MASTER'S THESIS

Customer Segmentation and Strategy Definition in Segments

Case Study: An Internet Service Provider in Iran

Rana Soudagar

Master of Arts (120 credits)
Business Administration

Luleå University of Technology
Department of Business Administration, Technology and Social Sciences

Customer Segmentation and Strategy Definition in Segments

Case Study: An Internet Service Provider in IRAN

By:

Rana Soudagar

Supervisor

Dr. Keramati

Advisor

Dr. Limayem

Lulea University of technology

Department of Business Administration and Social Sciences

Division of Industrial marketing and e-commerce

Abstract

Maintaining customer relationships is a key to business success in today's competitive environment. But all markets contain many subgroups of customers that behave differently, have different hopes, fears and ambitions, and have different purchasing behaviors. So, each subgroup must be behaved differently in order to build these relationships. On the road to this goal, customer segmentation is the first step.

The goal of a segmentation system is to identify groups in which the customers are as much alike as possible and greatly differentiated from customers in other segments. If the segmentation system is well designed, members of a segment have similar interests, attitudes and behaviors, and they will respond similarly to elements of the marketing mix such as pricing, promotion and sales channel. Properly developed, segmentation insights inform a strategic roadmap intended to take advantage of key profit driving opportunities within each unique customer group. This could be shortening customer purchase cycles, driving higher spend, building greater customer loyalty, deepening cross-product penetration or lowering service and support costs.

Internet service providers are one of the most active companies in today's business world and a key element of development. They provide various services and products for their customers who are growing so rapidly. The number of their customers is different in countries depending on the level of development of countries. But it can be said that in a close future, almost all of the people will be customers of Internet service providers. Furthermore, in today's world, where the market is highly competitive, customers face with various providers with different marketing strategies. These companies can be successful in the competitive environment by customer segmentation and designing proper strategies in each segment.

The goal of this project is to mine the customer data to perform customer segmentation and consequently defining proper and useful strategies to win in the competitive environment.

In this study Recency, Frequency and Monetary method which also known as RFM method has been used for customer segmentation in an Iranian internet service provider. By definition of some new variables in RFM method, two new RFM variant methods have been proposed which have some advantages with respect to simple RFM model. The results of applying these new methods show their effectiveness for customer segmentation and also their ability in identification of customer behaviors especially the risk of cancelling company services.

Keywords: Customer segmentation, RFM model, K-means clustering algorithm, EM clustering algorithm, Generalized Differential RFM method (GDRFM).

Table of Contents

List of tables

List of figures:

Chapter1: Introduction

Background of the study

Problem definition

Purpose of this study

Research questions

Research motivation

Research demarcation

Research outline

1.1Background of the study

Customers are regarded as important strategic resources of an enterprise, and gaining and retention of customers has become the most critical factor of an enterprise's success (Lai, 2009). By gaining an overall understanding of customers and then grouping them into categories, companies are able to better optimize marketing programs, satisfy customers and increase profits (Chen Y. and Li, 2009). Hence, for a company facing competitive environment, achieving efficient customer segmentation for applying high quality recommendation strategies is a key task. Traditionally, customer segmentation is achieved using statistics-based methods that compute a set of statistical measures from the customer data and then group customers into some segments by applying clustering algorithms in the space of these statistics (Jiang T. and Tuzhilin A., 2006).Customer segmentation is so common in real life. For instance, many business entities differentiate their customers by members and non-members. Also, many enterprises provide different service levels for different classes of customers. For example, customers can be divided into a couple of classes. The customers who pay more expensive shipping fee receive orders quicker than those who pay less expensive shipping fee (Chen Y. and Li, 2009).

Customer segmentation can effectively lower the marketing costs of a company and help it achieve more visible and profitable market penetration (Lai, 2009).It allows companies to design and establish different strategies to maximize the value of customers. (Cao et al, 2010).

The world around us changes continuously. For an internet service provider (ISP), knowledge about what is changing and how it has been changed is also essential. One of the most aspects of customer segmentation for an ISP is constructing an efficient strategy of behaving customers. Furthermore, in today's world where the market is highly competitive, customers face with various providers with different market strategies. In such a situation, managers must be aware of customer behaviors and customer situation in their segment. In such a market, it is necessary to mine customer data to reach this goal. But the most important key in the way of success in competitive situation is definition of proper strategies to interact with different customer groups appropriately. A little number of ISP companies in Iran have been tried to mine their customers' information by traditional ways. But there is no complete and comprehensive research or at least a published report on the application of this valuable and important issue in Iran.

1.2Problem Definition

In IRAN, approximately one third of population is using Internet services. There are so many companies which provide different internet services with different technologies in Iran and here we call all of them Internet Service Provider or ISP. Based on elementary survey, these companies have not the complete information about their customers and as a result there is no reported study on the application of customer segmentation in ISP companies of IRAN. The lack of such a study causes none efficiency and many shortages and problems for ISPs. Since they haven’t a clear view of their customers, they couldn’t adopt proper strategies and actions to gain competitive advantages in the market. They waste so much of their company resources and profit because they behave with all of their customers the same. One of the most well known problems in internet service providers is the fact that many of customers change their service providers frequently and so the companies have many churn customers. Certainly one of the main reasons of this phenomenon is the lack of different predefined strategies for different customer groups and also lack of customer segmentation in ISPs. The goal of this thesis is the application of customer segmentation in an ISP and providing different strategies in each segment using the gained results.

ATINET Company is the first ISP in Hamadan province. It works on different internet service categories such as Dial up, ADSL, wireless, broadband and etc. Now, the company is faced with the challenge of increasing competitions. There are various reasons behind it. First of all, according to the high demand for internet, every day services with higher speeds are required and requirements of users increase exponentially. In this situation, nobody knows about next year’s technology and service. This fast growth of internet enforces companies to switch to other services rapidly. In this situation, ISPs face the challenge of constantly evolving market where customer needs are changing all the time. Also, there are some powerful companies that make competition tighter for ATINET. So in such a market, the customer segmentation can help company to find some strategies to win the competition in this situation. Also, ATINET Company requires improving the customer satisfaction in order to improve the competitiveness to face these challenges. These goals can be reached by a set of actions that the first action among them is customer segmentation and definition of strategy for each segment.

1.3 Purpose of this study

A deeper understanding of customers has validated the value of focusing on them. It is now generally accepted that it costs about five times more to gain a new customer than to keep an existing one, and ten times more to get a dissatisfied customer back (Marcus C., 1998). Studies across numerous industries have also shown that a five-point increase in customer retention can increase profits by more than 25 percent (Marcus C., 1998).By looking more closely to these statistics, it is no wonder that managers are considering marketing as a powerful tool for their enterprises more and more. It is expected that the overall market for software and services using data mining technology will grow. By considering this fast growth of data mining technology and database marketing applications such as customer segmentation are taken into consideration.

According to (Lai, 2009) by analyzing traditional methods of customer segmentation, we can see that customer segmentation methods based on data mining are more advantageous in the following regards:

- The results of segmentation based on data mining are decided by the objectivity of the data, the subjectivity of the people who are processing them are avoided, resulting a more objective representation of the differences among different populations.
- It represents the categorization features among different customer categories more comprehensively, which facilitates marketing staff know their customers more thoroughly and in turn make more targeted and individualized marketing plans.
- The changes of customer behaviors can be tracked more easily by collocating clustering analysis models and updating the categorization of customers regularly.

In this study customer segmentation process is implemented to segment the customers and define some strategies for them. In order to reach this goal we need to find customer information and collect data in database. We need to collect as much data as possible about interactions between customers and the business, analyze this data to turn it into information and finally learn from it and take action (Bo¨ ttcher et al, 2009).This process is supported by techniques from data mining. As one of the most important techniques of data mining, clustering analysis is arisen method in customer segmentation. It aims to recognize a set of clustering rules and group the customers into several clusters. (Cao Et al, 2010). Nowadays, clustering analysis in the field of customer segmentation includes algorithms such as partitioned clustering, density-based clustering, grid-based clustering, fuzzy clustering and hierarchical clustering (Cao et al, 2010).

Based on the above consideration we can see that by analyzing the information obtained from the segmentation of customer behaviors, a company can provide its customers with products and services truly needed by them and also it can perform best efforts in order to maximize its customer retention and profitability. The purpose of this study is to apply customer segmentation method for an internet service provider in Iran and after that definition of proper strategies per segment. For doing so, some customer segmentation models which are suitable and applicable for out test case must be analyzed.

1.4 Research Questions

Based on problem discussed above, the purpose of study is to segment the customers of a company into some useful segments. In order to reach this goal, the research questions are as follows:

- Which factors are most important in segmentation?
- Which customer segmentation model is suitable for analysis?
- Which clustering algorithm must be selected?
- How many segments must be considered for segmentation?
- Which strategies must be defined for obtained segments?

1.5 Research Motivation

At present, market competition is becoming more and more drastic and products are more and more similar in quality. So, businesses have changed from product-driven to customer-driven (xiao-bin Zh. et al, 2009). Also, the customer need is changing all times. In this regards, it is important for business to know these changes and respond to these changes all the time. If business could not respond to customer needs, it will lose its customers. Today, customer segmentation can be used to solve this problem. Segmentation is a fundamental strategy to managing marketing efforts directed at customers (Xiao-bin, 2009). Customer segmentation can help company to identify who their best customers and help them apportion their marketing spend accordingly. Customer segmentation makes money for sellers by helping sellers define better value propositions, allocate resources, identify and effectively pursue opportunities, anticipate problems and find solutions, and think through situations.

Business cannot deploy marketing budget equally across all customer segments. By focusing marketing resource on the top customer segment they can improve overall revenue and also increase retention of the best customers. For an internet service provider like ATINET Company, segmentation can help it to improve company ability in facing with variation of services and competitors in the market. Customer segmentation will help ATINET Company to focus on the best actions to generate more profits, minimize downsides, and find and exploit upsides. This can increase profitability and help ATINET identify strengths and weaknesses in its overall business strategy.

1.6 Research Demarcation

This study focuses on customer segmentation and defining strategy in each segment. Segmentation has been done by data gathering from the database of an internet service provider in Iran. Most of the literature reviewed about customer segmentation has used RFM (Recency, Frequency, and Monetary) method or CLTV (Customer Life Time Value) method. In this project, based on available database for our case study we focused on RFM method and developed two new methods. These new methods consider the changes in purchase behavior of customers for segmentation. They can identify which customers are at risk of cancelling the company services.

1.7 Research Outline

This thesis consists of six chapters. The first chapter is introduction that gives a brief background about subject of study. Chapter 2 is a literature review on different methods and models of customer segmentation. Chapter3 is about our research methodology and introducing our new proposed methods for customer segmentation. Chapter4 is about the results of analysis and applying proposed methods. In chapter 5, the obtained customer segments will be explored more in detail and related strategies for each segment will be discussed. Finally the chapter5 is the last chapter includes conclusions and further research.

Chapter2: Literature Review

Customer segmentation is to provide enterprise a full range management perspective, enable to have a great chance for enterprises to communicate with customers, and to enhance the return rate of customers (Gong and Xia, 2009).

Customer segmentation needs a comprehensive understanding of companies' customers. Since the enterprises must make more scientific future decision, different methods to describe customer behavior exist in literatures. Among them, there are various types of applications based on data mining, RFM method, Customer Value matrix and CLV method.

Many applications of customer segmentation are based on personal customer attributes like sex, age, education, etc. Among them, there are various types of applications based on data mining. RFM analysis can be conducted by the use of data Mining methods specially clustering methods. Application of these data mining and clustering methods will result in exploitation of more useful information and analysis results. On the other side, customer value matrix is one of the methods that is so easy to implement and understanding.

The last and well known method among these applications is customer lifetime value methods which have been studied in many cases and by many enterprises.

It must be noted that there are many other customer segmentation methods in literatures which have not presented here because their application differ fundamentally from our application in this study. For example, Online purchasing behavior is one of them that can segment customers based on their purchasing sequences (Wang H. et al, 2006).

In this chapter we will review the above methods and related published studies briefly.

2.1Review of Customer Segmentation based on RFM method

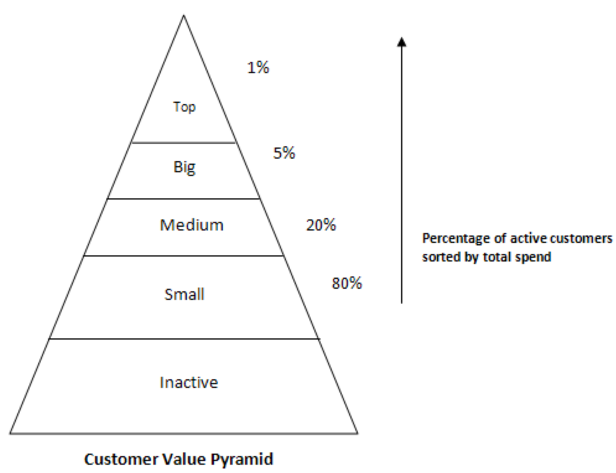
RFM is one of the most magnificent models for customer segmentation that identify customers' behavior by three dimensions which are Recency, Frequency and Monetary. This well-known method is used to identify customer behavior based on present customer behavior characteristics. (Madani, S., 2009 cited by Chan, H., 2008 and Sohrabi b. and Khanlari A., 2007). About more than 30 years, the direct markets are using RFM to identify customer behaviors. In RFM method for expressing customer profitability, all values concerning financial transactions are taken into consideration (Aggelis, Y., 2005). Moreover, the important factor that must be noticed in collecting demographic profiles of customers are the past purchases of consumers include customer's consumption interval, frequency and spent money. The RFM model was illustrated to distinguish important customer by these three variables. These variables are defined in the literature in the following:

- Recency (R): the latest purchase time.
- Frequency (F): the total number of purchases during a specific period.
- Monetary (M): monetary value spent during one specific period.

R stands for recency indicating the interval between the time when the latest consuming behavior occurs and the current time. F stands for frequency indicating the frequency of consuming behavior in a period of time. M stands for monetary indicating consumption amount of money in a period of time.

A large number of studies have considered RFM method. These previous studies in this area highlight the importance of RFM variables.

(Aggelis, Y., 2005) studied the RFM scoring of active e-banking users. This paper used clustering techniques as one of the methods of data mining to organize observed examples into clusters (groups) based on pyramid model which is shown in figure 2.1. K-means algorithm and two-step clustering method were selected as clustering algorithms. They provided the results for bank to identify easily the most important users-customers.



Source: (Aggelis, Y., 2005)

Figure 2.1pyramid model

In (Sohrabi and Khanlari, 2007), authors estimated customer lifetime value by calculating RFM variables and then they clustered the Bank's customers and proposed customer retention strategies for treating an Iranian private bank customers.

2.2 Review of customer segmentation based on Customer Value Matrix Model

By simplifying the RFM method and using the number of purchases and average purchase amount in a 2 × 2 matrix, we will arrive at a practical yet meaningful approach for customer segmentation.

The Customer Value Matrix was developed from RFM method for small-business retail environments. It introduced by Charles Edmundson (Marcus, C., 1998). Marcus noticed that the RFM in spite of its simple conceptual framework is too complex and time-consuming for small retailers. It is because of the fact that usually the results of segmentation based on RFM yield many segments and it causes one of the difficulties for marketers to understand which groups can be combined for a particular strategy.

Moreover, by examination of the RFM analysis, researchers seized the co-linearity of the Frequency of Purchase and the total Monetary Value variables. That is the reason of why Charles Edmundson suggested using Average Purchase Amount instead of the total Monetary Value of a customer. This work leaded to elimination of the co-linearity between these two variables and for more limpidity the Frequency of Purchase was converted to Number of Purchases (Marcos, C., 1998).

These changes represented refinements over conventional RFM analysis; however, they did not resolve the problem of ending up with too many segments to interpret and to work with (Marcus, C., 1998).

Frequency of Purchase and Average Purchase Amount are used for the segmentation of customers in to a 2 × 2 matrix. This method was used by Boston Consulting Group’s (BCG) Growth-Share (Marcus, C., 1998). One of the advantages of this matrix is using the easy-to-understand quadrant identifiers. It is up to business to add another value by considering what they want to do and what strategy would be adopted for each segment.

