Journal of Marketing and Consumer Research
ISSN 2422-8451 An International Peer-reviewed Journal
Vol.9, 2015

Data Mining for Marketing

Rafi Ahmad Khan
The Business School, University of Kashmir, Hazratbal,Srinagar, J & K 19000, India
mca_rafi@yahoo.com

Aiman Mushtaq
The Business School, University of Kashmir, Hazratbal,Srinagar, J & K 19000, India
aiman.mnm@gmail.com

Hina Kanth
The Business School, University of Kashmir, Hazratbal,Srinagar, J & K 19000, India
mirhina786@gmail.com

Abstract
This paper gives a brief insight about data mining, its process and the various techniques used for it in the field of marketing. Data mining is the process of extracting hidden valuable information from the data in given data sets .In this paper cross industry standard procedure for data mining is explained along with the various techniques used for it. With growing volume of data every day, the need for data mining in marketing is also increasing day by day. It is a powerful technology to help companies focus on the most important information in their data warehouses. Data mining is actually the process of collecting data from different sources and then interpreting it and finally converting it into useful information which helps in increasing the revenue, curtailing costs thereby providing a competitive edge to the organisation.

1. Introduction to Data Mining For Marketing

Marketing as a discipline involves researching and developing a product and facilitating its sale and distribution to the general public. The concept of marketing has existed since long and is changing as per the needs and purchasing behaviours of consumers. Thus, today's marketing is very different from what it used to be a few decades ago, mainly due to a rapidly changing world economy and the advancement in the technology, which together led to free and speedy knowledge distribution and exchange. With this, local markets were exposed to MNC's as the cost and complexity of operating overseas was reduced by globalisation thereby contributing to a wider competition. As markets have become more deregulated, there has been a major change in the way in which and the speed with which knowledge is disseminated. Almost everything we do, like taking a long walk in the woods, leaves little bits of electronic data behind. Every time the Internet or mobile phone is used, data is punched and there's a giant industry right behind sucking up all that data and using it to figure out how to sell you something. From selling toothpaste down to life insurance policies, every activity generates some amount of data, courtesy globalisation and technical advancements, which if analysed properly can provide a competitive edge in the international market by discovering hidden patterns and explicit relationships among large data sets. One such technique that can be employed to analyse such large amount of data is known as data mining.

Data mining is the process of finding correlations or patterns among dozens of fields in large databases. The overall goal of data mining is to extract knowledge from an existing data set and transform it into a human-understandable structure for further use.

Data mining is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data and data mining as the extraction of patterns or models from observed data. (Berzal, et al., 2001)

Data mining can be characterized as the technology which incorporates the statistical techniques and mathematical equations that are used in an attempt to identify the significant relationships between variables in the historical data, to forecast or perform analysis on the data; or determine any significant relationship within the data captured (Becker, 2002).

Data mining is the search for the relationships and global patterns that exist in large data bases but are hidden among vast amounts of data , such as the relationship between patient data and their medical diagonosis.This relationship represents valuable knowledge about the data base and the objects in the data base .(Pujari, 2002)

Data mining is an essential process where intelligent methods are applied in order to extract data pattern. (Han & Kamber, 2007)

Data mining is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. (Berry & Linoff, 2008)

Nowadays, large quantities of data are being accumulated. The amount of data collected is said to be almost doubled every 9 months. Seeking knowledge from massive data is one of the most desired attributes of Data Mining. Usually there is a huge gap between the stored data and the knowledge that can be constructed from it. This transition does not occur automatically, that's where Data Mining comes into picture. This technology is popular with many marketers because it allows them to learn more about their customers and make smart marketing decisions. The data mining business, as it is known, grows 10 percent a year as the amount of data produced is booming. The information thus produced from using data mining can help to increase return on investment (ROI), improve customer's relationships management (CRM) and market analysis, reduce marketing campaign costs, and facilitate fraud detection and customer retention.

Marketing is defined as putting the right product in the right place, at the right price, at the right time. It is a process of planning and executing the conception, pricing, promotion and distribution of ideas, goods and services to create exchanges that satisfy individual and organisational objectives (Kotler & Armstrong, 2006). A lot of hard work needs to go into finding out what customers want, and identifying where they do their shopping. Then it is to be figured out how to produce the item at a price that represents value to them, and get it all to come together at the critical time.

But one wrong element can lead to a disaster e.g. promoting a car with amazing fuel-economy in a country where fuel is very cheap or publishing a textbook after the start of the new school year, or selling an item at a price that's too high or too low to attract the people you're targeting. A good place to start off is the marketing mix where plans for a product or service can be thought over so as to avoid any mistakes. It is a general phrase used to describe the different kinds of choices organizations have to make in the whole process of bringing a product or service to market. The 4 Ps is one way of the best way of defining the marketing mix, and was first expressed in 1960 by E J McCarthy.

