2009:052

MASTER'S THESIS

Predicting Customer Churn in Telecommunications Service Providers

Ali Tamaddoni Jahromi

Luleå University of Technology

Master Thesis, Continuation Courses
Marketing and e-commerce
Department of Business Administration and Social Sciences
Division of Industrial marketing and e-commerce

2009:052 - ISSN: 1653-0187 - ISRN: LTU-PB-EX--09/052--SE

MASTER'S THESIS

Predicting Customer Churn in
Telecommunications Service Providers

Supervisors:

Dr. Mohammad Mehdi Sepehri (TMU)

Dr.Albert Caruana (LTU)

Prepared by:

Ali Tamaddoni Jahromi

Tarbiat Modares University Faculty of Engineering

Department of Industrial Engineering

Lulea University of Technology

Division of Industrial Marketing and E-Commerce

MSc PROGRAM IN MARKETING AND ELECTRONIC COMMERCE Joint

LULEÅ UNIVERSITY OF TECHNOLOGY

TARBIAT MODARRES UNIVERSITY

T.M.U.

April 2009

Abstract

Customer churn is the focal concern of most companies which are active in industries with low switching cost. Among all industries which suffer from this issue, telecommunications industry can be considered in the top of the list with approximate annual churn rate of 30%. Tackling this problem, there exist different approaches via developing predictive models for customers churn, but due to the nature of pre-paid mobile telephony market which is not contract-based, customer churn is not easily traceable and definable, thus constructing a predictive model would be of high complexity. Handling this issue, in this study, we developed a dual-step model building approach, which consists of clustering phase and classification phase. With this regard firstly, the customer base was divided into four clusters, based on their RFM related features, with the aim of extracting a logical definition of churn, and secondly, based on the churn definitions that were extracted in the first step, we conducted the second step which was the model building phase. In the model building phase firstly the Decision Tree (CART algorithm) was utilized in order to build the predictive model, afterwards with the aim of comparing the performance of different algorithms, Neural Networks algorithm and different algorithms of Decision Tree were utilized to construct the predictive models for churn in our developed clusters. Evaluating and comparing the performance of the employed algorithms based on “Gain measure”, we concluded that employing a multi-algorithm approach in which different algorithms are used for different clusters, can bring the maximum “Gain” among the tested algorithms.

Furthermore, dealing with our imbalanced dataset, we tested the cost- sensitive learning method as a remedy for handling the class imbalance. Regarding the results, both simple and cost-sensitive predictive models have a considerable higher performance than random sampling in both CART model and multi-algorithm model. Additionally, according to our study, cost-sensitive learning was proved to outperform the simple model only in CART model but not in the multi-algorithm.

Key words: Customer relationship management; customer churn; data mining

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1. Introduction

Acquisition and retention of new clients are one of the most significant concerns of businesses. While recipient companies concentrate on acquiring new customers, mature ones try to focus on retention of the existing ones in order to provide themselves with the opportunity of cross – selling. According to Freeman (1999) one of the most significant ways of increasing customers' value is to keep them for longer period of time.

In the new era emergence of electronic commerce has boosted the available information, and as Peppard (2000) believes, the internet channel has empowered the customers who are no longer stuck with the decisions of a single company and has led to exacerbation of the competition, while competitors are only one "click away", customer empowerment is likely to amplify the attrition rate of a company's customers (Lejeune, 2001). Facing with this threat companies should be equipped and armed with the most efficient and effective methods of examining their client's behavior predicting their possible future failure.

In accordance with (Lejeune, 2001) churn management consists of developing techniques that enable firms in keeping their profitable customers.

The study at your disposal aims at finding an efficient and accurate predictive model for customer churn in pre-paid mobile telephony market segment by utilizing machine learning techniques.

With the intention of making you more familiar with the research's realm and its importance we start the report by providing you with statistics regarding the customer churn magnification in telecommunications industry and afterwards we address our problem definition and the question of our research.

1.2. Churn magnitude in telecommunications industry

The mobile telephony market is one of the fastest-growing service segments in telecommunications, and more than 75% of all potential phone calls worldwide can be made through mobile phones and as with the any other competitive markets, the mode of competition has shifted from acquisition to retention of customers (Kim & Yoon, 2004). Regarding this, examining the existing statistics concerning churn magnitude and its costs in this realm would be beneficial for gaining an appropriate mental picture of the importance of this area of research:

- SAS (2000) reported that the telecommunications sector endures an annual rate of churn, ranging from 25 per cent to 30 percent this churn rate could still continue to increase in correlation with the growth of the market.
- Churn costs for European and US telecommunications companies are estimated to amount to US$4 billion annually (SAS Institute, 2000)
- The ratio (customer acquisition costs/ customer retention or satisfaction costs) would be equal to eight for the wireless companies (SAS Institute, 2000).

While the annual rate of customer churn in telecommunications sector is around 30 percent (Groth, 1999; SAS Institute, 2000) and it costs US$ 4 billion per year for European and US telecommunications companies, it would seem reasonable to invest more on churn management rather than acquisition management for mature companies especially when we notice that the cost of acquiring a new customer is eight times more than retaining an existing one (SAS Institute, 2000).

1.3. Problem definition

Customer churn is the focal concern of most companies which are active in industries with low switching cost. Among all industries which suffer from this issue, telecommunications industry can be considered in the top of the list with approximate annual churn rate of 30%. This means wasting the money and efforts or as Kotler and Keller (2006) mentioned, “it is like adding water to a leaking bucket”.

Consequently, in order to tackle this problem we must recognize the churners before they churn, so developing a model which predicts the future churners seems to be vital. This model has to be able to recognize the customers which tend to churn in close future. But, due to the nature of pre-paid mobile telephony market which is not contract-based, customer churn is not easily traceable and also definable, thus building a predictive model would be of high complexity. In order to achieve such goal in pre-paid market segment the initial step appears to be defining the churn and a churner and then predicting the churn. Furthermore, due to the nature of churn datasets in which the churn class is always suffering from rarity, handling such imbalance in the dataset can help to improve the model’s performance.

1.4. Research Purpose

The purpose of this research is to develop and design an effective and efficient model for customer churn prediction in telecommunication industry (Pre-paid mobile telephony market).

1.5. Research Question

1- How “customer churn” can be defined in pre-paid mobile telephony service providers?
2- What features can be utilized in order to build a predictive model for customer churn in pre-paid mobile telephony industry?
3- What are the remedies for data imbalance in churn data sets?

1.6. Thesis structure

This thesis report starts with defining and explaining the research problem and providing the readers with the magnification and importance of the problem and exploiting the definition of problem and purpose of the research, it provides you with the research questions.

The second chapter begins with defining and explaining different perspective towards customer relationship management and then it narrows its focus down to analytical and IT based CRM with a special look at predictive models. It reviews different existing models for churn prediction and ultimately based on what have been done previously, the appropriate features is selected also the methodology is specified – which comprises the 3rd chapter of this report. However, since the methodology of this research is hardly separable form its analysis its detail has been addressed in the 4th chapter. Chapter 4 mostly consists of the analysis and the results of it, although as mentioned before, it contains the detailed aspects of methodology part. Ultimately the 5th chapter comes, which contains the conclusion and the interpretations of it. Figure 1.1 illustrates different steps of this thesis report.

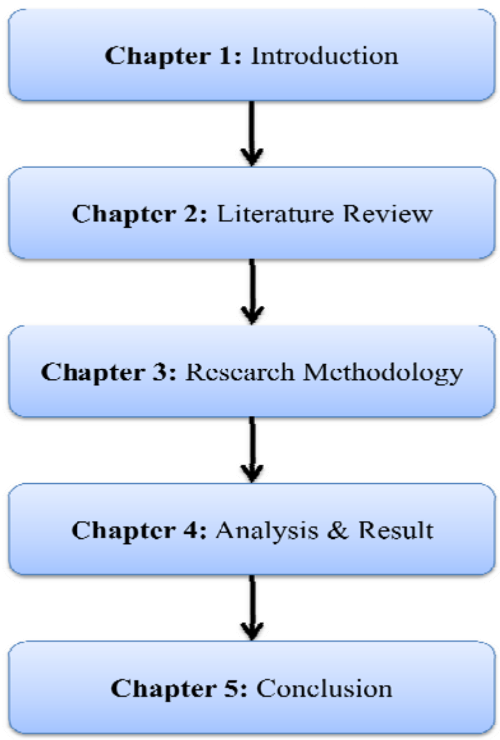

Figure 1.1: Outline of the thesis

Chapter 2 Literature Review

2.1. Introduction

The current chapter consists of three individual sections. The first section aims at introducing Customer Relationship Management (CRM) and its basic concepts while it also tries to depict the contribution of machine learning techniques (Especially Data Mining) to this realm. Section two is an introductory part to Data Mining and its significance role in CRM. The second section ends with addressing the most common and applicable Data Mining models and techniques in CRM which have also been utilized in this research's model building phase, and ultimately the third section represents the existing studies regarding the customer churn in different industries. Although the focus of this research is on machine learning predictive models for customer churn, this chapter has taken a look at churn literature from both explanatory and predictive point of view in order to broaden the visions toward all sides of churn issue.

2.2. Customer Relationship Management: Basic Concepts

Eagerness toward **C**ustomer **R**elationship **M**anagement (CRM) began to grow in 1990 (Ling & Yen, 2001; Xu, Yen, Lin, & Chou, 2002). A developed relationship with one's clients can finally result in greater customer loyalty and retention and, also profitability (Ngai, 2005).

Despite the fact that CRM has become widely recognized, there is no comprehensive and universally accepted definition of CRM.

Swift (2001) defined CRM as an" enterprise approach to understanding and influencing customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability. Kotler and Keller (2006) have defined Customer relationship management (CRM) as the process of managing detailed information about individual customers and carefully managing all customer "touch points" to maximize customer loyalty. Kincaid (2003) viewed CRM as "the strategic use of information, processes, technology, and people to manage the customer's relationship with your company (Marketing, Sales, Services, and Support) across the whole customer life cycle". Bose (2002) viewed CRM as an integration of technologies and business processes used to satisfy the needs of a customer during any given interaction more specifically from his point of view Customer relationship management (CRM) involves acquisition, analysis and use of knowledge about customers in order to sell goods or services and to do it more efficiently. Richards and Jones (2008) have defined CRM as "a set of business activities supported by both technology and processes that is directed by strategy and is designed to improve business performance in an area of customer management".

Having a glimpse to the above mentioned definitions of CRM one can understand that all above authors' emphasis is on considering CRM as a "comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer. It involves the integration of marketing, sales, customer service, and supply – chain functions of the organization to achieve greater efficiencies and effectiveness in delivering customer value." (Parvatiyar & Sheth, 2001).

Olafsson, Li, and Wu (2008) believe that a valuable customer is usually dynamic and the relationship evolves and changes over time. Thus, a critical role of CRM is to understand this relationship. This is achievable by studying the customer life-cycle, or customer lifetime, which refers to various stages of the relationship between customer and business (Olafsson, Li, & Wu, 2008). A typical customer life-cycle is shown in Figure 2.1.

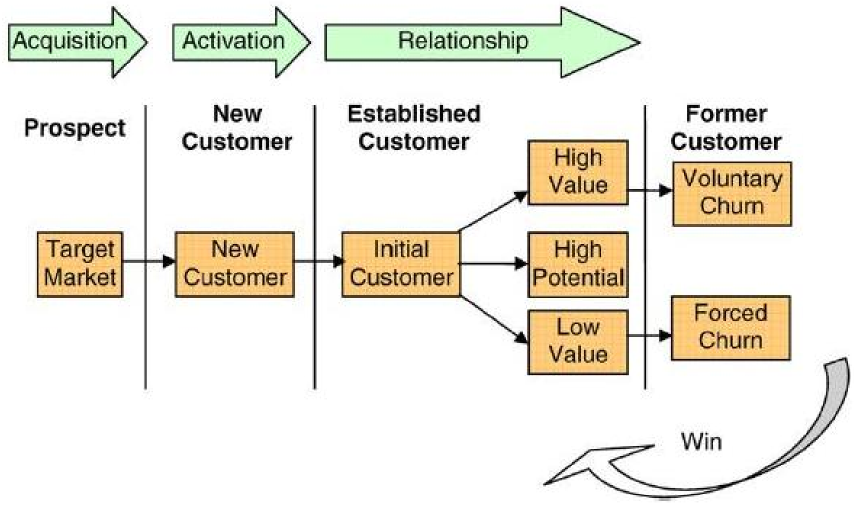

Figure 2.1: Illustration of a customer life-cycle (source: Olafsson, Li, & Wu, 2008)

As it is presented in the above figure, a prospect that responds to the marketing campaigns of the company in acquisition phase, becomes a customer and this "New Customer" becomes a established one once the relationship between him/her and the company has been established and this is the point that in which the company can benefit from its established customers by revenue that comes from cross – selling and up – selling, but the peril that threatens the company in this stage is that at some point established customers stop being customers (Churn) (Olafsson, Li, & Wu, 2008). Thus, in simple words, the main goal of customer relationship management is to create satisfaction and delight among customers in order to prevent customer churn which is the most important threat that threatens all companies. It has been shown that a small change in retention rate can result in significant changes in contribution (Van den Poel & Larivie're, 2004).

In accordance with Rayls CRM falls in two categories; attracting new customers what he calls offensive marketing, and keeping the existing customers, known as defensive marketing (Ryals, 2005). While acquiring new customers is the first step for any businesses to start growing, the importance of retaining customers should not be overlooked. Reinartz, Thomas & Kumar showed that insufficient allocation to customer-retention efforts will have a greater impact on long-term customer profitability as compared to insufficient allocation to customer-acquisition efforts (Reinartz, Thomas, & Kumar, 2005). As Chu, Tsai, and Ho have highlighted the cost of acquiring a new customer is five to ten times greater than that of retaining existing subscribers (Chu, Tsai, & Ho, 2007). Even if we put aside the existing studies, which mentioned that it costs more to acquire new customers than to retain the existing customers, we can consider that customer retention is more important than customer acquisition because lack of information on new customers makes it difficult to select target customers and this will cause inefficient marketing efforts.

The emergence of electronic commerce has increased the amount of available information and so offers new ways for companies to efficiently respond to clients' expectations. Meanwhile, customers can more easily get information about the market opportunities. They become more demanding and tend to switch from their previous supplier to another. This gave birth to the notion of churn (Lejeune, 2001).

During 1850s businesses were able to sell anything they made and generally the focus was on production. In early 1900 the customer empowerment forced firms to find reasons for people to buy their products. In the mid 20th century a paradigm shift occurred and firms started making what people wanted instead of trying to persuade them to buy whatever they had to sell. This new marketing orientation leaded to customer centric orientation in 21st century. A customer centric orientation is capable of treating all customers individually depending on customer preference (Bose, 2002). In fact today's variety of tastes and preferences among customers has made it impossible for the companies to group them into large homogenous populations to develop marketing strategies and what actually firms are facing with are customers who want to be served according to their individual and unique needs (Shaw, Subramaniam, Tan, & Welge, 2001).

This, gave birth to the need of IT and knowledge management in the realm of Customer Relationship Management. In fact in a broader view CRM can be presented in the form of customer management which requires the collection and treatment of a significant amount of data that enables companies to exploit them in acquisition, retention, extension, and also selection of their customers (Komenar, 1997).

In the IT realm, CRM means an enterprise wide integration of technologies such as data warehouse, website, intranet/extranet, etc (Bose, 2002). In fact CRM utilizes information Technology and Information Systems to gather data which can be used to develop required information to create a one-to-one interaction with the customers (Bose, 2002; Ngai, 2005).

In actual fact, turning the dream of one-to-one marketing would be impossible in the absence of IT contributions. Although there are some controversies among academics about the key components of IT success in one-to-one marketing (Bose, 2002; Wells, Fuerst, & Choobineh, 1999), most experts confirm the necessity of IT in this field.

In fact it is the above mentioned need of individual recognition of customers that let the Information technology to be combined with CRM and with this IT based perspective, CRM can be defined as the integration of technologies and business process in order to satisfy the customer needs in a given interaction. Thus, in new definition, CRM deals with acquisition, analysis, and use of knowledge about customers in order to increase the sales volume in the most efficient way (Bose, 2002).

There exists different categorization approaches toward CRM (Teo, Devadoss, & Pan, 2006; Ngai, 2005; He, Xu, Huang, & Deng, 2004; Xu, Yen, Lin, & Chou, 2002). From the architecture point of view, the CRM framework can be classified into operational and analytical (He, Xu, Huang, & Deng, 2004; Teo, Devadoss, & Pan, 2006). While operational CRM refers to the automation of business process, the analytical CRM refers to the analysis of customer characteristics and attitudes in order to support the organization's customer management strategies. Thus, it can help the company in more effective allocation of its resources (Ngai, Xiu, & Chau, 2009).

On the other hand Kincaid (2003), West (2001), Xu, Yen, Lin, & Chou (2002), and Ngai (2005) believe that CRM falls in the four following categories:

1) Marketing
2) Sales
3) Service and support
4) IT and IS

According to what experts believe, the role of Information Technology (IT) and Information Systems (IS) in CRM can't be denied (Kincaid, 2003; Ling & Yen, 2001). Using IT and IS will makes the companies capable of the collection of the necessary data to determine the economics of customer acquisition, retention, and life – time value. This means involving the use of database, data warehouse, and data mining (a complicated data search capability which is able to discover patterns and correlations in data by using statistical algorithms) to help organizations increase their customer retention rates and their own profitability (Ngai, 2005).

