CHAPTER 2
LITERATURE REVIEW

This chapter provides an overview of the various data mining techniques that have been adopted for solving business problems. The supervised and unsupervised learning techniques for customer relationship management are also discussed in brief. The literature support for classification and clustering techniques for customer segmentation are presented.

2.1 CRM AND DATA MINING

CRM related researches uses data mining techniques to understand and analyse the customer characteristics and behaviours. (Bortiz et al. 1995; Fletcher et al. 1993; Langley et al. 1995; Lau et al. 2003). The study on data mining for CRM showed that data mining techniques used to elicit untapped useful knowledge from a larger customer data. Data mining has the primary goals of describing, predicting, and building knowledge (Fayyad et al. 1996). CRM creates interaction of customers with the organization by using information technology (IT). Moreover, identifying customer’s need/interest better and treating them accordingly can increase their life times (Verhoef 2003). Customer segmentation is the grouping of customers into different groups based on their common attributes. It is the main part of CRM (Verhoef 2003).

Data mining is used to construct six types of models aimed at solving business problems: classification, regression, time series, clustering, association analysis, and sequence discovery (Thearling K. 1999). The first two, classification and regression, are used to make predictions. Association and sequence discovery are used to describe behavior. Clustering can be used for either forecasting or
One of the main types of predictive modeling tasks is classification (Dunham 2003; Weiss and Indurkhya 1998). In classification based supervised learning, data are mapped into classes (groups), which are redefined before the data are examined (Dunham 2003).

Ngai E.W.T. et al. (2009) analysis provides a roadmap to guide future research and facilitate knowledge accumulation and creation concerning the application of data mining techniques in CRM. Their analysis identified classification and association models are the two commonly used models for data mining in CRM.

The objective of descriptive data mining is to derive patterns of relationships in the data. Examples of descriptive data mining tasks are cluster analysis, association rules, summarization, and sequence discovery (Dunham, 2003). However, prediction uses variables or fields to predict future values of other variables (Chen 2001; Fayyad et al. 1996; Tan et al. 2006). The tasks of building a model for the dependent variable (target) as a function of the independent variables (explanatory) is called predictive modeling (Barbieri M.M and Berger 2004; Tan et al., 2006).

2.1.1 Supervised learning techniques for CRM

Each of the CRM elements is supported with different data mining models. As for as supervised learning techniques are concerned, background work is as follows:

Association modeling is usually adopted for market basket analysis and cross selling programs (Ahmed 2004, Jiao et al. 2006; Mitra et al. 2002).

Nan-Chen Hsieh et al. (2009) study presents a two-stage frame-work of consumer behavior analysis for analyzing bank databases. The key feature is a cascade involving Self-Organizing Map (SOM) neural network to divide customers into homogeneous groups of customers and a decision-tree simplified method to identify relevant knowledge. Identifying consumers by this approach is helpful
characteristic of customers and facilitates marketing strategy development. Once SOM was used to identify the profitable groups of customers, the statistical summarized data and the decision tree inducer were used to characterize the groups of customers. A tree simplified mechanism was used for finding relevant classification rules.

Siavash Emtiyaz et al. (2011) investigate the semi-supervised learning technique, for the management and analysis of customer-related data warehouse and information. The idea of semi-supervised learning is to learn not only from the labeled training data, but to exploit also the structural information in additionally available unlabeled data. The semi-supervised method is a model by means of which a feed-forward neural network is trained by a back propagation algorithm (multi-layer perceptron) in order to predict the category of an unknown customer (potential customers).

Rekha (2011) have used Naïve Bayesian classifier and Decision Tree-Based classifier to predict and analyze fraudulent in auto insurance claims. The performance of the model is examined with confusion matrix.

Keyvan (2012) applied three supervised classifiers namely decision tree, bayesian network and neural networks for prediction of customer behavior in an insurance industry in Iran. He concluded that decision tree evaluation methods show good performance compared with other methods.