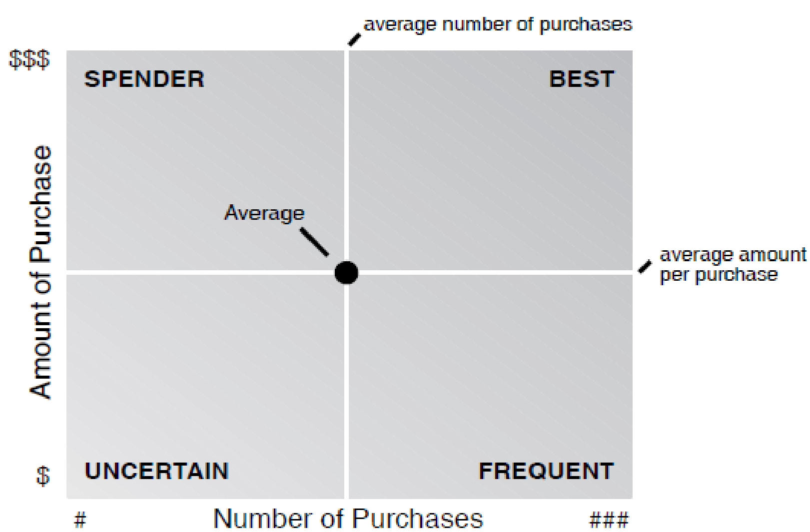
2.2.1 Methodology of the Customer Value Matrix

The first step is collection of data to create Customer Value Matrix. A customer identification (ID) number, the date of a purchase and the total amount of the purchase are the data that must be extracted from enterprise's database. The customer ID number is used to associate purchases with the appropriate customer and the total amount of each purchase is used to calculate the Average Purchase Amount (Marcus, C., 1998).

The next step is segmentation process. In the initial step of this process, the average values for the Number of Purchases and Average Amount Spent must be calculated. After that, each customer is allocated to one of the four resulting quadrants, which has been shown in figure2.1. Table 2.1 show the parameters needed for the segmentation must be calculated.

According to (Marcus, C., 1998), Average Number of Purchases is calculated by taking the total number of purchases for the customer base and dividing it by the total number of customers in the customer base. The Average Purchase Amount is derived by taking the total revenue and dividing it by the total number of purchases (see table2.1).

Comparing each customer's Average Number of Purchases and Average Purchase Amount with total average values is the next step of Customer Value Matrix process. Then each customer will be located to one of four quadrants based on whether customers are above or below the axis averages.



Source (Marcus, C., 1998)

Figure 2.2 Customer Value Matrix

Table2.1 Information table for customer value matrix

Source (Marcus, C., 1998)

Average number of purchase = Total Number of purchases/ Total number of customers
Total Number of purchases
Total number of customers
Average purchase amount = Total sales/ Total number of customers
Total sales
Total number of customers

(Madani, S., 2009) used customer value matrix to apply RFM for the small-business retail environment. She used three types of data includes, purchasing transaction data for extracting RFM, customer data and product data. In her study, RFM variables are extracted from purchasing transaction data to analyze the customer behavior. After segmentation, for describing customer behaviors, association rules used to build customer behavior patterns and their purchase behavior changes.

2.3Review of Customer Segmentation based on Data Mining

The traditional ways of customer segmentation are mainly categorizing methods based on experiences, statistics or simple partitioning (Xin-a Lai, 2009). They can't satisfy the requirements of some more complex analysis which enterprises faced them in recent years.

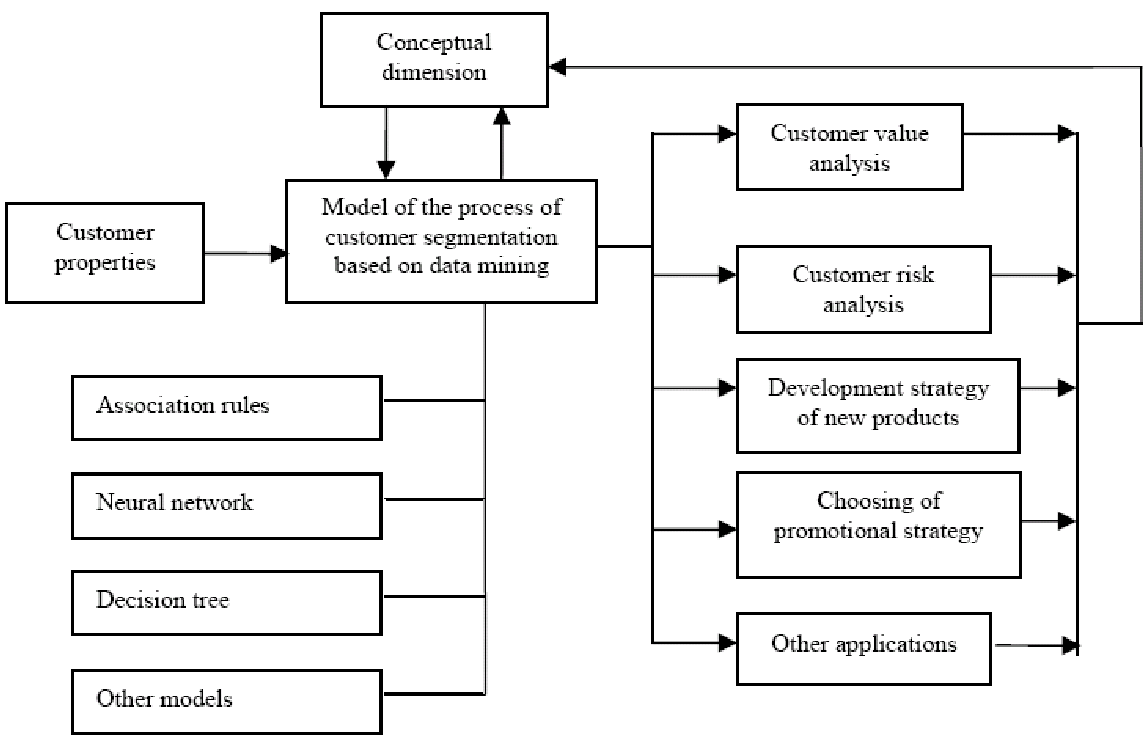
The new methods of customer segmentation are based on data mining. It is the best solution for extracting meaningful data and information from databases which have a huge amount of data. Within the raw data marketers can't understand expressive conclusions easily. The data mining and its related results are being used not only to increase revenue and improving communication between enterprises and their customers but also to reduce costs.

By growing amount of customers' data with the abundant use of management information systems, the traditional customer segmentation methods cannot undertake such a great amount of data. It is an inexplicable task to find valuable information in decision-making purpose. It needs to extract knowledge from large databases or data warehouses. Nowadays the proposed data mining methods make people finally recognize the true value of data, which is embedded in the data information and knowledge (Gong and Xia, 2009). By using data mining technology enterprises can sort and handle and also analyze a huge amount of sophisticated customer's data.

Data mining is the process of sorting through large amounts of data and picking out appropriate information and knowledge by using a series of modern techniques (Xin-a Lai, 2009). Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms (algorithms that improve their performance automatically through experience, such as neural networks or decision trees) and machine learning methods. Consequently, data mining consists of more than collecting and managing data; it also includes analysis and prediction (Cheng Li, 2008).

Data Mining includes association, sequence or path analysis, classification, clustering, and future activities.

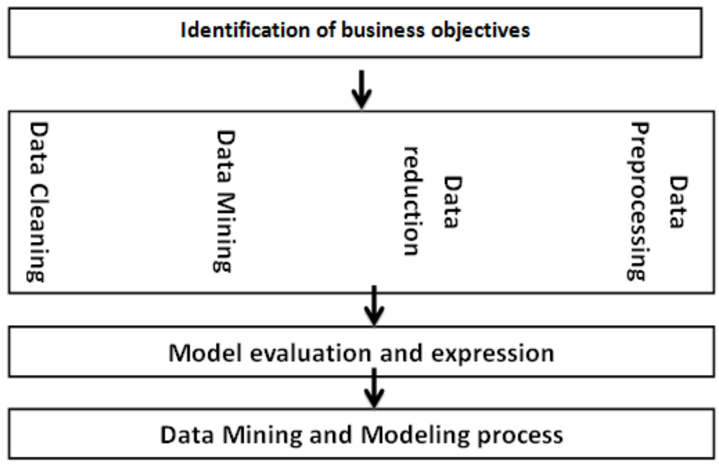
According to the characteristics of data mining and the requirements of an enterprise, process model of customer segmentation based on data mining can be presented as shown in figure2.2 (Lai, 2009).



Source :(Lai, 2009)

Figure 2.3 process model of customer segmentation based on data mining

The implementation of data mining system has a complete structure of flow, generally composed of four main stages: identification of business objectives, data preprocessing, data mining and modeling process, model evaluation and expression as shows in figure 2.4 (Gong and Xia, 2009).

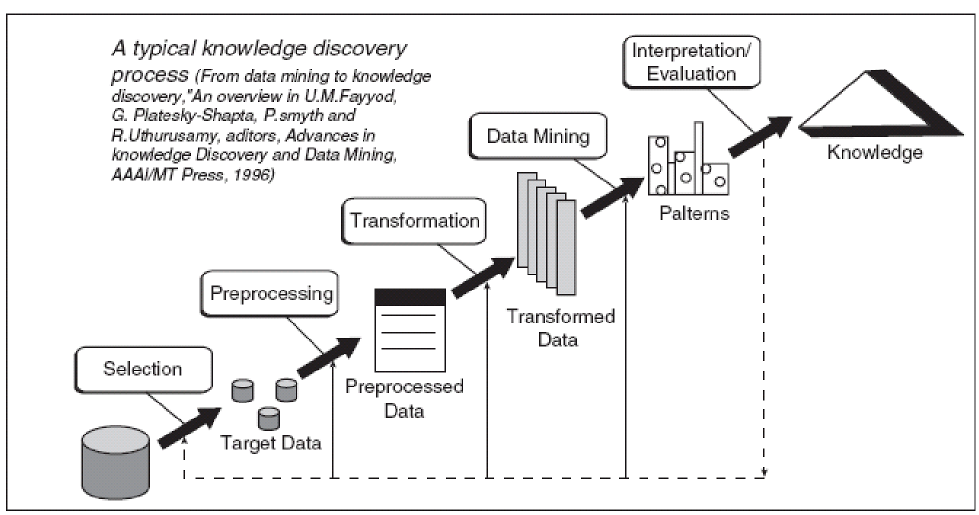


Source: (Gong and Xia, 2009)

Figure 2.4The flow chart of data mining

Data mining is the main step of the knowledge discovery in database (KDD) process.

As it has been depicted in figure 2.5, the KDD process consists of steps: data selection, data cleaning, data transformation, pattern searching (data mining), finding presentation, finding interpretation, and finding evaluation.



Source: (Sivanandam and Sumathi, 2006)

Figure 2.5 A Typical Knowledge Discovery Process

In the following, some studies which perform segmentation of customers based on data mining technology are presented.

Lai in (Lai, 2009) stated that the most frequently used customer segmentation technique in data mining is clustering analysis. Clustering analysis can be used to categorize customers based on the differentiating features of their address, ages, sexes, incomes, occupations, education levels, etc. Meanwhile, clustering analysis can generate the different levels of importance associated with different variables in the classifying process; those data can be used to assist decision-makers.

Gong and Xia in (Gong and Xia, 2009) studied specific implementation of data mining processes and technology for customer segmentation in a supermarket. The main aim of this work was to apply the methods of customer segmentation based on customer purchase behavior to formulate a model in order that enterprises can profoundly understand the customers and make more scientific future decision.

Data mining tasks are very distinct and diverse because many patterns exist in a huge database. Different methods and techniques are needed to find different kinds of patterns According to (Zaïane, 1999). The data mining functionalities and the variety of knowledge they discover are: Characterization, Discrimination, Association Analysis, classification, Prediction and Clustering.

Authors in (Chen et al, 2006) build customer segmentation function model based on data mining and summarize the advantages of customer segmentation function model based on data mining in customer relationship management (CRM).This segmentation model firstly segment customers according to the mapping relationship between customer's attributes and connection category and subsequently constructs the mapping relationship between attributes space and conception space

(Li, 2008) worked on binding data mining technology with customer segmentation theory in aviation freight.

2.4 Review of Clustering Methods

Clustering is similar to classification, but conversely, in clustering, class labels are unknown and the algorithms work to identify a limited set of categories or clusters not only to describe the data but also to determine acceptable classes.

Clustering analyzes data objects without consulting a known class label. Clustering can also facilitate taxonomy formation. Customer Analytics Taxonomy and customer behavior metrics will be explained in detail in the next chapter.

Clustering methods can be categorized into two different types of algorithms which are Hierarchical algorithms and\non-hierarchical or Partitional algorithms (Yuanli T. and Liangshan sh., 2010cited by Sag lam et al, 2006 and Zhongding et al, 2009).

By using previously established clusters, hierarchical algorithms (HC) can find successive clusters. It starts with a single cluster containing all instances and end when a predefined terminating criterion is achieved. Density-based clustering algorithms are arranged to predict arbitrary-shaped clusters in which a cluster is considered as a sphere in which the density of data objects exceeds a threshold (Yuanli T. and Liangshan sh., 2010).

In hierarchical algorithms, number of clusters is unknown in the beginning, which is a strong advantage of these algorithms over non-hierarchical methods. On the other hand, once an instance is assigned to a cluster, the assignment is irrevocable. Therefore, we can say that the output of hierarchical methods can be used to generate some interpretations over the data set and may be used as an input for a non-hierarchical method in order to improve the resulting cluster solution.

Non-hierarchical or\Partitional algorithms (NHC) typically determine all clusters initially, but they can also be used as divisive algorithms in the hierarchical clustering. In these algorithms usually, the data is divided into k clusters at once and the NHC algorithm iterates for all possible movements of data points between the formed clusters until a stop-ping criterion is met. In these methods, each cluster can be represented by the center of the cluster (K-Means) or by one instance located in the cluster center (K-Medoids). The NHC algorithms are sensitive to initial partitions and due to this fact, there exist too many local minima (Sag. lam et al, 2006).

Su-li in (Su-li, 2010) implemented customer segmentation in a commercial bank. He applied the unascertained clustering to divide the commercial bank customers. Although, the commercial bank concerns customer life cycle value, this study has been improved on the customer evaluation method. The new method calculates the currency value, non-currency value, current value and potential value adequately. It considers the customer currency as the mainly evaluation indicators, and the other indicators as the assistant indicators. Combining the quantitative evaluation and qualitative evaluation, the customer value has been synthetically evaluated. The unascertained clustering overcomes the deficiency of C-mean value clustering, and it has quantitative description to the sample characteristics. By applying unascertained clustering, the paper divides the commercial banks customer into quality customer, backbone customer, mass customer and low-class customer.

In the following sections we will review two of the most well-known and popular clustering methods.

2.4.1 K-means method

K-means is the simplest clustering algorithm because of its simplicity in implementation and fast execution and has been widely used in customer segmentation, pattern recognition and information retrieval (Qin et al, 2010). Clustering results are affected by the choice of initial point, and therefore the solutions obtained are always local optimum, not global optimum (Zhao et al, 2010). *K*is used as an input for the predefined number of clusters. An average location of all the members of a special cluster shows means. Each point has been donated to a cluster whose center (or centroid) is nearest. The center is the average of all the points in the cluster. K-Means algorithm calculates its centers iteratively (Qin et al, 2010).

