N° d'ordre : .....

Democratic and popular republic of Algeria
Ministry of Higher Education and Scientific Research
Faculty of Mathematics and informatics
Department of computer Science

A Dissertation in Fulfillment
For the Requirements of master degree in Computer Science
Option: Advanced Information Systems

A Data Mining Application for Managing the Higher Education sector
Case study: University of M'sila

publicly supported : 27/06/2012 The jury ::

Mr A.Moussaoui	**Université de M'sila**	**Président**
Mr K.Dehmeche	**Université de M'sila**	**Examiner**
Mr A.Attir	**Université de M'sila**	**Examiner**

Presented by:
GHODBANE Mebarek

Supervised by
Dr. BRAHIMI Mahmoud

class : 2011 /2012

List of Figures

EDM intersects data mining with pedagogy. Pedagogy contributes with the knowledge of learning processes, while data mining adds analysis and modeling techniques. Among the most used techniques are clustering, association rules and patterns analysis.

Educational Data Mining has specific requirements, which are not present in other domains, like pedagogical aspects of teachers, students and the system itself.

There are some problems in EDM, and we can start by the data mining tools, that are still rather difficult to use, especially for teachers [40]. These tools must suffer an improvement, to have a more intuitive and easy interface, with simple configuration and execution and with good visualization of the results.

There is also a need to standardize methods and data for EDM, since there are no general techniques to apply to different educational systems.

Another problem is the fact that it is difficult, or even impossible, to compare different methods or measures previously and decide which are the best [40]. It is necessary to experiment for knowing which is better, although the experimentation phase is difficult in the educational field because data is very dynamic, it can vary a lot between samples, and teachers just cannot afford the time or the technical experience to do these tests on each sample, especially in real time.

The lack of completeness from students is also a problem; they usually ask the questions and express their information needs incompletely. Consequently, the automatic classification of student questions, with the goal of ultimately providing automated tutorial responses is a challenging problem in EDM.

The problems of data redundancy and inherited correlation between many attributes, complicate the discovery of truly interesting patterns in the data. To avoid it, we could try to define appropriate interestingness measures for the patterns to be mined, integrate prior domain knowledge into the Data Mining techniques and adapt Data Mining technology towards the EDM needs in general.

Thus, our goal is to build mining modele to predict some educational situations, based on the information discovered.

The rest of the document is organized as follows: the First chapter has a small description of Data Mining and a detailed description of the work on Educational Data Mining. There is a revision of what was done in this area, with special attention to classification, as it is the focus of this work. Also some main applications and open issues in this area are presented. In chapter two, we present data warehouse techniques, including how to use it in Education. The theird chapted has a description the capabilities of data mining and its applications in higher education Institutions would like to know, for example, which students will enroll in particular course programs, and which students will need assistance in order to graduate. Are some students more likely to transfer than others? What groups of alumni are most likely to offer pledges? In addition to this challenge, traditional issues such as enrollment management and time-to-degree continue to motivate higher education institutions to search for better solutions. Finally, we conclude with developing an Analysis Services Project using Business SQL server Intelligence Development Studio 2008, we implement data warehouse and Mining Structures (Naïve Bayes, Decision Trees)., the results achieved and suggesting some guidelines for future research.

In this thesis, we have been able to demonstrate the process of designing and developing data-warehouse and data mining applications using SQL server business intelligence Development tools using a case study in an academic environment. It is to be noted however that this technique can be applied to any organization wishing to implement business intelligence as part of their strategic decision support operations. The power of Data-Warehousing in data analysis is tremendous and data-mining can discover hidden treasures in the data-warehouse.
Some of the main conclusions and contributions of the work are summarized, and some possible future development lines are commented.

As it has been emphasized, the most important oint of this research is the acquisition of knowledge from students' academic performance. The main purpose is to provide pridective mining modele for students.

We found that using data mining, it is possible to devlop a model representing the behavior of students in their way through different academic itineraries. This facilitates a prper vision of the behavior and performance of the groupe of students at certain university career and, at the same time, allows feeding the systeme to offer recommendation with the courses to be enrolling on.

One of the most attractive future works is to collect large data set from the university database and apply these classification methods on such data. Several other new classification methods can also be applied to test the most suitable method for student dataset i.e. lazy learners, Classification via regression etc. These classification methods can also be applied to predict other student outcomes such as dropouts or alumni pledges.

