

MASTER'S THESIS

Thesis Title **Mining Changes in Customer Purchasing Behavior: a Data Mining Approach**

Supervisors:

Prof. Mohammad Mehdi Sepehri
Prof. Peter Naude

Referee:

Dr. Chaharsooghi
Dr. Wallstrom

Prepared by:

Tarbiat Modares University Faculty of Engineering
Department of Industrial Engineering

Lulea University of Technology
Division of Industrial Marketing and E-Commerce

Joint MSc PROGRAM IN MARKETING AND ELECTRONIC COMMERCE



Abstract:

The world around us is changing all the time. For businesses, knowing what is changing and how it has changed is also crucial. One of the most important aspects of surviving in a dynamic market is to know and adapt to changes happening in customer behavior. In Fast Moving Consumer Goods (FMCG) Distribution Company, this issue has more importance. Because of the variety of FMCGs products, distribution companies and their different strategies, the purchasing behavior of customers may change many times during a period and the competition become tougher. The purpose of this study is to help Kalleh Company as a manufacturer and distributor of food products in Iran market to mine changes happening in their customer behavior.

Mining changes has several steps includes data collection, data preprocessing, customer segmentation, mining customer behavior patterns and change mining. For customer segmentation, we use Customer Value Matrix. For mining pattern of behavior, we use Apriori algorithm and maximal frequent itemsets. We have different kinds of changes based on the literature, added/perished rules, emerging pattern and unexpected changes. Also, there are two measures of similarity and unexpectedness to measure the change. In this study, one time we calculate changes based on these measures from the literature. Then, we modified these measures to calculate the difference between ordinal attribute to bring their information in the calculation of changes. Our contribution is modifying these change measure to bring more information and higher accuracy in change mining. The result presented in the chapter 4. Marketing managers can apply these detected changes to be responsive accurately and timely to the changes in the market. In addition, they can use it to evaluate different marketing campaigns to build stronger relationship with their customer and knowing the market better. There are many implications for mining changes in macro in micro aspects of businesses and also in marketing campaigns and manufacturing.

Acknowledgment:

First of all, I am heartily and deeply thankful to my supervisor, Dr. Sepehri, for his guidance, patience and support. I consider myself very fortunate for being able to work with a very considerate and encouraging professor like him.

I would also like to extend my thanks to Mr. Babak TeymourPour, PHD candidate, whose guidance and support from the initial to the final level enabled me to finish this work.

I would like to thank my parents and my brother for supporting and encouraging me to pursue this degree. And my final words go to my husband whose support and guidance is with me and without his understanding and relentless support it would have been impossible for me to complete this thesis.

Samira Madani

Abstract:	1
Chapter 1: Introduction	9
1.1 Background of the study:	9
1.2 Problem definition:	11
1.3 Purpose of this study:	11
1.4 Research question:	13
1.5 Research motivation:	13
1.6 Demarcation:	14
1.7 Research outline:	14
Chapter 2: Literature Review	15
2.1 Mining Customer Behavior:	16
2.2 Review of Data Mining	16
2.2.1 Data mining: in brief	16
2.2.2 Data mining Functions:	18
2.2.3 Classification in brief:	19
2.2.4 Clustering in brief:	19
2.2.5 Association Rules in Brief:	20
2.3 Association Rule Mining Review:	20
2.3.1 Association Rule mining problem:	20
2.3.2 Apriori Algorithm	20
2.3.3 Association Rule Mining Approaches: Apriori Approach	27
2.4 Mining Changes Literature Review:	31
2.5 Customer segmentation review:	38
2.5.1 Clustering Analysis	39
2.5.2 Customer Segmentation Model	39
2.5.3 RFM Model	39
2.5.4 RFM Scoring	40