The 4Ps are:

- Product (or Service)
- Price
- Place
- Promotion

Product

It is the tangible object or an intangible service that is getting marketed through the program. Tangible products may be items like consumer goods (Toothpaste, Soaps and Shampoos) or consumer durables (Watches, IPods). The first thing to start a business is a product. Therefore Product is also the first variable in the marketing mix. Product decisions are the first decisions you need to take before making any marketing plan. A product can be divided into three parts; core product, augmented product and tertiary product. Before deciding on the product component there are some questions which need to be pondered upon:

- What does the customer want from the product? What needs does it satisfy?
- What features does it have to meet these needs?
- Are there any features you've missed out?
- Are you including costly features that the customer won't actually use?
- How and where will the customer use it?
- What does it look like? How will customers experience it?
- What size(s), colour(s), and so on, should it be?
- What is it to be called?
- How is it branded?
- How is it differentiated versus your competitors?
- What is the most it can cost to provide, and still be sold sufficiently profitably?

Pricing

Pricing of a product depends on a lot of different variables and hence it is constantly updated. Major consideration in pricing is the costing of the product, the advertising and marketing expenses, any price fluctuations in the market, distribution costs etc. Many of these factors can change separately. Thus the pricing has to be such that it can bear the brunt of changes for a certain period of time. However, if all these variables change, then the pricing of a product has to be increased and decreased accordingly. Some of the questions that need to be kept in view while deciding on the pricing are:

- What is the value of the product or service to the buyer?
- Are there established price points for products or services in this area?
- Is the customer price sensitive?
- Will a small decrease in price gain you extra market share? Or will a small increase be indiscernible, and so gain you extra profit margin?

- What discounts should be offered to trade customers, or to other specific segments of your market?
- How will your price compare with your competitors?

Place
Place represents the location where a product can be purchased. It is often referred to as the distribution channel. This may include any physical store (supermarket, departmental stores) as well as virtual stores (e-markets and e-malls) on the Internet. This is crucial as this provides the place utility to the consumer, which often becomes a deciding factor for the purchase of many products across multiple product categories. Some of the things that need to be considered are:

- Where do buyers look for your product or service?
- If they look in a store, what kind? A specialist boutique or in a supermarket, or both? Or online? Or direct, via a catalogue?
- How can you access the right distribution channels?
- Do you need to use a sales force? Or attend trade fairs? Or make online submissions? Or send samples to catalogue companies?
- What do you competitors do, and how can you learn from that and/or differentiate?

Promotion
This represents all of the communications that a marketer may use in the marketplace to increase awareness about the product and its benefits to the target segment. Promotion has four distinct elements: advertising, public relations, personal selling and sales promotion. A certain amount of crossover occurs when promotion uses the four principal elements together (e.g. in film promotion). Sales staff often plays a major role in promotion of a product. Some of the questions that need to be answered before promoting a product are:

- Where and when can you get across your marketing messages to your target market?
- Will you reach your audience by advertising in the press, or on TV, or radio, or on billboards? By using direct marketing mail shot? Through PR? On the Internet?
- When is the best time to promote? Is there seasonality in the market? Are there any wider environmental issues that suggest or dictate the timing of your market launch, or the timing of subsequent promotions?
- How do your competitors do their promotions? And how does that influence your choice of promotional activity?

There are a number of applications of data mining in the field of marketing. One of them is called market segmentation, with which common customer behaviours are identified. Patterns among customers that seem to purchase the same products at the same time are looked for. Another application of data mining is called customer churn, which helps estimating such customers who are likely to stop purchasing products or services and go to on competitors. In addition, a company can use data mining to find out which purchases are most likely to be fraudulent. For example, by using data mining, a retail store may be able to determine which products are stolen the most so that protective measures are taken accordingly.

Further, while direct mail marketing is an old technique, the companies can, however, combine it with data mining for fantastic results. For example, data mining can be used to find out which customers will respond favourably to a direct mail marketing strategy. It also determines the effectiveness of interactive marketing. Some of the customers are more likely to purchase products online than offline, and, as such, there is a need to identify such customers.

While many marketers use data mining to help increase their profits, it can also be used to create new businesses and industries, and every such data-mining based industry is based on the automatic prediction of both behaviours and trends. For example, automatic prediction is used in data mining to look at past marketing strategies. Which one worked the best? Why did it work the best? Who were the customers that responded most favourably to it? Data mining answers these questions and thus helps avoid making any mistakes that were made in previous marketing campaign. Thus data mining for marketing helps organisations to gain competitive advantage over others and sustain in the international market.

2. Process
Data mining process analyses large amount of data stored in databases so as to discover hidden valuable information or knowledge. This process helps marketers make better decisions that ultimately lead to the required objective such as increase in revenue or process efficiency. To get a competitive edge over others various industries in the field of marketing is making use of data mining process. The process used must be reliable and repeatable by business people with little or no knowledge about data mining.

One such data mining process developed in 1990 is the cross industry standard procedure for data mining (CRISP-DM). It is an iterative process which consists of the following six phases:

- Defining Project Objectives
- Data Exploration
- Data Preparation
- Model Building
- Assessment of Models
- Implementation

