Identifying Purchasing Patterns of Arab and Malaysian Students Using Data Mining Technique

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

By

Anmar Fakhri Moh'd Abuhamdah

PUSAT PENGAJIAN SISWAZAH
(Centre for Graduate Studies)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

ANMAR FAKHRI MOH'D ABUHAMDAH

calon untuk Ijazah
(candidate for the degree of) **MSc. (Int. Sys.)**

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

IDENTIFYING PURCHASING PATTERNS OF ARAB AND MALAYSIAN SUTDENTS USING DATA MINING TECHNIQUE

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 dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the filed is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **DR. FAUDZIAH AHMAD**

Tandatangan
(Signature) : _____ Tarikh (Date): 22/10/06

Nama Penyelia Kedua
(Name of 2nd Supervisor): **MR. WAN HUSSIN WAN ISHAK**

Tandatangan
(Signature) : _____ Tarikh (Date): 22/10/06

PERMISSION TO USE

In presenting this project in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it feely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purposes 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 theses or parts there of 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 project.

Request for permission to copy or to make other use of materials in this project, in whole or in part, should be addressed to:

Dean of Faculty on Information Technology
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman

ABSTRACT

Currently, Universiti Utara Malaysia (UUM) has a significant number of international students. Since they are from various background and cultures, their preferences towards purchasing products are different. This study intends to identify purchasing patterns of 2 groups of students: Arab and Malaysian. The 2 groups have been chosen because they represent the major groups of the postgraduate students. A questionnaire has been constructed and used to collect data. The sample of data consists of postgraduate students from Arab and Malaysia. The total number of postgraduate students is 2122 and the total number of the sample data is 547 (30% of the population). Apriori Algorithmn, which is a popular data mining technique, has been used to identify the purchasing patterns. The study discovered that items such as Fruits, Vegetables, Drinks, and Pickled Food are frequently purchased by the Arabs. The Malaysians, however, prefer items such as Pickled Foods, Snack Foods, and Other Stuff. A more comprehensive work in the future is suggested so that results obtained can be generalized. The study has been successful in achieving all objectives. It is hoped that the results could be useful to UUM as the patterns identified could be used to strategize UUM's retailing businesses and at the same time provide adequate facilities in terms of selling preferred products to its consumers.

ACKNOWLEDGEMENTS

In the name of Allah

Most Beneficent and Most Merciful, Praise and thanks to Allah, first and last, lord and Cherisher of all the worlds who taught humankind everything they knew not. May his blessings and His Mercy be upon the holy prophet Muhammad S.A.W, the best of mankind.

I am very grateful to Universiti Utara Malaysia and especially to the Faculty on Information Technology for giving me the opportunity to complete my Master of Science (Intelligent System).

I would like to thank my supervisors Dr. Faudziah Bt Ahmad and Mr. Wan Hussain Wan Ishak for their guidance, critique and comments, and Ms Nooraina Yusoff, Mr Azizi bn Aziz and Mr Mohammad Shamri for their help. I am deeply indebted to them for their kindness and patience throughout the supervision and preparation of this project from the start until the final stage.

I express my true appreciation to those involved directly or indirectly in the writing of this project. Without them this project would not be a reality and only Allah will return the favor. My appreciation and thanks goes to the staff and librarians of Universiti Utara Malaysia, and all the authors whom I have quoted or to whom I have referred.

My heartfelt, special appreciation, thanks and love goes to the spirit of the late my father Fakhri Moh'd Abuhamdah, my mother Muasar Alahmad, brothers Ahmad, Monther, Firas, Ali, Alzarouq, Saleh, Hussain, Mustafa and all my relatives and friends, who have constantly support and motivate me to complete this study. I do and will continue to pray that Allah reward all of them abundantly.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

CHAPTER 1
INTRODUCTION

1.1 Background

In Universiti Utara Malaysia (UUM), the number of postgraduate has increased significantly since 2001. Currently, most of the postgraduate's students are international students with the majority of them coming from the Arab countries. The second highest number of students is from Malaysia. Thus, it can be seen that the postgraduate students come from many different background and cultures and due to this fact, the needs and preferences of these students vary. In terms of purchasing, their patterns also differ. The Arabs prefer bread and take less rice, while other group of students may prefer rice more than bread. Some studies on purchasing patterns have been done in the past. However, not many studies have been done on purchasing patterns using data mining techniques.