2,0,0 Customer Value Matrix Model	43
Chapter 3: Research Methodology	46
3,1 Research Methodology:	47
3,2 Research Design:	47
3,3 Research Purpose:	48
3,4 Research Approaches:	49
3,0 Research Strategy:	49
3,6 Research process:	50
3,7 Data Collection and Description:	51
3,8 Data Pre-Processing:	54
3,9 Customer Segmentation:	57
3,9,1 Customer Value Matrix	57
3,9,2 An effective analytical tool	58
3,9,3 Customer Value Matrix Methodology	58
3,10 Mining Customer Behavior:	61
3,10,1 Association Rule Mining:	61
3,10,2 Apriori algorithm:	62
3,11 Change Mining:	64
3,11,1 Change Mining:	64
Chapter 4: Results & Analysis	73
4,1 Data preprocessing result:	74
4,1,1 Data Cleaning	74
4,1,2 Data Transformation result:	74
4,2 Customer segmentation (in sql server 2008)	76
4,2,1 Customer Value Matrix Result:	76
4,3 Customer Behavior Mining:	79
4,3,1 Discretization Result:	79
4,3,2 Association Rule Mining Results:	81

ξ,ξ Change Mining:.....	۸۲
ξ,ξ,۱ Some examples of change pattern:.....	۸۲
ξ,ξ,۲ Association rules and changes based (Chen et al, ۲۰۰۵):.....	۸۴
ξ,ξ,۳ Rules with discrete variables in RHS:.....	۱۰۵
ξ,ξ,ξ Change mining with Manhattan distance.....	۱۰۶
Chapter۵: Conclusion, further research.....	۱۳۵
۵,۱ Conclusion:	۱۳۷
۵,۲ Our contribution:.....	۱۳۸
۵,۳ Limitation:.....	۱۳۹
۵,ξ Managerial Implication:	۱۳۹
۵,۵ Future works:	۱۴۰
References:	۱۴۰

List of tables

Table ۲,۱: Factors for classification of ARM.....	۲۶
Table ۲,۲: Mining in a changing environment timetable.....	۳۸
Table ۳,۱: Data collected from Kalleh Company.....	۵۳
Table ۳,۲: calculating variables for customer value matrix.....	۶۰
Table ξ,۱: RFM table fields.....	۷۳
Table ξ,۲: calculating variables for customer value matrix.....	۷۴
Table ξ,۳: calculating variables for customer value matrix.....	۷۴
Table ξ,ξ: segment information in for period ۱.....	۷۵
Table ξ,۵: segment information in for period ۲.....	۷۶
Table ξ,۶: R quantile.....	۷۷
Table ξ,۷: M quantile.....	۷۷

Table ξ, λ : F quantile.....	78
Table ξ, ρ : Area quantile.....	78
Table ξ, σ : Generated rule summary.....	79
Table ξ, τ : Generated Rules for period τ Cluster τ	81
Table ξ, ν : Generated Rules for period ν Cluster τ	82
Table ξ, ω : Generated Rules for period ω Cluster ν	82
Table ξ, π : Generated Rules for period π Cluster ν	86
Table ξ, ρ : Generated Rules for period ρ Cluster ν	89
Table ξ, σ : Generated Rules for period σ Cluster ν	90
Table ξ, τ : Generated Rules for period τ Cluster ξ	82
Table ξ, θ : Generated Rules for period θ Cluster ξ	98
Table ξ, ι : Cat τ quantile.....	103
Table ξ, κ : Cat ν quantile.....	103
Table ξ, λ : Cat ω quantile.....	104
Table ξ, μ : Cat π quantile.....	104
Table ξ, ν : Cat ρ quantile.....	104
Table ξ, ω : Cat σ quantile.....	105
Table ξ, π : Cat τ quantile.....	105
Table ξ, ρ : Generated Rules for period τ Cluster τ , Change mining by (Chen et al, 2008) measures & Manhattan distance.....	106
Table ξ, σ : Generated Rules for period ν Cluster τ , Change mining by (Chen et al, 2008) measures & Manhattan distance.....	107
Table ξ, τ : Generated Rules for period ω Cluster ν , Change mining by (Chen et al, 2008) measures & Manhattan distance.....	108
Table ξ, θ : Generated Rules for period π Cluster ν , Change mining by (Chen et al, 2008) measures & Manhattan distance.....	111