A review of the literature in CRM realm by Ngai (2005) reveals that since 1999, eagerness toward this issue has boosted and a total of 191 publications were found to be from 2000 to 2002 which represents 93 percent the total publications in this field from 1992 to 2002 (see figure.2.2). The research also depicts that a major part of the researches in CRM field is related to the application of IT and IS in CRM. Furthermore in IT and IS field the first role is played by data mining (Ngai, 2005) . This fact also has been confirmed in recent studies (Ngai, Xiu, & Chau, 2009). That's why Shaw, Subramaniam, Tan and Welge (2001) believe that True Customer Relationship Management is possible only by integrating the knowledge discovery process with the management and use of the knowledge for marketing strategies.

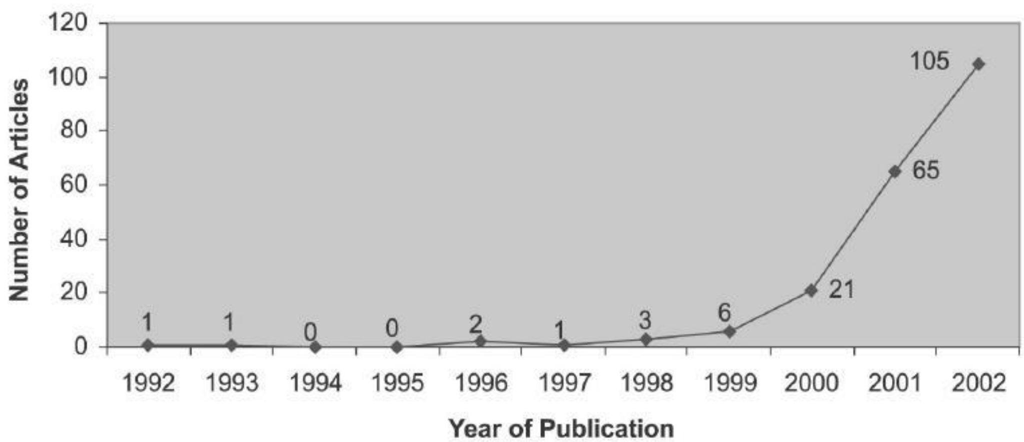

Figure 2.2: Distribution of articles by year (source: Ngai,2005)

Therefore one can conclude that the role of data mining in CRM process is fundamental and critical (Rygielski, Wang, & Yen, 2002) and it enables us to transform customer data, which is a company asset, into useful information and knowledge and exploit this knowledge in identifying valuable customers, predicting future behaviors, and make proactive and knowledge based decisions (Rygielski, Wang, & Yen, 2002) . In CRM context, data mining can be seen as a business driven process, aimed at discovery and consistent use of knowledge from organizational data (Ling & Yen, 2001).

Consequently, deep understanding of data mining and knowledge management in CRM seems to be vital in today's highly customer – centered business environment (Shaw, Subramaniam, Tan, & Welge, 2001).

2.3. Data Mining and Its Application in CRM

Nowadays lack of data is no longer a problem, but the inability to extract useful information from data is (Lee & Siau, 2001). Due to the constant increase in the amount of data efficiently operable to managers and policy makers through high speed computers and rapid data communication, there has grown and will continue to grow a greater dependency on statistical methods as a means of extracting useful information from the abundant data sources. Statistical methods provide an organized and structured way of looking at and thinking about disorganized, unstructured appearing phenomena. Figure 2.3 illustrates the different stages involved in the never – failing quest for more refined information (Lejeune, 2001).

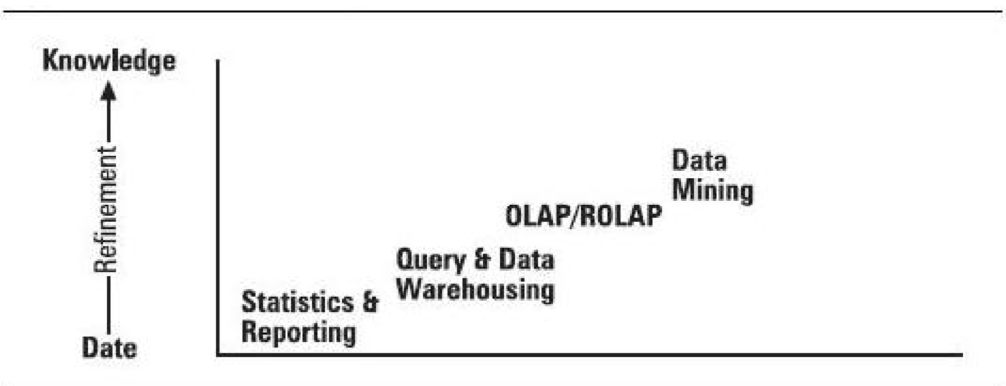

Figure 2.3: Evolution in the quest for information (source: Lejeune, 2001)

In fact the accelerated growth in data and databases resulted in the need of developing new techniques and tools to transform data into useful information and knowledge, intelligently and automatically. Thus, data mining has become an area of research with an increasing importance (Weiss and Indurkhya, 1998; cited by Lee & Siau, 2001). Data mining techniques are the result of a long term research and product development and their origin have roots in the first storage of data on computers, which was followed by improvement in data access

(Rygielski, Wang, & Yen, 2002). Table 2.1 depicts the evolutionary stages of data mining from user's point of view.

Table 2.1: Evolutionary stages of data mining (Source: Rygielski, Wang, & Yen, 2002)

Stage	Business question	Enabling technologies	Product providers	characteristics
Data collection (1960s)	"What was my average total revenue over the last five years?"	Computers, tapes, disks	IBM,CDC	Retrospective, Static data Delivery
Data access (1980s)	"What were unit sales in New England last March?"	Relational databases(RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Retrospective, dynamic data delivery at record level
Data navigation (1990s)	"What were unit sales in New England last March? Drill down to Boston"	On- line analytic processing (OLAP), multidimensional databases, data warehouses	Pilot, IRI, Arbor, Redbrick, Evolutionary Technologies	Retrospective, dynamic data delivery at multiple levels
Data mining (2000)	"What's likely to happen in Boston unit sales next month? Why?"	Advanced algorithms, multiprocessor computers, massive databases	Lockheed, IBM, SGI, numerous startups (nascent industry)	Retrospective, Proactive information delivery

Data mining is "the process of selecting exploring and modeling large amount of data to uncover previously unknown data patterns for business advantage" (SAS Institute, 2000). It also can be defined as:" the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules" (Berry & Linoff, 2004) and it involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and finally comprehensible information, from large databases (Shaw, Subramaniam, Tan, & Welge, 2001).What data mining tools do is to take data and construct a model as a representation of reality. The resulted model describes patterns and relationships, present in the data (Rygielski, Wang, & Yen, 2002).

The broad application of data mining falls in two major categories (Ngai, 2005):

1- **Descriptive data mining**: aims at increasing the understanding of the data and their content;
2- **Predictive or perspective data mining:** aims at forecasting and devising, at orienting the decision process.

Aiming at solving business problems, data mining can be used to build the following types of models (Ngai, Xiu, & Chau, 2009):

- Classification
- Regression
- Forecasting
- Clustering
- Association analysis
- Sequence discovery
- Visualization

Among the above mentioned models the first three one are prediction tools while association analysis and sequence discovery are used for description and clustering is applicable to either prediction or description.

The wide spread applications of data mining range from, evaluation of overall store performance, promotions' contribution to sales and determination of cross – selling strategies, to segmentation of the customer base (Gomory, Hoch, Lee, Podlaseck, & Schonberg, 1999). Moreover the data warehouse tools have enabled us to establish a customer data base which includes both traditional sources such as customer demographics data, and customer relationship data, and technical quality data (SAS Institute, 2000; Srivastava, Cooley, Deshpande, & Tan, 2000).

The application of data mining tools in CRM is an emerging trend in global economy. Since most companies try to analyze and understand their customers' behaviors and characteristics, for developing a competitive CRM strategy, data mining tools has become of high popularity (Ngai, Xiu, & Chau, 2009).

Beside the aforementioned roles for data mining in marketing, Rygielski, Wang, and Yen (2002) have identified a wide continuum of applications for data mining in marketing in different industries, from retailing to banking and telecommunications industry.

According to Rygielski, Wang, and Yen (2002) in retailing data mining can be used to perform basket analysis, sales forecasting, database marketing, and merchandise planning and allocation. Besides, data mining-based CRM in banking industry can be utilized in card marketing, cardholder pricing and profitability, fraud detection, and predictive life-cycle management. In addition to the above mentioned realms, data mining possesses a significant role in telecommunications industry. To be more specific, using data mining, companies would be able to analyze call detail records and identify customer segments with similar use patterns, and develop attractive pricing and feature promotions. Furthermore, data mining enables companies to identify the characteristics of customers who are likely to remain loyal and also determine the churners (Rygielski, Wang, & Yen, 2002).

With large volumes of data generated in CRM, data mining plays a leading role in the overall CRM (Shaw, Subramaniam, Tan, & Welge, 2001). In acquisition campaigns data mining can be used to profile people who have responded to previous similar campaigns and these data mining profiles is helpful to find the best customer segments that the company should target (Adomavicius & Tuzhilin, 2003). Another application is to look for prospects that have similar behavior patterns to today's established customers. In responding campaigns data mining can be applied to determine which prospects will become responders and which responders will become established customers. Established customers are also a significant area for data mining. Identifying customer behavior patterns from customer usage data and predicting which customers are likely to respond to cross-sell and up-sell campaigns, which are very important to the business (Chiang and Lin, 2000 cited by Olafsson, Li,and Wu, 2008). A review of literature from 2000 to 2006 shows that 54 out of 87 papers (62%) in field of data mining and CRM have focused on customer retention dimension of CRM. Besides, the authors have spotted an increasing trend toward this area of research that makes us to expect more publications in it (Ngai, Xiu, & Chau, 2009). Regarding former customers, data mining can be used to analyze the reasons for churns and to predict churn (Chiang et al., 2003; cited by Olafsson, Li,and Wu, 2008). Regarding this, there exist two different conceptions which have been developed by Ansari, Kohavi, Mason, & Zheng (2000) and Groth (1999). Ansari, Kohavi, Mason, & Zheng (2000) considered the importance of data, related to Recency, Frequency, and Monetary (RFM) attributes for evaluating customer churn, while Groth (1999) believes that considering the recency of purchase as a churn indicator may lead us to misrepresent the infrequent shoppers and as Lejeune (2001) noted such rules (RFM), neglect the purchasing behavior, that may significantly differ across segments and individuals.

Groth prefers to hire a methodology called "Value, Activity, and loyalty method (VAL)". From this point of view using descriptive data mining, one can divide the customer in the customer base into four classes on loyalty basis. According to Jones and Sasser (1995) customers fall in one of the following categories:

1. Loyalists and apostles
2. Hostages
3. Defectors
4. Mercenaries

After assigning the existing customers to one of the above mentioned classes by the use of descriptive data mining we would be able to use predictive data mining in order to specify the customers who are likely to churn (Lejeune, 2001). Thus the need for predictive data mining models arises.

Since in this research we utilized classification and clustering models in order to construct our predictive models, in next two sections we'll have a brief review of both model's definitions and their utilized techniques.

2.3.1. Classification

Classification is the most frequent learning model in data mining, especially in CRM field and it is capable of predicting the effectiveness or profitability of a CRM strategy through prediction of the customers' behavior (Ahmad, 2004; Carrier & Povel, 2003; Ngai, Xiu, & Chau, 2009). Classification can be defined as 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. The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is known) (Han & Kamber, 2006) or as Lee and Siau (2001) noted the classification process is the process of dividing a data set into mutually exclusive groups such that the members of each group are as “close” as possible to one another, and the members of different groups are as “far” as possible from one another. Also we can define the classification as “examining the features of a newly presented object and assigning it to one of the predefined set of classes” (Berry & Linoff, 2004). The objective of the classification is to first analyze the training data and develop an accurate description or a model for each class using the attributes available in the data. Such class descriptions are then used to classify future independent test data or to develop a better description for each class (Weiss and Kulikowski, 1991; cited by Olafsson, Li,and Wu, 2008).

Among all existing classification techniques Neural Network and Decision Tree are of high frequency of use respectively, but since the logic of Decision Tree is more understandable for business people than Neural Network, it should be a good choice for non-experts in data mining (Ngai, Xiu, & Chau, 2009; Wei & Chiu, 2002). As Olafsson, Li, and Wu, (2008) mentioned one of the main reasons behind their popularity appears to be their transparency, and hence relative advantage in terms of interpretability.

Decision tree

Decision Tree is a tree-shaped structure that represents sets of decisions and is able to generate rules for the classification of a data set (Lee & Siau, 2001) or as Berry and Linoff (2004) noted is a structure that can be used to divide up a large collection of records into successively smaller sets of records by applying a sequence of simple decision rules. Whatever the technique is, it has been proven to be one of the top 3 popular techniques of data mining in CRM (Ngai, Xiu, & Chau, 2009)

The Decision Tree technique is suitable for describing sequence of interrelated decisions or predicting future data trends (Berry & Linoff, 2004; Chen, Hsu, & Chou, 2003; Kim, Song, Kim, & Kim, 2005). The technique is capable of classifying specific entities into specific classes based on feature of entities (Buckinx, Moons, Van Den Poel, & Wets, 2004; Chen, Hsu, & Chou, 2003).

According to Tan, Steinbach, & Kumar (2006) each tree cosists of three types of nodes:

- Root Node
- Internal Node
- Leaf or Terminal Node

A record enters the tree at the root node. The root node applies a test to determine which internal node the record will encounter next. There are different algorithms for choosing the initial test, but the goal is always the same: To choose the test that best discriminates among the target classes. This process is repeated until the record arrives at a leaf node. All the records that end up at a given leaf of the tree are classified the same way, and each leaf node is assigned a class label (Tan, Steinbach, & Kumar, 2006; Berry & Linoff, 2004).

In fact decision tree is bale to solve a classification problem b asking a series of exact created questions about the characteristics of the test record. The following example provided by Tan, Steinbach, & Kumar (2006) can clarify the way a decision tree works:

Generally speaking vertebrates fall in two major categories: mammals and non-mammals. Now for classifying a newly doscovered species into one of these groups one way is to ask a series of questions about the attributes of the species.

1- Is the species cold blooded or warm blooded? Possible Answers: (Cold blooded: not mammal) or (Warm blooded: it is either a bird or a mammal so question two is necessary to be asked)
2- Do the females of the species give birth to their young? Possible answers: (Yes: mammals) or (No: nonmammal)

Figure 2.4 illustrates the decision tree shape of the later classification procedure.

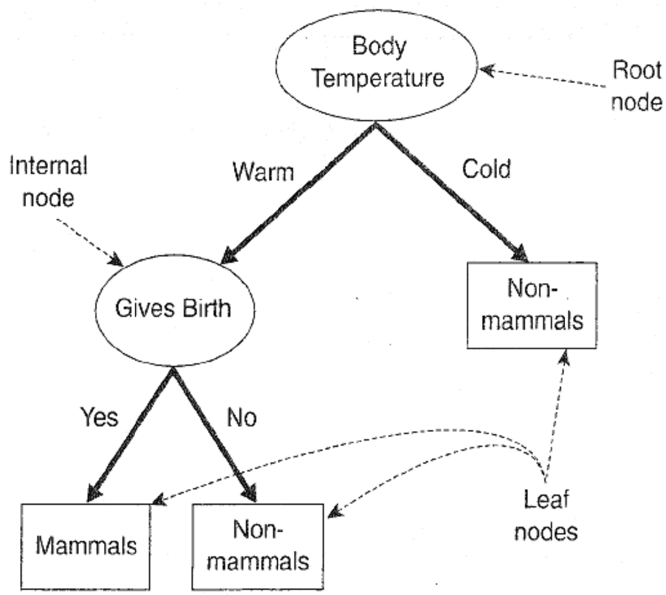

Figure 2.4: A Decision Tree for the mammals classification problem (Source: Tan, Steinbach, & Kumar, 2006)

Neural Networks

According to Berry and Linoff (2004) Neural networks— the "artificial" is usually dropped—are a class of powerful, general-purpose tools readily applied to prediction, classification, and clustering. A neural network consists of at least three layers of nodes. The input layer consists of one node for each of the independent attributes. The output layer consists of node(s) for the class attribute(s), and connecting these layers is one or more intermediate layers of nodes that transform the input into an output. When connected, these layers of nodes make up the network we refer to as a neural net (Olafsson et al 2006).

Neural networks have the ability to learn by example in much the same way that human experts gain from experience (Berry and Linoff, 2004).

Figure 2.5 shows the important features of the artificial neuron.

