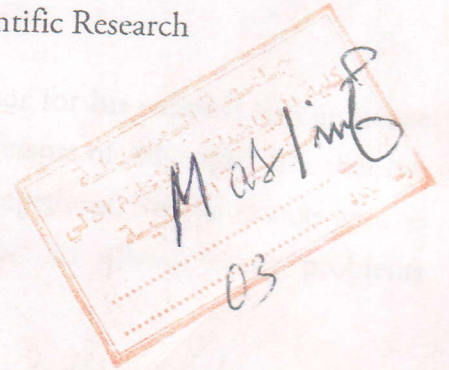


Acknowledgements

بسم الله الرحمن الرحيم

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 the Degree of Master in Computer Science
Option: Advanced Information System

Integrated data mining environment in CRM solution.
A case study of applied data mining techniques in CRM.

Presented by:
LOUNNAS Bilal

Supervised by:
MOKHTARI Rabeh

June 2011

Contents

	Introduction	1
1	General framework for business intelligence	4
1.1	What is business intelligence?	4
1.2	Why business intelligence?	4
1.3	Data, information and knowledge	5
1.4	Business intelligence architecture	6
1.4.1	Business intelligence pyramid	7
1.4.2	Development of a business intelligence system	9
2	Data warehousing	12
2.1	What is a Data Warehouse?	12
2.2	Differences between Operational Database Systems and Data Warehouses	13
2.3	Why Data Warehouse?	15
2.4	Data Warehouse Architecture	15
2.5	Data Warehouse design process	17
3	Data mining	19
3.1	What is data mining?	19
3.2	Why data mining is important?	19
3.3	Data mining task	20
3.3.1	Classification	20
3.3.2	Estimation	21
3.3.3	Prediction	21
3.3.4	Time series analysis	21
3.3.5	Association rules	21
3.3.6	Clustering	21
3.3.7	Summarization	22
3.4	Data mining algorithm	22
3.4.1	C4.0	22
3.4.2	K-Means	22
3.4.3	SVM	22
3.4.4	Apriori	22
3.4.5	EM	23
3.4.6	PageRank	23
3.4.7	AdaBoost	23

3.4.8	kNN	23
3.4.9	Naïve Bayes	24
3.4.10	CART	24
3.5	Data mining process	24
3.5.1	Business understanding	25
3.5.2	Data understanding	26
3.5.3	Data preparation	26
3.5.4	Modeling	26
3.5.5	Evaluation	26
3.5.6	Deployment	26
3.6	Architecture of data mining system	27
3.7	The kinds of data mining	29
3.7.1	Graph mining	29
3.7.2	Spatial Data Mining	29
3.7.3	Multimedia Data Mining	29
3.7.4	Text Mining	29
3.7.5	Mining the World Wide Web	30
3.8	Data mining application	30
3.8.1	Data Mining for Finance	30
3.8.2	Data mining for the Retail Industry	30
3.8.3	Data Mining for the Telecommunication Industry	31
3.8.4	Data Mining for Biology	31
3.8.5	Data mining for Intrusion Detection	31
3.8.6	Data mining for Education	32
3.9	Data mining tools	32
3.9.1	SAS Enterprise Miner	32
3.9.2	Intelligent Miner	32
3.9.3	Clementine, from SPSS	33
3.9.4	Statistica Data Miner from Statsoft	33
3.9.5	Oracle Data Mining (ODM)	33
3.9.6	Microsoft SQL Server 2008R2.....	33
3.9.7	Weka	33
3.9.8	RapidMiner	34

4	Data mining for customer relationship management	35
4.1	CRM Literature	35
4.1.1	What is CRM?.....	35
4.1.2	Why CRM?	35
4.1.3	CRM Components	35
4.1.4	CRM Architecture	37
4.1.5	CRM Success Factors	38
4.2	Data mining models for CRM	39
4.2.1	CRM dimensions	39
4.2.2	Data mining models	41
4.2.2.1	Customer segmentation	41

4.2.2.2	Direct marketing	42
4.2.2.3	One-to-one marketing	43
4.2.2.4	Market basket analysis	43
4.3	Integration DM environment in CRM solution	43
4.3.1	Architecture	43
4.3.2	How it works.....	45
Conclusions and Feature Work		47
	Conclusions	47
	Future Work	48
Appendix		49
	Glossary	49
	Framework tools	53
	Implementation issues.....	64
References		69

Introduction

In a 1958 article, IBM researcher Hans Peter Luhn used the term business intelligence. He defined intelligence as: "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal".

Several different definitions has been writing for business intelligence term, but all of them refer to the term BI as a "technology provides technics such as predictive analytics, business performance management... etc. to generate knowledge useful for better business decision-making".

We all know that every enterprise needs to know about their customers and what they need, so that, what it has to do to satisfy them. From this approach, enterprise uses small part of BI called CRM to deal with this kind of challenges. CRM stands for customer relationship management is an information industry term for business strategy design to reduce cost and increase profitability by solidifying customer satisfaction, loyalty, and advocacy.