Table ٤,٢٩: Generated Rules for period ١ Cluster ٣, Change mining by (Chen et al, ٢٠٠٢) measures & Manhattan distance.....	١١٤
Table ٤,٣٠: Generated Rules for period ٢ Cluster ٣, Change mining by (Chen et al, ٢٠٠٢) measures & Manhattan distance.....	١١٧
Table ٤,٣١: Generated Rules for period ١ Cluster ٤, Change mining by (Chen et al, ٢٠٠٢) measures & Manhattan distance.....	١١٩
Table ٤,٣٢: Generated Rules for period ٢ Cluster ٤, Change mining by (Chen et al, ٢٠٠٢) measures & Manhattan distance.....	١٢٧

List of figures:

Figure ٢,١: Knowledge Discovery in Database Processes.....	١٧
Figure ٢,٢ the major steps in data mining process.....	١٨
Figure ٢,٣: Classification of DM techniques.....	١٨
Figure ٢,٤: Classic Problem of association rule mining	٢١
Figure ٢,٥: Mining in a changing environment review.....	٣٧
Figure ٢,٦: Customer Value Matrix.....	٤٥
Figure ٣,١: Research design of this study.....	٤٧
Figure ٣,٢: Change mining process perspective.....	٥٠
Figure ٣,٣: Change mining process.....	٥١
Figure ٣,٤: Change mining process in detail.....	٥١
Figure ٣,٥: Product categories of Kalleh company.....	٥٣
Figure ٣,٦: customer value matrix.....	٦٠
Figure ٤,١: generalized product category.....	٧٢
Figure ٤,٢: The Customer Value Matrix.....	٧٥
Figure ٤,٣: R histogram.....	٧٧
Figure ٤,٤: M histogram.....	٧٧
Figure ٤,٥: F histogram.....	٧٨

Figure 4.7: Area histogram	78
Figure 4.8: Cat histogram.....	103
Figure 4.9: Cat histogram.....	103
Figure 4.10: Cat histogram.....	14
Figure 4.11: Cat histogram.....	14
Figure 4.12: Cat histogram	100
Figure 4.13: Cat histogram.....	10

Chapter 1: Introduction

Background of the study

Problem definition

Purpose of this study

Research question

Research motivation

Research demarcation

Research outline

1.1 Background of the study:

The world around us changes continuously. Knowing and adapting to changes is an important aspect of our lives. For businesses, knowing what is changing and how it has changed is also essential (Liu et al, ۲۰۰۰). One of the most important aspects of surviving in a dynamic market is to know and adapt to changes happening in customer behavior. Moreover, in recent years, there has been the explosive growth in the amount of information (Min, S., H., Han, I., ۲۰۰۵). In general, Fast moving consumer goods (FMCG) distribution companies collected huge amount of data from their customers and their purchasing transactions. In this gathered data, we can find interesting hidden information about the customers and their behaviors.

The traditional approach for marketing decision making for marketing promotions, campaigns and market research in FMCG distribution companies is to focus more on their internal expert opinions. These experts include the marketing managers and also sales managers who are in constant touch with their salespeople and merchandisers who bring them market information.

However, this kind of decision making process ignores the customer data and their behaviors. Furthermore, in today's world where the market is highly competitive and products are overwhelming, customers face with various products and various providers with different marketing strategies (Hossein Javaheri, S., ۲۰۰۸). In such a situation, customer behavior changes all the time due to such a dynamic market (Chen et al, ۲۰۰۵). When the marketing manager became aware of some changes in the market by sales team; he/she does not have any idea about how and where to start understanding these changes and their reasons. It results to design a wide time-consuming and costly market research which its result maybe did not reach on time to the marketing department to react to these changes. Also in such a market, there are many promotion campaigns by company itself and competitors that it is difficult to analyze the effectiveness of them in the market. So, in the competitive environment, there is a need to mine customer data and their transactions to find changes in customer purchasing behavior which is an effective and efficient way to respond to their needs timely and accurately. As a result, many FMCG distribution companies in Iran are trying to move away from traditional way for planning their marketing campaigns, promotions and market research by understanding changes happening in their customers' purchasing behavior. Change mining helps managers to make better marketing strategies.