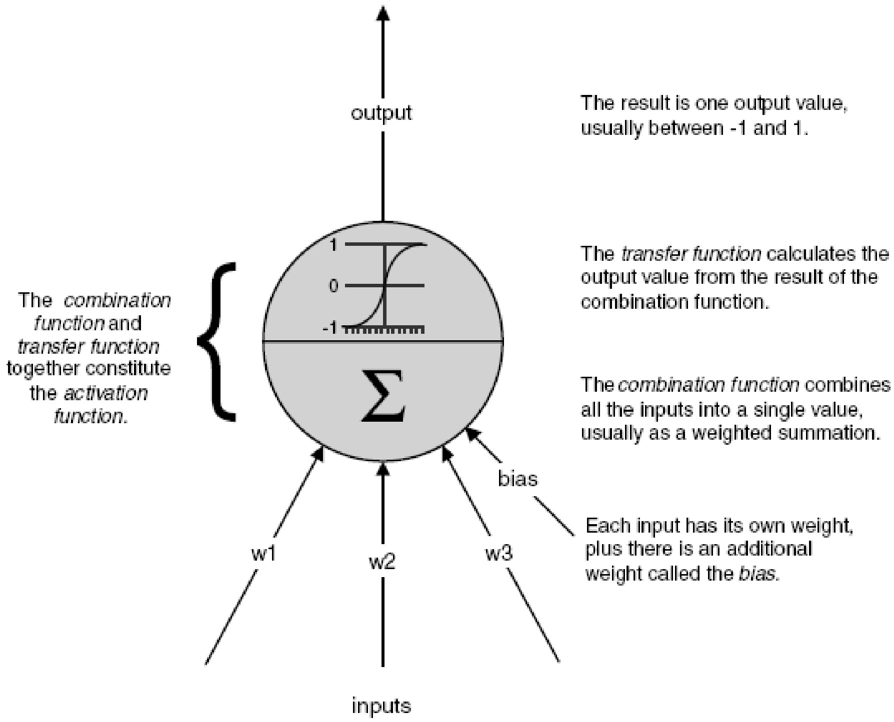

Figure 2.5 The unit of an artificial neural network is modeled on the biological neuron. The output of the unit is a nonlinear combination of its inputs. (source: Berry and Linoff, 2004).

2.3.2. Clustering

Cluster analysis is an approach by which a set of instances (without a predefined class attribute) is grouped into several clusters based only on information found in the data that describes the objects and their relationships (Wei & Chiu, 2002; Tan, Steinbach, & Kumar, 2006). “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 cluster” (Han & Kamber, 2006) .

While in classification the classes are defined prior to building the model, cluster analysis divides the data based on similarity them.

There exist different types of clustering from different point of view. The most common distinction among different types of clustering is to separate it two Partitional and hierarchical methods.

As Tan, Steinbach, & Kumar (2006) defined “Partitional Clustering” is the simple division of a set of data objects into non-overlapping segments such that each data object is in exactly one segment and if we permit clusters to have sub-clusters then we obtain a "Hierarchical Clustering".

Among existing clustering methods TwoStep Cluster technique is a clustering algorithm which has been designed to handle very large data sets (SPSS Inc, 2007).

TwoStep Cluster

TwoStep is a clustering technique that uses agglomerative hierarchical clustering method and as its name implies, involves two steps (SPSS Inc, 2007):

A. Pre-Clustering
B. Clustering

Pre-cluster

Using sequential clustering approach, the pre-cluster step scans the data records one by one and decides if the current record should be merged with the previously formed clusters or starts a new cluster based on distance criterion.

Cluster

This step takes the resulting pre-clusters from pre-cluster step and groups them into desired number of cluster.

TwoStep uses the hierarchical clustering method in the second step to assess multiple cluster solutions and automatically determine the optimal number of clusters for the input data (SPSS Inc, 2007).

2.4. Customer churn: Review of Literature

"The propensity of customers to cease doing business with a company in a given time period" can be defined as customer churn (Chandar, Laha, & Krishna, 2006).

Companies aim at getting more and more new customers. Nevertheless, the ratio (new customers/ churners) tends towards one over time. The impact of churn becomes then markedly more sensitive (Lejeune, 2001).

According to Lejeune (2001) the concept of churn is often correlated with the industry life-cycle. When the industry is in the growth phase of its life-cycle, sales increase exponentially; the number of new customers largely exceeds the number of churners, but for products in the maturity phase of their life- cycle, companies put the focus on the churn rate reduction.

Customer churn figures directly in how long a customer stays with a company and, in turn, the customer's lifetime value (CLV) to that company (Neslin, Gupta, Kamakura, Lu, & Mason, 2006), which is the sum of the revenues gained from company's customers over the lifetime of transactions after the deduction of the total cost of attracting, selling, and servicing customers, taking into account the time value of money (Hwang, Jung, & Suh, 2004).

Previous researches have examined the concept of customer churn from different points of view. According to Olafsson, Li, and Wu, (2008) there are two different types of churns. The first is voluntary churn, which means that established customers choose to stop being customers. The other type is forced churn, which refers to those established customers who no longer are good customers and the company cancels the relationship.

Burez and Van den Poel (2008) have divided the voluntary churners to two groups: commercial churners and financial churners. According to their research customers who voluntary leave the company can be divided into two groups: customers who do not renew their fixed term contract at the end of that contract, and others who just stop paying during their contract to which they are legally bound. The first type of churn can be considered commercial churn, i.e., customers making a studied choice not to renew their subscriptions. The second phenomenon is defined as financial churn, people who stop paying because they can no longer afford the service.

Nowadays Customer churn has become the main concern for firms in all industries (Neslin, Gupta, Kamakura, Lu, & Mason, 2006), and companies, regardless of the industry that they are active in, are dealing with this issue. Customer churn can blemish a company by decreasing profit level, losing a great deal of price premium, and losing referrals from continuing service customers (Reichheld & Sasser, 1990). A research by Reichheld (1996) revealed that an increase of 5% in customer retention rate can increase the average net present value of customer by 35% for software companies and 95% for advertising agencies.

Considering the churn rate of different industries, one can find that the telecommunications industry is one of the main targets of this hazard such that the churn rate in this industry ranges from 20 to 40 annually (Berson, Smith, & Therling, 1999; Madden, Savage, & Coble-Neal, 1999). Customer churn in mobile telecommunications (often refers to customer attrition in other industries) refers to “the movement of subscribers from one provider to another” (Wei & Chiu, 2002).

There exist two basic approaches to manage the customer churn. Untargeted approaches which rely on superior product and mass advertising to increase brand loyalty and retain customers and Targeted approaches which rely on identifying customers who are likely to churn, and then either provide them with a direct incentive or customize a service plan to stay.

The targeted approach falls in two categories: Reactive and Proactive. Adopting a reactive approach, a company waits until customers contact the company to cancel their (service) relationship. The company then offers the customer an incentive, for example a rebate, to stay. Adopting the proactive approach, the company tries to identify customers who are likely to churn at some later date in advance. The company then targets these customers with special programs or incentives to keep the customer from churning. Targeted proactive programs have potential advantages of having lower incentive costs (because the incentive may not have to be as high as when the customer has to be ‘‘bribed’’ not to leave at the last minute) and because customers are not trained to negotiate for better deals under the threat of churning. However, these systems can be very wasteful if churn predictions are inaccurate, because then companies are wasting incentive money on customers who would have stayed anyway. (Neslin, Gupta, Kamakura, Lu, & Mason, 2006; Coussement & Van den Poel, 2008)

In order to tackle this problem numerous attempts have been made to achieve an appropriate insight toward the churn concept. In general, researches in this field have been made with one of the following aims: finding the influential factors on customer churn, or model building for customer churn prediction which is still of high importance (Coussement & Van den Poel, 2009).

Despite the fact that the approach and focus of this research is on extracting and designing a predictive model for customer churn in telecommunications industry, we should bear in mind that due to the consistence nature of churning behavior of customers in almost all industries, attaining a true insight about customer churn in mobile telephony segment would be next to impossible in the absence of knowledge regarding the churn in other industries. Considering this fact, in this section the existing predictive models for churn in different industries have been studied. Additionally, in order to acquire insight into underlying factors of this problem in telecommunications industry, explanatory studies in this realm have been reviewed. In this regard numerous of exploratory and explanatory researches have been conducted with the aim of recognizing determinant factors that leads a customer to churn or to retain. Such researches have roots in the fact that service attributes and demographic attributes are of influential factors in defection of customers (Rust & Zahorik, 1993; Zeithaml, Leonard, & Parasuraman, 1996; Li S. , 1995; Bhattacharya, 1998). Among these researches that have been conducted in different industries some are about to find the churn drivers while the others was about to construct a predictive model using a statistical techniques.

In (2004) Kim and Yoon investigated the underlying elements of customer churn in mobile telecommunications service providers. From what they found we can understand that attrition of customers in this industry depends on the level of satisfaction with alternative specific service attributes including call quality, tariff level, handset, brand image, as well as income, and subscription duration, but only factors such as call quality, handset type, and brand image affect customer loyalty as has been measured by the positive word of mouth in the form of recommendation. In other words, according to Kim and Yoon (2004) determinants of churn clearly differ from those of loyalty and in order to decrease the churn rate in telecom industry the company is supposed to focus on boost the satisfaction level rather than loyalty.

Gerpott, Rams, and Schindler (2001) believe that retention, loyalty and satisfaction of customers in telecom industry are causally inter-correlated and that service price, perceived benefits, and also lack of number portability have strong effects on customer retention. They investigated the influential factors on bringing superior economic success for telecommunications network operators in German market and tested the hypotheses suggesting that Customer Retention (CR) Customer Loyalty (CL), and Customer Satisfaction (CS) should be treated as differential constructs which are causally inter-linked. The result shows that overall CS has a significant positive impact on CL which in turn influences a customer's intention to terminate / extend the contractual relationship (CR). It's also been revealed that mobile service price and personal service benefit perceptions as well as lack of number portability between various cellular operators' perceived customer care performance had no considerable effect on CR.

In 2006, Ahn, Han, and Lee conducted an exploratory research in which they aimed at finding the most influential factors on customer churn. In their research they considered a mediator factor named "Customer's Status", between churn determinants and customer churn in their model, and they've mentioned that "Customer's Status" (from active use to non – use or suspended) change is an early signal of total customer churn. In fact the main focus of this research is on finding determinants of churn and authors have found that call quality – related factors influence customer churn.

Figure 2.6 demonstrates four major constructs hypothesized by Ahn, Han, and Lee (2006) to affect customer churn and the mediation effects of customer's status that indirectly affect customer churn.

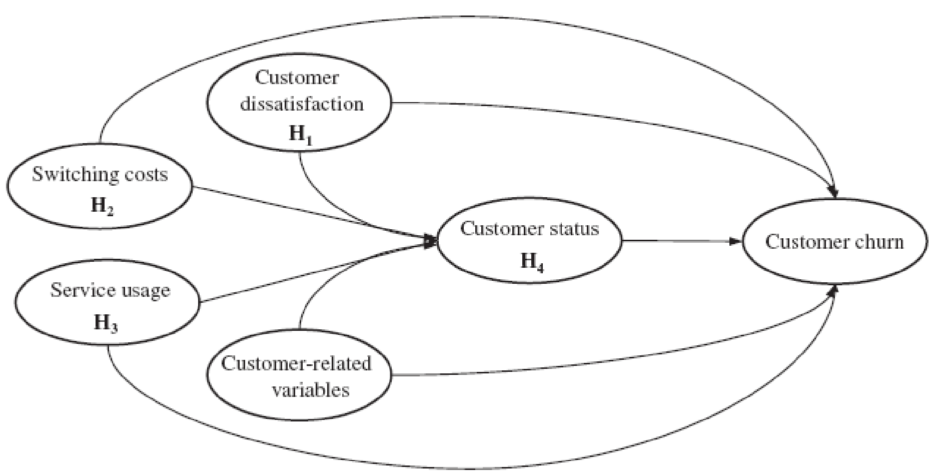

Figure 2.6: A conceptual model for customer churn with mediation effects (Source: Ahn, Han, & Lee, 2006)

In their research a mediator named "Customer Status" has been taken into account between churn determinants and customer churn, and it has been hypothesized that a customer's status change is an early signal of total customer churn.

Conducting their empirical analysis they draw a random sample of subscribers of a leading telecommunications service provider. The account had to be active during the time period between September 2001 and November 2001. For those customers, all accounts were tracked and examined for 8 month from September 2001 to April 2002, and "Churn" was defined as the event in which a subscription was terminated by the end of April 2002. In other words according to the above mentioned hypotheses churn happened during the period from December 2001 to April 2002. For churners 3-month, 2-month, and 1-month prior data was collected before the actual termination. For the non-churners, the most recent last 3 months of data was collected (from February 2002 to April 2002).

From the collected data they extracted the subscriber's usage and billing data and also the demographic data were added. The available data consisted of billed amounts, accumulated loyalty points; call quality-related indicators, handset-related information, calling plans, gender, etc.

In order to analyze the data and test the research questions three logistic regression adopted.

The results show that dissatisfaction indicators such as number of complaints and call drop rate have a significant impact on the probability of churn. Besides, it has been revealed that loyalty points such as membership card programs have a significant negative impact on the probability of customer churn. Moreover, surprisingly the findings showed that heavy users are more likely to churn and also customer status was found to have significant impact on the probability of churn. In addition they found out that customer status has a significant impact on the probability of churn. The customer's status changes from active use to either non-use or suspended increases the churn probability.

Delving into factors affecting customer churn Madden, Savage, and Coble-Neal (1999) investigated customer churn in Australian Internet Service Providers (ISPs). They designed a questionnaire asking Internet users about their Internet use and expenditure, pricing plan and Socio-demographic background, and at the end the respondents were asked about their intention to change their ISP within the next twelve months, and the reason of it. The results of the research show that probability of churn is positively associated with monthly ISP expenditure, but inversely related to household income. Furthermore the findings show that employing flat-rate pricing can decrease the churn tendency in compare with some form of timed usage charging structure. Besides, customers who use Internet for work related purposes and have an account with another ISP found to be at more risk of churn. Ultimately, the demographic factor, age, found to have significant effect on switching behavior of subscribers.

Furthermore in (2008) Seo, Ranganathan, & Babad investigated about retention factors in telecommunications industry by examining other features and variables. The focus of their study is on understanding the factors related to customer retention behavior i.e. both behavioral factors such as switching costs and customer satisfaction and demographic factors and its two goals are to understand (1) how factors that affect switching costs and customer satisfaction, such as length of association, service plan complexity, handset sophistication and the quality of connectivity, drive customer retention behavior, and (2) how customer demographics such as age and gender affect their choice of service plan complexity and handset sophistication, leading to differences in customer retention behavior.

The methodologies they used were a binary logistic regression model and a two – level hierarchical linear model.

The factors analyzed consisted of: complexity of service plan, handset sophistication, length of association, and connectivity. Customer demographics to be related to these factors are gender and age in figure 2.7.

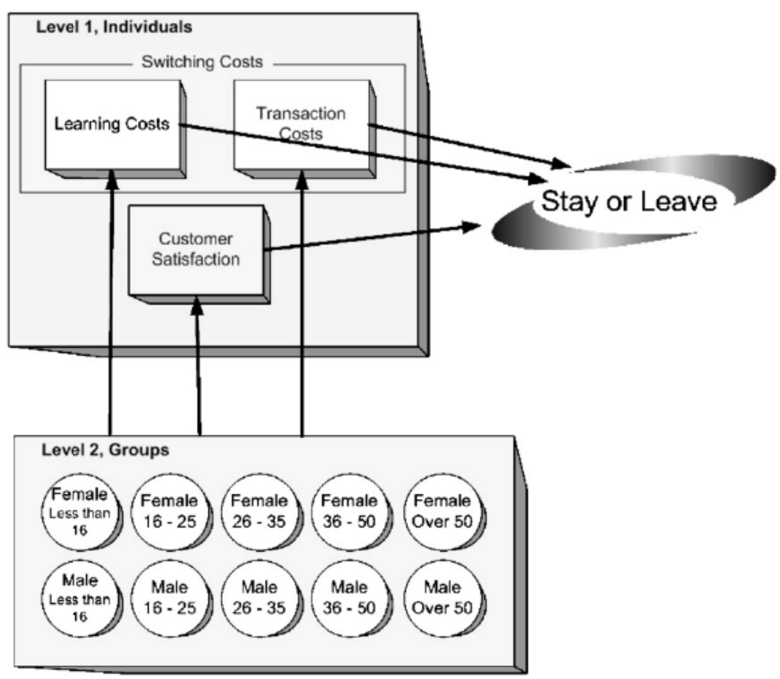

Figure 2.7: Conceptual model of customer retention behavior in wireless service (Source: Seo, Ranganathan, & Babad, 2008)

The results show that:

1. The more complex service plan, more sophisticated handset, longer customer association, higher connectivity quality of wireless is positively related to customer retention behavior.
2. Different age and gender groups revealed differences in wireless connectivity quality and service plan complexity, affecting their customer retention behavior, while they did not experience differences in terms of length of customer association and handset sophistication.

These results raise very interesting questions particularly that of asking why different age and gender groups would differ on the connectivity quality of wireless service and not on handset sophistication? So they divided the customer base into 10 groups according to their age and gender.

And they understood that the group of females over 25-years of age was most likely to stay with its current service provider, Customers under 26-year-olds, regardless of gender, were most likely to churn, and Customers in all groups preferred the most sophisticated handsets.

The most unpredicted result was that the different demographic groups do actually show a difference in connectivity quality (dropped-call ratio). This was surprising, because connectivity quality is not related to customer taste, but is a technical aspect of wireless service that should remain the same across different age and gender groups. However, the group of males over 25 years old had a much higher dropped-call ratio than all other groups, while males between 16 and 25 years old had the second highest dropped-call ratio. One possible conjecture is that males are more mobile than females. A dropped call happens most in handovers, when one cell-center hands over its users to another cell-center as they move from one area to another. This means that customers who are more mobile have a greater chance of experiencing dropped calls.