2.1.2 Unsupervised Data Mining techniques for CRM

Unsupervised learning does not rely on predefined classes and class-labeled training examples (Kamber M and Han J. 2008). Clustering is a form of learning by observation, rather than learning by examples.

Customer clustering is the most important data mining methodology used in customer relationship management. Customer clustering uses customer-purchase transaction data to track buying behavior and create strategic business initiatives.
As an unsupervised data mining technique, Self organization map clustering is a good tool for exploratory analysis, as is the case when no a priori classes have been identified. The SOM is a very visual tool and possesses strong capabilities for dealing with non-linear relationships, missing data and skewed distributions.

Ruey-Shun Chen et al. (2005) have employed data mining tools to effectively discover the current spending pattern of customers and trends of behavioral change. The author considered data for the empirical study as relating to credit card customer data and spending records in 2003 and 2004 from banks. The study focused on the spending behavior of customers, all purchases of the same customer using more than one credit card were combined under said customer. The author has classified the selected customers into clusters using RFM model to identify high-profit, gold customers. Subsequently, carried out data mining using association rules algorithm to measure the spending pattern of customers.

Horng et al. (2007) have focused clustering and association rule techniques to map star products to projected potential customers. Clustering analysis done to locate potential customers using key characteristics of loyal customers personal information. Association rule analysis used to detect potential customers' near-future interest in a star product using knowledge of loyal customers’ purchasing behavior.

Jayanthi Ranjan (2011) has presented the importance and significance of Data mining techniques and tools in managing Customer Relationship Management (CRM) by finding the hidden and unknown information from the real case data of insurance company. The author analysed customer satisfaction level for the profitability of the organization with clustering technique.

Dilbag Singh et al. (2012) work on the conceptual mapping of various task of insurance risk management to data mining outlined the data mining techniques for risk identification, risk analysis, risk prioritization, risk monitoring and risk planning in insurance parlance.
The author performed conceptual mapping of risk management tasks with data mining techniques such as association, clustering, classification and time series analysis.

2.2 CUSTOMER SEGMENTATION AND DATA MINING TECHNIQUES

Customer segmentation is defined as the practice of classifying customer base into distinct groups (Farn and Huang 2009). In other words, customer segmentation is also described as the process of dividing customers into homogeneous groups on the basis of shared or common attributes (Bounsaythip and Rinta-Runsala 2001). The goal of segmentation is to know the customer better and to apply that knowledge to increase profitability, reduce operational cost, and enhance customer service. Segmentation can provide a multidimensional view of the customer for better treatment strategy.

Segmentation can be defined as aggregating customers into groups with similar characteristics such as demographic, geographic or behavioral traits and marketing them as a group (Parvatiyar and Sheth 2001). Consequently, each member of the segment has similar needs and wants; however, they are not completely uniform. Segmentation requires the collection, organization and analysis of customer data. With proper segmentations of a customer’s data, it is possible to identify the reliability/loyalty of customers so as to increase the revenue of the organization. Segmentation is the process of developing meaningful customer groups that are similarly based on individual explanation characteristics and behavior (Trappey et al. 2009). Greengrove (2002) explained that there are two main segmentation approaches: the first type of segmentation is the process of segmenting the customers based on understanding the needs of the end user which is called needs-based segmentation. The second type of segmentation, characteristics-based segmentation, is the process of segmenting customers based on their characteristics, attitudes or behaviors.
Twedt (1964) has suggested the use of segmentation models based on volume of sales, meaning that marketing efforts should focus on customers engaged in a considerable number of transactions. This approach, called “heavy half theory”, highlighted that one half of the customers can account for up to 80% of total sales. During the 1970s, the validity of the multivariate approaches used to identify the variables that affect deal proneness was criticized (Green and Wind 1973), which motivated the development of enhanced theoretical models of consumer behavior (Blattberg et al. 1978). A decade later, Mitchell (1983) developed a generalizable psychographic segmentation model that divided the market into groups based on social class, lifestyle and personality characteristics. However, practical implementation difficulties of this complex segmentation model was widely noted (Piercy and Morgan 1993, Dibb and Simkin 1997).