۱,۲ Problem definition:

Kalleh Company is a private manufacturer and distributor of food product in Iran. It produces different categories of food product from dairy products to ice cream and meats and sauces. It has more than ۱۰ different categories and about ۸۰۰ products. Now, the company is faced with the challenge of increasing competition. There are some reasons behind it. First, according to the high variation of products, it should compete in different food market like dairy, ice cream and meat. It results to compete with many competitors with different product categories and different marketing strategies. Also there are some powerful governmental companies that make competition tougher for Kalleh. So in such a market, the customer behavior may change by the trend of companies' strategies in the market and also by changing their need by themselves.

Kalleh Company in order to answer to the changes in customer purchasing behavior timely and not being behind the customer needs and the competition need to mine changes in the customer purchasing behavior. The goal of Kalleh Company is to mine changes in purchasing behavior of the customers in different segments to respond to these changes timely and accurately to increase its return on investment (ROI).

۱,۳ Purpose of this study:

The purpose of this study is to mine changes in customer purchasing behavior. In order to reach this goal we need to building customer purchasing patterns of customers based on the customer, product and transaction data collected in databases.

Data mining techniques can help us to reach this goal. According to (Song et al, ۲۰۰۱), data mining is the process of exploration and analysis of large quantities of data in order to discover meaningful pattern and rules. Many of data mining Studies has focused on developing techniques to build precise models to predict customer's behavior, and to set up marketing strategies and customization. According to (Nemati & Barko, ۲۰۰۱; cited by Nemati, H.R., Barko, C. D., ۲۰۰۳), most of data mining applications (۷۲%) are centered on predicting customer behavior. Comparatively little attention has been paid to discover changes in databases collected eventually (Liu et al., ۲۰۰۰). From literature review, what is obvious is too much time spent on worrying about "absolute" numbers, like Lifetime Value. However, what they should really be observing is "relative" numbers – change over time. Highest potential ROI customers from a marketing viewpoint are Customers who are in the process of changing their

behavior either accelerating their relationship with you, or ending their relationship with a company (Novo, j., ٢٠٠٨). In many applications, mining changes can be more crucial than producing precise prediction models, which are in the center of existing data mining researches. Regardless of how the model is accurate, it is inactive by itself because it can only predict based on patterns mined in the old data. Acting based on the built model should not guide to actions that may change the environment because otherwise the model will stop to be correct (Liu et al., ٢٠٠٠). Prediction model building is more appropriate in areas where the environment is comparatively steady. However, in many business conditions, constant human interference to the environment is a fact. Businesses simply cannot let nature take its course. They constantly need to do actions in order to provide better services and products by finding the attractive changes and steady patterns in customer behaviors. Still in a comparatively steady environment, changes are also unavoidable due to internal and external issues (Liu et al., ٢٠٠٠).

From these viewpoints the question: ‘Which patterns exist?’ as it is responded by state-of-the art data mining technology, is replaced by the question: ‘How do patterns change?’ (Böttcher, M., et al, ٢٠٠٦). Actually, discovery of interesting and earlier unidentified changes in customer, product and transaction data, not only let the user monitor the influence of past business decisions but also to get ready today’s business for tomorrow’s needs (Böttcher, M., et al, ٢٠٠٦).

Major changes often need instant concentration and actions to modify the existing practices and/or to change the domain condition (Liu et al, ٢٠٠٠). By using change mining methodology, Kalleh Company can detect different kinds of changes happening in the customer purchasing behavior to build stronger relationship with the customers. Also, understanding changes in customer behavior can assist managers to set up effective and efficient promotion campaigns.

(Liu et al, ٢٠٠٠) mentioned that there are two main goals for mining changes in a business environment:

"To follow the trends": The main feature of this kind of applications is the word "follow". Companies like to know where the trend is going not to be left behind. They need to investigate customers' changing behaviors so as to provide products and services that suit the changing needs of the customers.

"To stop or to delay undesirable changes": In this kind of applications, the keyword is "stop". Companies like to know undesirable changes as soon as possible

and to plan corrective measures to stop or to delay the pace of such changes.