Additionally their research revealed that ales are more likely to have more complex service plans than females. Older customers tended to have more complex service plans as well, which sounds logical because heavy users like working people tend to have more complex plans.

The findings of Seo, Ranganathan, & Babad (2008)'s study contribute to the literature in three ways. First, they showed a strong relationship between switching costs and customer retention behavior. Accordingly, they understood that service plan complexity, reflecting price and wireless service usage, and handset sophistication can increase switching costs, which are positively related to customer retention behavior. Secondly, they confirmed once again the importance of technical performance in customer retention behavior. The fundamental quality characteristic of wireless service, connectivity quality, does affect customer retention behavior. Thirdly, the study reveals how age and gender demographics can affect customer retention behavior indirectly. These groups differ with respect to service plan complexity and connectivity of wireless service but are similar in terms of length of stay and handset sophistication, which lead to varying retention behavior.

Despite the efforts which have been made in order to utilize the statistical techniques for constructing the models for customer churn prediction, it is needless to say that model building for churn prediction is strongly dependent on machine learning techniques due to the better performance of machine learning techniques than the statistical techniques for non-parametric dataset (Baesens, Viaene, Van den Poel, Vanthienen, & Dedene, 2002; Bhattacharyya & Pendharkar, 1998)

Based on previous researches on churn prediction, Wei and Chiu (2002) developed a new model for customer churn prediction in telecommunication service providers by using data mining techniques. In that time, past researches on churn prediction in the telecommunications industry mainly had employed classification analysis techniques for the construction of churn prediction models and they had used user demographics, contractual data, customer service logs and call patterns extracted from call details (e.g. average call duration, number of outgoing calls, etc.), but Wei and Chiu believed that existing churn – prediction model had several disadvantages. They listed the disadvantages in two groups; first, use of customer demographics in churn prediction renders the resulting churn analysis at the customer rather than contract (or subscriber) level. In other words, tendency of each customer toward churning was calculated on a per-customer rather than contract basis. It is quite common that a customer concurrently holds several mobile service contracts with particular carrier, with some contracts more likely to be churned than others. In this regard, customer – level – based churn prediction is considered inappropriate. Second, information on some of the input variables (features) was not readily available and this unavailability of customer profiles, had been limited the applicability of existing churn – prediction systems.

In response to the described limitations of existing churn – prediction systems in that time, Wei and Chiu exploited the use of call pattern changes and contractual data for developing a churn – prediction techniques that identifies potential churners at the contract level. They claimed that subscribers' churn is not an instantaneous occurrence that leaves no trace. Before an existing subscriber churn, his/her call patterns might be changed (e.g. the number of outgoing calls gradually gets reduced). In other words, changes in call patterns are likely to include warning signals pointing toward churning. Such call pattern changes can be extracted from subscribers' call details and are valuable for constructing a churn prediction model based on a classification analysis technique. In their investigation they used two types of available data: Contractual data including length of services, payment type, contract type, and Call details such as Minutes of Use (MOU), Frequency of Use (FOU) and Sphere of Influence (SOI: refers to the total number of distinct receivers contacted by the subscriber over a specific period) in order to develop a churn prediction technique.

Using the data set Wei and Chiu (2002) randomly selected a prediction period (P) in order to generate an evaluation data set and also determine the churn status. According to them churn status of a subscriber was the connected or disconnected status of the subscriber within the prediction period P, and subscribers who disconnected his/her mobile service during P were considered as churner while the ones who disconnected the service before P were not included in their evaluation data set. Furthermore subscribers who were still connected to the service provider at the end of P classified as non-churner.

After determining the prediction period, the authors considered a retention period (R) immediately prior to P and the call records from this period were not used for churn prediction model construction. Moreover prior to R, an observation period (T) was specified and the required data for extracting the call pattern changes were employed from this period. Anyone whose contract started no earlier than the observation period T was excluded from the evaluation dataset. In brief their aim can be defined as the employing the call details of subscriber usage in observation period T to predict their churn status in prediction period P.

Representing call pattern changes of a subscriber during a specific observation period (T), the authors divided the T period into several sub-periods of equal duration. Then they modeled the call pattern change of a subscriber by considering the change rate of each measure between any two consecutive sub-periods. The variable used to signify the call pattern changes of a subscriber consist of:

1. MOU of a subscriber in the first sub-period ($MOU_{initial}$)
2. FOU of a subscriber in the first sub-period ($FOU_{initial}$)
3. SOI of a subscriber in the first sub-period ($SOI_{initial}$)
4. ΔMOU_S: The change in MOU of a subscriber between the sub-period s-1 and s (for s=2,3,…..,n) and is measured by $\Delta MOU_s = (MOU_s - MOU_{s-1} + \delta)/(MOU_{s-1} + \delta)$ where $MOU_1 = MOU_{initial}$ and $\delta = 0.01$.
5. ΔFOU_s: The change in FOU of a subscriber between the sub-period s-1 and s (for s=2,….,n) and is calculated as $\Delta FOU_s = (FOU_s - FOU_{s-1} + \delta)/(FOU_{s-1} + \delta)$
6. ΔSOI_s: The change in SOI of a subscriber between the sub-period s-1 and s (for s=2,….,n) and calculated as $\Delta SOI_s = (SOI_s - SOI_{s-1} + \delta)/(SOI_{s-1} + \delta)$.

As it is clear, the number of sub-periods and the duration of each sub period are reversely related to each other and the increase of each one causes the decrease of the other one. Thus choosing the appropriate number of sub-periods was one of the major concerns of authors.

Developing the churn prediction model they considered a set of subscribers as training instances and described them by the above mentioned input variables and labeled them to indicate the user's churn status.

Employing decision tree as their modeling technique and Detection Error Tradeoff (DET) curve as their evaluation criteria Wei and Chiu (2002) took their steps toward building their churn prediction model.

In their model building phase they tested the role of different variables such as desired class ratio, number of sub-periods in observation period, and length of retention period on accuracy of model. The initial result showed that the desired hit ratio equal to 1:2 and the number of sub-period equal to 2 can leverage the model accuracy to its optimum level. Moreover they built two models based on hit ratio=1:2 and number of sub-periods = 2. With two different lengths for retention period (i.e. 7 and 14 days for model 1 and model 2 respectively) in order to test the effect of Retention period on model's accuracy.

3. Model 1: R=7days

- ✓ Identified 10.03% of the subscribers that contained 54.33% of true churners (Lift factor =5.42)
- ✓ Identified 20% of the subscribers that contained 64.72% of true churners (Lift factor = 3.24
- ✓ Identified 29.68% of the subscribers that contained 72.16% of true churners (Lift factor = 2.43)

4. Model 2: R=14

- ✓ Identified 10.03% of the subscribers that contained 46.95% of true churners (Lift factor = 4.68)
- ✓ Identified 19.65% of the subscribers that contained 57.58% of true churners (Lift factor = 2.93)
- ✓ Identified 28.32% of the subscribers that contained 65.07% of the true churners (Lift factor = 2.30)

As it is presented above the first model out performs the second one and clearly both models have better performance in compare with an untargeted effort. (See figure 2.8)

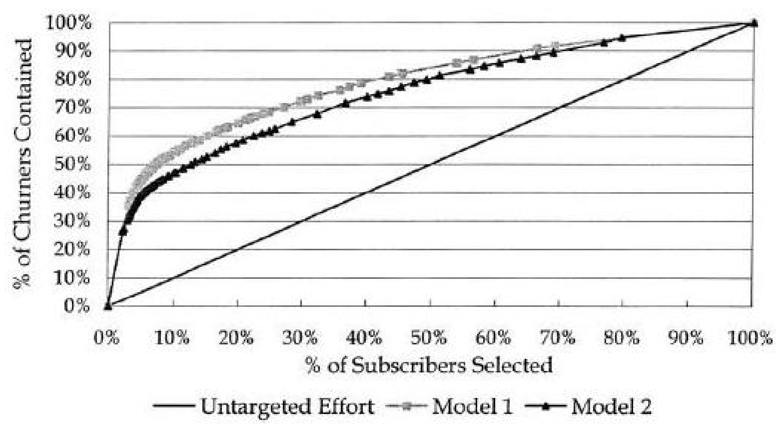

Figure 2.8: Lift chart attained by the proposed churn – prediction technique (source: Wei and Chiu, 2002).

As another approach Yan, Fassino and Baldasare (2005) tried to construct a predictive model for customer churn in pre-paid customer segment in mobile telephony market and due to the limited availability of data in prepaid customer segment, they exploited Call Detail Record (CDR). In order to construct their predictive model, the authors extracted the calling links i.e. who called whom as inputs to a neural network model and achieved an acceptable accuracy in their predictive model.

Using the CDR, they defined two categories of calling links as follows:

1. **Direct calling neighbor:** A person who calls the customer or whom the customer calls.
2. **Indirect calling neighbor:** A person who calls the same numbers as the customer does.

Utilizing these neighbors they discovered the calling community of each customer and hypothesized that people from a calling community behave in a similar way. So they supposed that if a customer's most frequently called parties churned from the same service provider, the customer may eventually churn also.

With the intention of building the churn predictive model they used the CDR data of July and August so that predict the churn in December. As it is clear they considered a 3 month gap between the observation and prediction period. In addition, they were provided with churn labels i.e. who churned, in both November and December. In fact their research's task was to develop a churn prediction model, with churn in December as the dependent variable (Prediction Target) and two independent variable including: the CDR data in July and August and the churn information in November.

Then they analyzed the data by using decision tree and neural networks and understood that for the neural network, if the customer service representatives contact the 10% of customers with the highest scores from the model, they are able to correctly identify 20% of the churners. By random sampling, the lift curve is the diagonal line. Also they understood that the neural networks outperform the decision tree, which performs even worse than random sampling for a higher contact rate (figure 2.9).

The evaluation of the model was based on lift curve with the following axis:

- Y-Axis: True Positive Rate
- X-Axis: Customer Contact Rate

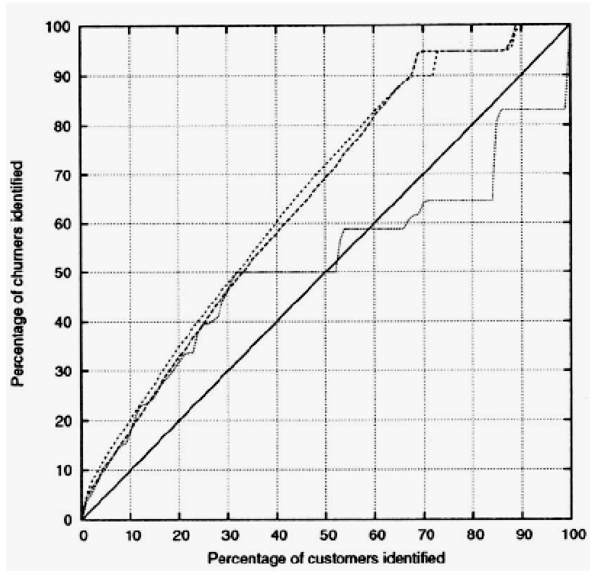

Figure 2.9: Lift curves of chum prediction. The neural network model of the long-dashed line used only features of first order distance, while the short-dashed line is for the neural network model using features based on both first and second order distances. The dotted line is based on boosting decision trees. (Source: Yan, Fassino, & Baldasare, 2005)

As another effort on predicting customer churn in telecommunications companies Hung, Yen, and Wang (2006), compared different data mining techniques that can be utilized in order to build a model for churn prediction. Using the lift factor as the criterion model performance evaluation, the authors compared the performance of Decision- Tree without segmentation, Decision-Tree with segmentation, and Neural-Network in building a model for churn prediction.

The study concentrates on post-paid subscribers who were activated for at least 3 months prior to July 1, 2001, and the "churner" was defined as a subscriber who is voluntary to leave and a non-churner is the one who is still using the specified operator service.

The authors used the latest 6 months (July-June) transaction data of each subscriber to predict the churn probability in the following month. As the input variables of their model they extracted the following variables from other researches and interviews with experts:

- **Customer Demography**
  - *Age:* analysis shows that the customers between 45 and 48 have a higher propensity to churn than population's churn rate.
  - *Tenure:* customers with 25 – 30 months tenure have a high propensity to churn. A possible cause is that most subscription plans have a 2-year contract period.
  - *Gender:* churn probability for corporate accounts is higher than others. A possible cause is that when employees quit, they lose corporate subsidy in mobile services.

- **Bill and payment analysis**
  - *Monthly fee:* the churn probability is higher for customers with a monthly fee less than $100 NT or between $520 and $550.
  - *Billing amount:* the churn probability tends to be higher for customers whose average billing amount over 6 months is less than or equal to $190 NT.
  - *Count of overdue payment:* the churn probability is higher for customers with less than four counts of overdue payments in the past 6 months. In Taiwan, if the payment is 2 months overdue, the mobile operator will most likely suspend the mobile service until fully paid. This may cause customer dissatisfaction and churn.

- **Call detail records analysis**
  - *In-net call duration:* customers who don't often make phone calls to others in the same operator's mobile network are more likely to churn. In-net unit price is relatively lower than that of other call types. Price-sensitive subscribers may leave for the mobile operator his/her friends use.
  - *Call type:* customers who often make PSTN or IDD calls are more likely to churn than those who make more mobile calls.

- **Customer care/service analysis**
  - *MSISDN change count:* customers who have changed their phone number or made two or more changes in account information are more likely to churn.
  - *Count of bar and suspend:* customers who have ever been barred or suspended are more likely to churn. In general, a subscriber will be barred or suspended by the mobile operators due to overdue payments.

Using the above mentioned variables, Hung, Yen, and Wang (2006) adopted the following three approaches toward model building for customer defection prediction:

a) **Decision-Tree with segmentation:**

   by the use of K-Means Cluster technique and variables such as bill amount, tenure, MOU (outbound call usage), MTU (inbound call usage), and payment rate, they clustered the customer base to five clusters and the Decision-Three was constructed in each of these five customer segments

b) **Decision-Tree without segmentation:**

   The tree was built for all customers of a single cluster

c) **Neural Network (Back Propagation Network, BPN)**

The results depicted that the Decision-Tree model without segmentation outperforms the Decision-Tree model with segmentation. Besides, the outcomes show that BPN based model posses a better performance than the two other models.

As it has been mentioned before in chapter one, the RFM model which is proposed and developed by Ansari, Kohavi, Mason, & Zheng (2000), is one of the major approaches toward predcting the probability of churn and retention. In accordance with Fader, Hardie & Lee (2005), a customer past behavior is an important predictor for one's future behavior and indeed RFM model has considered to be the model based on past behavior and as you may considered, up to this point most churn prediction models were basically based on input data from RFM plus some additional information such as demographic or transactional data. In other words most built models were the same but the utilized techniques differentiated them from each other. In contrast with most of the above mentioned predictive models, Coussement & Van den Poel (2008) developed a predictive model for customer churn by adding the "Voice of Customers" (VOC) to the independent variables of their model.

They used data from a large Belgian newspaper publishing company in a time period from January 2002 to September 2005, and extracted two renewal time between July 2004 and July 2005. Furthermore they defined a churner as a person who did not renew his/her contract in a 4-week period after maturity date.

Conducting this study, the authors extracted the information from emails as the Voice of Customers by the use of text mining (a process of deriving high-quality information from text) and used it as a feature, in addition to other structured marketing information i.e. all transactional and marketing related information, in order to build their the prediction model.

Thus, the built model exploited two types of data as its independent variables. The first type of data includes, the information from the structured marketing database such as Client/Company interaction variables, Subscription related variables, renewal specific variables, and Socio-demographics. The second type of independent data consists of all information sent by the subscriber via email during the last period of his/her subscription.

Using Logistic Regression as the data mining technique and lift factor and Area Under the receiving operating Curve (AUC), as the evaluation criteria, Coussement & Van den Poel (2008a) conducted their model building phase and the analysis results came out to show that combining the unstructured information from emails with other RFM (Recency, Frequency and Monetary) features can cause an increase from 73.80 to 77.75 in AUC and from 2.69 to 3.07 in lift factor in the first decile.

Continuing their previous research Coussement & Van den Poel (2009) tested the performance of different classifiers on the similar data in order to choose the best performing classification technique, in addition to testing the model enhancement by relying on customer information. Indeed, the aim of this study was to contribute to the literature by finding the proof that adding emotions in client/company emails increases the predictive performance of an RFM churn model and also compare the performance of three classification techniques i.e. Logistic Regression, Support Vector Machines (SVM), and Random Forests to distinguish churners from non- churners.

Thus, by defining “Extended RFM” (eRFM) model as a RFM model which also includes other information such as demographic or other transactional data. Coussement & Van den Poel (2008b) put one step ahead and extended the eRFM model by adding client/company interaction email data which includes the emotional aspect of clients toward the company and called it “eRFM-EMO”.

Using data from a news paper company and the time window same as their previous research (Coussement & Van den Poel, 2008) for observation and prediction and Percentage Correctly Classified (PCC) and the Area Under the receiving operating Curve (AUC) as their evaluation criteria, they applied SVM, Logistic Regression, and Random Forests on the data.