According to Bounsaythip and Rinta-Runsala (2001), segmentation is also viewed as a method to have more targeted communication with the customers; and the process of segmentation describes the characteristics of the customers groups (called segments or clusters) within the data. The diversity of customer needs and buying behavior, influenced by lifestyle, income levels or age, makes past segmentation approaches less effective. Therefore, current models for marketing segmentation are often based on customer behaviour inferred from transaction records or surveys. The resulting data is then explored with data mining techniques, such as cluster analysis. Kiang et al. (2006) surveyed the applications of data mining for segmentation purposes.

In most studies, customer segmentation is mentioned as the ideal way to obtain customer profitability through careful customer targeting (Chan, 2005; Chung et al., 2004; Hwang et al. 2004; Jones et al. 2006; Kim and Street 2004; Kim et al. 2005; Kuo et al. 2006; Shin and Sohn 2004; Woo et al. 2005).

Helsen and Green (1991) have also identified market segments for a new computer, system based on the use of cluster analysis with data from a customer survey. The segmentation was supported by the rate of importance given to the product attributes. Min and Han (2005) clustered customers with similar interests in
movies based on data containing explicit rating information for several movies provided by each customer. The rating information allowed inference of the perceived value of each movie for each customer.

Chu Chai et al. (2008) have identified customer behavior using a Recency, Frequency and Monetary (RFM) model and then used a customer Life Time Value (LTV) model to evaluate proposed segmented customers. He proposed an intelligent model that uses Genetic Algorithm (GA) to select customer RFM behavior using a LTV evaluation model. Customer life time value is taken as the fitness value of GA. If the proposed methodology is applied, high-value customers can be identified for campaign programs. Another advantage of the proposed methodology is that it considers the correlation between customer values and campaigns. Valuable customers thus can be identified for a campaign program.

Razieh et al. (2012) performed customer segmentation using RFM technique and clustering algorithms based on customer’s value, to specify loyal and profitable customers. The author considered the model for a grocery store’s data. The author used a combination of behavioral and demographical characteristics of individuals to estimate loyalty. The author performed customer segmentation using RFM technique and clustering algorithms based on customer’s value, to specify loyal and profitable customers.

2.2.1 Classification techniques for customer segmentation

Classification analysis is also known as supervised classification. Classification analysis is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown (Han and Kamber 2008). By using classification, it is possible to organize data in a given class. The classification uses given class labels to organize the objects in the data collection in an orderly manner. Classification model is one of the most commonly used supervised modeling techniques. In classification, a user needs to divide data into segments and then make distinct non-overlapping groups. For dividing data into
groups, a user needs to have certain information about the data to be divided into segments. Classification problems aim to identify the characteristics that indicate the group to which each case belongs. This pattern can be used both to understand the existing data and to predict how new instances will behave. Data mining creates classification models by examining already classified data (cases) and inductively finding a predictive pattern (Westphal C. and Blaxton T. 2005).

Classification approaches normally use a training set where all objects are already associated with known class labels. Then the classification algorithm learns from the training set and builds a model. The model is used to classify new objects. In other words, classification is a two-step process, first a classification model is built based on training data set and then the model is applied to new data for classification (David Hand et al. 2008).

Classification tasks have been carried out for various purposes in CRM domain. Kim et al. (2006) adopted decision tree to classify the customers and develop strategy based on customer life time value.

Baesens (2004) identified the slope of the customer lifecycle based on Bayesian network classifier. The author illustrated Bayesian network classifiers as a useful tool in the toolbox of CRM analysts in application of identifying the slope of the customer lifecycle of long-life customers.

Sheu et al. (2009) adopted decision tree to explore the potential relationship between important influential factors and customer loyalty. The findings of these studies inspire us to adopt the decision tree to explore the relationship between customers’ purchase amounts and customers’ demographic and behavioral characteristics, with special attention to the characteristics of high-and low-spending customers.