The K-means algorithm calculates cluster centers iteratively as shown in the steps of the K-means algorithm are given in Figure 2.6

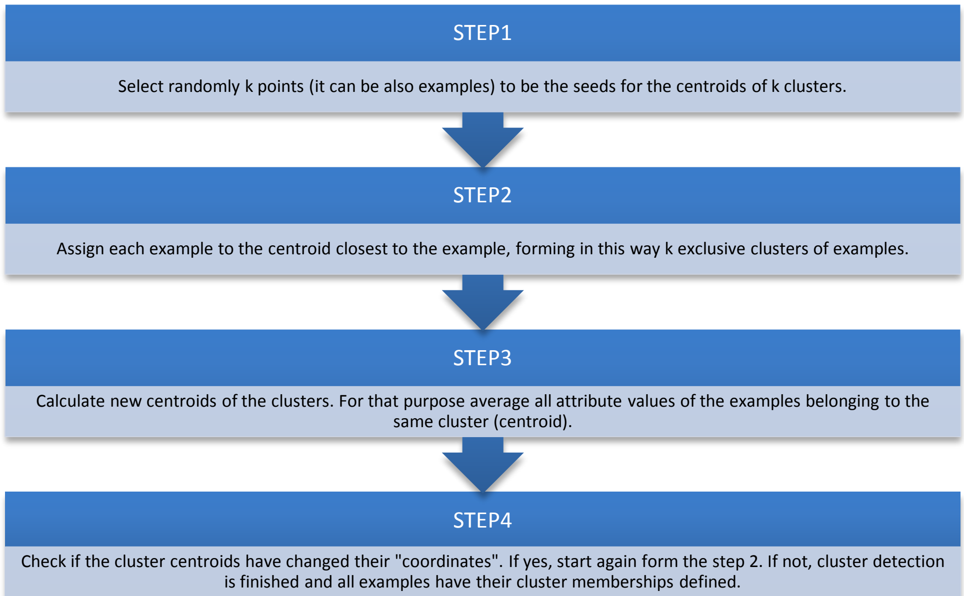


Figure2.6 Classic K-means algorithm

This algorithm is proper for large amount of data. By considering the simplicity and the speed, it can be concluded that it is faster than hierarchical clustering and also in globular clusters K-Means may produce tighter clusters than hierarchical clustering. But there are some disadvantages in using this technique. In each run it does not show the same result, since the final clusters depend on the first random assignment. Another disadvantage of this algorithm is that comparing quality of the clusters produced is so difficult. It also is not useful and appropriate for non-globular clusters.

2.4.2 EM (Expectation Maximization) Clustering Method

Expectation-maximization (EM) algorithm is a method similar to the k-Means algorithm. This algorithm is based on two different steps iterated until there are no more changes in the current hypothesis (Batista, P. and Silva M. J., 2004). Expectation refers to computing the probability that each datum is a member of each class; maximization refers to altering the parameters of each class to maximize those probabilities. Eventually it converges, though not necessarily correctly. One important feature of the EM algorithm is that it can be applied to problems in which observed data provide "partial" information only. Describing the principles and details of applying this method can be found in Wikipedia website and many other resources and has been omitted from here.

2.5 Review of Customer segmentation Models based on CLV

Customer lifetime value (CLV) analysis has been used for decades in many marketing companies. (Hwang et al, 2004) defines Lifetime Value or LTV as 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.

Authors in (Berger and Nasr, 1998), defined CLV as an excess to attracting, selling, and servicing the person, household, or company whose revenues over time exceed. According to (Gupta and Lehmann, 2003), CLV is the present value of all future profits generated from a customer.

From other point of view Customer lifetime value for a firm is the net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm (Jain and Sing h, 2002).

Moreover, many researchers focused on customer segmentation based on the lifetime value.

(Gloy et al, 1997) used the customer lifetime value model in making decisions in the rural petroleum market. For this aim, 11 segments had been developed among 3,281 customers. Then, the single period results were used to make projections of future revenues and costs so that a CLV can be calculated. Finally, the CLV's were used to evaluate alternative marketing strategies and analyzing the profitability of two of the firm's customer segments.

CLV models have a variety of usages in all kinds of business organizations. Particular use of such models, however, will depend upon the type of products and customers a firm has. Firms having large number of customers with small sales to each customer might benefit from models that help segment customers based on lifetime value (Berger and Nasr, 1998).

Authors in (Jain and Singh, 2002) noted that most of enterprises believed that long-lifetime customers are more profitable for a firm.

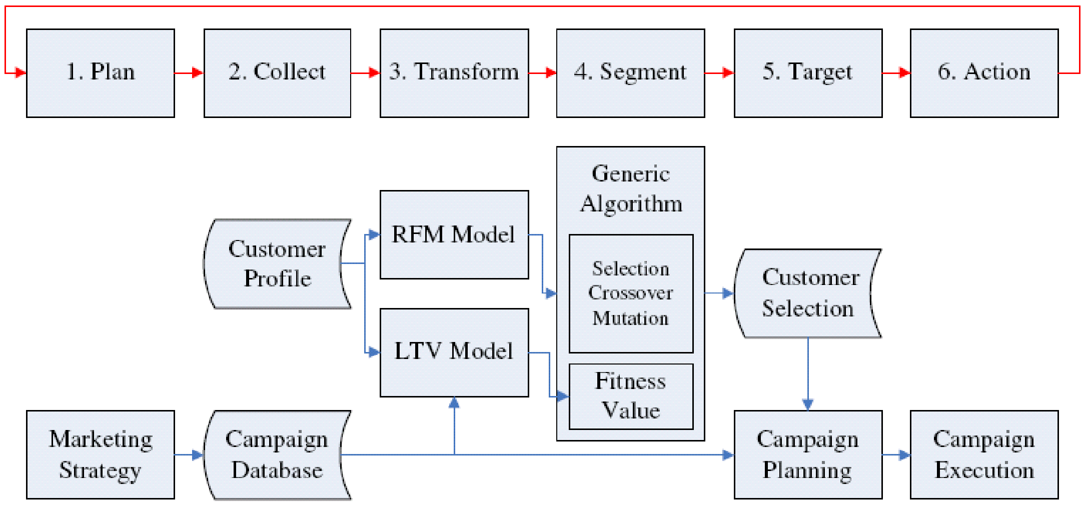
(Kim et al, 2006) suggested a new Life Time Value (LTV) model and also segment customers based on their value. After segmenting customers, they proposed marketing strategies according to customer segments in their case study which was a wireless telecommunication company. This study includes three phases. The data of this study consists of 6-month service data of a wireless communication company in Korea. Phase 1 is data preparation and setting up marketing strategies. The dataset that has been worked in this study is composed of 200 data fields and 16,384 records of customers. After preparation step, the customer value has been evaluated from three viewpoints, current value, potential value and customer loyalty. In phase II, segment analysis has been performed. Phase III analyzes the characteristics of each segment and this part presents the procedure of building strategies based on these three customer values. The method for segmentation analysis is Decision Tree used for mining the characteristics of customers.

(Hwang et al, 2004) and (Kim et al, 2006) suggested a new Life Time Value (LTV) model. They segmented customers considering past profit contribution, potential benefit, and defection probability of a customer for a wireless communication company.

These papers measure the leaving probability for each customer to calculate the churn rate, using data mining techniques; they take several models (decision tree, neural network, and logistic regression) and then select an optimal model among them, based on the result of comparative test.

(Ruiz et al, 2004) studied a segmentation of customers based on their activities. They used clustering algorithm. The algorithm used in the study is P-median method.

(Henry Chan, 2008) presented a novel approach that combines customer targeting and customer segmentation for campaign strategies. This investigation identifies customer behavior using a recency, frequency and monetary (RFM) model and then uses a customer lifetime value (LTV) model to evaluate proposed segmented customers. For selecting more appropriate customers for each campaign strategy, this work proposed using generic algorithm (GA). This paper performed an empirical study of a Nissan automobile retailer to segment over 4000 customers to demonstrate the efficiency of the proposed method. As it has been shown in figure 2.10, this work has been implemented in six phases.



Source: (Henry Chan, 2008)

Figure 2.7 the framework

(Haining et al, 2010) established an index system of dynamic customer segmentation based on customer lifetime value in the China Telecom's database mining. In this paper they introduced the evaluation indices for the telecom industry. Achieving dynamic customer segmentation and increasing the objectivity of this index system in describing customer behavior are studied.

Table 2.2 shows the brief view of literature that was studied in this thesis.

Title	Authors/ year	Major method	Case study	Purpose	Methodology	Conclusion
Intelligent value-based customer segmentation method for campaign management	Chu Chai Henry Chan (2008)	RFM model, LTV model and generic algorithm (GA)	Nissan Automobile retailer	Presenting a novel approach that combines customer targeting and customer segmentation for campaign strategies	-Gathering data and establishing a basic customer profile. -Building RFM model -Then the LTV model calculates current customer value and predicts potential customer value. -Finally, applying GA to select the optimum of customer segmentation for each marketing strategy.	This study suggests an intelligent model that uses GA to select customer RFM behavior using a LTV evaluation model. If the proposed methodology is applied, high-value customers can be identified for campaign programs and it considers the correlation between customer values and campaigns. Therefore, Valuable customers can be identified for a campaign program.
Mining Changes in Customer Purchasing Behavior	Samira Madani (2009)	CLV, Association rules, Apriori algorithm	Kalleh Company	Mining changes happening in customer behaviors of a company	-Data preprocessing -Customer segmentation based on Customer Value Matrix -Using apriori algorithm for recognizing mining pattern of behavior.	The results shows different kinds of changes include added/perished rules, emerging pattern and unexpected changes. Also, two measures of similarity and unexpectedness has been identified.

Table 2.2 shows the brief view of literature that was studied in this thesis.

Title	Authors/ year	Major method	Case study	Purpose	Methodology	Conclusion
Research on Segmentation implementation process of air cargo customer based on Data Mining	Cheng Li (2008)	Data mining, Clustering analysis	Air cargo	Implementing customer segmentation for aviation cargo based on data mining. Describing the hierarchical design idea and functions of different levels, which will have some reference value for the airlines to start CRM.	-Data preprocessing -Customer segmentation based on customer value (current value and value-added). -Forecasting model of customer value in the air cargo industry -Definition of marketing strategies	This work analysis the segmentation in freight customers and connection with mining theory can help air cargo business to find out customers with the real value, and analyze their features so as to maintain them.
Customer Clustering using RFM analysis	VASILIS AGGELIS (2005)	RFM , k-means and Two Step Cluster as clustering algorithms.	E-banking	Calculation of RFM scoring for active e-banking customers for the evaluation of the customer's behavior such as: better Decision making, forecasting future revenue or conservation of the most important customers.	-Collecting data -Calculating RFM variables -Clustering by K-means and two-step cluster methods	This work shows that the knowledge of RFM scoring of active e-banking users can rank them according to the pyramid model. This result was highlighted by the use of 2 clustering methods. Thus, the e-banking unit of a bank may easily identify the most important customers.

Table 2.2 shows the brief view of literature that was studied in this thesis.

Title	Authors/ year	Major method	Case study	Purpose	Methodology	Conclusion
Improved K-Means algorithm and application in customer segmentation	XQin et al (2010)	RFM model, K-Means algorithm, clustering algorithm	Mobile communication company	Improvement of K-means algorithm for making customer segmentation faster and more accurate.	-Calculating RFM variables -Applying improved K-means algorithm	The experimental results show that the improved method lead to lower time consumption, and therefore more effective for large-scale dataset.
Customer Lifetime Value (CLV) Measurement Based on RFM Model	B. Sohrabi and A. Khanlari (2007)	K-Means clustering, CLV and RFM	Iranian private bank	This paper aims at suggesting a new CLV model and customer segmentation considering RFM model. It also proposed customer retention strategies after segmenting customer base	-RFM variables calculation -Building CLV model -Clustering customers by K-means algorithm -Proposing customer retention strategies	This paper suggested a CLV model considering the RFM at the same time. It clusters customers into segments according to their lifetime value expressed in terms of RFM.

Table 2.2 shows the brief view of literature that was studied in this thesis.

Title	Authors /year	Major method	Case study	Purpose	Methodology	Conclusion
Improved K-Means Cluster Algorithm in Telecommunications Enterprises Customer Segmentation	J. Zhao et al (2008)	K-Means algorithm, clustering algorithm	Telecommunications enterprises	The aim of this paper is introducing an improved K-Means algorithm and designing a model of telecommunications enterprises customer segmentation.	-Finding a set of data objects that reflect the data distribution and take it as the cluster center. -Performing Clustering	By comparison with original algorithm in terms of time of iterations and accuracy, improved K-Means was more stable and also more advance. The segmentation results obtained can be used as the data basis in differentiated services for customers and have positive significance for product design and phone packages recommendation.
A mixed-integer programming approach to the clustering problem with an application in customer segmentation	B. Sag̃lam et al (2006)	Data mining	A satellite broadcasting company (Digiturk)	Proposing a mixed-integer programming model to partition the data set into exclusive clusters. The objective function of the model is to minimize the maximum diameter of the generated clusters with the goal of obtaining evenly compact clusters.	-Presenting a mathematical formulation for the clustering problem with the objective of minimizing the maximum cluster diameter -Presentingseed finding algorithm. -Appling the proposed algorithm on a set of 81 data points. -The performance and accuracy of the proposed model and the proposed clustering algorithm are examined on a real data set	The MIP-Diameter model forms clusters by minimizing the maximum diameter of the generated clusters. The run time of the proposed MIP-Diameter model is improved drastically with linearization and the proposed seed finding algorithm. In addition, the reassignment of the instances leads to better solutions.

Table 2.2 shows the brief view of literature that was studied in this thesis.

Title	Authors/ year	Major method	Case study	Purpose	Methodology	Conclusion
Joint optimization of customer segmentation and marketing policy to maximize long-term profitability	J. Jonker et al (2008)	RFM	Dutch charitable organization	Presenting a joint optimization approach addressing two issues: (1) the segmentation of customers into homogeneous groups of customers, (2) determining the optimal policy towards each segment.	Determining segmentation. The optimal marketing policy is determined for the given segmentation In order to find new candidate segmentation, this paper proposes to adopt a local search method. Appling proposed method in a direct mailing framework.	The results show that their model leads to a significant improvement over CHAID, a model that determines an optimal strategy given segmentation. They also see that the best segmentations proposed by their method are almost identical. This indicates that our method does not converge to various different local optima.
A practical yet meaningful approach to customer segmentation	Claudio Marcus (1998)	RFM	-	The purpose of this article is to introduce a simple yet powerful approach to customer segmentation. It is called the Customer Value Matrix.	Data gathering Calculating Average Number of Purchases and Average Purchase Amount Segmentation by proposed method (Customer Value Matrix) Defining some strategies and tactics.	The result shows that the Customer Value Matrix provides an affordable, easy to implement segmentation methodology that delivers substantial value relative to the amount of effort involved.

Chapter3: Research Methodology

Research Purpose

Research Approaches

Research Strategy

Data mining process

Data Collection Method

Data Pre- Processing

Data cleaning

Data Transformation

Customer Segmentation based on RFM Model

Frequency, Monetary and Purchase Change rate (FMC) Model

Generalized Differential RFM method (GDRFM)

Data Clustering and Customer Segmentation

Strategy Definition per Segment

3.1 Research Purpose

According to (Zhahang et al, 2006) research purpose is to express what should be achieved by leading research and how the results of the research can be used. It can be classified by its purpose as exploratory, descriptive, explanatory and predictive. The aim of the exploratory research is looking for patterns, ideas or hypotheses in a new light rather than testing or supporting a hypothesis. Furthermore, exploratory research can be conducted using a literature search, surveying expert about their experiences, conducting focus groups, and case studies.

In contrast, descriptive research identifies and obtains information on accurate profile of a person or the characteristics of a particular issue. The descriptive research is often used when a problem is well structured and there is no intention to investigate cause-effect relationship (Xi Zhang X. and Tang Y, 2006).

Analytical or explanatory research is to understand phenomena by searching and analyzing casual relationship between cause and effect. This is a continuation of descriptive research. Predictive research goes further by predicting the similar condition. The goal of this research is to generalize from the analysis by forecasting certain event on the basis of hypothesized. Table 3.1 shows the differences among these three aspects of research

Source (Wang C. and Wang Zh., 2006)

Table 3.1 the differences among these three aspects of research

Type of research purpose	Description	General Research Question
Exploratory	-To satisfy researcher's desire for more clear and better understanding of the problem to be studies. -To test feasibility of undertaking a more extensive study.	What, why and how one variable produces changes in another
Descriptive	-To describe and document existing circumstances and events	What are the visible event action, beliefs, social structure and process occurring in this phenomena
Analytical or explanatory	-To understand phenomena by searching and analyzing casual relationship between cause and effect	Why and how one variable causes changes in another variable
Predictive	-To generalize from the analysis by predicting certain event on the basis of hypothesized	

The purpose of this thesis is descriptive. The descriptive data will be collected and analyzed.

3.2 Research Approach

There are two main research approaches to choose from when conducting a scientific research: quantitative and qualitative (Madani, S., 2009). The approaches that must be used depend on characteristics of the gathered information and the data types. Indeed, the most important difference between two approaches is how data and statistics are used (Wang C. and Wang Zh., 2006) and also it is related to purpose of study and research questions. Quantitative research deals in numbers, logic and the objective. It is based on measurement of variables, the delivery of findings in numerical form and also analysis conducted through the use of diagram and statistic.

On the other hand, qualitative research focuses on non-numerical data collection or explanation based on the attributes of the graph, analysis conducted through the use of conceptualization.

Based on purpose and research questions, the chosen approach for this thesis is the quantitative approach.