What is data mining? Data Mining, as written is literatures are defined in many ways. Data mining as described by (Seifert, 2004) is the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). With the advance in technology, data mining tools have incorporated new features. Thus, data mining tasks is not just collecting and managing data, it also includes analysis and prediction.

Data mining techniques have been widely applied to solve problems in the industry. Lau and Gao (2005) proposed a data mining approach for performance management in the banking industry. Meanwhile, in science Shi and Jaja (2002) considered the problem of organizing large scale earth science raster data to efficiently handle queries for identifying regions whose parameters fall within certain range values specified by the queries. Banks *et al.* (2004) have developed an undergraduate data

The contents of the thesis is for internal user only

REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases.* Proc. ACM SIGMOD Int. Conf. Management of Data, Washington, pp. 207-216.

Agrawal. R., & Srikant. R. (1994). *Fast algorithms for mining association rules in large databases.* In: Proc. of the 20th Intel. Conf. on, Sep 1994, pp. 487-499.

Ambwani, T. (2003). *Multi class support vector machine implementation to intrusion detection.* Proceedings of the IEEE International Joint Conference on 20-24 July 2003, 3, pp. 2300 - 2305.

Antonie, M-L., Zaiane, O. R., & Coman, A. (2001). *Application of Data Mining Technologies for Medical Image Classification.* Proceedings of The Second International Workshop on Multimedia Data (MDM/KDD'2001).

Banks, D. L., Dong, G., Liu, H., & Mandvikar, A. (2004). *Teaching undergraduates' data mining in engineering programs.* Frontiers in Education, 2, 1-6.

Brin, S., Motwani, R., & Silverstein, C., (1997). *Beyond Market Basket: Generalizing Association Rules To Correlations.* ACM SIGMOID-'97.

Chan, R., Yang, Q., & Shen, Y., (2003). *Mining High Utility Itemsets. Data Mining, ICDM 2003.* Third IEEE International, pp. 19 – 26

Chen, M. S., Jan, J., & Yu, P. S. (1996). *Data Mining: An Overview from a Database Perspective.* IEEE Transactions on Knowledge and Data Engineering, pp. 866-883.

Chen, Q., Jianghong, H., He, W., Mao, K., & Yungang Lai. (2005). *Utilize Fuzzy Data Mining to Find the Travel Pattern of Browsers.* Computer and Information Technology, 2005. 1148, pp. 228-232.

Doddi, S., Marathe, A., Ravi, S. S. & Toney, D. C., (2002). *Discovery of Association Rules in Medical Data*. Online: http://www. c3. lanl. gov.

Delavari N, & Beikzadeh M. R. (2005). *A New Analysis Model for Data Mining Processes in Higher Educational Systems*. MMU International Symposium on Information and Communications Technologies, July 7 – 9, 2005.

Dunham, M. H. (2003). *Data Mining Introductory and Advanced Topics*. New Jersey: Prentice Hall.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence, 1997, pp. 37-54.

Fernandez, M, C., Menasalvas, E., Marban, O., Pena, J, M., and Millan, S. (2001). *Minimal Decision Rules Based On The Apriori Algorithm*. IEEE, 3, pp. 691-704.

Grossman, R., Kasif, S., Moore, R., Rocke, D., & Ullman, J. (1998) *A Report of three NSF* Workshops on Mining Large, Massive, and Distributed Data,

. (1998).

Holt, J. D., Chung, S. M. (2002). *Mining Association Rules in Text Databases Using Multipass with Inverted Hashing and Pruning*. Proceedings of the 14th IEEE International, 2002, pp. 49 – 56.

Kuonen, D. (2003). *Challenges in Bioinformatics for Statistical Data Miners*. Issue of the "Bulletin of the Swiss Statistical Society, 2003, 46, pp. 10-17.