2.2.2 Clustering techniques for customer segmentation

Clustering can be defined as the process of grouping a set of physical or abstract objects into classes of similar objects (Han and Kamber 2008). Clustering is
also called unsupervised classification, because the classification is not dictated/ordered by given class labels. There are many clustering approaches, all based on the principle of maximizing the similarity between objects in a same class (intra-class similarity) and minimizing the similarity between objects of different classes (inter-class similarity).

Clustering is similar to classification, but classes are not predefined and it is up to the clustering algorithm to discover acceptable classes. Often, it is necessary to modify the clustering by excluding variables that have been employed to group instances because, upon examination, the user identifies them as irrelevant or not meaningful. After clusters are found that reasonably segment the database, these clusters are then used to classify new data. Some of the common algorithms used to perform clustering include Kohonen feature maps and K-means. Clustering is different from segmentation. Segmentation refers to the general problem of identifying groups that have common characteristics whereas clustering is a way to segment data into groups that are not previously defined (Two Crows Corporation 1999). Clustering is useful for finding natural groups of data which are called clusters. A cluster is a collection of data that are similar to one another. Clustering can be used to group customers with similar behavior and to make business decisions in industry. (David Hand et al. 2008).

Clustering studies are also referred to as unsupervised learning. Unsupervised learning is a process of classification with an unknown target, that is, the class of each case is unknown. The aim is to segment the cases into disjoint classes that are homogenous with respect to the inputs. (Han and Kamber 2008). Clustering studies have no dependent variables. Clustering is one of the most useful tasks in data mining process for discovering groups and identifying interesting distributions and patterns in the underlying data. Clustering problem is about partitioning a given data set into groups (clusters) such that the data points in a cluster are more similar to each other than points in different clusters (Guha et al. 1998).
Clustering analysis is widely used to establish object profiles on the basis of objects’ variables. Objects can be customers, web documents, web users, or facilities. Hruschka (1996), Ozer (2001), Weber (1996) use clustering technique to segment customers and markets. The K-means clustering algorithm and the Kohonen self-organizing map are the two most popular clustering techniques.

Samira et al. (2007) have applied segmentation of customers of Trade Promotion Organization of Iran using a proposed distance function which measures dissimilarities among export baskets of different countries based on association rules concepts. Later, in order to suggest the best strategy for promoting each segment, each cluster is analyzed using RFM model. Variables used for segmentation criteria are “the value of the group commodities”, “the type of group-commodities” and “the correlation between export group-commodities”.

Pramod et al. (2011) elaborates the use of clustering to segment customer profiles of a retail store. The study concluded that the K-Means clustering allows retailers to increase customer understanding and make knowledge-driven decisions in order to provide personalised and efficient customer service.

Huang et al. (2009) applied K-means method, Fuzzy C-means clustering method and bagged clustering algorithm to analyze customer value for a hunting store in Taiwan and finally concluded that bagged clustering algorithm outperforms the other two methods.


Migueis V.L et al. (2012) proposed a method for customers segmentation, given by the nature of the products purchased by customers. This method is based on clustering techniques, which enable segmenting customers according to their lifestyles. The author segmented customers of an European
retailing company according to their lifestyle and proposed promotional policies tailored to customers from each segment, aiming to reinforce loyal relationships and increase sales. The author used the VARCLUS algorithm, integrated in SAS software, to cluster the products. The methodology also involved the inference of the lifestyle corresponding to each cluster of products, by analyzing the type of products included in each cluster.

2.3 PREDICTION ANALYSIS OVER CUSTOMER DATA

In the supervised modeling, whether for the prediction of an event or for a continuous numeric outcome, the availability of a training dataset with historical data is required (Tsipitsis K et al. 2009). Supervised algorithms that need the control of a human operator during their execution (Pieter et al. 2011). Predictive models are built, or trained, using data for which the value of the response variable is already known. This kind of training is sometimes referred to as supervised learning, because calculated or estimated values are compared with the known results. In predictive models, the values or classes we are predicting are called the response, dependent or target variables. The values used to make the prediction are called the predictor or independent variables. There are two major types of predictions: numeric prediction and class label prediction. The first type of prediction that is numeric prediction predicts some unavailable data values or pending trends, and the second type of prediction (the one which is tied to classification) predicts a class label for some data.