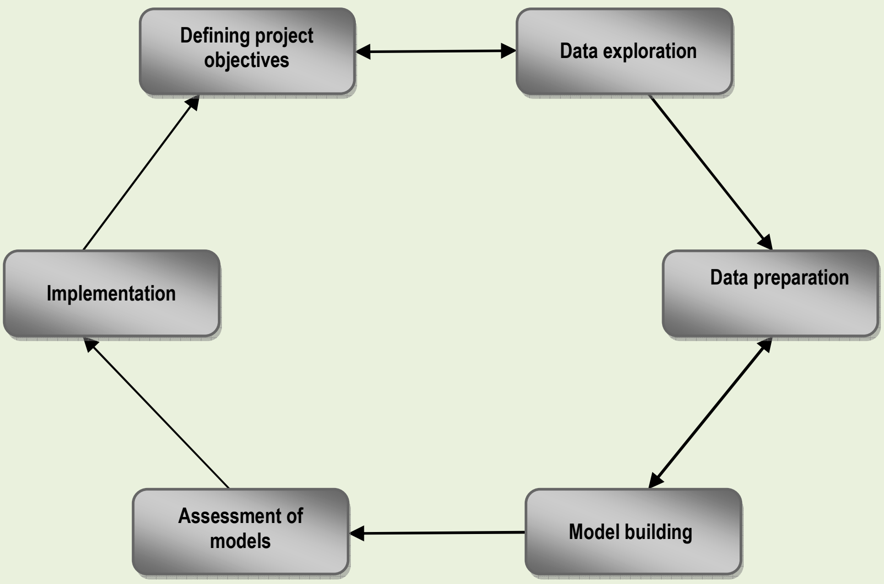

Fig.1 cyclic process of data

2.1 Defining Project Objectives

This phase focuses on understanding and defining the project objectives from the perspective of the business. The knowledge about these objectives is then converted into data mining problem and a plan is designed so as to fulfil these objectives. The steps involved in this phase are:

2.1.1 Determining Business Objectives

This step requires the data analyst to thoroughly understand the objectives and requirements of the project, which the client wants to accomplish. Any important factors that can influence the outcome of the project need to be uncovered.

2.1.2 Assess Situation

This task involves detailed fact finding about the resources, constraints, assumptions and requirements.

2.1.3 Determining Data Mining Goals

Once the objectives are uncovered, these need to b stated in business terms. Data mining goal states the project objectives and requirements in technical terms

2.1.4 Project Plan

To accomplish the required goals a plan is devised which states the necessary steps to be performed, including an initial selection of data mining tools and techniques.

2.2 Data Exploration

This phase involves the collection of required data and proceeds through activities for getting familiar with the data. Assessment of sources, quality and characteristics of the data are accomplished in this phase along with understanding the meaning of metadata to form hypotheses for the hidden information. The various steps involved in this phase are:

2.2.1 Collecting Initial Data

This task deals with collecting or getting access to the data listed in the resources from the objective definition phase.

2.2.2 Describe Data

The surface properties of the collected data are examined and the results are reported.

2.2.3 Explore Data
Initial data exploration involves provides first insights into the data .Meaning of metadata is understood and any questions that can be addressed using quering, visualisation and reporting are tackled. These analyses may even address the project goals directly.
2.2.4 Verify Data Quality
This task involves examining the quality of the collected data. It addresses questions such as: is the data complete? Is the data correct? Are there any missing values? How common are they and how are they represented?

2.3 Data Preparation
This phase covers all the activities required to construct the final data set that is to be fed into the modeling tools from the initial raw data .The tasks such as table, record and attribute selection as well as the transformation and cleaning of data for modeling tools in the data preparation phase are tweaked multiple times in no prescribed order. The various steps involved in this phase are:
2.3.1 Data Selection
This step involves deciding on the data that is relevant to data mining goals, quality and technical constraints to be used for analyses .Selection of attributes as well as of records is accomplished in this step.
2.3.2 Clean Data
Any problems that occur due to missing data, empty values, non-existent values and incomplete data are addressed so as to raise the data quality to the level required by the selected analysis technique. Also the data is converted to a format suitable for model building.
2.3.3 Construct Data
In this task new derived attributes e.g. average value, entirely new records or transformed values for existing attributes are created.
2.3.4 Data Integration
The constructed data is integrated to create new values e.g. two tables containing different information about same product are integrated to generate new values.
2.3.5 Data Formatting
Using a modeling tool syntactic modifications are made to data without changing its meaning.

2.4 Model Building
This phase involves selecting and applying various modeling techniques as well as calibrating the parameters to optimal values .Several techniques can address same data mining problem .Some of these techniques require a specific form of data .Therefore interaction with the data preparation phase is often required. The various steps involved in this phase are:
2.4.1 Selecting Modeling Technique
This step involves selecting the actual modeling technique to be used e.g. building decision trees or generating neural networks.
2.4.2 Generate Test Design
To test the quality and validity of the model a test design is generated.
2.4.3 Building Model
Models are developed so as to use the predictions for effective decision making. The model developed should be stable i.e. it should be able to make predictions that will hold true even when applied to unseen data. Also the problem of over fitting should not occur.
2.4.4 Assess Model
In this step the assessment of the developed model for accuracy and generality takes place.

2.5 Assessment of Models
This phase involves the thorough evaluation of model before implementation. It assess the degree up to which the developed model meets the business objectives and requirements and also seeks to determine if there is any reason that the chosen model is deficient. Two main questions need to be addressed:
- Does the model achieve the business objective?
- Have all the business issues been considered?

2.6 Implementation
This phase involves incorporating the data mining results into the day-to-day decision making process. Depending on the requirement this phase can be as simple as giving a report or as complex as implementing a repeatable data mining process. Implementation is followed by review of project to assess what went right and what went wrong, what has been done well and what needs to be improved.

