. pp. 866-883.

Garcia, E., Romero, C., Ventura, S. & Calder, T. (2007). Drawbacks and solutions of applying association rule mining in learning management systems. *Proceeding of International Workshop on Applying Data Mining in e-Learning (ADML'07).* pp. 13-22.

Gargano, M. L. & Raggad, B. G. (1999). Data mining-a powerful information creating tool. *Proceeding of OCLC Systems & Services.* 15(2), pp. 81-90.

Hamalainen, W., Laine, T. H. & Suitinen, E. (2004). *Data Mining in Personalizing Distance Education Courses*. University of Joensuu, Finland.

Lloyd, N. M., Heffernan, N. T. & Ruiz, C. (2007). Predicting student engagement in intelligent tutoring systems using teacher expert knowledge. *Proceeding of Educational Data Mining Workshop.* Marina del Rey, CA. USA. pp. 40-49.

Luan, J. (2001). Data mining as driven by knowledge management in higher education: Persistence clustering and prediction. *Proceeding of Keynote for SPSS Public Cpnference, UCSF*. pp. 1-16.

Ma, Y., Liu, B., Wong, C. K., Yu, P. S. & Lee, S. M. (2000). Targeting the Right Students Using Data Mining.

Merceron, A. & Yacef, K. (2007). Revisiting interestingness of strong symmetric association rules in education data. *Proceeding of International Workshop on Applying Data Mining in e-Learning (ADML'07).* pp. 3-12.

Minaei-Bidgoli, B., Kortemeyer, G. & Punch, W. F. (2003). Optimizing classification ensembles via a genetic algorithm for a web-based educational system. East Lansing, USA. pp.1-9.

Minaei-Bidgoli, B., Tan, P. N. & Punch, W. F. (2004). Mining interesting contrast rules for a web-based educational system. East Lansing, USA. pp. 1-8.

Perera, D., Kay, J., Yacef, K. & Koprinska I. (2007). Mining learners' traces from an online collaboration tool. *Proceeding of Educational Data Mining Workshop*. University of Sydney, Australia. pp. 60-69.

Ravi, S., Kim, J. & Shaw, E. (2007). Mining on-line discussions: Assessing technical quality for student scaffolding and classifying messages for participation profiling. *Proceeding of Educational Data Mining Workshop.* Marina del rey, CA. USA. pp. 70-79.

Raghavan, V. & Hafez, A. (2000). *Dynamic Data Mining*. University of Louisiana, USA. pp. 1-10.

Reza, B. & Naeimeh, D. (2004). *A New Analysis Model for Data Mining Processes in Higher Educational Systems*. MMU, Cyberjaya. Malaysia. pp. 5-8.

Seifert, J. W. (2004). Data mining: An overview. *Proceeding of CRS Report for Congress*. pp. 1-16.

Tanimoto, S. T. (2007). Improving the prospects for educational data mining. Pp. 1-6.

Tsantis, L. & Castellani, J. (2001). Enhancing learning environments through solution-based knowledge discovery tools: Forecasting for self-perpetuating systemic reform. pp. 1-35.

Yudelson, M. V., Medvedeva, O., Legowski, E., Castine, M., Jukic, D. & Crowley, R. S. (2006). Mining student learning data to develop high level pedagogic strategy in a medical ITS. University of Pittsburgh, pp. 1-8.