3.3 Research Strategy

Research strategy will be a general plan of how researchers are going to respond to the research questions (Madani, S., 2009). It will comprise clear objective come from research questions. It specifies the sources from which researcher attempt to collect data and consider money, time, location and ethical issues. According to (Nosrati, 2008), identifying the type of research questions is the most important condition for differentiating among the various research strategies.

There are five research strategies in social science, i.e. experiment, survey, archival analysis, history, and case study. Table 3.2 shows each strategy in the three conditions and shows how these are related to the five types of strategies.

Source: (Yin, 2003, p.5)

Table3.2 Different Type of Research purpose

Strategy	Form of research questions	Requires control over behavior event	Focuses on contemporary
Experiment	How, Why?	Yes	Yes
Survey	Who, what, how many, how much?	No	Yes
Archival analysis	Who, what, how many, how much?	No	Yes /No
History	How, Why?	No	No
Case study	How, Why?	No	Yes

The case study strategy is a common strategy in business research that is usually associated by quantitative approach.It is based on an in-depth investigation of a single individual, group, or event. A fundamental difference between case studies and these alternative methods is that the case study researcher may have less a priori knowledge of what the variables of interest will be and how they will be measured (Benbasa et al, 1987).

The focus of this study is customer segmentation and the data has been collected from an Internet service provider database. Therefore, it uses case study as the research strategy. The characteristics of case studies have been shown at table3.3.

Source: (Benbasa et al, 1987)

Table3.3. Key Characteristics of Case Studies

1. Phenomenon is examined in a natural setting.
2. Data are collected by multiple means.
3. One or few entities (person, group, or organization) are examined.
4. The complexity of the unit is studied intensively.
5. Case studies are more suitable for the exploration, classification and hypothesis developments tags of the knowledge building process; the investigator should have a receptive attitude towards exploration.
6. No experimental controls or manipulation are involved.
7. The investigator may not specify the set of independent and dependent variables in advance.
8. The results derived depend heavily on the integrative powers of the investigator.
9. Changes in site selection and data collection methods could take place as the investigator develops new

hypotheses.

10. Case research is useful in the study of "why" and "how" questions because these deal with operational links to be traced over time rather than with frequency or incidence.

11. The focus is on contemporary events.

3.4 Data mining process

Data Mining (DM) is a technology to discover and extract implicit and useful information from large databases or data warehouses. It is a highly valued field of application in database research. It can extract potential valuable knowledge, and even feasible models or rules under a large amount of data to help companies find business trends so that they can make better prediction (Huaping Gong, Qiong Xia, 2009). In this project, our aim is to perform customer segmentation and this can be done by process of data mining.

3.4.1 Data Collection Method

Data is the base of customer segmentation, so it is necessary to collect relative and appropriate data. If the collected data is not complete and accurate, the follow-up steps are totally useless (Cheng Li, 2008).

Data are categorized as secondary data and primary data. Secondary data are collected from secondary sources such as publication; personal record and census. Primary data are collected through observation, interview and questionnaires. (Nosrati L., 2008) States that, conducting case study as the research study, there are several common sources of data collection that can be used. Documentation, interviews, direct observation, participant observation and questionaries' are among the device to record row data. Table 3.5 shows strength and weakness of them.

Source (Yin, 2003, p.86)

Table 3.4: Six Sources of Evidences: Strengths and Weaknesses

Source of evidence	Strengths	Weakness
Documentation	+ Stable: can be reviewed repeatedly +Unobtrusive: Not created as a result of the case +Exact: Contains exact names, references, and details of an event +Broad Coverage: Long span of time, many events, and many setting	- Retrievability: Can be low - Biased Selectivity: If collection is incomplete - Reporting bias: Reflects (unknown) bias of author - Access: May be deliberately blocked
Archival records	+(Same as above for Documentation) +Precise and quantitative	- (Same as above for Documentation) -Accessibility due to privacy blocked
Interviews	+Targeted: Focuses directly on case study topic +Insightful: Provides perceived casual inferences.	- Bias due to poorly constructed questions -Response bias -Inaccuracies due to poor recall -Reflexivity: Interviewee says what interviewer wants to hear
Direct observations	+Reality: Cover events in real life +Contextual: Covers context of event	- Time consuming -Selectivity: Unless broad coverage -Reflexivity: Event may proceed differently because it is being observed -Cost: Hours needed by human observers
Participant Observations	+(Same as above for direct observations) +Insightful into interpersonal behavior and motives	-(Same as above for direct observations) -Bias due to investigator's manipulation of events
Physical Artifacts	+Insightful into cultural features +Insightful into technical operations	-Selectivity -Availability

Many studies use questionnaires for data collection. The questionaries' questions were rarely specified and, when they were, it was in a very general form. Sometimes the researchers mentioned that they used documents and observations, but they did not provide any more detail about them (Benbasa et al, 1987).

The data needed to perform customer segmentation in our case study were provided by the company under study. The customer identification (ID) number, the date of a purchase and the total amount of the purchase and other related fields came from the accounting program of the ATINET Company.

3.4.2Data Pre- Processing

Much of the raw data collected in database are imperfect and noisy. So, these data are not appropriate for data mining. It is necessary to perform preprocessing.

The data preprocessing includes data cleaning, integration, selection and transformation.

3.4.2.1Data Cleaning and Integration

Data cleaning is one of the most important phases in the data mining process. Sometimes it may be time-consuming and frustrating but it is essential for quantitative research. Generally, if this phase of project doesn't be considered as substantial as other phases, it shows the weakness of research. In this stage, errors must be detected, missing values must be filled, bad designed optional fields or useless attributes must be removed and abnormal or out of bounds or ambiguous items must be checked.

3.4.2.2Data Transformation

In this step, string variables must be converted into numeral or numeric categorical variables and some codes must be interpreted or replaced by text. The other tasks in this phase are data aggregation and data generalization. In this study, total purchase data of a customer in a period of time must be aggregated for performing consequent processes. In data generalization, low- level data will be substituted by higher level ones.

3.4 Customer Segmentation based on RFM Model

In this thesis, we will use RFM model for customer segmentation. This decision is based on data available for analysis in our case study. RFM model also will be used as a foundation for developing new models and approaches for customer segmentation. These new proposed models will be described in following sub-sections.

3.4.1 Frequency, Monetary and Purchase Change rate (FMC) Model

A different viewpoint on customer segmentation can be the answer of this question: "Is a customer at high risk of canceling the company's services?" One of the most common indicators of high-risk customers is a drop off in usage and purchase of the company's service and also increasing the recency parameter of that customer. For example, in an Internet service provider company this could be signaled through a customer's decline in using his or her internet service credit.

One of the shortcomings of available customer segmentation models is that they do not consider behavioral changes of customers during the period of analysis or at last they do not consider it by a direct and defined separate parameter. Although the recency parameter is one of the indicators of this behavior, it suffers from transient behavior of customer and also it is only based on last purchase date of customer. So, considering a new parameter seems to be helpful.

For a company, each customer has different average values of purchase during each season of year or predetermined periods. These average values change based on purchase behavior of customer. If these average purchase amounts decrease continuously, it can be concluded that this customer is on the line of canceling its services or at least falling from beneficial customer segment to non-beneficial ones. Similar conclusion also can be derived for a customer who has an increasing average purchase value during the period of analysis. Such a customer can become a profitable customer for company.

So, these different customers with different reflected behavior must be treated differently. In order to convert this idea in to a computable parameter, all of the purchase amounts of customers in each period of analysis are required. Then a parameter named change rate of purchase amount in each time section can be defined as follow:

$$ChangeRateofPurchaseAmount(k+1) = \left\{\begin{matrix} \frac{purchaseamount(k+1) - purchaseamount(k)}{purchaseamount(k)} \times 100\% & if purchaseamount(k) \neq 0 \\ 100\% & else \end{matrix}\right\}$$

(3-1)

in which $purchaseamount\ (k)$, indicated the total purchase amount of customer in k^{th} time snapshot of analysis period.

If the time period of analysis is divided into n+1 time section, there will be n change rate of purchase amount for whole period. The minimum amount of n is 2, in order to have at least 2 parameters for detecting changes in customer behavior. But for making the final indicator of change rate independent from transient and timely behaviors of customer, it is better to increase the number of time sections. Now, there is a sequence of rate changes which can be used to explore the overall purchase behavior of the customer during the analysis time period. This phase is so important to assign each sequence of change rate values to a distinct and unique value.

In the simplest approach, if the change rate values of the last two or three time sections have the same sign (negative or positive), then the average of these values will be used as the final change rate parameter of customer. Other customers who have the change rates with different signs in each time section are assigned zero final change rate value. So, there will be customers with positive, negative and zero final change rate values in dataset.

The second approach in computing final change rate parameter is averaging all of the change rate values of all time sections for each customer.

The third approach is extracting and recognizing change patterns of customers using intelligent algorithms such as neural networks.

Since the third approach is hard to implement and needs so many considerations in practice, it is not proper for small and mid-sized companies. The second approach also suffers from one negative fact that can be better understood by an example. When a customer has two positive and large change rate values at the beginning of period and after that has four small and negative change rate values which are not comparable with respect to the two first change rate values, by averaging all values definitely a positive final change rate parameter will be obtained. But, this positive value doesn't really reflect the fact that this customer is at risk of canceling the company services or at least is not so profitable for company.

So, it seems that the first method is more appropriate for customer segmentation than other methods. But, there are two other approaches that can mitigate the weakness of mentioned methods.

The first solution is to compute the slope of purchasing amount line in time axis. For doing so, application of linear regression is proposed.

The computation of the slope of purchase amount in time is based on a best-fit regression line plotted through the known x-values (which are time of purchase) and known y-values (which are purchase amount in each time section).The equation for the intercept of the regression line, a, is:

$$a = \bar{y} - b\bar{x} \tag{3-2}$$

where the slope, b, is calculated as:

$$b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \tag{3-3}$$

and where x and y are the sample means AVERAGE(known_x's) and AVERAGE(known_y's).

Figure 3.1 shows the concept of slope computation for a sample customer data. In this graph the purchase amounts of a customer are shown in blue points while red line shows the best-fit regression line for the main sample data. The slope of this line indicates the value change rate of this customer's purchase.

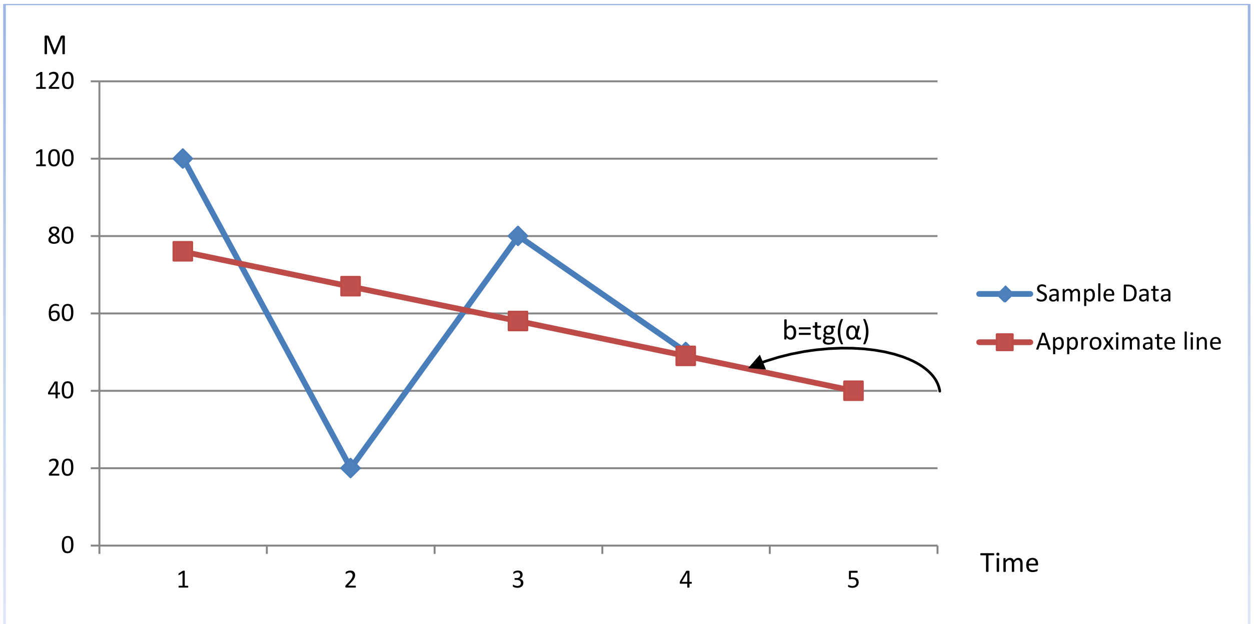


Figure 3.1 illustration of slope computation for a sample customer data

The slope of this line says that this customer has a decreasing purchase amount behavior equal to *b*.

The next new approach proposed in this project is to compute a new parameter which we named it discounted purchase amount slope (DPS). In this project we use a definition of slope that is slightly more complex conceptually. The additional concept that we need is that of *discounting*. According to this approach, the purchase amount slope of each customer is computed by the sum of the discounted slopes of purchase amount in all time sequences. The formula of the DPS is as follow:

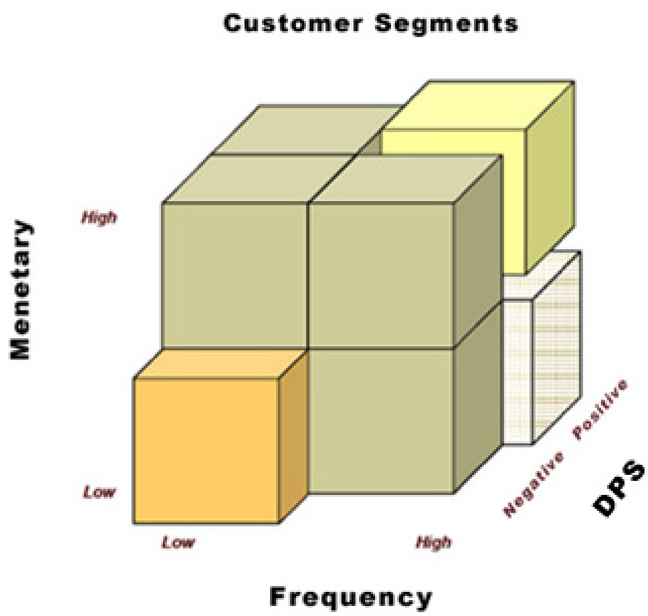
$$DPS = \sum_{i=1}^{n} \gamma^{n-i} S_i \quad (3\text{-}4)$$

where n is the number of time segments, S_i is the slope of purchase amount in i^{th} time segment and γ is a parameter, called the *discount rate*.

The discount rate determines the present value of past slopes. A slope of purchase amount in ***k*** time steps in the past is worth only γ^k times what it would be worth if it were received immediately. By defining this parameter, we reinforce the effect of recent purchase behaviors of customer in computation of total purchase amount slope while mitigating the importance of previous purchase slopes by inserting a discount factor. For example, for a customer with 4 time segments, if we set discount rate equal to 0.7, the last slope is multiplied by 1, the 3rd one is multiplied by 0.7, the second one is multiplied by 0.49 and the 1st slope will be multiplied by 0.343. Considering this discount factor in computation of total purchase amount slope, leads to decrease the effect of primary visited slopes in DPS parameter.

As impleapproach of clustering which is easy to implement is combining customer value matrix method with the new DPS parameter. By this method, we will have eight different segments in F, M and DPS plane as indicated in figure 4.15.

It must be noted that in comparison with customer value matrix, this method results in two sub-segments in each cluster.



Source (McGuirk M., 2007)

Figure 3.2 Customer segments based on FMC values

It means that, for example, best segment will be divided into two segments with different DPS sign. One of these clusters has a positive DPS(discounted purchase change rate or slope) and the other cluster has a negative DPS. So, company must be careful about those customers of best segment with negative DPS that are at risk of falling to other segments with less profit. Armed with this knowledge, the company can quickly communicate with these customers and attempt to offset this anticipated decline in shopping behavior with targeted offers or incentives.

By defining this variable and specifying different clusters with different DPS signs, these two groups of customers will be treated differently and targeted plans and strategies can be better adopted and designed based on their purchase behaviors.

3.4.2 Generalized Differential RFM method (GDRFM)

The next variant of RFM method which is proposed in this project is based on the idea of value change rate stated in the above RFMC method. If we generalize the computation of purchase amount change rate to R, F and M parameters we can distinguish at risk customers and customer segments more adequately.