From the point of data view, CRM technology collect very large amount of information about customer "personal information, sales information, marketing information, etc.", these kinds of information need to be stored on data base repository.

Problem statement

Many business owners don't realize the importance of the data they collect from their customers. While most of these businesses will store this information, this alone is not enough. The information that they store is not as important as how they use it.

Knowing how to collect and use the information about customers, can be valuable, collecting information about customer can be very simple task comparing to how to use these information and how to extract knowledge from it. This is exactly what can data mining do, data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories, and summarizing it into useful information that can be used to predict the future.

While the most important benefits of data mining is increased profits for business owners, there are a number of other important advantages as well. You will be able to cross-sell products at a reduced cost. Because you will know the

shopping patterns of your customers, you will be able to market to them in a way which will increase their retention. Data mining will also allow you to find the customers that are the most valuable. Based on the information you have, you will be able to alter your marketing strategy in a way that will allow you to reach more customers. As with anything, you will want to start with small amounts of data before attempting to tackle data that is more complex.

Research & objectives

As the nature of research in CRM and data mining are difficult to confine to specific disciplines, the relevant materials are scattered across various journals. Business intelligence and knowledge discovery are the most common academic discipline for data mining research in CRM.

In general data mining has been a new and most interesting technology for the recent years, this technology integrate almost for every science, education, medical, security,, etc. but the most application area that applied data mining techniques is in the banking and CRM.

The objective of CRM data mining is to enable you to manage customer needs and expectations more effectively leading to increased profits, by extract knowledge from customer information collected through many CRM task and try to use it on customer need.

Most CRM solution focus about how can collect information from their customer, many of them integrate a part that use data mining technique for this information as we discuss before, by using some data mining models developed by the CRM solution owner. But the problem is: we suppose that the enterprise grow up, more services, different product produce, more sales territory,, etc. even it has a new look for its customers and needs more data mining models for the new services and politics, what can they do? Should we call a CRM solution developer and ask him for design a new data mining models? How much time can the developer take to develop a new CRM solution with the new data mining models? From this view our approach is to develop a CRM solution with a data mining service part in CRM solution that can create, use, modify, and delete data mining models by miner working for the enterprise.

Dissertation outline

The first chapter is a general framework for business intelligence including BI literature, BI architecture, and the development of BI system.

In the second chapter, we are going to introduce the first part of BI, which is a data warehouse describing how it work, the deference between DW and data base administration, the architecture of DW and how we can design it.

For the next part of BI architecture we find data mining and that what we will discuss in the third chapter by showing a literature view of DM, DM task, most successful DM algorithm, also some DM application areas, and finally data mining tools.

The chapter four includes CRM literature and how data mining can be useful for CRM solution by describing some DM models for CRM and the architecture of our approach showing how it work together with data mining engine.

In addition, we put conclusion and future work, references at the end of our dissertation.

References

- [1] *Business Intelligence*, Wikipedia, Available: http://en.wikipedia.org/wiki/Business_intelligence, [Accessed 26-02-2011].
- [2] Carlo V, *Business Intelligence: Data Mining and Optimization for Decision Making*, John Wiley & Sons Ltd Publication, 2009.
- [3] *Data*, Wikipedia, Available: <http://en.wikipedia.org/wiki/Data>, [Accessed 26-02-2011].
- [4] *Business intelligence architecture*, Search Business Analytics, Available: <http://searchbusinessanalytics.techtarget.com/definition/business-intelligence-architecture>, [Accessed 27-02-2011].
- [5] Jiawei H & Micheline K, *Data Mining: Concepts and Techniques, Second Edition*, Morgan Kaufmann Publishers 2006.
- [6] Dunham, M.H, *Data mining introductory and advanced topics*, Prentice Hall Publishers 2003.
- [7] Julia P & Michal Z, *Elaboration of a method for comparison of Business Intelligence Systems which support data mining process*, Master thesis, Blekinge Institute of Technology, 2009.
- [8] Oded M & Lior R, *Data Mining and Knowledge Discovery Handbook Second Edition*, Springer Publisher 2010.
- [9] *Knowledge Discovery in Databases*, Course in computer science department, available: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/2_tasks.html, [Accessed 07-03-2011].
- [10] Diego S & Alvaro S, *Data Mining Web-Tool Prototype Using Monte Carlo Simulations*, Master thesis, Blekinge Institute of Technology, 2008.
- [11] *Data Mining*, ArabTeam, available: <http://www.arabteam2000-forum.com/index.php?showtopic=211435>, [Accessed 07-03-2011].
- [12] Xindong W & Vipin K, *The top ten algorithms in data mining*, Data mining & knowledge discovery series, Chapman & Hall/CRC, 2009.