REFERENCES:

[1] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery: An overview. Advances in Knowledge Discovery and Data Mining, MIT Press, pages 1-36.

[2] Avaiblabe http://www.educationaldatamining.org/index.html [Acceessed 24-02-2012]

[3] Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., and Segovia, J. (2005). Web usage mining project for improving web-based learning sites. In Web Mining

[4] Han, J. and Kamber, M. (2001). Data Mining:Concepts and Techniques. Morgan Kaufmann Publishers.

[5] Ye, N. (2003). The Handbook of Data Mining. Lawrence Erlbaum Associates, Inc., Publishers, New Jersey.

[6] Russell, S. and Norvig, P. (2003). Arti_cial Intelligence: A Moddern Approach. Pearson Education, Inc., New Jersey.

[7] Hansen, P. and Jaumard, B. (1997). Cluster analysis and mathematical programming. Math. Program, pages 191-125.

[8] Xu, R. andWunsch, D. I. (2005). Survey of clustering algorithms. IEEE Transactions on Neural Networks, 16(3):645-678.

[9] Baker, R.S., Corbett, A.T. and Koedinger, K.R. 2004, Detecting StudentMisuse of Intelligent Tutoring Systems, in Proceedings of the 7th International Conference on Intelligent Tutoring Systems, Maceio, Brazil, 531-540.

[10] Beck, J.E. and Mostow, J. 2008, How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of student, in Proceedings of the Proceedings of the 9th International Conference on Intelligent Tutoring Systems, 353-362.

[11] Antunes, C. Acquiring Background Knowledge for Intelligent Tutoring Systems, in 1st International Conference on Educational Data Mining (EDM08). 2008. Montreal, Canada.

[12] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2001 [13] [12 p.1]

[14] Surajit Chaudhuri, Umeshwar Dayal, Venkatesh Ganti, "Database Technology for Decision Support Systems", IEEE Computer, vol: 34, no: 12, p. 48-55, December 2001

[15] Arun Sen and Atish P. Sinha, "A Comparison of Data Warehousing Methodologies", Communications of the ACM, vol. 48. no.3, March 2005

[16] Thomas Conolly and Caroly Begg., Database Systems, 3th Edition, Addison- Wesley, 2002

[17] W.H.Inmon, Building the Data Warehouse, 3rd Edition, John Wiley, Chap.2, p. 36, 2002

[18] Surajit Chaudhuri, Umeshwar Dayal, Venkatesh Ganti, "Database Technology for Decision Support Systems", IEEE Computer, vol. 34, no. 12, p. 48-55, December 2001

[19] Paul Gray and Hugh J. Watson, "Present and Future Directions in Data Warehousing", The DATA BASE for Advances in Information Systems-
Summer vol.29, no.3, 1998

[20] Ananth Srinivasan, David Sundaram and Joseph Davis, Implementing Decision Support Systems, McGraw-Hill, 2000

[21] Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell,Eunsaeng Kim, Ann Valencic, "Data Modeling Techniques for Data Warehousing", International Technical Support Organization

[22] Thomas Conolly and Caroly Begg., Database Systems, 3th Edition, Addison- Wesley, 2002

[23] H.Galhardas, "Declarative Data Cleaning Model, Language and Algorithms", Proc. VLDB Conf., Morgan Kaufmann, San Francisco, pp.371-380, 2001

[24] Data Analysis and Assessment Tools, Technology and Learning, 10536728,
vol.25, Issue 11, Jun 2005

[25] Microsoft, "Data Driven Decision Making in Higher Education, Improving decision making across the campus", February 2004

[26] Andreas Breiter, Daniel Light, "Data for School Improvement: Factors for designing effective information systems to support decision-making in schools", Educational Technology and Society, vol.9, no.3, p. 206-217, 2006

[27] R. Remes, "Learning Management System", WDS'05 Proceedings of Contributed Papers, Part I, pp. 207–212, 2005

[28] Thelma G.Lussier and Beverley Doern, "Improving Decision Making Support by Building a Data Warehouse", 5th Annual Conference of The Canadian Institutional Research and Planning Association, Calgary, Alberta, October 6-8, 1996

[29] Johnstone, 1976; Johnstone, 1981; Wako, 1988).

[30] Han and Kamber, 2001; Two Crows Corporation, 1999; Chen et al., 1996).

[31] Data mining and knowledge management in higher education – potential applications. In Proceedings of AIR Forum, Toronto, Canada

[32] Data Mining with CRCT Scores. Office of information technology, Geogia Department of Education.