Prediction has attracted considerable attention by giving the potential implications of successful forecasting in a business context. Once a classification model is built based on a training set, the class label of an object can be forecast based on the attribute values of the object and the attribute values of the classes (Two Crows Corporation 1999). Prediction is more often referred to the forecast of missing numerical values or increase/ decrease trends in time related data. The major idea in prediction is to use a large number of past values to consider probable future values. Two important distinct kinds of tasks in predictive modeling depend on whether Y is categorical or has real values. For categorical Y, the task is called
classification and for real valued Y the task is called regression. (David Hand et al. 2008).

Afaq Alam et al. (2010) have addressed the use of demographic, billing and usage data of Internet service users to identify the best churn predictors. In addition, the study evaluates the accuracy of decision tree, logistic regression and neural network data mining techniques for predicting customers of high risk of churning.

Young et al. (2001) have examined the characteristics of the knowledge discovery and decision tree and logistic regression data mining algorithms to demonstrate how they can be used to predict health outcomes and provide policy information for hypertension management using the Korea Medical Insurance Corporation database.

Lejeune (2001) has presented a CRM framework that is based on the integration of the electronic channel. He has concluded that churn management consists of developing techniques that enable firms to keep their profitable customers and aims at increasing customer loyalty.

Bolton (1998) has suggested that service organisations should be proactive and learn from customers before they defect by understanding their current satisfaction levels. He has also suggested that service encounters act as early indicators of whether an organization’s relationship with a customer is flourishing or not. He concluded that the customers who have longer relationships with the firm have higher prior cumulative satisfaction ratings and smaller subsequent perceived losses that are associated with subsequent service encounters.

Debahuti Mishra et al. (2010) have done an extensive study on various predictive techniques with all its future directions and applications in various areas such as CRM, clinical decision support systems, product level prediction and direct marketing. The author also presented an overview of some of the notable data
mining techniques such as Bayesian analytics, Classification, Clustering, Dependency or association analysis and Regression analysis for prediction.

Bolton et al. (2000) have suggested that it is theoretically more profitable to segment and target customers on the basis of their (changing) purchase behaviors and service experiences, rather than on the basis of their (stable) demographics or other variables. The author used Logistic Regression (LR) and t-tests for loyalty programmed membership and conclude that loyalty rewards programmers help build stronger relationships with customers.

Reza Allahyari Soeini et al. (2012) have analyzed customer churn prediction in the Insurance industry with clustering and decision tree techniques. They have concluded that decision tree is more suitable to understand the reasons for customer churn and clustering technique is more suitable to identify the characteristics of customers.

Kanwal Garg et al. (2008) applied clustering and decision tree techniques for identifying the trend of customer investment behavior in life insurance sector in India. This paper analyzed the prediction of customer buying preference over newly launched policies.

Burez et al. (2007) concluded that, using the full potential of their churn prediction model and the available incentives, the pay-TV company’s profits from the churn prevention programme would double when compared to its current model.

Hadden et al. (2008) have compared neural networks and decision trees in predicting customer churn. He concluded that the decision tree outperformed neural technique.

Hung and Yen (2006) have reported that both the decision tree and neural network techniques can deliver accurate churn prediction models by using customer demographics, billing information, the contract/service status, call detail records and service change logs.
Hung and Yen (2006) have applied the decision tree and back propagation neural network on a wireless telecom company’s customer data and concluded that the performance of building a predictive model on individual segments is more accurate than the one built on the entire customer population.

Thakur et al. (2012) have outlined the implementation of Bayes classifier using Bayes theorem for prediction of online vehicle Insurance system. The author considered the problem of predicting whether a customer will go for manual insurance or online insurance. To illustrate the task of predicting whether a customer i.e. a vehicle owner will go for online insurance, author has considered the attributes such as vehicle_ownership, qualification, and age. Customers interested in online insurance are classified as yes or No, based on conditional probability of dependent attributes.