The overall procedure consists of several steps. In the literature, there are some methods for change mining in the dynamic situation. According to (Song et al, ۲۰۰۱), the majority of data mining techniques like association rules and neural networks cannot be used alone because they cannot manage dynamic situation well. (Song et al, ۲۰۰۱) and (Chen et al, ۲۰۰۵) developed a methodology for mining changes. They used association rule to detect interesting association relationships among a large set of data items which introduced by (Agrawal et al., ۱۹۹۳). The methodology detects all kinds of changes. According to (Chen et al, ۲۰۰۵), Change mining has several steps including data preprocessing, customer segmentation, mining association rule and change mining. In the first customers are segmented based on their behavioral variables, recency, frequency and monetary (RFM). Then by building association rule with customer behavioral variable (RFM), customer data and transaction data, we describe the customer purchasing behavior in two different time snapshots, and in the end we compare generated rules for each segment to mine changes in the customer purchasing behavior. To mine changes, various algorithms and techniques should be used. In order to implement these algorithms and techniques, an extensive programming is needed. Finally, we combined all of the algorithms to build a change mining package.

۱,۴ Research question:

Based on the problem discussion that we have above, the purpose of this study is to mine changes in customer purchasing behavior. In order to reach this purpose, the research question will be as followed: How businesses can be responsive to the changes of customer behavior in dynamic market. In addition, how businesses can detect and access to the changes happened in the customer behavior pattern to be responsive accurately and timely.

۱,۵ Research motivation:

Recently, we have watched an explosion of data produced and collected by individuals and organizations. This fast growth in data and databases made the problem of data overload (Li, X. B., ۲۰۰۵). More recently, increased computing power has led to greater elasticity in the models one can use and the amount of data that can be stored and processed (Bolton, R. J., ۲۰۰۴) and as a result, data mining techniques have come out and flourished in the past several years to encounter this demand (Li, X. B., ۲۰۰۵). Organizations are starting to understand the importance of data mining in their marketing strategies.

In this situation, businesses currently face the challenge of a constantly evolving market where customer needs are changing all the time (Chen et al, ۲۰۰۵). In such an environment, knowing the changes and responding rapidly and correctly to them, has a high importance. While customer needs change over time, if businesses could not meet their need, they would lose their customers who are their ROI resources. Some works have been done in change mining in retailing. One of the businesses that change mining can help it to improve, is FMCG distribution business that face a dynamic markets by huge variation of products and competitors in the market. The purpose of the change mining is following the trends that are happening in the customer purchasing pattern, detecting the changes and respond to them timely to satisfy customers more and meet their needs.

۱,۶ Demarcation: This study focus on mining changes in customer purchasing behavior based on the customer and purchasing transaction stored in a database. Change mining has been done by data gathered from a database of FMCG Distributor Company in Iran. Most of the literature reviewed is about mining changes in customer purchasing behavior. Our work focus on building customer behavior patterns by association rule mining and the comparison of these built rules. These patterns just based on their previous transactions.

۱,۷ Research outline: This thesis consists of five chapters. The first chapter is introduction that gives a brief background about subject followed by research question, objectives, and motivation. Chapter ۲ is a literature review, consists literature review on data mining, association rule, change mining and customer segmentation. Chapter ۳ is about our research methodology including data preprocessing, market segmentation, mining customer behavior and change mining. Chapter ۴ is about the results and analysis. Chapter ۵ is the last chapter that contains conclusion, limitation, and further research.

Chapter 7: Literature Review

Review of Mining Customer Behavior

Review of Data Mining

Review of Association Rule Mining

Review of Change Mining

Review of Customer Segmentation

2. 1 Mining Customer Behavior: Different methods to describe customer behavior exist in the literature. Among them, there are various types of conjunctive rules to build customer behavior pattern including association rules and classification rules (Agrawal R. et al, 1996 & Breiman L., et al., 1984 cited on Adomavicius, G., Tuzhilin, A., 2001)

Using rules to describe customer behavior has certain advantages. Besides being descriptive way to portray behaviors, a conjunctive rule is a well-studied concept and it is used widely in data mining, expert systems, and many other areas. In addition, researchers have proposed many rule discovery algorithms in the literature, especially for association rules (Adomavicius, G., Tuzhilin, A., 2001). To discover rules that describe the behavior of customers, we can use various data mining algorithms, like Apriori for association rule mining.