The results show that an eRFM-EMO model always (with all three tested techniques) has a higher predictive performance in compare with the eRFM model. It has also revealed that implementing Random Forests is an opportunity to improve the predictive performance and its performance is always significantly higher than the performance of Logistic Regression and SVM. Furthermore, the study found no significant relationship between positive expressed emotions in information requests and someone’s churn. Besides, negative expressed emotions in information requests seems to be influential on customers' churning behavior. To be more specific, according to this research, one can say that the more negative emotional words are used in emails other than complaints, the lower the chance that the customer will churn and also it has been concluded that the more complaints a customer has in her/his emails sent to the company, the more certain he/she stays with the company.

In the same year another research conducted by Pendharkar, for churn prediction in telecommunications industry, using Genetic Algorithm (GA) based Neural Network (Pendharkar, 2009). The authors designed two GA based Neural Network model. One by using cross entropy based criterion and the other one with direct approach.

They compared these two proposed model with a statistical z-score model and concluded that both above mentioned models outperform the statistical z-score model. Furthermore it's been proven that the cross entropy based criterion may be more resistant to overfitting outlier in training dataset.

Conducting the process of model building, the pendharkar (2008) used the following features:

- Subscriber ID Number
- Billing Month
- Subscription Plan
- Monthly Total Peak Usage in Minutes
- Promotional Mailing Variable
- Churn Indicator

For his Neural Network classification model, he excluded Subscriber ID Number and Billing Month and considered subscription plan, monthly total usage in minutes, , and promotional mailing variable as inputs and churn variable as the output variable. Regarding this, he split the original set of 195,956 examples into five train and test pair (70%-30% respectively) randomly and for Neural Network model and pair of datasets they performed three different tests with different number of nodes in the hidden layer (i.e. Three, six, and nine).

The final results showed that Neural Networks models dominated the z-score model in all aspects while both Neural Network models have the same performance. Furthermore the study revealed the point that medium sized Neural Network (i.e. the one with 6 nodes in hidden layer) posses the optimum performance (Pendharkar, 2009).

Mining with rarity

Considering all researches conducted by having a focus on churn prediction, one can discover a common problem among them. The problem with churn analysis derives from the specific nature of churn prediction (Xie, Li, Ngai, & Ying, 2009). As Zhao, Li, Li, Liu, & Ren (2005), Au, Chan, & Yao (2003) and Shah (1996) have noted, we can name three major characteristics for churn prediction as follow:

1. The data is usually imbalanced and the number of churners constitutes only a very small minority of the data
2. Large learning applications will have some type of noise in the data
3. Churn prediction requires the ranking of subscribers according to their likelihood to churn

Among these three, the problem of imbalanced data is becoming the focal point of most studies in this realm during recent years (Burez & Van den Poel, 2009).

Since the customer churn is often a rare event in service industries, nearly all datasets by which the predictive models are built are imbalanced (i.e. the number of churners is considerably lower than the non-churners) (Burez & Van den Poel, 2009). And due to this issue six mining problems may arise (Weiss, 2004):

1. Improper evaluation metrics:
2. Lack of data (absolute rarity)
3. Relative lack of data (relative rarity)
4. Data fragmentation
5. Inappropriate inductive bias
6. Noise

Coping with these problems different approaches have been adopted by experts. As mentioned before Wei & Chiu (2002) used multi-classifier class combiner their approach to tackle the relative rarity problem and they showed that under sampling the data and working with data with hit ratio of 1:2 (churner : non-churner) can help to improve the model's accuarcy.

According to Weiss (2004) there are ten solutions to the aforementioned problems:

1. Using more appropriate evaluation metrics
2. Non-greedy search techniques
3. Using a more appropriate inductive bias
4. Knowledge/Human interaction

5. Segmenting the data
6. Learn only the rare class
7. Accounting for rare item
8. Sampling
9. Cost-sensitive learning
10. Other methods such as boosting, placing rare cases into sepaaret classes, and two phase rule induction

Based on the study by Weiss (2004) , Burez & Van den Poel (2008b) have put 4 of the above mentioned solutions into practice and tested the performance of them for handling the imbalance in customer churn prediction. They used appropriate evaluation metrics such as AUC and lift curve as their evaluation metrics, cost-sensitive learning such as Weighted Random Forests, basic and advanced smpling methods such as under sampling, over sampling, and CUBE ((Deville & Tille, 2004) and boosting in order to buid a model with better performance.

Reults revealed that, regarding the evaluation metrics, both AUC and lift curve showed acceptable performance but since AUC has the advantage of being dependent on the churn rate, it would be more appropriate to be used for evaluation of churn prediction models. Furthermore resluts dipicted that under-sampling can lead to improved predictive accuracy especially when evaluated with AUC but the advaned sampling techniques CUBE found to cause no increase in predictive performance. Additionally, according to Burez & Van den Poel (2009)'s findings the Weighted Random Forests, as a cost-sensitive learner, has a significantly better performance compared to Random Forests.

Moreover and as another attempt regarding the handling the data imbalance, Xie, Li, Ngai, & Ying, (2008) used a combination of wighted and balaced Random Forests ,called improved balanced Random Forests, and they concluded that their proposed technique significantly outperforms the other standard methods, namely Artificial Naural Network, Decision Tree, and Support Vector Machine (See figur 2.10 and figure 2.11).

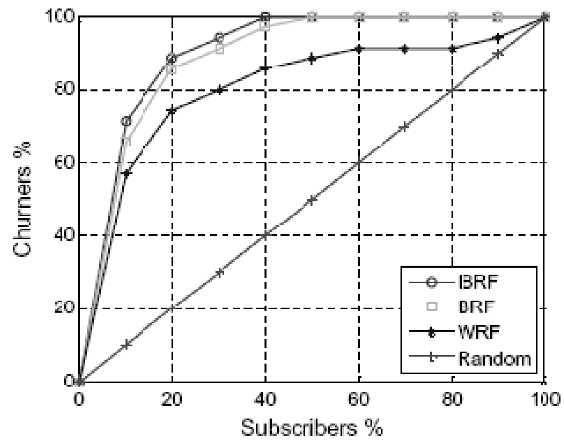

Figure 2.10: Lift curve of different random forests algorithms (Source: Xie, Li, Ngai, & Ying, 2009)

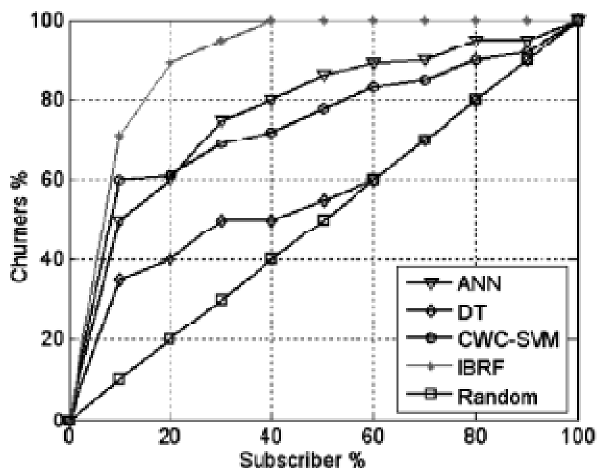

Figure 2.11: Lift curve of different algorithms (Source: Xie, Li, Ngai, & Ying, 2009)

As mentioned before recently applying cost-sensitive methods, has emerged among experts as a remedy for handling the class imbalance in churn datasets (Burez & Van den Poel, 2009; Xie, Li, Ngai, & Ying, 2009).

"Cost-sensitive learning methods can take advantage of the fact that the value of correctly identifying the rare class outweighs the value of correctly identifying the common class. For two-class problems this is done by associating a greater cost with false negatives than with false positives which leads to improving the model's performance with respect to the rare class (Weiss, 2004).

According to Ling and Sheng (2008) different costs such as costs of false positive (actual negative but predicted as positive; denoted as FP), false negative (FN), true positive (TP) and true negative (TN), in cost-sensitive learning can be given in a cost matrix similar to table 2.2

Table 2.2: An example of cost matrix for binary classification.

	Actual negative	**Actual positive**
Predict negative	C(0,0), or TN	C(0,1), or FN
Predict positive	C(1,0), or FP	C(1,1), or TP

Where $C(i, i)$ is considered as the benefit and the rare class is regarded as the positive class and it is often more expensive to misclassify an actual positive example into negative, than an actual negative example into positive. In other words the costs imposed by FN or $C(0,1)$ is always larger than the costs imposed by FP or $C(1,0)$. As Ling and Sheng (2008) mentioned according to the cost matrix an example should be classified into the class with the minimum expected cost. This is the minimum expected cost principle. The expected cost R(i|x) of classifying an instance x into class i (by a classifier) can be expressed as:

$$\mathrm{R(i|x)} = \sum_{\mathrm{j}} \mathrm{P(j|x)\ C(j, i)}$$

Where P(j|x) is the probability estimation of classifying an instance into class j. That is, the classifier will classify an instance x into positive class if and only if:

$$P(0|x)C(1,0) + P(1|x)C(1,1) \leq P(0|x)C(0,0) + P(1|x)C(0,1)$$

This is equivalent to

$$P(0|x)(C(1,0) - C(0,0)) \leq P(1|x)(C(0,1) - C(1,1))$$

Thus, the decision (of classifying an example into positive) will not be changed if a constant is added into a column of the original cost matrix. Thus, the original cost matrix can always be converted to a simpler one by subtracting C(0,0) to the first column, and C(1,1) to the second column. After such conversion, the simpler cost matrix is shown in Table 2.3. Thus, any given cost-matrix can be converted to one with C(0,0) = C(1,1) = 0. In the rest of the paper, we will assume that C(0,0) = C(1,1) = 0. Under this assumption, the classifier will classify an instance x into positive class if and only if:

$$P(0|x)C(1,0) \leq P(1|x)C(0,1)$$

Table 2.3: A simpler cost matrix with an equivalent optimal classification

	True negative	**True positive**
Predict negative	0	C(0,1) – C(1,1)
Predict positive	C (1,0) – C (0,0)	0

As P(0|x) = 1 – P(1|x), we can obtain a threshold P^* for the classifier to classify an instance x into positive if P(1|x) ≥ P^*, where

$$P^* = \frac{C(1,0)}{C(1,0) + C(0,1)}$$

Thus, if a cost-insensitive classifier can produce a posterior probability estimation p(1|x) for test examples x, we can make it cost-sensitive by simply choosing the classification threshold according to (2), and classify any example to be positive whenever P(1|x) ≥ P^*. This is what most of cost-sensitive learning methods based on.

2.5. Summary

In the later chapter after reviewing the basic concepts regarding the Customer Relationship Management (CRM) and locating the IT based or analytical CRM in its scale and clarifying its importance, we introduced data mining as an advanced machine learning approach which is applicable in Customer relationship management realm. Afterwards the significant studies regarding the customer churn from both descriptive and predictive point of view were reviewed. At the end the issue of data imbalance in churn datasets was discussed and remedies for it were extracted from the previous studies. As it was obvious, almost all predictive models that have been developed in this realm were utilized all or some of the RFM variables as their input variables for model building (Wei & Chiu, 2002; Coussement & Van den Poel, 2008; Hung, Yen, & Wang, 2006; Coussement & Van den Poel, 2009). Conducting our study we also followed our antecedents' procedure and utilized the RFM features of our customer base as the input variables in clustering phase and afterwards we tailored the behavioral variables proposed by Wei & Chiu (2002) in order to build our predictive model.

Chapter 3 Research Methodology

3.1. Introduction

The research design is a framework for conducting marketing research (Malhotra, 2007). Thus it's the basic plan for conducting the data collection and analysis phase. In the current chapter firstly the design of the research will be explained and scrutinized and afterwards the process of executing the designed research will be illustrated and explained.

3.2. Research Design

3.2.1. Research purpose

Studies generally fall into the following three categories: Descriptive, Explanatory (causal), and Exploratory (Saunders, Lewis, & Thornhill, 2000).

The primary purpose of exploratory research is to shed light on the nature of a situation and identify any specific objectives or data needs to be addressed through additional research. Exploratory research is most useful when a decision maker wishes to better understand a situation and/or identify decision alternatives. Exploration is particularly useful when researchers lack a clear idea of the problems they will meet during the study. The object of descriptive studies is to describe market characteristics or functions (Malhotra, 2007). Describe is to make complicated things understandable by reducing them to their component parts. Descriptive research could be in direct connection to exploratory research, since researchers might have started off by wanting gain insight to a problem, and after having stated it their research becomes descriptive (Saunders, Lewis, & Thornhill, 2000). Explanatory studies establish causal relationship between variables. In these studies the emphasis is on studying a situation or a problem in order to explain the relationships between variables (Saunders, Lewis, & Thornhill, 2000).

Based on the definition given for data mining, such approach aims at describing the process of discovering knowledge from databases stored in data warehouses. The purpose of data mining is to identify valid, novel, useful, and ultimately understandable patterns in data. Data mining is a useful tool, an approach that combines exploration and discovery with confirmatory analysis. Since the focus of this study is data mining, thus the purpose of this research is exploratory.

3.2.2. Research approach

3.2.2.1. Qualitative Vs. Quantitative Research

Quantitative research is an inquiry into an identified problem, based on testing a theory, measured with numbers, and analyzed using statistical techniques. The goal of quantitative methods is to determine whether the predictive generalizations of a theory hold true. On the other hand qualitative research is often a broad term that describes research focusing on how individuals and groups view and understand the world and construct meaning out of experiences. Some researchers consider it simply to be research whose goal is not to estimate statistical parameters but to generate hypothesis that can be tested quantitatively.

Malhotra (2007) has briefly compared the quantitative and qualitative research approaches as it is illustrated in table 3.1

Table 3.1: Qualitative research Vs. Quantitative research (Source: Malhotra, 2007)

	Qualitative Research	**Quantitative Research**
Objective	To gain a qualitative understanding of the underlying reasons and motivations	To quantify the data and generalize the results from the sample to the population of interest
Sample	Small number of non-representative cases	Large number of representative cases
Data Collection	unstructured	structured
Data Analysis	Non-statistical	Statistical
Outcome	Develop an initial understanding	Recommend a final course of action

With this regard the current study falls in quantitative category which uses data mining (a series of sophisticated statistical algorithms in order to build the predictive model for customer churn in telecommunications industry.

3.2.2.2. Inductive vs. Deductive research

Regarding the approach, researches can be divided to two groups: inductive (Bottom – up) research and deductive (Top – Down) research. While in a deductive approach the research strategy is designed to test the hypotheses (or hypothesis) based on a Pre-developed theory, the role of an inductive research is the production of a theory from specific observations (Saunders, Lewis, & Thornhill, 2000). In other words deductive research works from the general to the specific whereas inductive research works from specific observations to broader generalizations and theories.

This research pursues inductive/deductive approach by studying, developing a model for customer churn prediction in telecom service providers for a specific telecom service provider by the use of train data and then testing and revising the initial model, using test data which both test and train data have been generated by the users.

3.2.3. Research strategy

Research strategy is a general plan of how researcher will go about answering research question (Saunders, Lewis, & Thornhill, 2000).

This strategy contains clear objectives, derived from the research question(s), sources from which the data are intended to be collected and the possible constraints (Saunders, Lewis, & Thornhill, 2000)

Generally two categories of data can be utilized in researches: primary and secondary data. Primary data are originated by a researcher for the specific purpose of addressing the problem at hand. Secondary data are data that have already been generated and collected for purposes other than the problem at hand. Secondary data include data generated within an organization, information made available by business and government sources, commercial marketing research firms and computerized databases. Secondary data can be classified as internal and external. The internal data are those generated within the organization for which the research is being conducted and the external data are those generated by sources outside the organization (Malhotra, 2007). Internal Secondary data were gathered from Talia Co.'s database for this study. Since the aim of this thesis is data mining and data were collected from service providers's databases, the strategy, which fits this study is a secondary data analysis.