3. Techniques

The various techniques employed in data mining are as follows:

3.1 Classification

It is a classic data mining technique and is based on machine learning. It can be defined as the process of predicting the class label of the various data present in the data set, and is used to map the data item into one of the several predefined classes. In classification approach we normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and then builds a model which is used to classify the objects. For example, a typical classification problem is to divide a database of companies into groups that are as homogeneous as possible with respect to a creditworthiness variable with values "Good" and "Bad." Other examples of classification technique may include predicting tumor cells as benign or malignant, classifying credit card transactions as legitimate or fraudulent, classifying secondary structures of protein as alpha-helix, beta-sheet, or random coil, categorizing news stories as finance, weather, entertainment, sports, etc.

Classification is learning function that maps (classifies) a data item into one of several predefined classes (Weiss, Brusic, & Zeleznikow, 1991)

The classifier-training algorithm uses pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

Classification is the separation or ordering of objects (or things)into classes.if the classes are created without looking at the data (non-emperically),the classification is called apriori classification.if however the classes are created emperically (by looking at the data),the classification is called posteriori classification. (Gupta, 2006)

Classification involves finding rules thar partition the data into disjoint groups .the input for the classification is the training data set ,whose class labels are already known. (Pujari, 2002)

The commonly used methods for data mining classification tasks can be classified into the following groups:

3.1.1 Logistic regression (LOG)

It extends the ideas of multiple linear regression to the situation where the dependent variable, y, is discrete. In logistic regression (Hosmer & Lemeshow, 2000) no assumptions are made concerning the distribution of the independent variables.

3.1.2 K-Nearest Neighbors

The methodology of the k-nearest neighbor algorithm (KNN) is very intuitive. It considers the k labelled samples nearest to sample x to be classified and assign x to the most common class of these k neighbors (Aha & Kibler, 1991) (Han & Kamber, 2007). An appropriate k value can be determined by using k-fold cross validation. The parameter k is supposed to be odd in order to avoid ties. This algorithm can produce non-linear classification boundaries, while it is very computationally expensive and may have the effect of overtraining when the samples are overlapping.

3.1.3 Naïve Bayes

The Naïve Bayes technique (NB) is one simplified form of the Bayesian network for classification. A Bayes network can be seen as a directed acyclic chart with a joint probability distribution over a set of discrete and stochastic variables (Pearl 1988). It has the advantage of incorporating domain knowledge and prior distribution information of the observed samples. (Han & Kamber, 2007)

3.1.4 Tree Augmented Naïve (TAN)

Bayes technique is an extension of the NB approach. Since the conditional independence of variables might be unrealistic in real-life, TAN relaxes the assumption by allowing dependence between condition variables. From a tree-diagram viewpoint, the class variable has no parents and the condition variable has the class variable and at most one other variable as parents. (Friedman, Geiger, & Goldszmidt, 1997)have presented an algorithm to learn the structure of a TAN and its corresponding parameters.

3.1.5 Decision Trees

A decision tree (DT) classifier has a tree-like structure, which contains two types of nodes – internal nodes and leaf nodes. An internal node corresponds to a test for samples on individual or several attributes and a leaf node assigns class labels to samples based on the class distribution it records. A decision tree classifies samples in a top-down manner, starting from the root node and keeping moving according to the outcome of the tests at internal nodes, until a leaf node is reached and a class label is assigned. (Breiman, Friedman, Olshen, & Stone, 1984)

3.1.6 Artificial neural networks (ANN)

They are mathematical representations based on the understanding of the structure and mechanism of the human brain (Ripley, 1996) (Haykin, 1998). The characteristics of ANN are subject to their topologies and corresponding weights. The learning procedure in an ANN is to tune weights intelligently after their topologies are designed. From a marketing perspective, neural networks are a form of software tool used to assist in decision making. Neural networks are effective in gathering and extracting information from large data sources and have the ability to identify the cause and effect within data . These neural nets through the process of learning identify relationships and connections between data bases. Once knowledge has been accumulated, neural networks can be relied on to provide generalisations and can apply past knowledge and learning to a variety of situations

Neural networks help fulfil the role of marketing companies through effectively aiding in market segmentation and measurement of performance while reducing costs and improving accuracy. Due to their learning ability, flexibility, adaption and knowledge discovery, neural networks offer many advantages over traditional models. Neural networks can be used to assist in pattern classification, forecasting and marketing analysis.

3.1.7 Associative Classification

Deriving association rules were first proposed by Agrawal and his associates in order to discover the concurrence of items in transaction datasets.(Agrawal & Imielinski , 1993) (Agrawal & Srikant, 1994) One of the most popular associative classification techniques is CBA (Liu, Hsu, & Ma, 1998) It first generates, by using an adapted Apriori algorithm (Agrawal & Srikant 1994), all class association rules that satisfy minimum support and confidence. These rules are then ranked and sorted in descending sequence. Another associative classifier is ADT (Wang & Zhou 2000) and it organizes the rule sets in the tree structure according to its defined relations. Three of the most recent associative classification algorithms are CAEP (Dong, Zhang, Wong & Li 1999), CMAR (Liu, Han & Pei 2001) and CPAR (Yin & Han 2003). Moreover, another new approach to associative classification is GARC , which incorporates information gain, excluded sets, and certain redundancy/conflict resolution strategies into the classification rule generation process, so as to significantly reduce the number of rules generated while keeping the accuracy satisfactory.