References

- [13] Pete Chapman & ALL, *CRISP-DM 1.0 Step-by-step data mining guide*, CRISP-DM consortium: NCR, DaimlerChrysler AG, SPSS Inc, OHRA Verzekeringen 2000, available: <http://www.crisp-dm.org/CRISPWP-0800.pdf>, [Accessed 09-03-2011].
- [14] *Educational data mining*, Wikipedia, Available: http://en.wikipedia.org/wiki/Educational_data_mining, [Accessed 10-03-2011].
- [15] *STATISTICA Data Mining, Text Mining and Predictive Analytics Software*, StatSoft, available: <http://www.statsoft.com/products/data-mining-solutions/>, [Accessed 10-03-2011].
- [16] *RapidMiner*, Wikipedia, available: <http://en.wikipedia.org/wiki/RapidMiner>, [Accessed 10-03-2011]. RapidMiner web site: <http://rapid-i.com/>.
- [17] *CRM (customer relationship management)*, Search CRM, available: <http://searchcrm.techtarget.com/definition/CRM>, [Accessed 12-03-2011].
- [18] *What Is CRM?*, Destination CRM, available: <http://www.destinationcrm.com/Articles/CRM-News/Daily-News/What-Is-CRM-46033.aspx>, [Accessed 12-03-2011].
- [19] *Customer relationship management: why CRM?*, Business Link, available: <http://www.businesslink.gov.uk/bdotg/action/detail?itemId=1075422931&type=RESOURCES>, [Accessed 12-03-2011].
- [20] Mohammad A, *CRM Success Factors Taxonomy*, European and Mediterranean Conference on Information Systems 2008, available: [http://www.iseing.org/emcis/emcis2008/Proceedings/Refereed%20Papers/Contributions/C%2019/CRM%20Success%20Factors%20Taxonomy%20\(%20PAPER\).pdf](http://www.iseing.org/emcis/emcis2008/Proceedings/Refereed%20Papers/Contributions/C%2019/CRM%20Success%20Factors%20Taxonomy%20(%20PAPER).pdf), [Accessed 12-03-2011].
- [21] Kristin A & Carol K, *Customer Relationship Management*, McGraw-Hill 2002.
- [22] Darius Z, *CUSTOMER PERFORMANCE MEASUREMENT, Analysis of the Benefit of a Fuzzy Classification Approach in Customer Relationship Management*, Master thesis University of Fribourg, Switzerland, Faculty of Economics, 2007, available:

References

http://diuf.unifr.ch/main/is/sites/diuf.unifr.ch.main.is/files//file/studentprojects/M-2007_Darius_Zumstein.pdf, [Accessed 13-03-2011].

[23] E.W.T. Ngai, Li Xiu & D.C.K. Chau, *Application of data mining techniques in customer relationship management: A literature review and classification*, an international journal *Expert Systems with Applications* 36 (2009) 2592–2602.

[24] Konstantinos T & Antonios C, *Data Mining Techniques in CRM: Inside Customer Segmentation*, John Wiley and Sons, Ltd., Publication 2009.

[25] Michael J & Gordon S, *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management Second Edition*, John Wiley and Sons, Ltd., Publication 2004.

مخلص. بنى العملاء مع أكثر الأشياء المهمة في التسويق والصناعة. مع أجل هذه الأهمية أصبح لها دعم كبير بواسطة التكنولوجيا. إدارة علاقة العملاء تهدف إلى زيادة أداء وفعالية التجارة مع خلال تخصيص بنى العملاء والإبقاء على ولائهم. إدارة علاقة العملاء أصبح الآن أكثر شعبية في مناهج الإدارات في الصناعة. المبيعات، التسويق، المالية. مع خلال هذه النظرة السابقة نوضح كم هو مهم تقيد البيانات لإدارة علاقة العملاء.

كلمات مفتاحية: ذكاء الأعمال، إدارة علاقة العملاء، تقيد البيانات، المبيعات، التسويق، تعريفات أدبية.

Abstract. Supported by technologies of Customer Satisfaction, Information Technology, and Data mining, etc., CRM aims to enhance the effectiveness and performance of the businesses by improving the customer satisfaction and loyalty. CRM is now becoming a popular management methodology in manufacturing, sales, marketing, and finance. Based previous literatures we illustrate how data mining can be very useful tools for CRM solution by shown data mining models in this kind of application.

Key word: BI, CRM, Data mining, Sales, Marketing, Literature view.

Résumé. Soutenu par les technologies de satisfaction de la cliente, technologie de l'information, et le data mining, etc, CRM vise à améliorer l'efficacité et le rendement des entreprises en améliorant la satisfaction et la fidélité de leur cliente. CRM maintenant la plus populaire méthode pour la gestion de la fabrication, vent, marketing, finance. Pour cette vue nous illustrons comment le data mining peuvent être des outils très utiles pour la solution CRM.

Mot clé: BI, CRM, Data mining, Vent, Marketing, général vue.