3.3. Research Process

Figure 3.1 illustrates the flow chart depicts the procedure of the research (Li & Ruan, 2007)

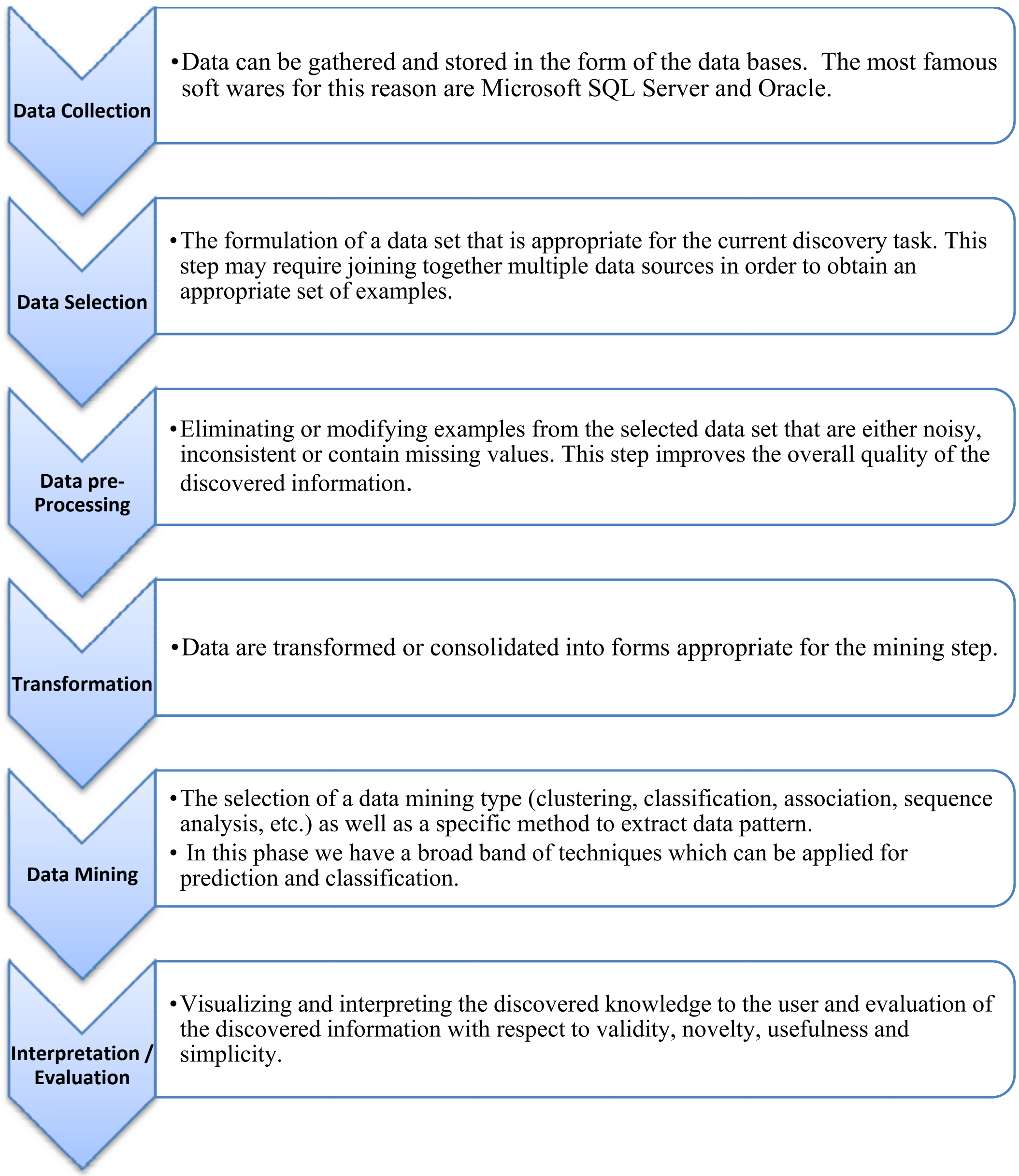

Figure 3.1: The flow chart of Knowledge Discovery in Databases

3.3.1. Data collection

The examined data of this research is the call records of Talia Co., which has been gathered and stored in Oracle data base software. Since telecom company's data base is a dynamic one that is being updated and extended every second, the call records' data base is a huge data base with millions and thousands million of records.

3.3.2. Data selection

The working data of this research is extracted from the Talia data base in a period of 3 months from 1 November 2007 to 31 January 2008 and it contains the call records of 34523 customers of Talia Company. The number of records is 19500504. Among various data types that are being saved and gathered in Talia's database we extracted the following one in order to utilize them in building the required and targeted features:

- ✓ Date of Call
- ✓ Time of call
- ✓ Duration of call
- ✓ Incoming call / Outgoing call

3.3.3. Data Pre-Processing

Since the working data has been produced and recorded by machine and contains call records information we did not face with the problem of noisy data or missing value in our dataset. The only pre-processing phase that we went through was data integration. Since the data was given in three individual (TXT) files in the pre-processing phase we needed to integrate them in one single database so, data integration was conducted and the whole data was gathered in a single database in Microsoft SQL Server.

3.3.4. Data Transformation

In this stage the raw data that had been extracted from the data base (as discussed in data selection and data processing part) was exploited in feature building.

The problematic side of this calculation was different fees that are being applied to phone calls in different times of a day and also in different days of a week. Based on costs that the Talia Company has considered for its services, each phone call that be made in the time period between 9 p.m and 8 a.m or on Fridays and holidays costs 536 Rials per Minute and each phone call that be made in the time period from 8 a.m to 9 p.m costs 670 Rials per Minute. Thus in order to calculate the cost of each outgoing phone call we considered the time of it and the day of it (whether it is holiday or not) and made the "Cost" variable.

The features that we built in the process of transforming the data were recognized and selected in accordance with the previous literature of churn prediction in telecommunications service providers (Wei & Chiu, 2002; Coussement & Van den Poel, 2009; Hung, Yen, & Wang, 2006) and also RFM related features. The reason behind choosing these studies as our foundation of feature construction is that due to the nature of pre-paid service providers our focus was on constructing features that are able to reflect the changes in usage behavior and among the reviewed researches the abovementioned ones (especially the one conducted by Wei & Chiu, 2002) appropriately satisfies this need.

3.3.5. Data Mining

The data mining phase was conducted in two steps and in these steps both descriptive and predictive data mining techniques were utilized.
Firstly, the customer base was clustered based on their usage behavioral feature (RFM). In order to cluster our customer base according to their usage behavior we utilized TwoStep Cluster method.

In the second step and for executing the predictive data mining approach, classification technique was utilized. Classification 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. The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is known), (Han & Kamber, 2006). Also we can define the classification as "examining the features of a newly presented object and assigning it to one of a predefined set of classes" (Berry & Linoff, 2004). The objective of the classification is to first analyze the training data and develop an accurate description or a model for each class using the attributes available in the data. Such class descriptions are then used to classify future independent test data or to develop a better description for each class (Weiss and Kulikowski, 1991; cited by Olafsson, Li, & Wu, 2008).

Many techniques have been adopted for classification and prediction, including decision tree induction, support vector machines (SVM), neural networks, and Bayesian networks. Among the existing classification method we chose to use Decision Tree due to its ease of interpretation and more understandable logic (Wei & Chiu, 2002; Ngai, Xiu, & Chau, 2009).

3.3.6. Interpretation/ Evaluation

According to the model's output we can measure its accuracy. There exists a broad choice of evaluation metrics for the predictive models which have been built by data mining techniques and each one possesses its pros and cons. Since our data suffers from class imbalance and the field of our research is marketing, in this research we have evaluated our developed models by using Gain Chart (Lift Curve) based on what Burez & Van den Poel, (2008b) proposed. It gives a graphical interpretation of what percentage of customers one has to target to reach a certain percentage of all churners. An example given by Berry & Linoff (2004) may helps to explain better. Suppose that we are building a model to predict who is likely to respond to a direct mail solicitation. As usual, we build the model using a pre-classified training dataset. Now we are ready to use the test set to calculate the model's lift (Gain). The classifier scores the records in the test set as either "predicted to respond" or "not predicted to respond." If the test set contains 5 percent actual responders and the sample contains 50 percent actual responders, the model provides a lift of 10 (50 divided by 5). The gain charts (Figure 3.2) is created by sorting all the prospects according to their likelihood of responding as predicted by the model. As the size of the mailing list increases, we reach farther and farther down the list. The X-axis shows the percentage of the population getting our mailing. The Y-axis shows the percentage of all responders we reach.

If no model were used, mailing to 10 percent of the population would reach 10 percent of the responders, mailing to 50 percent of the population would reach 50 percent of the responders, and mailing to everyone would reach all the responders. This mass-mailing approach is illustrated by the line slanting upwards. The other curve shows what happens if the model is used to select recipients for the mailing. The model finds 20 percent of the responders by mailing to only 10 percent of the population. Soliciting half the population reaches over 70 percent of the responders.

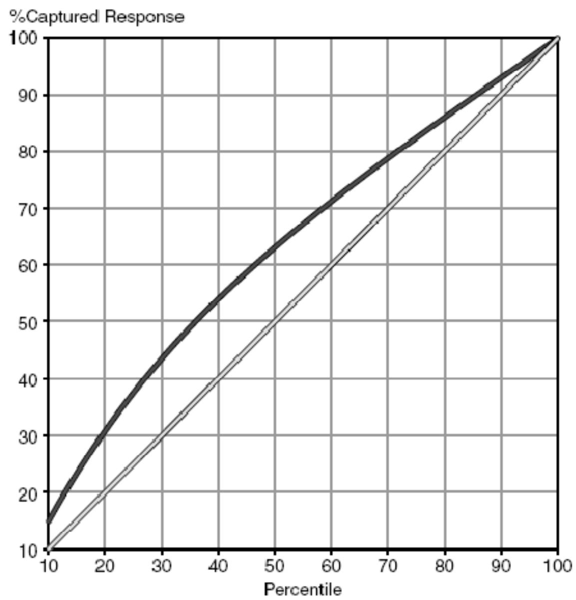

Figure 3.2: Cumulative response for targeted mailing compared with mass mailing

Chapter 4 Analysis & Result

4.1. Introduction

In this chapter the detailed procedure of model building has been and its results in each step has been brought. Furthermore in order to improve the model's performance I have tested the methods proposed in the literature for handling the class imbalance and reported its outcome.

4.2. A Dual-Step Model for Churn Prediction

With the intention of constructing a classification model which is capable of predicting the churners in pre-paid mobile telephony market segment the first task is to define the "Churn" and "Churner" so that we can label churners from non-churners. Due to the nature of this segment of market which is pre-paid and non-contract based giving an appropriate definition for churn is the initial step prior to the model building phase. Thus, regarding this fact with the purpose of model construction under such circumstances two steps have to be taken:

- Step 1: Defining the churn
- Step 2: Constructing the Predictive Model

4.2.1. Step 1: Churn Definition

The major hurdle that we faced with prior to the model building phase was to giving a logical definition for churn. In almost all studies that we reviewed in the literature review phase, the customers of the service provider were its subscribers who had a contract with the company. Consequently, "Churn" in such conditions could be defined as the terminating the contract from customer's side or not renewing it after its expiry date, But circumstances would be different about pre-paid telecommunications service providers. In such companies there is no contract between the company and the clients. Anyone can simply purchase the SIM Card and become a user. On the other hand, any customers at any time can just stop using the provided services by the company, and become a churner without leaving an immediate trace. In other words churn in such cases happens with no tracking point such as terminating the contract or not renewing it and its recognition becomes complicated.

To shed light on the issue, imagine a data base of customers consisting a number of customers with different calling behaviors some of them use their cell phone every day, but the others use it every 2, 3,… or 20 days. Now if we define a churner as *"A person who has not used his/her cell phone for 7 days"* a considerable part of our customers who use their cell phone occasionally (i.e. every 8, 9,….., 20 days) would be considered as a "churner", mistakenly. On the other hand if we take a longer time span for prediction period and define a churner as *"A person who hasn't used his/her cell phone for 25 days"* our model may suffer from inability in recognizing the real churners.

The above discussed wrong signals would increase the number of False Negatives (FN) and False Positives (FP) and consequently lower the level of model's accuracy.

Tackling this problem, I tailored the RFM features defined by Ansari, Kohavi, Mason, & Zheng (2000) and constructed the following set of 12 RFM related variables, which has been extracted from customers' calling records data by the use of Microsoft SQL Server software, in order to segment the customer base, based on their calling behavior:

1. **Call Ratio:** proportion of calls which has been made by each customer with more than one day time distance to his/her total number of calls.

2. **Average Call Distance:** the average time distance between one's calls
3. **Max Date:** the last date in our observed time period in which a call has been made by a specific customer
4. **Min Date:** the first date in our observed time period in which a call has been made by a specific customer
5. **Life:** the period of time in our observed time span in which each customer has been active
6. **Max-Distance:** the maximum time distance between two calls of an specific person in our observed period
7. **No-of-days:** number of days in which a specific customer has made or received a call
8. **Total-no-in:** the total number of incoming calls for each client in our observed period
9. **Total-no-out:** the total number of outgoing calls for each client in our observed period
10. **Total-cost:** the total money that each customer has been charged for using the services in the specific time period under study
11. **Total-duration-in:** the total duration of incoming calls (in Sec) for a specific customer in our observed time span
12. **Total-duration-out:** the total duration of outgoing calls (in Sec) for a specific customer in our observed time span

By the use of TwoStep Cluster technique the customer based was divided into 6 individual clusters with the following specifications (see table 4.1):

Table 4.1: characteristics of 6 initially extracted clusters of customers

Cluster No.	**Call Ratio**		**Average Call Distance**		**Life**		**Max Distance**		**Max date**		**Min date**		**No of Days**		**Total No In**		**Total No out**		**Total cost**		**Total Duration In**		**Total Duration Out**	
	Mean	***Std***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***
1	0.039	0.194	0.0	0.0	88.777	6.229	3.874	5.3	90.447	5.148	1.67	3.072	84.398	12.474	1200.262	1130.186	663.854	475.644	558612.0	553774.7	131101.2	97157.98	55047.57	55745.66
2	0.139	0.53	0.0	0.0	89.317	5.507	3.874	4.266	91.03	4.536	1.713	3.082	84.039	11.003	411.191	236.888	350.891	138.659	253962.5	104169.2	39083.19	29434.37	24614.9	10211.00
3	3.972	6.31	0.07	0.281	27.049	19.628	5.662	6.562	54.641	30.327	27.592	27.985	19.523	15.378	66.087	111.277	43.526	67.199	33185.7	60918.67	6344.509	12506.34	3269.822	5855.495
4	7.898	6.401	0.071	0.257	85.16	8.274	16.328	12.521	88.655	6.933	3.496	5.287	46.625	13.227	90.647	92.204	55.81	55.508	40815.3	46030.1	9144.61	14972.46	3963.771	4482.45

5	31.885	13.452	3.309	4.023	65.052	22.806	24.88	17.192	76.414	18.499	11.361	15.409	14.801	10.905	11.806	12.293	7.822	8.521	5696.602	10269.78	788.183	995.136	639.885	1009.705
6	1.576	1.919	0.0	0.0	90.134	2.956	3.902	2.514	91.553	2.296	1.419	1.883	80.119	8.243	228.534	149.117	105.002	64.578	64265.7	43362.5	20492.08	19057.06	6193.124	4182.128

Among all utilized features in clustering phase, we found Max-Distance to be a suitable representative for normal usage frequency of the customers. This feature for each cluster addresses that, in what time distance the majority of cluster members uses their mobile phone and thus provides us with the regular usage behavior of customers. Consequently by having the routine manner of clients we would be able to spot any deviation from the standard and determine a definition for deviation with an acceptable approximation. Table 4.2 has summarized the Max-Distance feature for all extracted clusters:

Table 4.2: Average Max-Distance of each developed cluster

Cluster Label	Max-Distance
Cluster-1	3.87 Days
Cluster-2	3.87 Days
Cluster-3	5.66 Days
Cluster-4	16.33 Days
Cluster-5	24.88 Days
Cluster-6	3.90 Days

By looking at the extracted clusters from Max-Distance point of view we reach to 4 clusters out of these 6 clusters. So we redesigned the clusters as depicted in table 4.3, and by considering the "Prediction Period" as twice the Max-Distance, we defined the Prediction, Retention, and Observation as illustrated in table 4.4 for each cluster based on Wei & Chiu (2002)'s definition.

Table 4.3: Combining the 6 initially developed into 4 clusters based on Max-Distance measure

Old Cluster Label	New Cluster Name	Max-Distance
Cluster-1 **Cluster-2** **Cluster-6**	Cluster-1	3.88 Days
Cluster-3	Cluster-2	5.66 Days
Cluster-4	Cluster-3	16.33 Days
Cluster-5	Cluster-4	24.88 Days

Table 4.4: : Model building time periods for each cluster

Cluster Label	Max-Distance	Prediction Length	Retention Length	Observation Length
Cluster-1	3.88 days	7 days	7 days	30 days
Cluster-2	5.66 days	11 days	7 days	30 days
Cluster-3	16.33 days	32 days	7 days	30 days
Cluster-4	24.88 days	49 days	7 days	30 days

4.2.2. Step 2: Constructing the Predictive Model

As the second step, by considering two sub-periods of 15 days in the observation period (Wei & Chiu (2002) in their study tested different number of sub-periods and concluded that the predictive accuracy of the model would be in its highet level when the number of sub-periods is equal to two), we made the following features for every single cluster based on Wei & Chiu (2002)'s paper:

1. $\boldsymbol{IMOU_{initial}}$: Incoming MOU of a subscriber in the first sub-period
2. $\boldsymbol{IFOU_{initial}}$**:** Incoming FOU of a subscriber in the first sub-period
3. $\boldsymbol{OMOU_{initial}}$: Outgoing MOU of a subscriber in the first sub-period
4. $\boldsymbol{OFOU_{initial}}$**:** Outgoing FOU of a subscriber in the first sub-period
5. $\Delta\, \boldsymbol{IMOU_2}$**:** The change in Incoming MOU of a subscriber between the sub-period 1 and 2 and is measured by $\Delta IMOU_2 = (IMOU_2 - IMOU_1 + 0.01)/(IMOU_1 + 0.01)$ where $IMOU_1 = IMOU_{initial}$.
6. $\Delta \boldsymbol{IFOU_2}$**:** The change in Incoming FOU of a subscriber between the sub-period 1 and 2 and is calculated as $\Delta IFOU_2 = (IFOU_2 - IFOU_1 + 0.01)/(IFOU_1 + 0.01)$ where $IFOU_1 = IFOU_{initial}$.
7. $\Delta\, \boldsymbol{OMOU_2}$**:** The change in Outgoing MOU of a subscriber between the sub-period 1 and 2 and is measured by $\Delta OMOU_2 = (OMOU_2 - OMOU_1 + 0.01)/(OMOU_1 + 0.01)$ where $OMOU_1 = OMOU_{initial}$.
8. $\Delta \boldsymbol{OFOU_2}$**:** The change in Outgoing FOU of a subscriber between the sub-period 1 and 2 and is calculated as $\Delta OFOU_2 = (OFOU_2 - OFOU_1 + 0.01)/(OFOU_1 + 0.01)$ where $OFOU_1 = OFOU_{initial}$.
9. **Churn:** binary churn labels for each client according to their churn status in prediction period

Using the first 8 features as the input features to the tree, the "churn" feature as the output of the tree, and the hit ratio= 1:2 (churner : non-churner), we built different predictive models for each cluster. At first, based on 75% of data (Training Dataset) and by utilizing Decision Tree (CART algorithm) the predictive models were constructed on each of our four developed clusters.