3.1.8 Support Vector Machines

A support vector machine (SVM) is based on the statistical learning theory (SLT) and structural risk minimization (SRM) developed by Vapnik and his co-workers since the 1960's (Burges 1998; Cristianini & Shawe-Taylor 2000; Vapnik 1995). It exerts a deliberate trade-off between complexity and accuracy of the classifier on the training set in order to achieve better generality ability.

3.1.9 Genetic Algorithms (GAs)/Evolutionary Programming (EP)

Genetic algorithms and evolutionary programming are algorithmic optimization strategies that are inspired by the principles observed in natural evolution. Of a collection of potential problem solutions that compete with each other, the best solutions are selected and combined with each other. In doing so, one expects that the overall goodness of the solution set will become better and better, similar to the process of evolution of a population of organisms. Genetic algorithms and evolutionary programming are used in data mining to formulate hypotheses about dependencies between variables, in the form of association rules or some other internal formalism.

3.1.10 Fuzzy Sets

Fuzzy sets form a key methodology for representing and processing uncertainty. Uncertainty arises in many forms in today's databases: imprecision, non-specificity, inconsistency, vagueness, etc. Fuzzy sets exploit uncertainty in an attempt to make system complexity manageable. As such, fuzzy sets constitute a powerful approach to deal not only with incomplete, noisy or imprecise data, but may also be helpful in developing uncertain models of the data that provide smarter and smoother performance than traditional systems.

3.1.11 Rough Sets

A rough set is determined by a lower and upper bound of a set. Every member of the lower bound is a certain member of the set. Every non-member of the upper bound is a certain non-member of the set. The upper bound of a rough set is the union between the lower bound and the so-called boundary region. A member of the boundary region is possibly (but not certainly) a member of the set. Therefore, rough sets may be viewed as with a three-valued membership function (yes, no, perhaps). Rough sets are a mathematical concept dealing with uncertainty in data. They are usually combined with other methods such as rule induction, classification, or clustering methods.

Application of classification in marketing:

Classification method can help the marketers in the following four dimensions:

(1) Customer Identification; Elements for customer identification include target customer analysis and customer segmentation. Target customer analysis involves seeking the profitable segments of customers through analysis of customers' underlying characteristics, whereas customer segmentation involves the subdivision of an entire customer base into smaller customer groups or segments, consisting of customers who are relatively similar within each specific segment (Woo, Bae, & Park, 2005)

(2) Customer Attraction

After identifying the segments of potential customers, organizations can direct effort and resources into attracting the target customer segments. An element of customer attraction is direct marketing. Direct marketing is a promotion process which motivates customers to place orders through various channels. (Cheung, Kwock, Law, & Tsui, 2003) (He, X, & J, 2004) (Liao & Chen, 2004) (Prinzie & Poel, 2005)For instance, direct mail or coupon distributions are typical examples of direct marketing.

(3) Customer Retention

Elements of customer retention include one-to-one marketing, loyalty programs and complaints management. One-to-one marketing refers to personalized marketing campaigns which are supported by analysing, detecting and predicting changes in customer behaviours (Chen, Chiu, & Chang, 2005) (Jiang & Tuzhilin, 2006 (Kim & Moon, 2006)

(4) Customer Development.

Elements of customer development include customer lifetime value analysis, up/cross selling and market basket analysis. Customer lifetime value analysis is defined as the prediction of the total net income a company can expect from a customer (Drew, Mani, Betz, & Datta, 2001) (Etzion, Fisher, & Wasserkrug, 2005) (Rosset, Neumann, Eick, & Vatnik, 2003)Up/Cross selling refers to promotion activities which aim at augmenting the number of associated or closely related services that a customer uses within a firm. (Prizie & Poel, 2006) Market basket analysis aims at maximizing the customer transaction intensity and value by revealing regularities in the purchase behaviour of customers

These four dimensions can be seen as a closed cycle of a customer management system that share the common goal of creating a deeper understanding of customers to maximize customer value to the organization in the long term. Classification techniques, therefore, can help to accomplish such a goal by extracting customer characteristics and behaviours from large databases and then grouping them classifying them based on their characteristics such as the buying pattern, buying frequency, etc.

Beside the above mentioned points the Classification technique can help us in identifying frequency of purchases, size of purchases, regency of purchases and in identifying typical customer groups. The characteristics of each group can be obtained by class identification or concept description. For example, a profile indicating that the customer has purchased a new house may lead to the marketer offering a special deal for home furnishings. Knowing the customer and targeting the right deal gets a far better response rate than a general message

3.2 Sequencing

Sequencing or time series analysis methods relate events in time, based on a series of preceding events. It consists of sequences of events obtained over repeated measurements of time. The items are typically measured at equal time intervals (e.g. Hourly, daily, weekly) through this analysis various hidden trends, often highly predictive of future events, can be discovered. Data is mined to anticipate behaviour patterns and trends. Sequences are often analysed as they relate to a specific customer or group of customers. Using this information a catalogue containing specific product types can be target mailed to a customer associated with a known sequence of purchases. E.g. an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes. Also a known buying sequence can be that parents tend to buy promotional toys associated with a particular movie within two weeks after renting the movie.