Lau, K., & Gao, C. (2005). *A data mining approach to performance measurement in the banking industry*. Proceedings of IEEE International Conference Services Systems and Services Management, 2005, 2, pp. 1009-1012.

Lee, K. B., & Suh, S. C. (1998). *The Efficient Algorithm in The Volume of Market Basket Data for Association Rules*, Expersys-98, The 10th International Conference.

Minaei-Bidgoli, B., A. Kashy, D., Kortemeyer, G., & F. Punch, W. (2003). *Predicting Student Performance: An Application of Data Mining Methods with the Educational Web-Based System LON-CAPA*. Proceedings of 33rd ASEE/IEEE Frontiers in Education Conference, 5-8.

Mitkas, P. A., Symeonidis, A. L., Kehagias, D., & Athanasiadis, I. (2003). *Application of Data Mining and Intelligent Agent Technologies to Concurrent Engineering*. Proceedings of 10th International Conference on Concurrent Engineering (CE-2003), Madeira, Portugal.

Pirttikangas, S., Riekki, J., & Röning, J. (2004) Routine Learning: *Analyzing Your Whereabouts*. Information Technology: Coding and Computing, 2004. Proceedings ITCC International Conference, 2004. 2, pp. 208-213.

Rygielski, C., Wang, J. C., & Yen, D. C. (2002). *Data mining techniques for customer relationship management*. Technology in Society, 2002. 24, pp. 483-502.

Shang, W., Zhu, H., & Huang, H. (2004). *WebCom Miner - a system of trends analysis for company products*. Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2004, 5, pp. 4084 – 4088.

Sharagai, A., & Schneider, M. (2001) *Discovering quantitative associations in databases*. IFSA World Congress and 20th NAFIPS International Conference, 2001. 1, pp. 423 – 428.

Shi, Q., & Jaja J. F. (2002). *Efficient Techniques for Range Search Queries on Earth Science Data. Proceedings of the 14th International Conference on Scientific and Statistical Database Management*, IEEE. 2002, 2, pp. 1099-3371.

Varde, A. S., Takahashi, M., Rundensteiner, E. A., Ward, M. O., Maniruzzaman, M., & Sisson Jr, R. D. (2004) *Apriori Algorithm and Game-of-Life for Predictive Analysis in Materials Science*.

Thiesing, F. M., Middelberg, U., & Vornberger, O. (1995). *Short Term Prediction of Sales in Supermarkets. Proceedings ICNN'95*, IEEE, 1995. 2, pp. 1028-1031.

Toshev, A., Bremond, F., & Thonnat, M. (2006). *An APRIORI-based Method for Frequent Composite Event Discovery in Videos.* IEEE International Conference. 2006, pp. 10 – 10.

Walker, P. R., Smith, B., Qing, Y. L., Famili, A. F., J. Valdes, J., Liu, Z., & Lach, B. (2004) *Data Mining of Gene Expression Changes in Alzheimer Brain.* Artificial Intelligence in Medicine. 2004, 31, pp.. 137-154.

Williams, G. J., & Huang, Z. (1996). *Modelling the KDD Process.* Australian Government's Cooperative Research Centres Program, 1996, pp. 1-8.

Xu, Y., Zhou, Sen-Xin., & Gong, Jin-Hua. (2005). *Mining Association Rules with New Measure Criteria. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, 2005, 18-21.

Yacoben, K., Carmichael, L. (1997). *Applying the Knowledge Discovery in Databases (KDD) Process to Fermilab Accelerator Machine Data.* Fermi National Accelerator Laboratory, 1997, pp. 1-5.

Yan, H., Ma, R., & Tong, X. (2005). *A Novel Fuzzy Comprehensive Evaluation Approach Based Apriori Algorithm for Unit's Bidding Ability Assessment.* Natural Science Foundation of China

Yang, P., & Liu, S. S. (2004). *Fault diagnosis for boilers in thermal power plant by data mining.* Proceedings of IEEE control, automation, robotics and vision conference, 2004, 3, pp. 2176-2180.