It is based on the fact that customer behaviors such as decrease in the purchasing amount, decrease in the number of purchases, decrease in number of product categories purchased by customer and also increase in the length of time between shopping can be useful in predicting a potential decline in retention of customers.

These indicators can be addressed just by computing the change rates of RFM variables. In another word, not only considering RFM values is necessary for segmenting customers but also computation of derivatives of recency, frequency and monetary amount of customers with respect to time can be useful for obtaining better and more adequate results in segmentation.

The process of computing average derivatives of RFM parameters is the same as the process stated in DPS method.

If $\frac{dR_i}{dt}, \frac{dF_i}{dt}$ and $\frac{dM_i}{dt}$ represent the derivatives of R,F and M in the i^{th} time step, then average of these derivative by considering different discount rate for each parameter can be calculated by following formulas:

$$\left(\frac{dR}{dt}\right)_{avg} = \sum_{i=1}^{n} {\gamma_R}^{n-i} \left(\frac{dR_i}{dt}\right)$$

$$\left(\frac{dF}{dt}\right)_{avg} = \sum_{i=1}^{n} {\gamma_F}^{n-i} \left(\frac{dF_i}{dt}\right)$$

$$\left(\frac{dM}{dt}\right)_{avg} = \sum_{i=1}^{n} {\gamma_M}^{n-i} \left(\frac{dM_i}{dt}\right)$$

(3-5)

where γ_R, γ_F and γ_M are discount rates of recency, frequency and monetary parameters respectively. $\frac{dR_i}{dt}$, $\frac{dF_i}{dt}$ and $\frac{dM_i}{dt}$ can be calculated easily just by computing differences of parameters at two consequent time steps:

$$\frac{dR_i}{dt} = \frac{R_{i+1} - R_i}{t_{i+1} - t_i}$$

$$\frac{dF_i}{dt} = \frac{F_{i+1} - F_i}{t_{i+1} - t_i}$$

$$\frac{dM_i}{dt} = \frac{M_{i+1} - M_i}{t_{i+1} - t_i}$$

(3-6)

It must be noted that, γ_R, γ_F and γ_M can have different values based on decision of analyst and the type of case study.

The advantage of this method with respect to simple RFM method is on the fact that in GDRFM method, changes in behaviors of customer during the time is considered. Therefore changes in frequency and recency of purchase for a customer are taken in to account with change slope of purchase amount simultaneously. So, customers with positive monetary change slope and positive frequency change slope will be treated differently from customers with negative or different frequency and monetary change slopes.

3.5 Data Clustering and Customer Segmentation

After computing RFM and newly proposed parameters for each customer, they can be fed to segmentation algorithm for performing consequent processes.

3.6 Strategy Definition per Segment

The last and most important phase of research is definition of proper and useful strategies for each customer segments. This must be performed by analyzing and studying the customer behaviors of each segment adequately. These strategies must be in direction of increasing profits of company or other goals that specified by company before running customer segmentation.

Chapter4: Results & Analysis

Data preprocessing

Data Cleaning

Data integration

Data Transformation

RFM Construction

Customer segmentation

    Customer Value Matrix

RFM Method Results

FMC Method Results

GDRFM Method Results

Chapter summary

In this chapter, firstly the results of data pre-processing phase of analysis which has been performed in Microsoft SQL Server 2005 [Ref SQL] are presented. Secondly, based on three customer segmentation models which are RFM model, customer value matrix and a new method proposed by author, the desired meaningful attributes of customers have been generated. After that by using some well known and proper clustering algorithms, the resulted data and customers have been clustered into different groups. These algorithms consist of K-means and EM method. The above phases altogether form the customer segmentation process.

4.1 Data preprocessing

The base of customer segmentation is data, so it is necessary to collect proper data before any other process. The collected data must be complete and accurate to make follow-up steps useful and reliable.

This analysis is based on customers' data of ATINET Company during 8 months, from October 2010 to May 2011. It must be noted that these information are related only to home and non-official users of company services. It is because of the fact that the number of major customers of company who are almost official and governmental organizations and have a great amount of financial transactions is limited and also these customers have a different behavior in comparison with other customers.

Customer transaction data and demographic data are gathered to construct a basic customer profile.

4.1.1 Data Cleaning

In this phase, noisy data or incomplete information has been removed from database. Before analyzing data and data cleaning phase, there was 630 customers' profiles. After data cleaning, 84 of them were recognized as noisy data so they were removed from database. So, the total number of customer after removing incomplete information becomes 546 records.

4.1.2 Data integration

The database contains two tables: customer demographic information table (which consists of customer-ID, name, family name, e-mail, telephone number, mobile number, birthday, sex, education, job and age), and also transaction table (which consists of all transactions of customer in detail). In order to meet the requirements of data mining, the information of two tables must be merged to obtain a customer sale table, which has integrity information for data mining.

4.1.3. Data Transformation

In this phase, the string variables must be converted to numeric variables and numbers. The missing values were checked and deleted or replaced by default values or mean values of each parameter. Total purchase amount and other values which are necessary for RFM method were aggregated in this phase.

4.1.4. RFM Construction

This study uses a RFM model to identify and represent customer behavior. The well-known RFM method, models three dimensions of customer transactional data, namely recency, frequency and monetary, in order to classify customer behavior.

As illustrated in chapter 2, the first dimension is recency (R), which indicates the date of the user’s last transaction.

Meanwhile, the second dimension is Frequency (F), which is defined as the count of financial transactions the user conducted within the period of interest.

Finally, monetary (M) value is the total value of financial transactions the user made within the above stated period.

It must be noted that, in using RFM model, we assume that future patterns of customer trading is similar to past and current patterns. The calculated RFM values are summarized to clarify customer behavior patterns.

In this phase, RFM variable were calculated. The recency of each customer was defined by calculating the interval between the last date of purchase and the last date of the period.

For frequency and monetary, the transaction data was aggregated to calculate the total number of purchases and total amount spent during this period. The final data that is ready to the next step has the format as illustrated in table 4.1.

Table 4.1 RFM table fields

Customer Code(ID)
Recency (days)
Frequency
Monetary

A sample of the data set on which data mining methods are applied lies in Table 4.2.

Table 4.2 Sample Data

Customer ID	Recency (Days)	Frequency	Monetary (thousandTomans)
User1	28	12	142
User2	92	4	52
User3	8	16	84
….	…	…	…

4.2 Customer segmentation

After completion of the above phases, the next step is customer segmentation and clustering based on available information in database. For this purpose, we have four methods, which have some differences in comparison with each other. These methods are, segmentation based on RFM metrics, Customer Value Matrix and our newly proposed methods (FMC and GDRFM) described in chapter 3.

In the following, the calculation steps and results of these methods are presented.

4.2.1 Customer Value Matrix Results

According to Customer Value matrix method, we have two axes: average purchasing of each customer and the Number of Purchases.

We divide total purchase amount by total number of purchases to calculate average amount of each purchase in the selected time period.

After computation of average purchase amount and number of purchases for each customer, the segmentation of customers can be done by use of different clustering methods. The simplest one is the method described in (Madani, S., 2009(Marcus, C., 1998). Based on this method, we can divide the customers into four clusters, which are uncertain, frequent, spender and best. This can be done by computation of two total average values of number of purchase and amount of purchase. We must divide total number of purchases in the selected period by the total number of customers in database to calculate the total average number of purchases. For calculating total average purchase amount, we must compute total purchase amount of all customers to the number of customers in database. Table 4.3 shows these variables and their calculation.

Source (Madani, S., 2009 and Marcus, C., 1998).

Table 4.3 calculating variables for customer value matrix

Average number of purchase = Total Number of purchases/ Total number of customers
Total Number of purchases
Total number of customers
Average purchase amount = Total sales/ Total number of customers
Total sales
Total number of customers
Average purchase amount

By comparing average values of each customer with total average values, customers will be divided into four mentioned segments. The total average values provide the base of separation of high and low values on each axis. Figure 4.1, shows this concept.

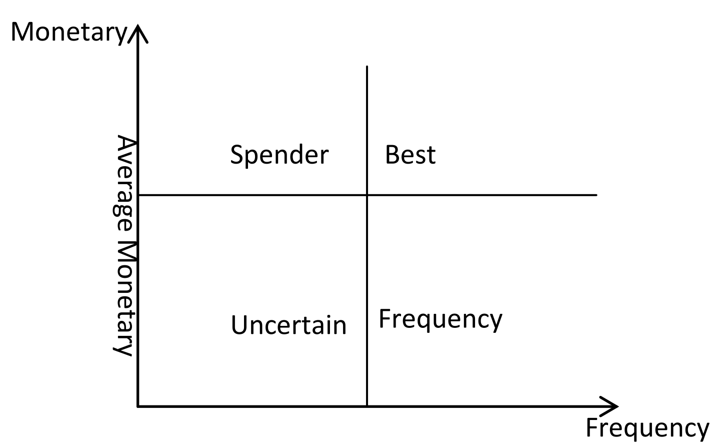


Figure4.1segmentation based on customer value matrix

Table 4.4 shows the computed values of customer value matrix in our study and test case.

Table 4.4calculated variables for customer value matrix in this study

Average Recency=14706/544=27 Days
Total Number of purchases= 3980
Total number of customers=544
Average number of purchases=3980/544=7.3
Total sales=477699000Rials
Total number of customers=544
Average purchase amount = 477699000/544=878000

Each customer's averages must be compared with total average values. So, each customer will be allocated exclusively to one of the four segments mentioned above. The output of this step is a matrix as shown in figure 4.2.

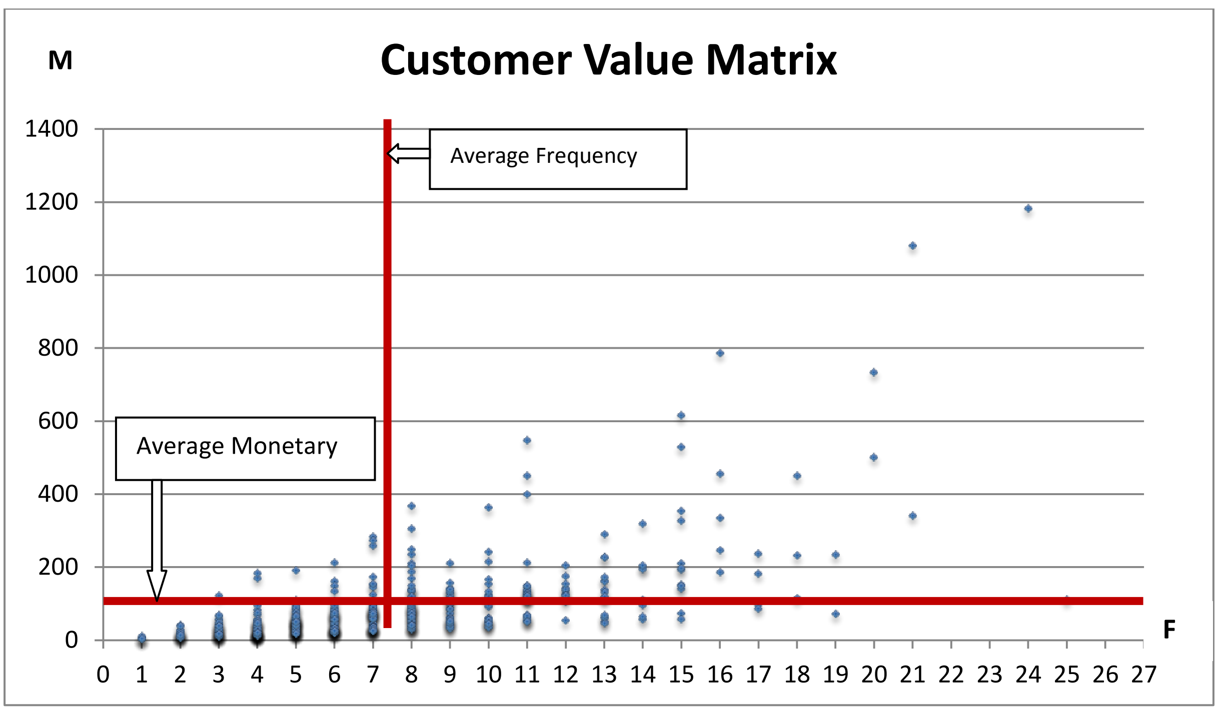


Figure 4.2distribution of customers in the FM plane.

Based on the customer value matrix, there are four clusters. When average purchase amount of a customer is less than the total average purchase and also average purchase frequency of customer is less than total average frequency, it said that this customer belongs to uncertain segment. Customers belong to this segment are uncertain about using services of company, so they purchase sometimes and spend a little money. The customers of this segment must be treated specially, because they maybe exit from our customer list and absorbed with other companies. On the other side, customers with average values greater than total average values lie in the best customer segment. These customers are so valuable and profitable for company. Up-Selling and Cross-Selling are two main actions that must be adopted for treating these customers.

When average purchase amount of a customer is greater than the total average purchase and average purchase frequency of customer is less than total average frequency, we say this customer belongs to spender segment. These customers are also valuable for company because they buy services sometimes but in a large amount or expensive items. Our strategy about these customers must be in the direction of increasing their frequency. By this strategy, they can become our most beneficial customers.

At last, when average purchase amount of a customer is less than the total average purchase and average purchase frequency of customer is greater than total average frequency, we say this customer belongs to frequent segment. These customers buy cheap items but frequently. Company must introduce them new products or services to increase their average purchase amount.

The percentages of customers who belong to these segments in our research are shown in table 4.5 and figure 4,3 respectively.

Table 4.5 percentages of customers who are arranged in each segment

Segment	Number of Customers	Percentage
Uncertain	294	54.04
Frequent	83	15.25
Spender	32	5.882
Best	135	24.81

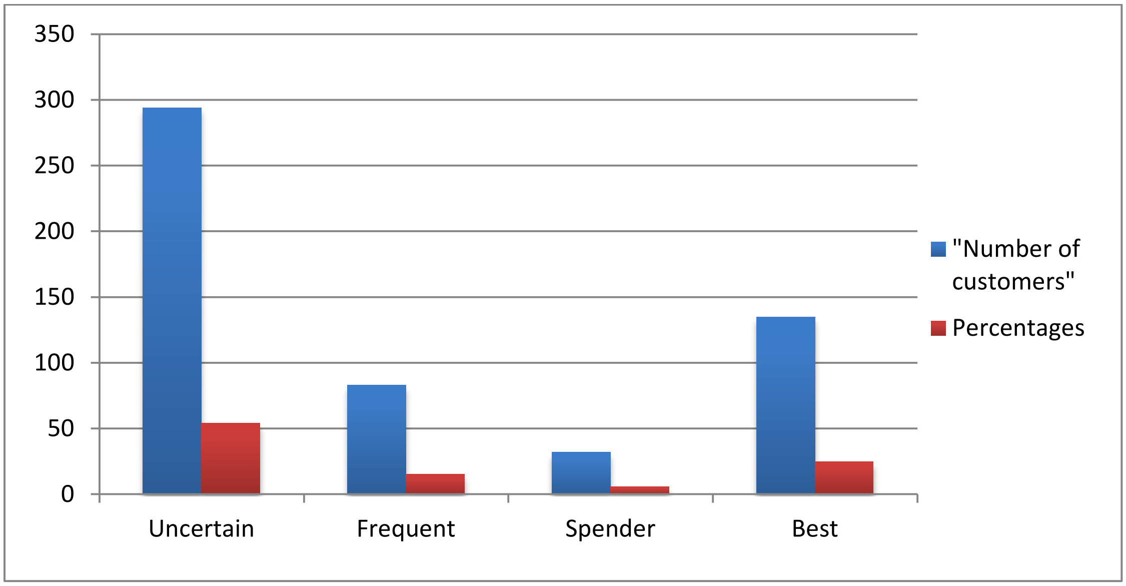
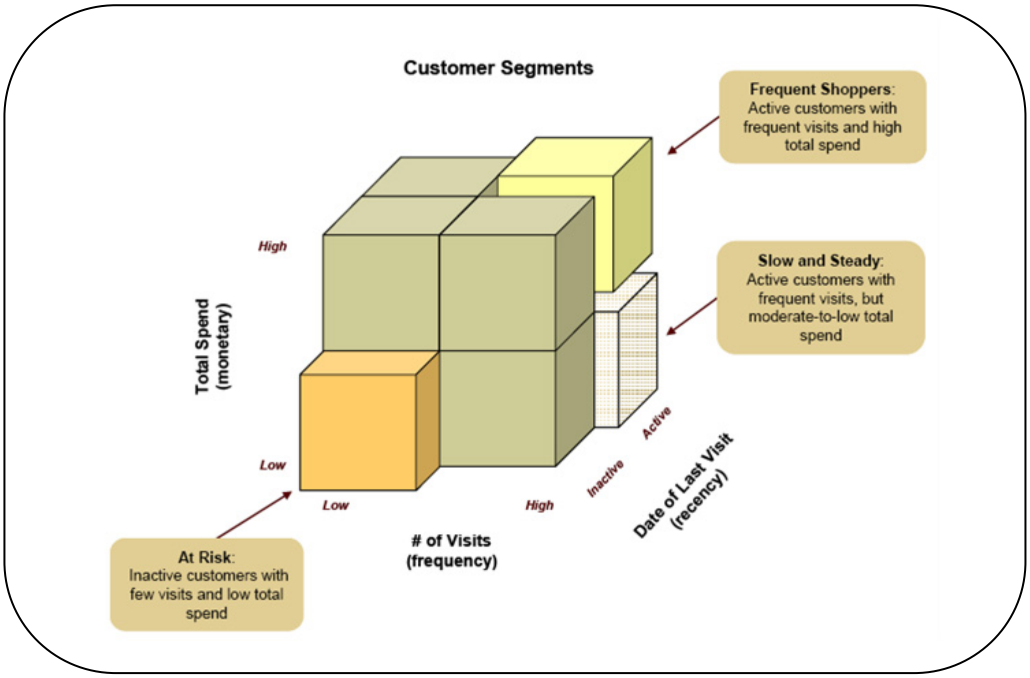


Figure 4.3.Percentage of customers arranged in each segment based on customer value matrix

4.2.2 RFM Method Results

As mentioned earlier, RFM values for each customer calculated. Now, it is the time to perform customer clustering based on these three variables. For doing so, we have two options. The first one is to compare customers' RFM values with total RFM average values and the other is to apply a clustering algorithm to our data.