3.3 Clustering

The process of grouping a set of physical objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group. We first partition the set of data into groups based on data similarity (example using clustering) and the assign labels to the relatively small number of groups. E.g. business clustering can help marketers discover distinct groups in their customer bases and characterise customer groups based on purchasing patterns. The customer of a given geographic location and of a particular job profile demand a particular set of services, like in banking sector the customers from the service class always demand for the policy which ensures more security as they are not intended to take risks, likewise the same set of service class people in rural areas have the preference for some particular brands which may differ from their counterparts in urban areas. This information will help the organization in cross-selling their products, This technique will help the management in finding the solution of 80/20 principle of marketing, which says: Twenty per cent of your customers will provide you with 80 per cent of your profits, then problem is to identify those 20 % and the techniques of clustering will help in achieving the same.

Clustering is the process of grouping objects into clusters such that the objects from the same cluster are similar and objects from different clusters are dissimilar. (Soman, Diwakar, & Ajay, 2006)

Clustering is the task of segmenting a heterogeneous population into a number of more homogeneous sub groups or clusters. (Berry & Linoff, 2008)

Clustering is a technique for the discovery of data distribution and patterns in the underlying data. (Pujari, 2002) Clustering is a collection of methods that assists the user in putting different objects from a collection of objects into different groups. (Gupta, 2006)

3.3.1 Approaches to Clustering

3.3.1.1 Hierarchical Clustering

The hierarchical clustering techniques do a sequence of partitions in which each partition is nested into the next partition in this sequence. It creates a hierarchy of clusters from small to big or big to small. The hierarchical techniques are of two types agglomerative and divisive clustering techniques. Agglomerative clustering techniques start with as many clusters as there are records, with each cluster having only one record. At each stage, the pairs of the clusters that are merged are the ones nearest to each other. Divisive clustering techniques take the opposite approach from agglomerative techniques this starts with all the records in one cluster, and then try to split that cluster into small pieces.

3.3.1.2 Partitioning Clustering

The partitioning clustering techniques partition the data base into a pre defined number of clusters. They attempt to determine partitions that optimize a certain criterion on function. Partitioning technique partitions a data base of N objects into a set of k clusters. The objective of partitioning is to make clusters in order to reach the optimisation.

3.3.1.3 Density-Based Clustering

These algorithms group objects according to specific density objective functions. Density is usually defined as the number of objects in a particular neighbourhood of a data objects. In these approaches a given cluster continues growing as long as the number of objects in the neighbourhood exceeds some parameter. This is considered to be different from the idea in partitional algorithms that use iterative relocation of points given a certain number of clusters. Three major algorithms in this category are Density-Based Spatial Clustering of Applications with noise (DBSCAN). Ordering Points To Identify the Clustering Structure (OPTICS) (Hinneburg and Keim, 1998).

3.3.1.4 Grid-Based Clustering

The main concept of these algorithms is the quantization of the data space into a number of cells. STING (Statistical Information Grid), (Wang et al, 1997) Wave Cluster (Sheikholeslami, Chatterjee, & Zhang, 1998) and CLIQUE (Clustering In Quest) (Agrawal et al,1998) are three representatives of this family of algorithms. The focus of these algorithms is spatial data, i.e., data that model the geometric structure of objects in space, their relationships, properties and operations. The objective of these algorithms is to quantize the data set into a number of cells and then work with objects belonging to these cells. They do not relocate points but rather build several hierarchical levels of groups of objects. In this sense, they are closer to hierarchical algorithms but the merging of grids, and consequently clusters, does not depend on a distance measure but it is decided by a predefined parameter.

3.3.1.5 Model-Based Clustering

These algorithms find good approximations of model parameters that best fit the data (Blimes,1998). They can be either partitional or hierarchical, depending on the structure or model they hypothesize about the data set and the way they refine this model to identify partitioning. They are closer to density-based algorithms, in that they grow particular clusters so that the preconceived model is improved. However, they sometimes start with a fixed number of clusters and they do not use the same concept of density.

3.3 .1.6 Categorical Data Clustering

These algorithms are specifically developed for data where Euclidean, or other numerical-oriented, distance measures cannot be applied. In the literature, we find approaches close to both partitional and hierarchical methods.

Application in marketing

Clustering may typically be used in marketing for advertisements when the customers are yet to be segmented. After running a cluster analysis, the clusters may be examined for characteristics based on which advertisement campaigns may be directed at the customer base. After segmentation, based on the characteristics of the clusters, product positioning, product repositioning and product development may be done, to improve its fit with the targeted customers. Cluster analysis may also be done to selecting test markets. Also clustering may be used in customer relationship management (CRM). Customer clustering would use customer-purchase transaction data to track buying behaviour and create strategic business initiatives. Companies want to keep high-profit, high-value, and low-risk customers. This cluster typically represents the 10 to 20 per cent of customers who create 50 to 80 per cent of a company's profits. A company would not want to lose these customers, and the strategic initiative for the segment is obviously retention. A low-profit, high-value, and low-risk customer segment is also an attractive one, and the obvious goal here would be to increase profitability for this segment. Cross-selling (selling new products) and up-selling (selling more of what customers currently buy) to this segment are the marketing initiatives of choice.