In data mining, a decision tree is a predictive model which can be used to represent both classifiers and regression models. Decision tree algorithm is a data mining induction technique that recursively partitions a dataset of records using either depth-first greedy approach or breadth-first approach until all the data items belong to a particular class (Han and Kamber 2008). Decision trees constitute a way of representing a series of rules that lead to a class or value (Two Crows Corporation 1999) and it is a powerful and popular tool for classification and prediction. Decision trees are also useful for exploring data to gain insight into the relationships of a large number of candidate input variables to the target variable. Since decision trees combine both data exploration and modeling, they are a powerful first step in modeling process even when building the final model using some other techniques (David Hand et al. 2008).

Decision trees are part of the Induction class of DM techniques. An empirical tree represents a segmentation of the data that is created by applying a series of simple rules. Each rule assigns an observation to a segment based on the value of one input. Rules are applied successively one after another, resulting in a hierarchy of segments within segments. In predictive modeling, the decision is simply the predicted value.
When a decision tree is used for classification tasks, it is more appropriately referred to as a classification tree and when it is used for regression tasks, it is called the regression tree. The attractiveness of decision trees is due to the fact that, in contrast to neural networks, decision trees represent rules. Rules can readily be expressed so that humans can understand them or even directly use in a database access language like SQL so that records falling into a particular category may be retrieved.

Hajizadeh et al. (2010) have also described the decision tree as a model which consists of a set of rules for dividing a large heterogeneous population into smaller, more homogeneous groups with respect to a particular target variable. The target variable is usually categorical and the decision tree model is used either to calculate the probability that a given record belongs to each of the categories, or to classify the record by assigning it to the most likely class.

The decision tree technique enables the creation of decision trees that classify observations based on the values of nominal, binary, or ordinal targets, predict outcomes for interval targets, or predict the appropriate decision when users specify decision alternatives. Tree techniques provide insights into the decision-making process, which explains how the results come about (G.K.Gupta 2006). They require no prior assumption of probability distribution, are computationally inexpensive, robust to noise, and easy to understand (Dunham 2003; Tan et al. 2006).

Jian et al. (2008), have applied the decision tree model to construct a classification (prediction) model to estimate the likelihood of a customer being a high-value customer. A collection of credit card customer data from a leading commercial bank in China was used. Three customer attributes: annual income, financial asset and education background were selected for analysis. The author showed that, with only three customer attributes, the model correctly identify nearly 80% of high-value customers by selecting only half of the candidate customers, which was more than 50% better than a model selecting customers at random.
Yongqiang Chen et al. (2005) have made an attempt to apply data mining techniques in CRM systems and propose a data mining model for the customers' Value and customers’ classification. Customer’s data collected from the questionnaires are analyzed with SAS Enterprise Miner.

Szymon et al. (2007) have presented two cross-selling approaches: one based on classifiers and the other based on Bayesian networks. The first approach is based on constructing a Bayesian network representing customer’s behavior and using this network to predict which customers are most likely to pick each service offered. This gives not only a cross-selling model, but also allows the analyst to gain insight into the behavior of customers. The second approach uses a separate classifier model for each service offered. Each model predicts, which customers are most likely to buy a specific product.

Raymond Chi-wing Wong et al. (2005) have proposed a method for actionable recommendations from itemset analysis and investigated an application of the concepts of association rules - Maximal- Profit Item Selection with cross-selling effect (MPIS). The problem is about choosing a subset of items which can give the maximal profit with the consideration of cross-selling effect.

2.4 SUMMARY

The various classification techniques that have been applied at CRM system for customer data analysis are presented in this chapter. From the literature, it is evident that more work has been done to apply clustering techniques in particular for the customer data analysis. Relatively less work has been done for the application segmentation methodology to service sector like insurance, banking etc. In our work, the segmentation methodology in terms of socio-demographic and life-stage, value and behaviour segmentation for customer data analysis is being derived. The generated segments utilized for prediction analysis of customer preferences over products. Further, empirical study for life insurance and health insurance company is being dealt in this thesis.