Considering the first method in our application we can develop eight customer segments. Customers are separated along the following key dimensions: recency of last visit, frequency of visits and monetary amount. The total average values of whole customers' RFM parameters are considered as the base of this segmentation. Figure 4.4, shows these segments in RFM plane.



Source (McGuirk M., 2007)

Figure 4.4 Customer segments based on average RFM values

In the above figure, three main segments are highlighted: Frequent (Best), At Risk and Slow and Steady.

Frequent or best segment indicates the customers who purchase regularly and spending a large amount per purchase. These customers are active and most beneficial customers of a company. The Slow and Steady segment contains active customers who purchase frequently but with a small amount of purchase each time. At Risk customers are customers who purchase rarely and with a small amount each time. The recency of these customers is large which shows that they may be absorbed by other companies if we do not adopt a proper strategy for them. The other segments which have not been highlighted in figure 4.4 can be illustrated and explored similarly. Figures 4.5-4.7show the histogram of distribution of RFM values of our dataset in this project.

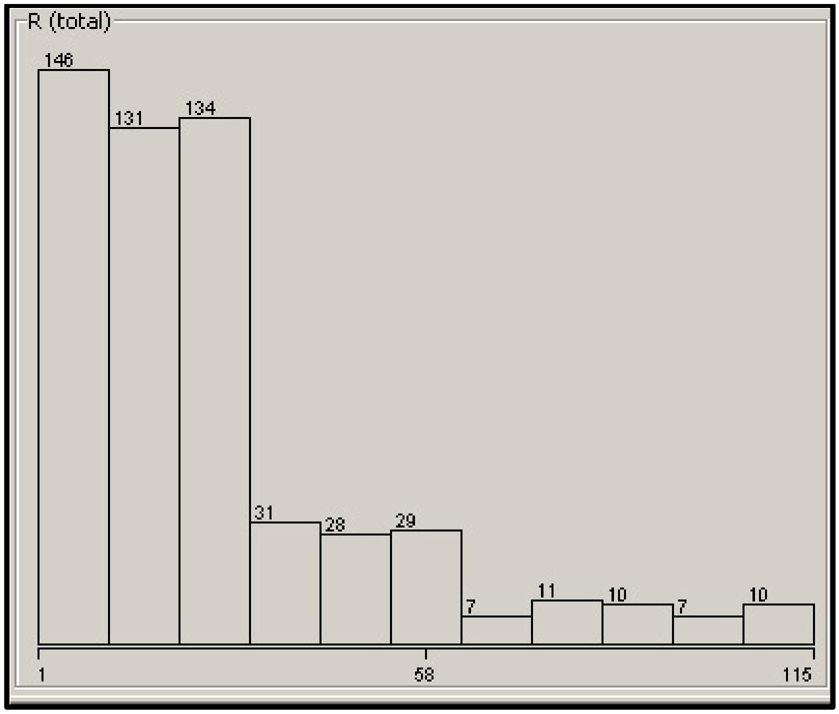


Figure 4.5 Histogram of recency (R)

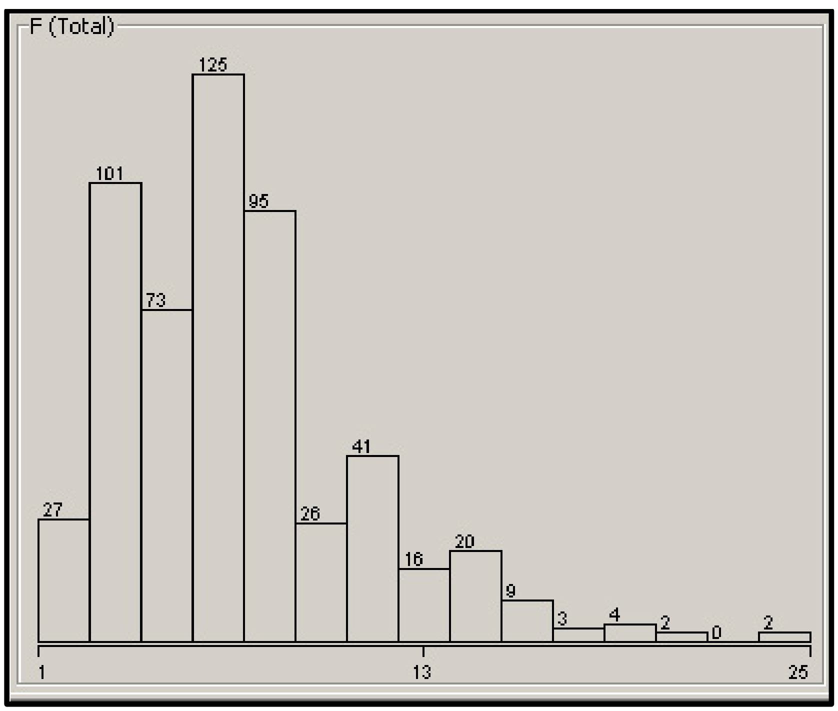


Figure 4.6 Histogram of Frequency (F)

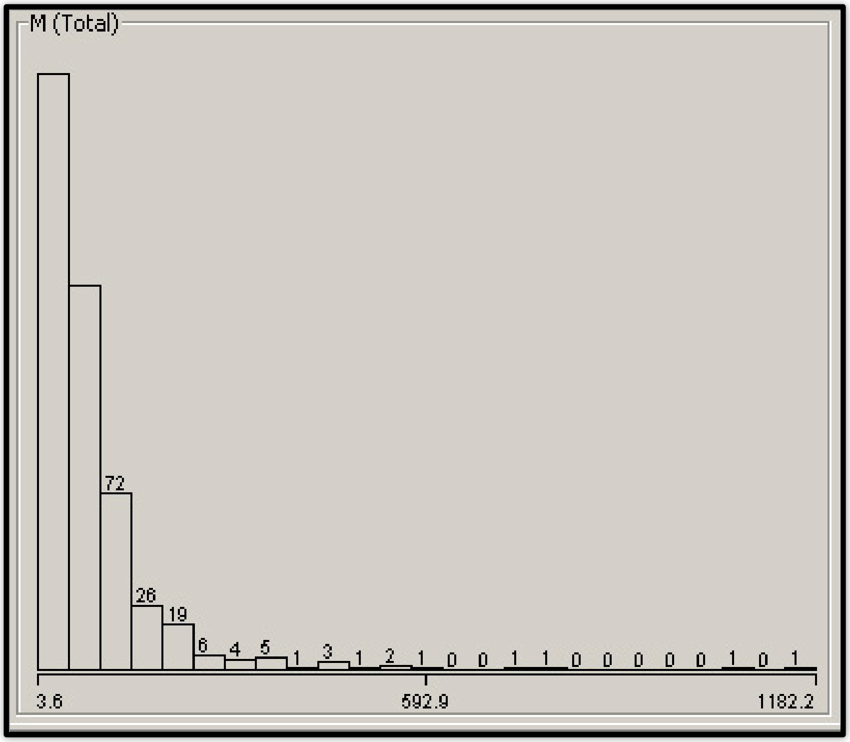


Figure 4.7 Histogram of Monetary (M)

Table 4.6 and figure 4.8 show the results of clustering based on average RFM values as mentioned above. Figure 4.9 shows the distribution of customers in RFM plane.

Table 4.6percentage of customers arranged in each segment by average RFM

Cluster Number	Number of Customers	Percentage
1	155	28.49
2	57	10.48
3	23	4.22
4	117	21.50
5	139	25.55
6	26	4.78
7	9	1.65
8	18	3.31
Total	544	100

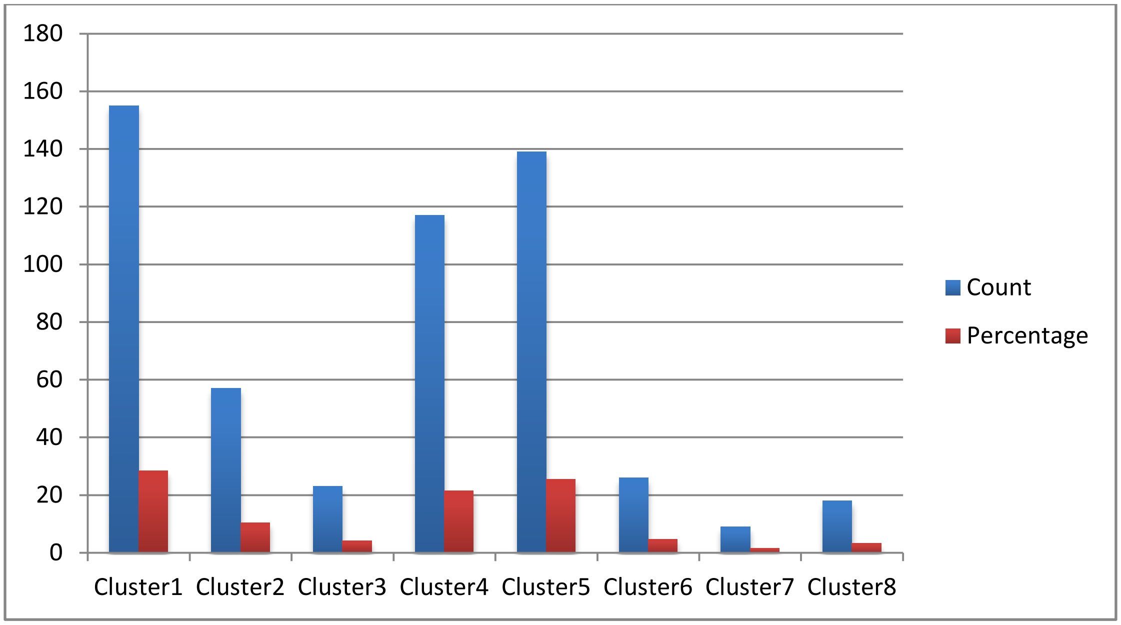


Figure 4.8 percentage of customers arranged in each segment by average RFM

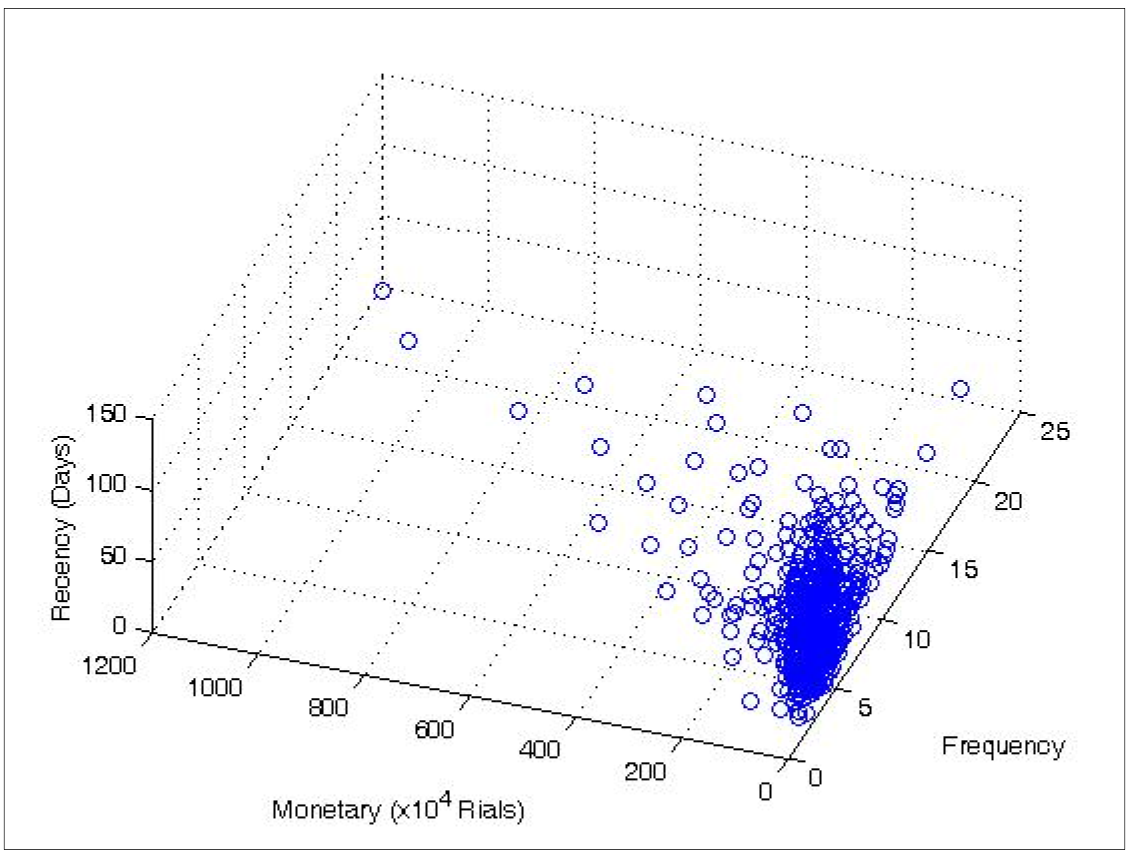


Figure 4.9 distribution of customers in RFM plane

An important consideration must be highlighted here which is the fact that since Recency, monetary and frequency parameters lie in different ranges and specifically monetary value of customers are so much greater than other parameter values, in order to obtain better and more reliable clustering results with some of the algorithms which use distance measures, it is better to scale all of the values to a similar ranges. We perform it by scaling all of the values in to the range between 0 and 1 which means that we must divide all of the values to the maximum values of each parameter. The formula of this normalization is as follow:

$$X_s = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4\text{-}1)$$

in which *X*, indicated the parameter that is under normalization process.

After that, customer classification was performed using the K-means and EM clustering algorithms in WEKA software [weka]. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from user specified Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. Weka is open source. [weka]

Application of K-means algorithm results in the 8 clusters. The number of each cluster members and also the average values of each variable in clusters are shown in table 4.7, 4.8 and figure 4.10.