3.4 Association

Association rule mining also known as affinity analysis is the study of "what goes with what". The task of affinity grouping is to determine which things go together. Association rule mining finds interesting associations and/ or correlation relationships among large sets of such data items. Association rules show attributes value conditions that occur frequently together in a given data sets. Association rules show strong associations between items that occur frequently in a given data set. These rules are commonly used to analyse the purchasing patterns of customers in a store, such analysis is useful in many decision making processes such as product placement and catalogue design. This technique helps the marketers in finding patterns which help in guiding the organization to make decisions regarding pricing, selling and to design the strategies for marketing. The association may be direct or in direct. Direct such as purchasing a pen and paper, That means when the customer buys paper then he/she will buy the pen also, this association will help the organization in designing the layout of store, by placing these two products adjacent to each other, which will lead to convenience to the customer and organization can use these results for designing the pricing decision and can give offers based on this study. The organization can find that which customer buys which product most of the times together and hence can provide discounts based on the results

Association analysis is the discovery of association rules sharing attribute-value conditions that occur frequently together in a given set of data. (Rao, 2003)

Association rules shows attribute value conditions that occur frequently together in a given data set. (Soman, Diwakar, & Ajay, 2006)

Other types of association mining

- Contrast set learning is a form of associative learning. Contrast set learners use rules that differ meaningfully in their distribution across subsets.
- Weighted class learning is another form of associative learning in which weight may be assigned to classes to give focus to a particular issue of concern for the consumer of the data mining results.
- K-optimal pattern discovery provides an alternative to the standard approach to association rule learning that requires that each pattern appear frequently in the data.
- Mining frequent sequences uses support to find sequences in temporal data.
- Generalized Association Rules hierarchical taxonomy (concept hierarchy)
- Quantitative Association Rules categorical and quantitative data
- Interval Data Association Rules e.g. partition the age into 5-year-increment ranged
- Maximal Association Rules
- Sequential Association Rules

Application in marketing:

Pattern association may be extensively used to predict customer preferences when very little data about the customer is available to the marketer. Tools for pattern association would help a marketer to predict which product or advertisement the customer may be interested in solely by the current buying behaviour of the customer and matching it with the buying behaviour of similar customers (who bought similar products) even when no information is available for the customer.

4. Future Trends

The diversity of data, data mining tasks and data mining approaches poses many challenging research issues in data mining .the development of efficient and effective data mining methods and systems, the construction of interactive and integrated data mining environments, the design of data mining language and the application of data mining techniques to solve large application problems are important tasks for data mining researchers and data mining system and application developers. Some of the rends in data mining that reflect the pursuit of these challenges are: application exploration, scalable and interactive data mining methods, integration of data mining with database systems, data warehouse systems and web data base systems, standardisation of data mining language, visual data mining, new methods for mining complex types of data, biological data mining, data mining and software engineering, web mining, distributed data mining, real-time data mining etc. (Han & Kamber, 2007)

5. Conclusion

This paper gives us a brief insight about data mining and the various techniques used to mine the data in the field of marketing and how this will help the marketers to fulfil their organisational goals. With growing volume of data every day, the need for data mining is also increasing day by day. It is a powerful technology to help companies focus on the most important information in their data warehouses. In this paper the concept of marketing mix i.e. 4 P's (product, place, promotion and price) have been deliberated upon which help the organisations to describe the different kinds of choices they can make in the process of bringing the product or service to market. This helps the marketers to increase return on investment (ROI), improve customer relationship management (CRM), helps in market analysis, reduces marketing campaign costs, facilitates fraud detection and also helps in customer retention thereby giving them a competitive edge over others.