Association rules were initially applied for market basket analysis to find the relationships between product items purchased by customers at retail stores (Agrawal, Imielinski, & Swami, 1993; Srikant, Vu, & Agrawal, 1997 cited by Chen et al, 2005). In a research of customer behavior, we can apply association rule to find the correlations between customer demographic variables, purchased product and product databases (Song et al, 2001).

In this chapter, we will have a review of data mining, then association rules. Then the next topic will be the change mining of customer behavior in the literature. And following by that finally we will have a brief review of customer segmentation.

2. 2 Review of Data Mining

2. 2.1 Data mining: in brief

Today, size of databases can be very large. Within this data you can find hidden strategic information. But when you have a huge amount of data, inducing meaningful conclusions is not easy. The novel answer is data mining being used both to increase revenues and to reduce costs. Many people use data mining as a synonym for another popular word, Knowledge Discovery in Database. In rotation other people define Data Mining as the core process of KDD.

The KDD processes are shown in Figure 2.1 (Han, J., & Kamber, M., 2006). Usually KDD has three processes. First one is preprocessing executed before data mining techniques applied to the right data. The preprocessing includes data cleaning, integration, selection and transformation. The main process of KDD is the data mining process. In this process different algorithm are applied to produce hidden knowledge.

The last process is post-processing comes evaluating the mining result according to users' requirements and domain knowledge.

Regarding the evaluation results, if the result is satisfactory the knowledge can be presented; else we have to run some or all of those processes again till we get the satisfactory result (Han, J., & Kamber, M., 2006).

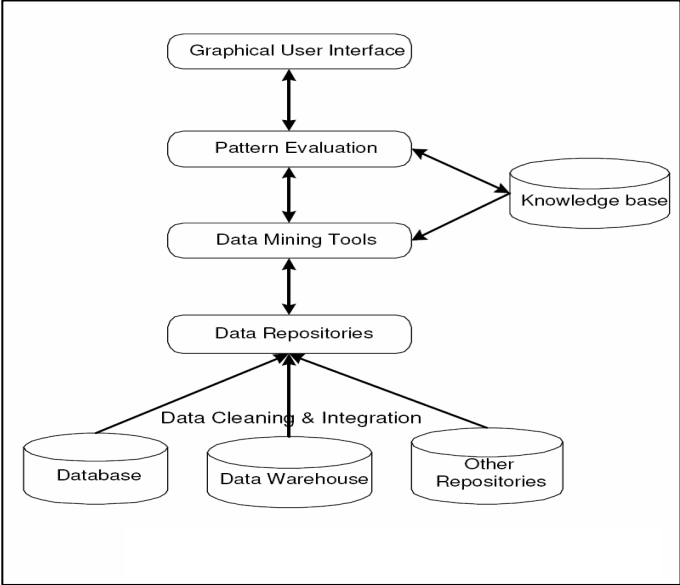
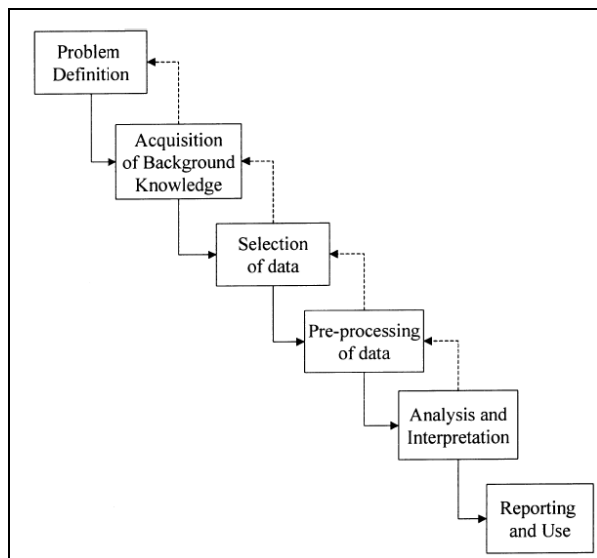