Table 4.5 depicts the performance of the developed churn predictive models for each cluster based on Gain measure.

Table 4.5: Performance of developed predictive models based on Gain measure

Cluster No.	%Gain for percentile = 10	%Gain for percentile = 20
1	46.3	77.8
2	54	66.7
3	15	30
4	17	45.5

Furthermore due to the nature of our data set which suffers from class imbalance, we tested the effect of cost-sensitive learning methods on the performance of our developed models. The results as are presented in table 4.6 confirm the positive effect of cost-sensitive learning on the performance of the models.

Table 4.6: Performance of Cost-sensitive predictive models based on Gain measure

Cluster No.	%Gain for percentile = 10	%Gain for percentile = 20
1	56.1	77.8
2	54	66.7
3	15	30
4	36	63

Figures 4.1 to 4.4 illustrate the performance of simple and cost-sensitive developed models for each of four clusters.

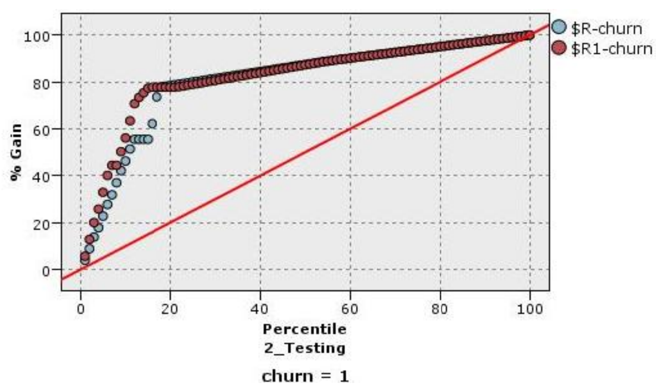

Figure 4.1: Gain chart of simple (blue points) and cost-sensitive (red points) models for cluater.1

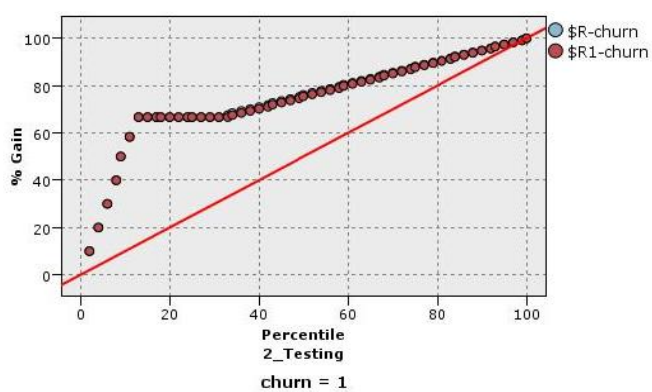

Figure 4.2: Gain chart of simple (blue points) and cost-sensitive (red points) models for cluater.2

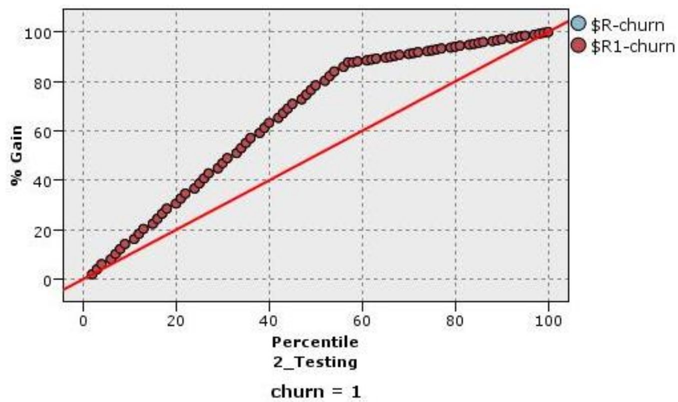

Figure 4.3: Gain chart of simple (blue points) and cost-sensitive (red points) models for cluater.2

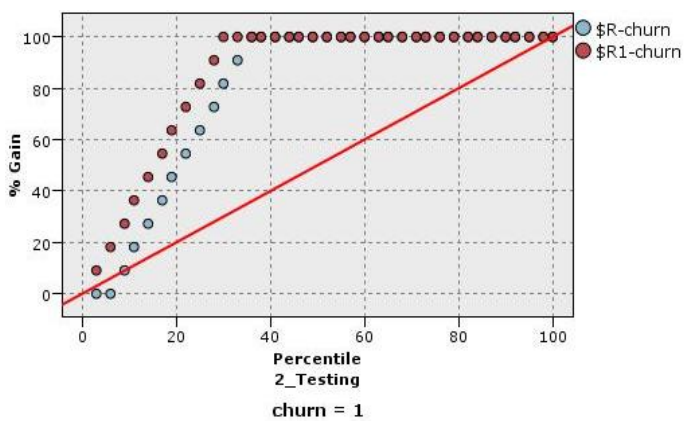

Figure 4.4: Gain chart of simple (blue points) and cost-sensitive (red points) models for cluater.2

Regarding the above figures both simple and cost-sensitive predictive models have a considerable better performance than the random sampling (diagonal line). Additionally, the cost-sensitive learning method has been proven to have contribution in model building with imbalanced data and it outperforms the simple model (see figure 4.1 and figure 4.4).

Tables 4.7 to 4.10 illustrate the accuracy of the cost-sensitive learnt developed model by utilizing CART algorithm.

Table 4.7: The accuracy measure of revised predictive model for cluster 1

Correct	4556	84.43%
Wrong	823	15.57%
Total	5379	

Table 4.8: The accuracy measure of revised predictive model for cluster 2

Correct	563	87.5%
Wrong	81	12.5%
Total	644	

Table 4.9: The accuracy measure of revised predictive model for cluster 3

Correct	800	91.03%
Wrong	79	8.97%
Total	879	

Table 4.10: The accuracy measure of revised predictive model for cluster 4

Correct	299	72.22%
Wrong	115	27.78%
Total	414	

Afterwards, in a comparative approach, the Neural Networks technique and also different algorithms of the Decision Tree technique were utilized to find the algorithm with the significant performance for model building on our dataset. With this regard CART, C5.0, and CHAID algorithms among Decisions Tree algorithms were applied and their performance was compared with the Neural Networks based constructed model. After constructing the model by the use of training dataset we applied the constructed model on the remaining 25% of data (Testing Dataset) with the aim of validating the model. Our adopted validation method (Single split Model Validation), has been proven to be an accurate validation method (Montgomery, Li, Srinivasan, & Liechty, 2004; Swait & Andrews, 2003; Burez & Van den Poel, 2009).

Tables 4.11 to 4.14 represent the performance of our constructed models with different algorithms on our clusters based on gain measure for top %10 and %20 clients of each cluster.

Table 4.11: Performance of developed Decision Tree (C5.0) predictive models based on Gain measure

Cluster No	%Gain for percentile = 10	%Gain for percentile = 20
1	77.9	80.4
2	10	29.6
3	15	30.6
4	22.5	63.6

Table 4.12: Performance of developed Decision Tree (CHAID) predictive models based on Gain measure

Cluster No	%Gain for percentile = 10	%Gain for percentile = 20
1	59	79.4
2	36	67.4
3	15	30.6
4	40	80

Table 4.13: Performance of developed Decision Tree (CART) predictive models based on Gain measure

Cluster No	%Gain for percentile = 10	%Gain for percentile = 20
1	46.3	77.8
2	54	66.7
3	15	30
4	17	45.5

Table 4.14: Performance of Neural Networks predictive models based on Gain measure

Cluster No	%Gain for percentile = 10	%Gain for percentile = 20
1	72.2	77.8
2	33.3	33.3
3	15	30.6
4	25	60

Comparing the performance of our developed models based on gain measure, one can find that Decision Tree algorithms outperform the Neural Networks algorithm. Furthermore, examining the gain which has been brought by each of the Decision Tree based constructed models, one can find that maximum performance will be gained by utilizing different algorithms of Decision Tree technique for different clusters. Table 4.15 depicts the most appropriate algorithm among the tested algorithms, for model building in each cluster.

Table 4.15: The Appropriate Algorithm for Model Building in Each Cluster

Cluster No.	Appropriate Algorithm
Cluster 1	Decision Tree C5.0 Algorithm
Cluster2	Decision Tree CART Algorithm
Cluster3	All Tested Algorithms
Cluster 4	Decision Tree CHAID Algorithm

Consequently, by applying this multi-algorithm approach the gain factor for each cluster would be in accordance with table 4.16.

Table 4.16: Performance of the Multi-algorithm Model Building Approach on Our Developed Clusters Based on Gain Measure

Cluster No	%Gain for percentile = 10	%Gain for percentile = 20
1	77.9	80.4
2	54	66.7
3	15	30.6
4	40	80

While the gain factor of random sampling is %20 for the top %20 of the customer base in all clusters, table 7 depicts that the developed model is able to bring the gain factor of %80.4, %66.7, %30.6, and %80 for the top %20 of the customer base of our four developed clusters, respectively. This implies that by applying the developed multi algorithm predictive model, choosing a sample size of only %20 of each cluster's customer base is enough for identifying %80.4, %66.7, %30.6, and %80 of the total number of churners in each of our four clusters, respectively.

Figures 4.5 to 4.8 illustrate the gain diagram of each developed model and as it can be understood from the figures all developed models have considerable better performance than the random sampling (diagonal line).

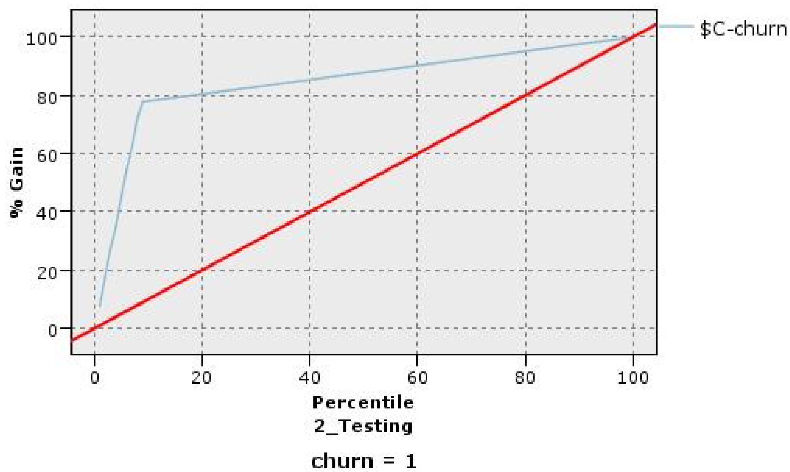

Figure 4.5: Gain chart of simple learnt Decision Tree C5.0 algorithm for cluster 1

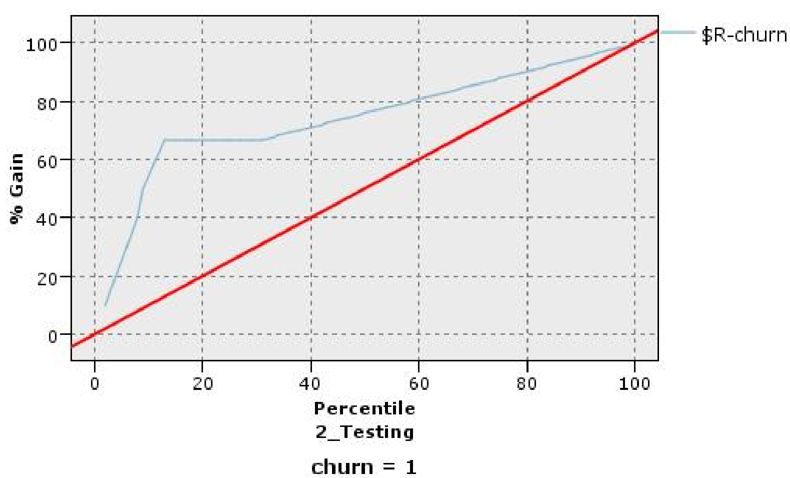

Figure 4.6: Gain chart of simple learnt Decision Tree CART algorithm for cluster 2

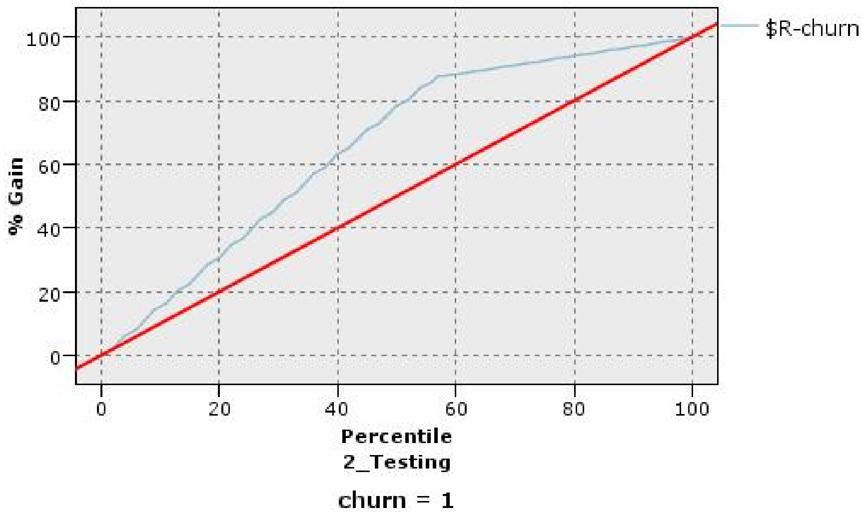

Figure 4.7: Gain chart of simple learnt Decision Tree CART algorithm for cluster 3

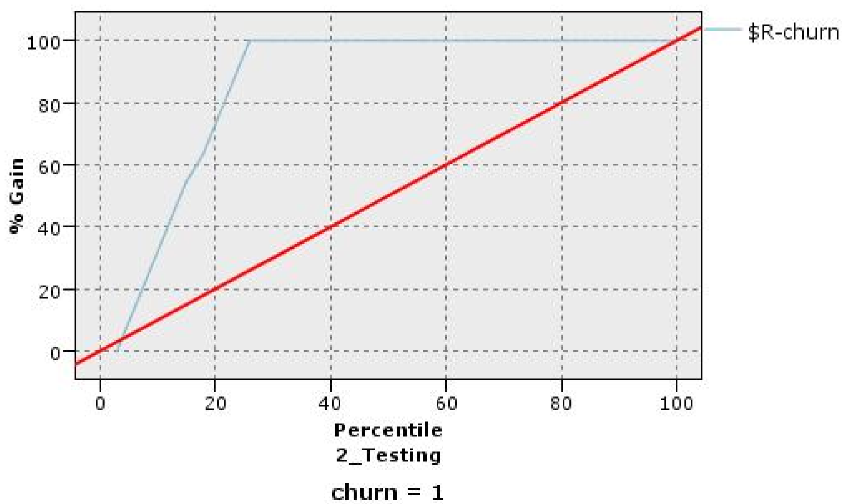

Figure 4.8: Gain chart of simple learnt Decision Tree CHAID algorithm for cluster 4

The accuracy of our multi-algorithm model on each of four developed clusters has been presented in tables 4.17 to 4.20.

Table 4.17: The accuracy measure of multi-algorithm predictive model for cluster 1

Correct	5016	92.96%
Wrong	380	7.04%
Total	5396	

Table 4.18: The accuracy measure of multi-algorithm predictive model for cluster 2

Correct	506	78.57%
Wrong	138	21.43%
Total	644	

Table 4.19: The accuracy measure of multi-algorithm predictive model for cluster 3

Correct	816	91.03%
Wrong	81	8.97%
Total	897	

Table 4.20: The accuracy measure of multi-algorithm predictive model for cluster 4

Correct	425	94.87%
Wrong	23	5.13%
Total	448	

With the intention of handling the class imbalance and similar to our single algorithm approach we tested the effect of cost-sensitive learning method on the performance of our developed models on each cluster and surprisingly we found out that this remedy has negative or no effect on the performance of our multi-algorithm model.

Chapter 5 Conclusion and Further Research

5.1. Introduction

In the previous chapters after addressing the purpose of the research shedding light into its importance and magnification, and reviewing its existing literature, we hired our methodology for conducting the current study. By putting the extracted methodology into practice we performed our analysis phase which results are presented in chapter 4. In the current chapter we have discussed the outcome of our research and derived the proper interpretations from them. Furthermore we have shared the limitations that we faced with in the way of conducting this research and also the implications of this research for businesses especially in telecommunications market segment and ultimately we proposed the research gaps in this area which can be filled by future researches.

5.2. Conclusion

The problem that Talia Telecommunications Co. was dealing with was to recognize the customers with high probability of churn in close future and target them with incentives in order to convince them to stay, but due to the absence of an accurate model for monitoring their clients' behavior, the company was unable to distinguish the churners from non-churners. In such case, the company had two ways; whether to send all customers the incentives, which was clearly the waste of money or to quit the churn management program and focus on acquisition program which is considerably more costly than the retention approach.