[33] Mehta, M., Agrawal, R. and Rissanen, J. (1996). SLIQ: A Fast Scalable Classifier for Data Mining. IBM Almaden Research Center.34 (Luan et al., 2004)

[35] Luan, J. (2002b). Data Mining Application in Higher Education. SPSS Executive Report.

[36] UNESCO Nairobi Cluster (2006a). Analysis and Data Requirements of Core Indicators for Monitoring Education for All Goals. Kenya.

[37] Cave, M., Kogan, M. and Hanney, S. (1990). The scope and effects of performance measurement in British higher education. In F.J.R.C. Dochy, M.S.R. Segers and W.H.F.W. Wijnen (Eds.), Management Information and Performance Indicators in Higher Education. Van Gorcum and Comp, B.V., Assen/Maastricht, 48–49.

[38] UNESCO (2006b). National Education Sector Development Plan. A result-based planning handbook, January.

Van Petegem, P., Vanhoof, J., Daems, F. and Mahieu, P. (2004). Benchmarking the quality of school education:

enhancing the impact of indicators. Accepted for publication in Assessment in Education.

[39] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide.

[40] Lee, S., Cho, S. and Wong, M.P. (1998). Rainfall prediction using artificial neural networks. Journal of

Geographic Information and Decision Analysis, 2(2), 233–244.

[41] Agrawal, R.T. and Imielinski, A.S. (1993). Mining association rule between sets of item in large database. In

Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D. C., 207–216.

[42] Available http://www.mssqltips.com/sqlservertip/2565/ssas--best-practices-and-performance-optimization--part-1-of-4/ [Accessed 24-05-2012]

[43] Philo Janus and Guy Fouché, Pro SQL Server 2008 Analysis Services , Apress 2010,

Abstract

In recent years, decision support systems otherwise called business Intelligence [BI] have become an integral part of organization's decision making strategy. Data warehousing and Data Mining are now playing a significant role in strategic decision making. It helps companies make better decisions, streamline work-flows, provide better customer services, and target market their products and services.. This work is all about developing data warehouse and data mining Methods for management using the University of M'sila Student database as a case study. It describes the process of data warehouse design and development data mining structures using Microsoft SQL Server Analysis Services. It also outlines the development of a data cube as well as application of Online Analytical processing (OLAP) tools and Data Mining tools in data analysis.

Keywords: Data warehouse, Data Mining, Dimensional modeling, OLAP

résumé

ces dernières années, les systèmes de soutien des décisions autrement appelé Business Intelligence [BI] sont devenus une partie intégrante de la décision de l'organisation de stratégie de prise. L'entreposage de données et Data Mining, jouent désormais un rôle important dans la prise de décision stratégique. Il aide les entreprises à prendre de meilleures décisions, de rationaliser les flux de travailler, de fournir de meilleurs services à la clientèle, et le marché cible de leurs produits et services .. Ce travail consiste à créer des entrepôts de données et méthodes d'exploration de données pour la gestion de l'aide de l'Université de base de données des étudiants de M'sila comme étude de cas. Il décrit le processus de l'entrepôt de données de conception et de développement des structures d'exploration de données en utilisant Microsoft SQL Server Analysis Services. Il décrit également le développement d'un cube de données ainsi que l'application de traitement analytique en ligne (OLAP) des outils et des outils d'exploration de données.

Mots-clés: entrepôt de données, Data Mining, la modélisation dimensionnelle, OLAP

ملخص

في السنوات الأخيرة، نظم دعم القرارات والمعروفة باسم ذكاء الأعمال [BI] أصبح لها دور هام في استراتيجية عملية صنع القرار .تخزين البيانات واستخراج البيانات تلعب الآن دورا هاما في عملية صنع القرار الاستراتيجي .أنها تساعد الشركات على اتخاذ قرارات أفضل، وتبسيط سير العمل، وتقديم أفضل الخدمات للعملاء، وتسويق المنتجات والخدمات ..الهدف من هذا العمل هو تطوير مستودع البيانات واستخراج البيانات باستخدام أساليب لإدارة قاعدة بيانات جامعة مسيلة للطلاب كدراسة حالة .ويصف عملية تصميم مستودع البيانات و استخراج البيانات باستخدام Microsoft SQL Server Analysis Service

كلمات البحث: مستودع البيانات، واستخراج البيانات، النماذج ثلاثية الأبعاد، OLAP