6. Bibliography

Agarwal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast Discovery of sssociation Rules . *In Advances in Knowledge Discovery and Data* , 307-328.
Agrawal, R., & Imielinski, T. S. (1993). Mining Associaton Rules Between Sets of items in large Data Bases. *Sigmore Confrence*, (pp. 207-216).
Agrawal, R., & Srikant, R. (1994). Fast Algorithms For mining Association Rules in Large Data Bases. In J. B. Bocca, M. Jarke, & C. Zaniono (Ed.), *20th International Conference on Very Large Data Bases*, (pp. 487-499). Santiago,Chile.
Aha, D., & Kibler, D. (1991). Instance based Learning Algorithms. *Machine Learning* , 6,37-66.
Berry, M. J., & Linoff, G. S. (2008). *Data Mining Techniques For Marketing ,Sales and Customer Relationships Management* (2nd ed.). Wiley Publishing,Inc.
Berzal, Fernando, Cubero, Carlos, J., Nicolas, M., Serrano, et al. (2001). An Efficient Method for Association Rule Mining in Relational Databases. 47-64.
Blimes, J. (1998). *A Gentle tutorial of the EM Algorithm and Its Application to Parameter Estimation for gaussian Mixture and Hidden Markow Models.* California,Berkley: University of California.
Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees.* Monterey,CA,USA: Wadsworth and Woods.
Burges, C. J. (1998). A Tutorial On Support Vector Machines for Pattern Recognition. *Data Mining And Knowledge Discovery* , 121-167.
Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines.* Cambridge University Press.
Dong, G., Zhang, X., Wong, L., & Li, J. (1999). Classification By Aggregating Emerging Patterns. *International Confrence on Discovery Science.* Tokyo.
D'souza, R. (2012, February 6). Data Mining Future Trends Predicted.
Fayyad, U., & Irani, K. (1993). Multi-interval Discretization of Continuous- Valued Attributes for Classification Learning. *Thirteenth International Joint Conference on Artificial Intelligence.* San Francisco, CA: Morgan Kaufmann.
Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers:Machine Learning. , 29, 131-163.
Gupta, G. K. (2006). *Introduction to Data Mining With Case Studies.* Prentice Hall of India Pvt Ltd.
Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques* (2nd ed.). Morgan Kaufmann Publishers.
Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation.* Prentice Hall.
Hinneburg, A., & Keim, D. (1998). An Efficient Approach to Clustering Large Multedia DataBbases With Noise. *Knowledge Discovery and Data mining*, (pp. 58-65). New york.
Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression.* New York: Wiley.
Jackson, J. (2002). Data Mining :A Conceptual Overview. *Communications of the Association for Information Systems , VIII*, 267-296.
Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster.* John Wiley & Sons.
Liu, B., Hsu, W., & Ma, Y. (1998). Integrating Classification and Association Rule Mining. *4th International Conference on Discovery and Data Mining.* New York,U.S. A.
Micci-Barreca, D., & Ramachandran, S. *Improving Tax Administration with Data Mining.* Elite Analytics and SPSS.
Moeller, R. A. (2001). Distributed Data Warehousing Using Web Technology. *American Management Association (AMACOM)* .
Pearl, J. (1988). *Probabilistic reasoning in Intelligent Systems: Networks for Plausible Inference.* San Francisco, CA, U.S.A: Morgan Kaufmann.
(2010). *Performing a Data Mining Tool Evaluation.* IBM SPSS.
Pujari, A. K. (2002). *Data Mining Techniques.* Hyderabad: Universities Press.
Rao, I. K. (2003). Data Mining and Clustering Techniques. *Symantic Web.* Banglore.
Raymond, T. N., & Jiawei, H. (1994). Efficient and Effective Clustering Methods for Spatial Data Mining. *International Conference onVery Large Data Bases* (pp. 144–155). Santiago, Chile: Morgan Kaufmann.
Ripley, B. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press.
Shawkat, A. S., & Sawant, H. K. (2012). Content Based Data Retrieval on KNN-Classification and Cluster

Analysis for Data Mining. *Global Journals of Computer Science And Technology , XII* (5).
Sheikholeslami, G., Chatterjee, S., & Zhang, A. (1998). WaveCluster: a Multi-Resolution Clustering Approach for Very large Spatial Databases. *International Conference onVery Large Databases (VLDB-98)*, (pp. 428-439). New York, USA.
Soman, K. P., Diwakar, S., & Ajay, V. (2006). *Insight Into Data Mining.* Prentice Hall of India.
Turban, E., Sharda, R., Delen, D., Aronson, J. E., Liang, T. P., & King, D. *Decision Support and business Intelligence Systems* (9th ed.). Pearson.
Wang, K., & Zhou, S. (2000). *Growing Decision Trees on Support-less Association Rules, in KDD.* Boston, MA, U. S. A.
Yin, X., & Han, J. (2003). CPAR: Classification Based on Predictive Association Rules. *SIAM International Conference on Data Mining* . San Francisco, CA, U.S.A.
Chen, M. C., Chiu, A. L., & Chang, H. H. (2005). Mining Changes In Customer Behaviorin Retail Marketing. *Expert Systems with Applications,* , 773–781.
Cheung, K. W., Kwock, J. T., Law, M. H., & Tsui, K. C. (2003). Mining Customer Product Ratings For Personalized Marketing. *Decision Support Systems* , 231–243.
Drew, J. H., Mani, D. R., Betz, A. L., & Datta, P. (2001). Targeting Customers With Statistical And Data-mining Techniques., 3. *Journal Of Service Research* , 205-220.
Etzion, O., Fisher, A., & Wasserkrug, S. (2005). A Modeling Approach ForCustomer Lifetime Evaluation In E-commerce Domains, With An Application And Case Study For Online Auction. *Information Systems Frontiers* , 421-434.
He, Z. X., X, H., & J, Z. D. (2004). Mining Class Outliers: Concepts,Algorithms And Applications In CRM. *Expert Systems with Application* , 681–697.
Jiang, T., & Tuzhilin, A. (2006). Segmenting Customers From Population To Individuals: Does 1-to-1 keep Your Customers Forever. *Transactions On Knowledge And Data Engineering* , 1297-1311.
Kim, Y. H., & Moon, B. R. (2006). Multicampaign Assignment Problem. *Transactions on Knowledge and Data Engineering* , 405–414.
Liao, S. H., & Chen, Y. J. (2004). Mining Customer Knowledge For Electronic Catalog Marketing. *Expert Systems With Applications* , 521–532.
Prinzie, A., & Poel, D. V. (2005). Constrained Optimization Of Data-mining Problems To Improve Model Performance: A Direct-Marketing Application. *Expert Systems With Applications* , 630–640.
Prizie, A., & Poel, D. V. (2006). Investigating Purchasing-Sequence Patterns ForFinancial Services Using Markov, MTD And MTDG Models. *European Journal of Operational Research* , 710–734.
Rosset, S., Neumann, E., Eick, U., & Vatnik, N. (2003). Customer Lifetime Value Models For Decision Support. *Data Mining and Knowledge Discovery* , 321–339.
Woo, J. Y., Bae, S. M., & Park, S. C. (2005). Visualization Method For customerTargeting Using Customer Map. *Expert Systems With Applications* , 763–772.