Table 4.7percentage of customers arranged in each segment based on k-means algorithm and RFM

Cluster Number	Number of Customers	Percentage
1	11	2
2	94	17
3	94	17
4	113	21
5	74	14
6	53	10
7	37	7
8	68	13
Total	544	100

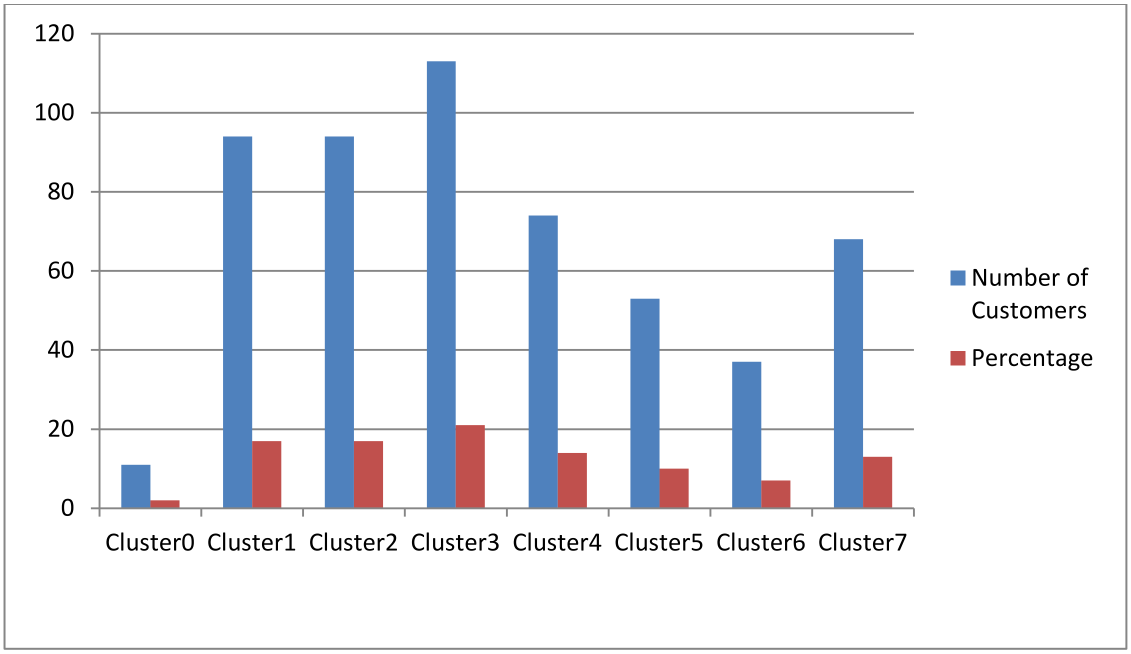


Figure 4.10 percentage of customers arranged in each segment based on k-means algorithm and RFM

Table 4.8Attributes of parameters for each segment based on k-means algorithm in RFM method

	0	1	2	3	4	5	6	7
	(11)	(94)	(94)	(113)	(74)	(53)	(37)	(68)
R	0.1507	0.2144	0.2256	0.0666	0.0625	0.1382	0.8132	0.4654
F	0.7045	0.3178	0.1445	0.1869	0.3519	0.5613	0.0721	0.182
M	0.5543	0.0827	0.0316	0.0434	0.0914	0.1374	0.0193	0.0348

To determine which clustering algorithms are good and for certifying the existence of different customer clusters it is better to run more than one algorithm and then analyze and compare the results carefully.

As suggested above, the EM clustering algorithm was used in order to compare the results. This method yielded the eight clusters of. The related values are listed in tables 4.9and 4.10 for this clustering algorithm.

Table 4.9 percentage of customers arranged in each segment based on EM algorithm and RFM

Cluster Number	Number of Customers	Percentage
0	39	7
1	114	21
2	27	5
3	118	22
4	88	16
5	36	7
6	91	17
7	31	6
Total	544	100

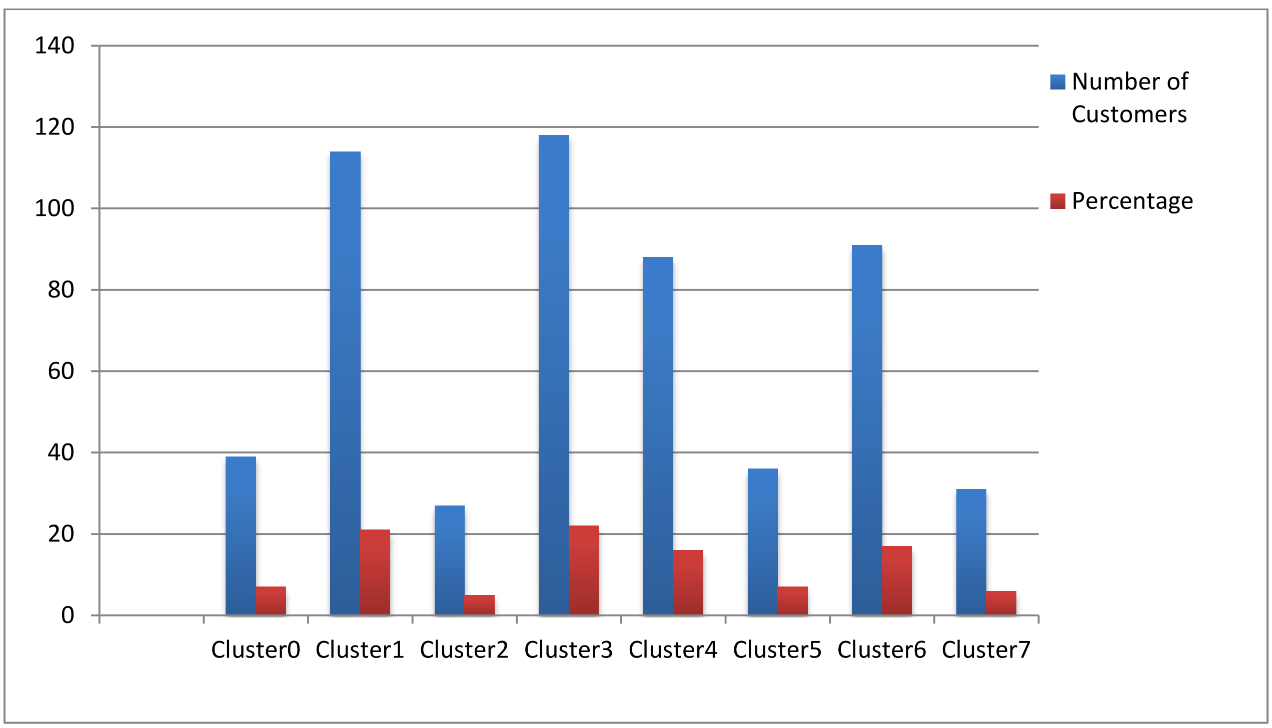


Figure 4.11 percentage of customers who are arranged in each segment based on EM algorithm

Table 4.10. Attributes of parameters for each segment based on EM algorithm in RFM method

	Cluster	0	1	2	3	4	5	6	7
	Attribute	(0.07)	(0. 2)	(0.06)	(0.23)	(0.16)	(0.06)	(0.17)	(0.06)
R	**Mean**	0.4397	0.311	0.1475	0.1427	0.0721	0.0135	0.2107	0.8306
	Std. dev.	0.054	0.1971	0.0798	0.0692	0.0413	0.0116	0.0405	0.1159
F	**Mean**	0.239	0.1269	0.6024	0.3841	0.221	0.3237	0.2381	0.0624
	Std. dev.	0.0935	0.0488	0.1848	0.1116	0.0609	0.1462	0.0746	0.0469
M	**Mean**	0.0532	0.0175	0.3376	0.1106	0.0407	0.0696	0.0454	0.0211
	Std. dev.	0.0253	0.0071	0.2181	0.0444	0.0168	0.0289	0.0171	0.028

Graphs that have been drawn in figures 4.12 to 4.14 show the distribution of customers in various RFM axes. The points belong to different clusters are shown in different colors.

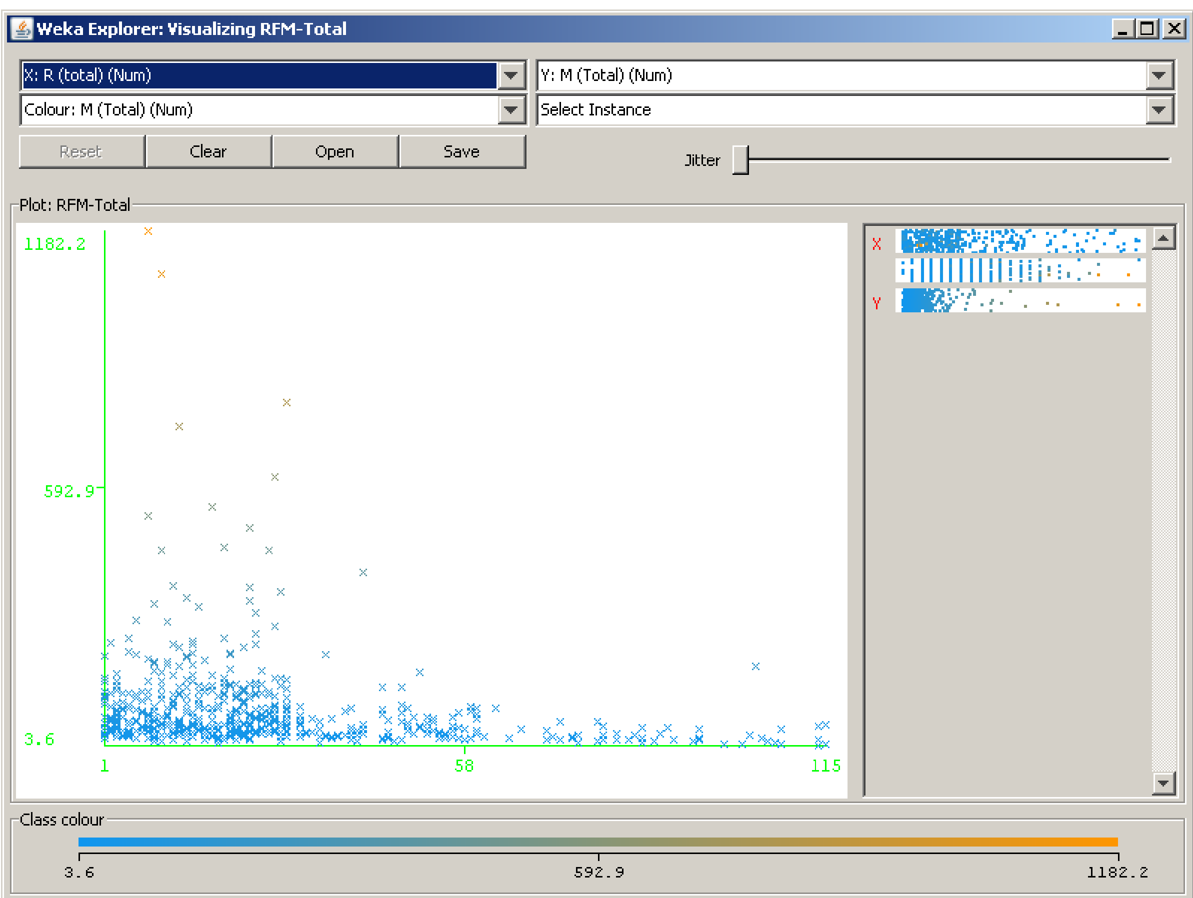


Figure 4.12 distribution of customers in RM plane and their corresponding clusters using EM algorithm

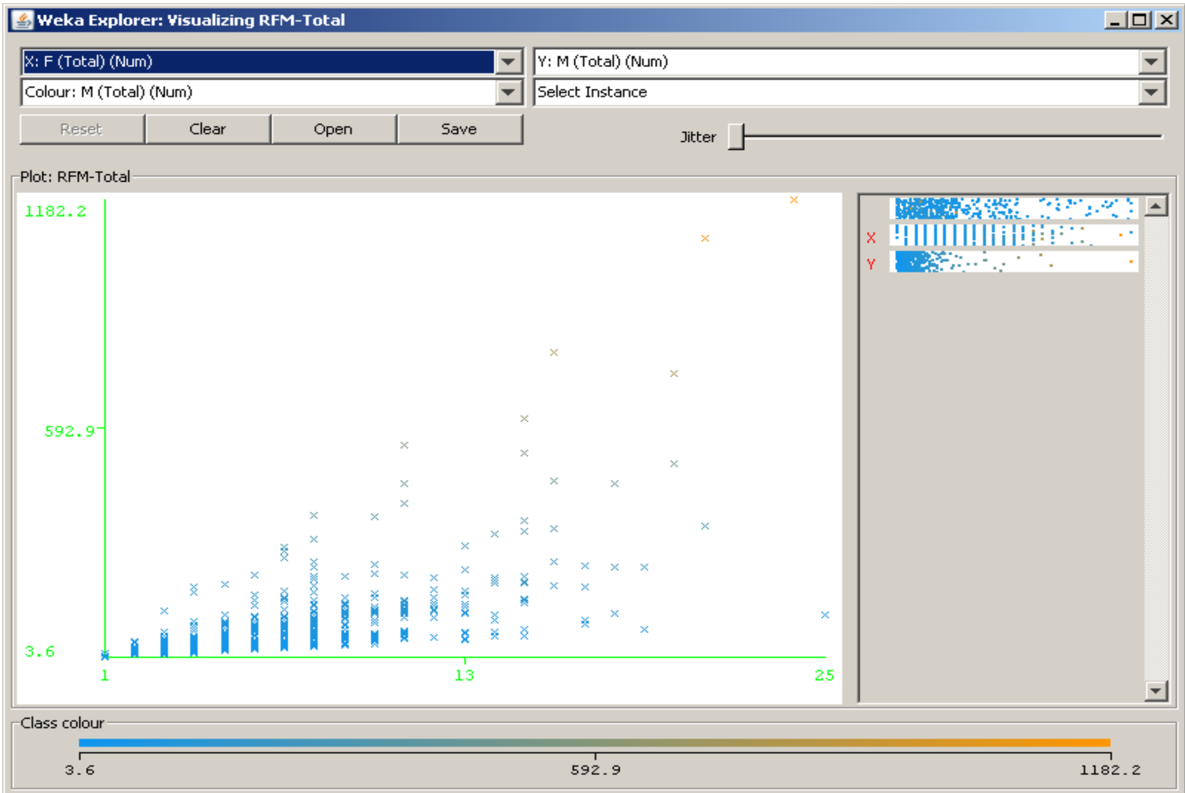


Figure 4.13 distribution of customers in FM plane and their corresponding clusters using EM algorithm

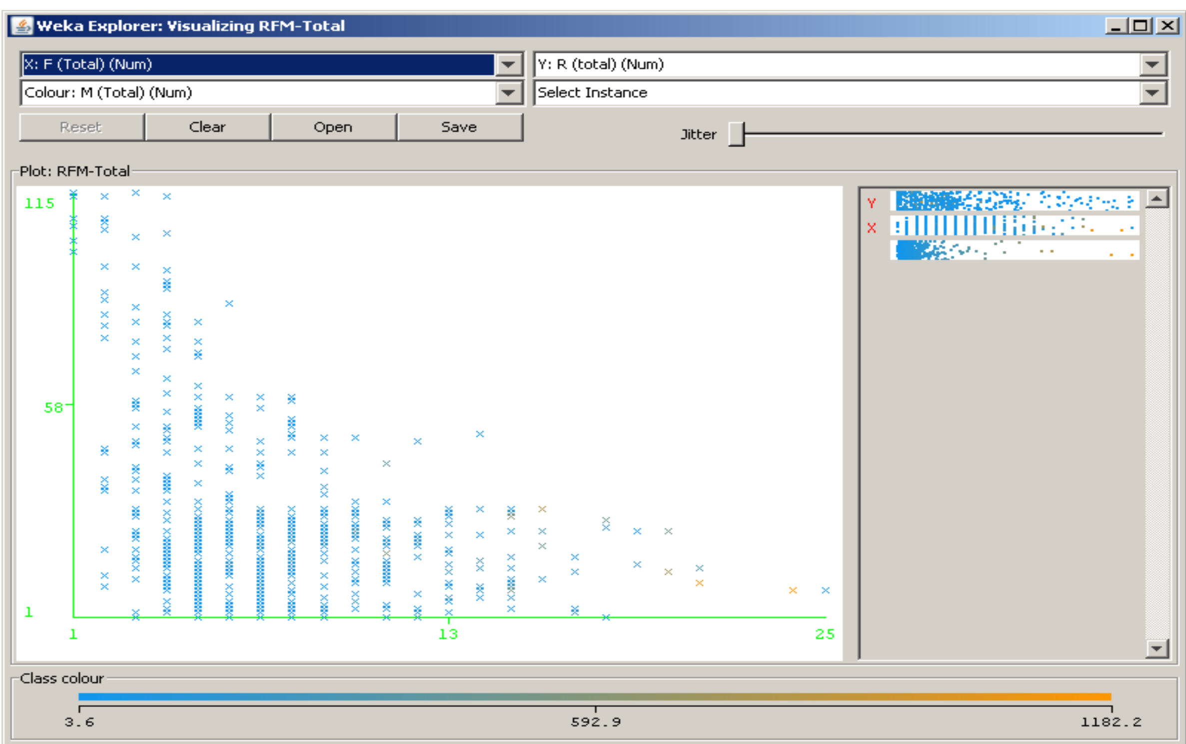


Figure 4.14 distribution of customers in FR plane and their corresponding clusters using EM algorithm

4.2.3 FMC Method Results

After the above analysis, the next approach to be implemented is FMC or frequency, monetary and purchase amount change rate. Based on details of method described in chapter 3, there are two approaches for computing purchase amount change rate. The first approach was computation of slope of purchase amount in time using linear regression while the second approach was computation of new parameter named DPS or discounted purchase amount slope. In this part, the first approach is used. The purchase amounts of each customer during the 8 months were divided in to 4 parts based on definition of 2 months for time step. So, the slope of best fitted line in the time-monetary plane was computed for each customer. Then frequency, monetary and change rate parameter were prepared for segmentation. For simplicity, we used a method which is similar to customer value matrix approach. The total average values of F and M parameter were calculated for whole customers in order to distinguish customers with values greater or smaller than average values. The purchase amount change rate values were classified based on positive or negative sign of purchase amount slope. The resulted segments will be eight segments which described in chapter 3.