Under such circumstances the company decided to find a way in order to distinguish the churners from non-churners. So that the company becomes able to target the right person with the incentives. In this case, not only the model helps the company to distinguish the real churners, but also it prevents the waste of money due to the mass marketing.

The churn predictive modeling is always formulated as a binary classification modeling and customers are divided into two groups of churners or non-churners. There exist different data mining techniques, such as Decision tree, Random Forests, and Neural Networks which have been utilized by experts to construct the predictive model, and due to the interpretability and more understandable logic of Decision Tree we chose this technique for our model building.

As it has been addressed in previous chapters the research process of Knowledge Discovery in Databases consists of different steps which include collecting the data, selecting the data, processing the data, transforming the data, mining the data, and evaluating the results.

The data was collected from Talia Co. data base with the approximate size of 1.5 GB in the form of three txt files which contained 19500504 transactional records, produced by 34523 customers, in the time period between 1 November 2007 and 31 January 2008. All tree files imported as one single database into Microsoft SQL Server. Due to the fact that the data was produced by machine there were no missing values in our data set and so the other preprocessing operations of data such as dat cleaning were unnecessary.

In transformation step we constructed RFM related features by the use of the raw transactional data for each customer in the time period between 1 November 2007 and 31 January 2008, and we used the extracted features in clustering the customer base, while the second group of features (the ones that have been used in model building phase) we considered the observation period for feature construction, which was extracted after clustering phase for each single cluster, individually.

As it has been mentioned before the whole process of model building consisted of two steps. In the first step we clustered the customer base, in order to come to a reasonable definition of churn.

After achieving the insight toward the churn definition, in the second step we constructed the churn predictive model for each cluster which enabled us to spot the future churners based on their prior calling behavior.

With the intention of building the predictive model for each cluster I divided the customer base of each cluster into two data sets: Train (70%) and Test (30%) and by considering the hit ratio of 2:1 (non-churner : churner) based on Wei & Chiu (2002)'s study I utilized the under sampling in order to handle the data imbalance and then conducted the model building phase. With this regard a decision tree were grown for each cluster. The trees depict that the significant features for churn prediction are different form a cluster to another. (See table 5.1)

Table 5.1: The most significant features in building the predictive model for each cluster

Cluster Number	Determinant features
1	$\Delta IFOU_2$, $\Delta\ OMOU_2$, $IMOU_{initial}$
2	$IFOU_{initial}$, $OFOU_{initial}$, $\Delta\ IMOU_2$
3	$IFOU_{initial}$, $OFOU_{initial}$, $\Delta IFOU_2$
4	$OFOU_{initial}$, $\Delta\ IMOU_2$

The performance of the Decision Tree models was evaluated by the use of Lift / Gain Chart. The gain chart was created the developed model of each cluster. With this regard, it sorts the customers based on likelihood to churn in the descending order.

The lift/gain chart showed that if the top 20% of customers in cluster 1, 2, 3, and 4 would be extracted as a sample of customers for sending the incentives, 77.8%, 66.8%, 30%, and 45.5% of the real churners in each cluster would be targeted, respectively, while this gain measure was 20% for all clusters in random sampling approach. Furthermore we adopted the cost-sensitive learning approach as another tools of handling the data imbalance and its promising outcome raised the gain measure in the first and the fourth cluster.

Consequently the developed models are considerably able to distinguish the churners form non-churners and help the Talia Co. to conduct a more efficient retention campaign. In fact by utilizing this approach the Company would be able to not only reduce the marketing cost and churns rate simultaneously.

5.3. Research Limitations

This research, like all other researches, was not without its limitations. The follow are some of the limitations that we faced with during our research:

- One the major limitations of this research was data classification and data confidentiality in Talia Co. that prevented us to have access to a part of customers data such as billing and credit data. This forced us to calculate the monetary features manually and deprived us form involving the credit features into our model building.
- Lack of demographic data of customers was also our other limitation in conducting this research. Due to this we were unable to involve such factors in our clustering phase which was probably able to improve the accuracy and also interpretability of clusters.

5.4. Managerial implications

The finding of this research has important application for companies that are active in mobile telecommunications market (Especially the pre-paid ones). Besides the idea of developing the dual-step model for extracting the churn definition prior to model building phase, can also be applied in baking industry in regard with building a predictive model for Debit Card’s customers’ churn.

Managing the great deal of data produced by customers in companies and organizations, can provide them with precious knowledge regarding their customers which can be exploited in developing new products, conducting retention campaigns, and also in cross selling and up-selling the products and services of a company. This demonstrates the significance of application of data mining in marketing. In fact mining the raw data produced by customers in their touch points with company can provide the company with a better insight toward their customers which helps them to conduct more efficient and also more effective marketing investments.

The objective of this research was to develop a predictive model for customer churn in pre-paid mobile telephony market which is able to distinguish between customers who are likely to churn in close futures and the ones who are stuck with the company. The contribution of such model for the company is that it would prevent the waste of money due to the mass marketing approaches and it enables the companies to target the real churners by extracting the customers with high probability of churn. Besides, as discussed in chapter 1 the cost of acquiring a new customer is 8 times more than retaining an existing one, thus since the churn predictive model is capable of indicating the future churners, the companies that are intended to maintain their customer base can focus on retention approaches instead of acquisition approaches which is clearly less costly.

Summing up the above discussion, in regard with the finding of this research, we can suggest the companies to utilize the data mining techniques in order to transform the existing customer data in their databases to exploitable knowledge that can help them in their marketing plans. Moreover, it would be beneficial for them to build a predictive churn model by the use of data mining which plays the role of an alerting system for the companies and also it can help them to spend their retention budget efficiently.

5.5. Suggestions for Further Research

After conducting the current research, still some interesting areas exist that worth to be worked on. Additionally the limitations of this study can provide us with the ideas for future researches:

The followings are the further studies which can be done in the realm of current research:

- Developing our model we utilized the TwoStep Cluster in the 1^{st} step as our clustering technique and Decision Tree classifier as the classification technique for building the predictive model. It was not our intention to compare the performance of different clustering techniques and different classifiers in this research, but for further research it can be suggested to apply different classification or clustering algorithms and compare the outcome of them.
- Handling the data imbalance we applied two of methods proposed by Weiss (2004) (i.e. under sampling and cost-sensitive learning) and we found the results promising. Further research is suggested to utilize the other approaches of handling the data imbalance proposed by Weiss (2004) and test the applicability of them.
- Since the data base of Talia was suffering from lack of demographic information of customers, we were unable to use such features in our model building phase. It is suggested to conduct the current dual-step modeling for churn prediction on a data base with demographic features and measure the effect of demographic variables involvement in clustering and classification accuracy.

References

Adomavicius, G., & Tuzhilin, A. (2003). *Recommendation technologies: Survey of current methods and possible extensions.* Working paper, Stern School of Business,New York University, New York.

Ahmad, S. (2004). Applications of data mining in retail business. *Information Technology , 2*, 455-459.

Ahn, J., Han, S., & Lee, Y. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy , 30*, 552-568.

Ansari, S., Kohavi, R., Mason, L., & Zheng, Z. (2000). Integrating e-commerce and data mining: architecture and challenges. *WEBKDD'2000 Workshop on Web Mining for E-Cpmmerce Challenges and Opportunities* .

Au, W., Chan, K., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transaction on Evolutionary Computation , 7* (6), 532-545.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., & Dedene, G. (2002). Bayesian neural network learning for repeat purchase modeling in direct marketing. *European Journal of Operational Research , 138* (1), 191-211.

Berry, M., & Linoff, G. (2004). *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management* (2nd Edition ed.). Indianapolis: Wiley Publishing Inc.

Berson, A., Smith, S., & Therling, K. (1999). *Building data mining applications for CRM.* New York: McGraw-Hill.

Bhattacharya, C. (1998). When customers are members: customer retention in paid membership contexts. *Journal of the Academy of MArketing Science , 26*, 31-44.

Bhattacharyya, S., & Pendharkar, P. (1998). Inductive, evolutionary and neural techniques for discrimination: Acomparative study. *Decision Sciences , 29*, 871-900.

Bose, R. (2002). Customer relationship management: key components for IT success. *Industrial Management & Data Systems , 102* (2), 89-97.

Buckinx, W., Moons, E., Van Den Poel, D., & Wets, G. (2004). Customer-adapted coupontargeting using feature selection. *Expert Systems with Applications , 26*, 509-518.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert System with Applications , 36*, 4626-4636.

Burez, J., & Van Den Poel, D. (2008). Separating financial from commercial customer churn: A modelingstep towards resolving the conflict between the sales and credit department. *Expert Systems with Applications , 35*, 497-514.

Carrier, C., & Povel, O. (2003). Characterising data mining software. *Intelligent Data Analysis , 7*, 181-192.

Chandar, M., Laha, A., & Krishna, P. (2006). Modeling churn behavior of bank customers using predictive data mining techniques. *National conference on soft computing techniques for engineering applications (SCT-2006).*

Chen, Y., Hsu, C., & Chou, S. (2003). Constructing a multi-valued and multi-labeled decision tree. *Expert Systems with Applications , 25*, 199-209.

Chu, B., Tsai, M., & Ho, C. (2007). Toward a hybrid data mining model for customer retention. *Knowledge-Based Syetems , 20*, 703-718.

Coussement, K., & Van den Poel, D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications , 36*, 6127-6134.

Coussement, K., & Van den Poel, D. (2008). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management , 45*, 164-174.

Deville, J., & Tille, Y. (2004). Efficient balanced sampling: the cube method. *Biometrika , 91*, 893-912.

Fader, P., Hardie, B., & Lee, K. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research , 62*, 415-430.

Freeman, M. (1999). The 2 customer lifecycles. *Intelligent Enterprise , 2* (16), 9.

Gerpott, T., Rams, W., & Schindler, A. (2001). Customer retention,loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy , 25*, 249-269.

Gomory, S., Hoch, R., Lee, J., Podlaseck, M., & Schonberg, E. (1999). *E-commerce intelligenece:measuring, analyzing and reporting on merchandising effectiveness of online stores.* IBM Research Report, NY.

Groth, R. (1999). *Data Mining:Building Competitive Advantage,.* Santa Clara, CA: Prentice Hall.

Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques* (2nd Edn ed.). San Francisco: Elsevier Inc.

He, Z., Xu, X., Huang, J., & Deng, S. (2004). Mining class outliers:Concepts, algorithms and applications in CRM. *Expert systems with applications , 27*, 681-697.

Hung, S., Yen, D., & Wang, H. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications , 31*, 515-524.

Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Exper Systems with Applications , 26*, 181-188.

Jones, T., & Sasser, W. (1995, Nov-Dec). Why satisfied customers defect? *Harvard Business Review* , 88-99.

Kim, H., & Yoon, C. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy , 28*, 751-765.

Kim, J., Song, H., Kim, T., & Kim, H. (2005). Detecting the change of customer behavior based on decision tree analysis. *Expert Systems with Applications , 22*, 193-205.

Kincaid, J. (2003). *Customer Relationship Management: Getting it Right.* NJ: Pretice-Hall PTR.

Komenar, M. (1997). *Electronic Marketing.* New York: Wiley Computer Publishing.

Kotler, P., & Keller, L. (2006). *Marketing Management* (12th edn ed.). New Jersey: Pearson Prentice Hall.

Lee, S., & Siau, K. (2001). A review of data mining techniques. *Industrial Management and Data Systems , 101* (1), 41-46.

Lejeune, M. A. (2001). Measuring the impact of data mining on churn management. *Internet Research: Electronic Networking Applications and Policy , 11* (5), 375-387.

Li, S. (1995). Survival analysis. *Marketing Research , 7*, 17-23.

Li, T. & Ruan, D. (2007). An extended process model of knowledge discovery in databases. *Journal of Enterprise Information Management.* 20 (2), 169-177.

Ling, C., & Sheng, V. (2008). Cost-sensitive learning and the class imbalance problem.Inc.Sammut, Encyclopedia of Machine Learning. Springer.

Ling, R., & Yen, D. (2001). Customer relationship management: An analysis framework and implementation strategies. *Journal of Computer Information Syetems , 41* (3), 82-97.

Madden, G., Savage, S., & Coble-Neal, G. (1999). Subscriber churn in the Australian ISP market. *Information Economics and Policy , 11*, 195-207.

Malhotra, K. (2007). *Marketing research -An applied orientation* (5th Edn ed.). New Jersey: Pearson Education.

Montgomery, A., Li, S., Srinivasan, K., & Liechty, J. (2004). Modeling online browsing and path analysis using click stream data. *Marketing Science , 23* (4), 579-595.

Neslin, S., Gupta, S., Kamakura, W., Lu, J., & Mason, C. (2006). Defection Detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research , XLIII*, 204-211.

Ngai, E. (2005). Customer relationship management reserach (1992-2002): An academic literature review and classification. *Marketing Intelligence & Planning , 23* (6), 582-605.

Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications , 36*, 2592-2602.

Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research , 187*, 1429-1448.

Parvatiyar, A., & Sheth, J. (2001). Customer relationship management: emerging practice, process and discipline. *Journal of Economic & Social Research , 3* (2), 1-34.

Pendharkar, P. (2009). Genetic algorithm based neural network approaches for predicting churn in cellular wireless networks service. *Expert Systems with Applications , 36*, 6714-6720.

Peppard, J. (2000). Customer relationship management (CRM) in financial services. *European Management Journal , 18* (3), 312-27.

Reichheld, F. (1996). *The loyalty effect: The hidden force behind growth,profits and lasting value.* Harvard Business School Press.

Reichheld, F., & Sasser, W. (1990). Zero defection: Quality comes to services. *Hravard Business Review , 68* (5), 105-111.

Reinartz, W., Thomas, J., & Kumar, V. (2005). Balancing acqusition and retention resources to maximize profitability. *Journal of Marketing , 69* (1), 63-79.

Richards, K. A., & Jones, E. (2008). Customer relationship management: Finding value drivers. *Industrial Marketing Management , 37*, 120-130.

Rust, R., & Zahorik, A. (1993). Customer Satisfaction, Customer Retention, and Market share. *Journal of Retailing , 69*, 193-215.

Ryals, L. (2005). Making customer relationship management work: The measurement and profitable management of customer relationships. *Journal of Marketing , 69* (4), 252-261.

Rygielski, C., Wang, J., & Yen, D. (2002). Data mining techniques for customer relationship management. *Technology in Society , 24* (4), 483-502.

SAS Institute. (2000). *Best practice in churn prediction.* A SAS Institute White Paper.

Saunders, M., Lewis, P., & Thornhill, A. (2000). *Research methods for business students* (2nd Edn ed.). Prentice Hall.

Schindler, C. (2003). *Business Research Methods.* New York: Mc Graw-Hill.

Seo, D., Ranganathan, C., & Babad, Y. (2008). Two-level model of customer retention in the US mobile telecommunications service market. *Teleccommunications Policy , 32*, 182-196.

Shah, T. (1996). Putting a quality edge to digital wireless networks. *Cellular Business , 13*, 82-90.

Shaw, M., Subramaniam, C., Tan, G., & Welge, M. (2001). Knowledge management and data mining for marketing. *Decision Support Systems , 31* (1), 127-137.

SPSS Inc. (2007). *Clementine 11.1 Algorithms Guide.* Integral Solutions Limited.

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. (2000). Web usage mining:discovery and applications of usage patterns from web data. *SIGKDD Explorations , 1* (2), 12-23.

Swait, J., & Andrews, R. (2003). Enriching scanner panel models with choice experiments. *Marketing Science , 22* (4), 442-460.

Swift, R. (2001). *Accelerating Customer Relationships Using CRM and Relationship Technologies.* NJ: Prentice-Hall PTR.

Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining.* Boston: Pearson Education.

Teo, T., Devadoss, P., & Pan, S. (2006). Towards a holistic perspective of customer relationship management implementation: A case studyof the housing and development board singapore. *Decision Support Systems , 42*, 1613-1627.

Van den Poel, D., & Larivie're, B. (2004). Customer attritionan analysis for financial service using proportional hazard models. *European Journal of Operational Research , 157* (1), 196-217.

Wei, C., & Chiu, I. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications , 23*, 103-112.

Weiss, G. (2004). Mining with rarity: A unifying framework. *SIGKDD Explorations , 6* (1), 7-19.

Wells, J., Fuerst, W., & Choobineh, J. (1999). Managing information technology (IT) for one-to-one custoner interaction. *Information & Management , 35*, 53-62.

West, J. (2001). Customer relationship management and you. *IIE Solutions , 33* (4), 34-37.

Xie, Y., Li, X., Ngai, E., & Ying, W. (2008). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications , 36*, 5445-5449.

Xu, Y., Yen, D., Lin, B., & Chou, D. (2002). Adopting customer relationship management technology. *Industrial Managemenet and Data Systems , 102* (8/9), 442-452.

Yan, L., Fassino, M., & Baldasare, P. (2005). Predicting Customer Behavior via Calling Links. *Proceedings of International Joint Conference on Neural NEtworks*, (pp. 2555-2560). Montreal.

Zeithaml, V., Leonard, L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing , 60*, 31-46.

Zhao, Y., Li, B., Li, X., Liu, W., & Ren, S. (2005). Customer churn prediction using improved one-class support vector machine. *Lecture Notes in Computer Science , 3584*, 300-306.