Figure 2.1: Knowledge Discovery in Database Processes

(Song et al, 2001) defines data mining as a process of exploration and analysis of large quantity of data to discover meaningful patterns and rules. (Feelders et al, 2000) define the process of data mining as follows:

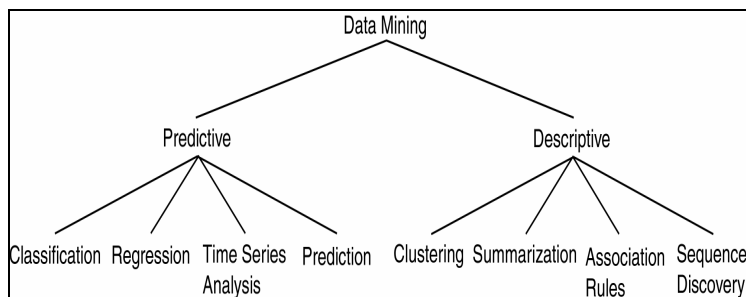


Source: (Feedlers et al, ٢٠٠٠)

Figure ٢,٢ the major steps in data mining Process

The data mining returns potential is immense. Innovative organizations worldwide are already using data mining to attract higher-value customers, to configure their product offerings differently to increase sales, and to minimize losses due to mistakes or fraud.

٢. ٢,٢ Data mining Functions: (Dunham, ٢٠٠٢) categorizes data mining to two categories, one is descriptive and the other one is predictive (Figure ٢,٣).



Source: (Dunham, ٢٠٠٢)

Figure ٢,٣: Classification of DM techniques

The first and simplest analytical step in data mining is to describe the data-summarize its statistical attributes such as means and visual review like charts and

graphs, and correlations among variables. The most important step is right data selection, data gathering and data exploration. Sometimes data description alone cannot provide an action plan. You must build a predictive model based on patterns determined from known results, and then examine that model with a new sample data. A good model should never be the same as reality, but it can be a useful guide to know your business. And after all we should empirically verify the model (Twocrows.com, 2005). In the next section, we explain briefly three important data mining techniques.

2.2.3 Classification in brief:

Based on (Han and Kamber, 2006), Classification is automatic model building that can classify a class of objects so as to predict the classification or missing attribute value of future objects whose class may not be known. The process has 3 steps. In the first step, a model is built to describe the characteristics of a set of data classes or concepts based on the collection of training data set. Because data classes or concepts are predefined, this step is also known as supervised learning. In the second step, the model is used to predict the classes of future data or objects. There are several techniques for classification (Han and Kamber, 2006). In Classification by decision tree many researches are done and plenty of algorithms have been designed, Murthy did a extensive survey on decision tree induction (Murthy, 1998; cited by Han, J., & Kamber, M., 2006). Bayesian classification is another technique that can be found in (Duda and Hart, 1993 cited by Han, J., & Kamber, M., 2006). Nearest neighbor methods are also talked about in many statistical texts on classification, such as (Duda and Hart, 1993, cited by Han, J., & Kamber, M., 2006) and (James, 1980, cited by Han, J., & Kamber, M., 2006). Besides, there are many other machine learning and neural network techniques used to help building the classification models.

2.2.4 Clustering in brief: As we mentioned before, classification can be taken as supervised learning process, clustering is another mining technique similar to classification. However clustering is an unsupervised learning process. "Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects" (Han, J., & Kamber, M., 2006), so that objects within the same cluster must be similar to some extent, also they should be dissimilar to those objects in other clusters. In classification each record belongs to a predefined class, while in clustering there is no predefined class. In clustering, objects are grouped together based on their similarities. (Han, J., & Kamber, M., 2006) Similarities between objects are explained by some similarity functions; usually similarities are quantitatively defined as distance or other measures by corresponding domain experts. (Han, J., & Kamber, M., 2006) Most clustering applications are used in market segmentation. When they cluster their