The number of each cluster members and also the average values of each variable in clusters are shown in table 4.11-4.12 and figure 4.15.

Table 4.11 Percentage of customers arranged in each segment based on FMC method and EM algorithm

Cluster Number	Number of Customers	Percentage
0	4	1
1	153	28
2	127	23
3	224	41
4	36	7
Total	544	100

Table 4.12 Attributes of parameters for each segment based on FMC method and EM algorithm

	Cluster	0	1	2	3	4
	Attribute	(0.01)	(0.28)	(0.25)	(0.4)	(0.07)
F	Mean	20.2395	4.1017	10.1041	6.5021	13.1216
	Std. dev.	2.8634	1.5688	2.8841	1.9783	4.6396
M	Mean	944.4523	23.2833	122.5965	57.3578	292.6166
	Std. dev.	191.5395	8.8141	40.2226	18.4839	124.1838
Slope	Mean	65.7687	-1.6007	-0.1234	-1.9607	2.4561
	Std. dev.	12.8668	1.9418	6.9905	4.9802	22.0708

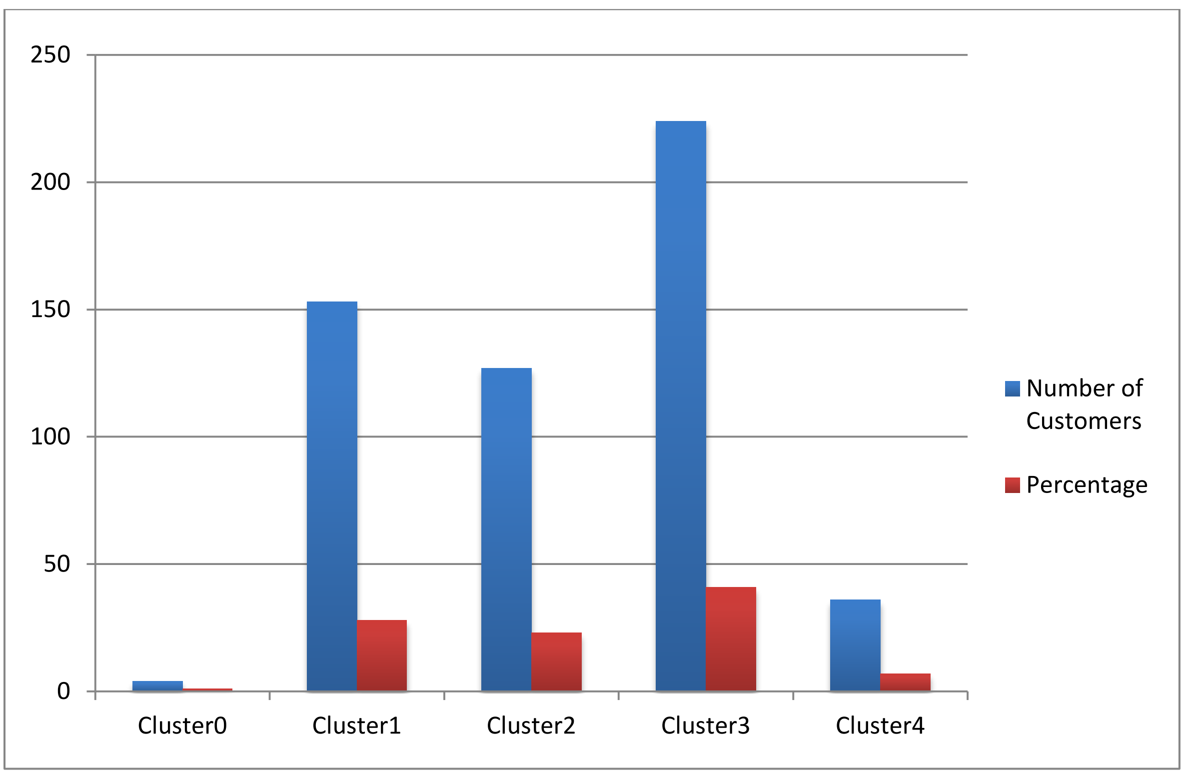


Figure 4.15Percentage of customers arranged in each segment based on FMC method and EM algorithm

4.2.4 GDRFM Method Results

In this section the results of applying GDRFM method for customer segmentation are presented.

As described before, the values of Recency, frequency and monetary of each customer during the 8 months were divided in to 4 parts based on definition of 2 months for time step. After that the average values of all parameter derivatives with respect to time were calculated based on the following formulas:

$$\left(\frac{dR}{dt}\right)_{avg} = \sum_{i=1}^{n} {\gamma_R}^{n-i}\left(\frac{dR_i}{dt}\right)$$

$$\left(\frac{dF}{dt}\right)_{avg} = \sum_{i=1}^{n} {\gamma_F}^{n-i}\left(\frac{dF_i}{dt}\right)$$

$$\left(\frac{dM}{dt}\right)_{avg} = \sum_{i=1}^{n} {\gamma_M}^{n-i}\left(\frac{dM_i}{dt}\right)$$

$$\frac{dR_i}{dt} = \frac{R_{i+1} - R_i}{t_{i+1} - t_i}$$

$$\frac{dF_i}{dt} = \frac{F_{i+1} - F_i}{t_{i+1} - t_i}$$

$$\frac{dM_i}{dt} = \frac{M_{i+1} - M_i}{t_{i+1} - t_i}$$

For simplicity, we set discount rates of all parameters equal to 0.7.

At last, a table of frequency of purchase (F), monetary (M) and recency together with their derivatives for customers of company will be obtained. These data must be fed to clustering algorithm to find customer segments.

The results of clustering using EM algorithm are shown in tables 4.13-4.14 and figure 4.16.

Table 4.13Attributes of parameters for each segment based on GDRFM method and EM algorithm

	Cluster	0	1	2	3	4	5	6	7
	Attribute	(0.1)	(0.12)	(0.09)	(0.15)	(0.08)	(0.18)	(0.11)	(0.16)
R	**Mean**	47.3314	17.9669	18.1688	25.7261	5.6294	9.2165	13.6689	66.159
	Std. dev.	13.9547	6.101	8.8584	2.9534	3.6057	5.6736	7.2581	25.6842
F	**Mean**	7.0167	5.0095	13.7717	7.8637	5.6993	8.7905	8.8133	3.5085
	Std. dev.	2.4575	1.6779	4.914	2.5194	1.4425	3.2498	2.433	1.3822
M	**Mean**	70.1207	28.7895	331.7204	78.8601	47.7121	93.0743	105.4878	24.3246
	Std. dev.	33.0806	10.1967	231.0975	37.3134	18.7298	51.1145	41.974	11.8809
dR	**Mean**	14.3338	-1.3037	0.3871	9.8096	-13.4145	-8.9076	-3.1023	12.1143
	Std. dev.	5.3628	6.953	9.4598	5.7117	5.7827	6.3693	8.0716	6.523
dF	**Mean**	-2.2831	0.2328	0.4928	0.4166	-0.3928	1.1137	-1.008	-1.2378
	Std. dev.	0.7601	0.5073	2.2952	0.9316	0.4107	0.7115	0.544	0.4174
dM	**Mean**	-23.3924	1.2513	15.6385	4.6776	-3.7472	12.2282	-12.5209	-8.7524
	Std. dev.	12.8599	2.8463	65.6489	10.1994	3.2437	9.367	7.3511	4.5615

Table 4.14 Percentage of customers arranged in each segment based on GDRFM method and EM algorithm

Cluster Number	Number of Customers	Percentage
0	53	10
1	13	12
2	46	9
3	87	16
4	46	8
5	97	18
6	55	11
7	88	16
Total	544	100

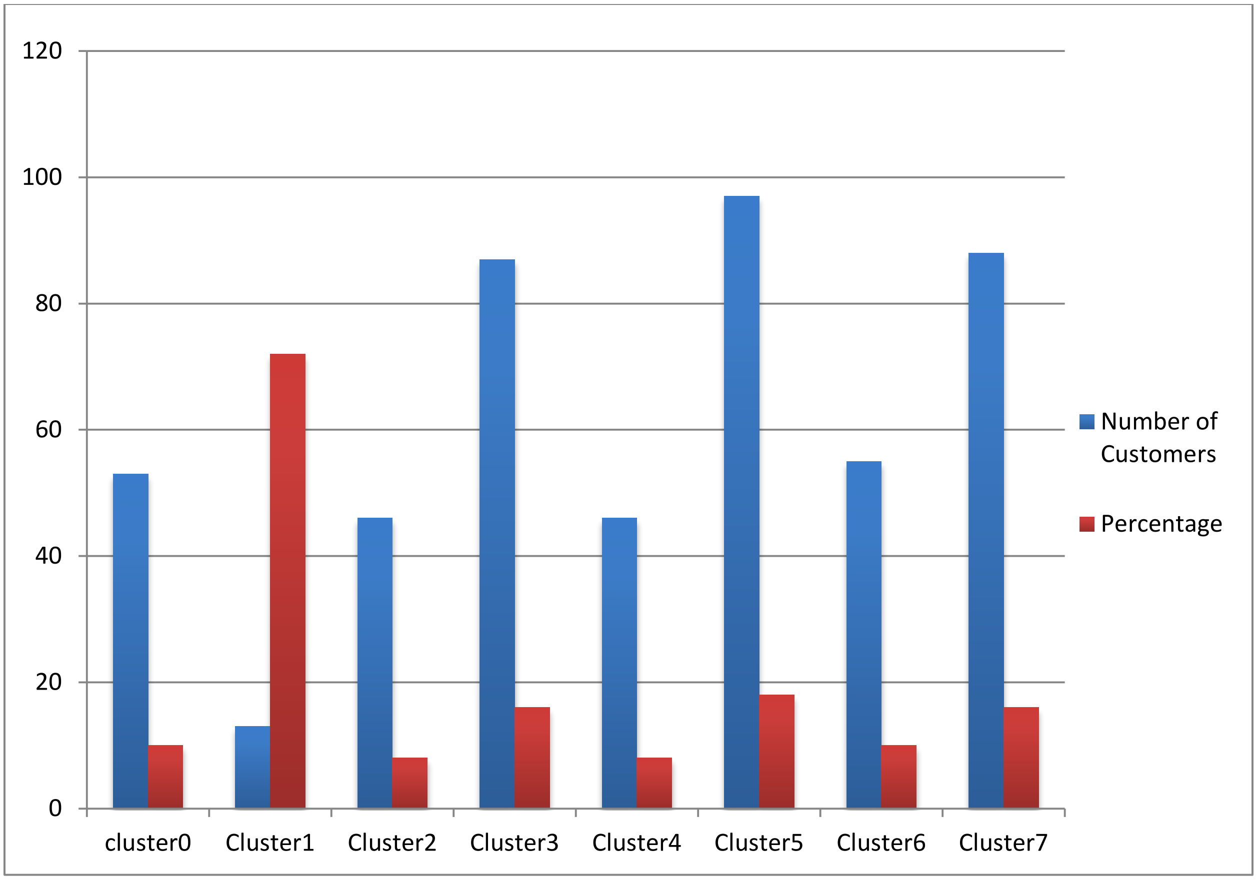


Figure 4.16percentage of customers arranged in each segment based on GDRFM method and EM algorithm

As shown in Table 4.15, cluster 2 is the most beneficial segment because it is superior to the others in terms of all inputs, R, F, and M. Its average recency value is 18 which is smaller than the total average recency value (Smaller recency parameter is better than a larger one). It also has the greatest value of dM (positive slope value in purchase amount), has a large positive dF (positive slope value in purchase frequency) and approximately zero slope for recency parameter. So, the customers of this segment are very valuable customers during the period of analysis. The difference of this segment with respect to "Best" segment described in simple RFM model is the fact that the customers of this segment have an incremental behavior in their buying and the number of buying. This useful information derived from the capability of GDRFM method in identifying the change rate of purchase for all customers. In simple RFM model, the customers of "Best" segment are treated the same. But in GDRFM, company must design different and proper strategies for each sub-segment of "Best" segment based on the sign of change rate slope of frequency, recency and monetary.

Cluster 6 is the group of customers which are valuable customers with approximately great purchase amount and frequency. But these customers have negative large value of dF and dM. So, these customers are at risk of falling from beneficial segment to a non-beneficial segment. On the other word, this segment is a sub-segment of “Best” segment described in simple RFM model. The customers of this segment must be treated on the way of transferring to cluster 2.

Cluster 5 is similar to cluster 6 in terms of R, F and M but has a large positive value of dF and dM. So, this group of customers can be conducted easily to cluster 2.

Cluster 7 is inferior to others in terms of frequency, monetary and recency. It has also negative large value of dF and dM. So, these customers are at risk of cancelling the services of our company. These customers has are not so beneficial customers for company. So, the strategies for these customers must be adopted carefully.

Cluster 0 has the largest negative value of dM and positive value of dR among other clusters and also has a negative value of dF. So these customers are tending incrementally to cancel their services from our company.

Cluster 1 has a small value of frequency, monetary and recency. It also has a positive dF, dM and negative dR. So, these customers can be treated on the way that transferring from non-beneficial segment to other beneficial segments.

Cluster 3 and 4 can be interpreted similarly.

4.3 Chapter summary

In this chapter, the results of applying RFM method and its variants for customer segmentation have been shown. The application of newly proposed method for customer segmentation for an internet service provider shows that GDRFM and FMC methods can be so useful not only for segmenting customers with different values of recency, frequency and monetary but also they can indicate easily which customers are at the high risk of cancelling services of company or falling from beneficial segment to other non-beneficial segments. In the next chapter the proper strategies for each segment will be proposed in detail.

Chapter5: Strategy Definition

Best Ascending Segment

Best Descending Segment

Best Frequency Descending Segment

Best Monetary Descending Segment

Spenders Segment

Frequent Segment

Uncertain Segment

Chapter Summary

Each of the customer segments found in previous chapter is further explored to provide better understanding and identifying opportunities and risks exist in each segment. After that we must develop targeted programs and strategies for each segment separately.

The strategies and tactics can be divided into two categories: segment-specific strategies and cross-segment strategies.

In this chapter of all the segment-specific strategies which are related specifically to each segment will be defined and explained in detail. The cross-segment strategies which are common for all customers include tactics such as customer retention, service affinity, special services for loyal customers and setting membership fee for customers. The cross-segment strategies haven't been investigated in this project.

The main segments found are as follow:

5.1 Best Ascending Segment

Customers in this segment are the most beneficial customers of company. As shown in Table 5.1, this segment is superior to the others in terms of all inputs, R, F, and M (Low recency is better than high recency). It also has the greatest value of dM (positive slope value in purchase amount), has a large positive dF (positive slope value in purchase frequency) and negative or at least zero slope for recency parameter. As mentioned in previous chapter, the difference of this segment with respect to "Best" segment described in simple RFM model and customer value matrix method is the fact that the customers of this segment have an incremental behavior in their buying and the number of buying. This useful information derived from the capability of GDRFM method in identifying the change rate of purchase for all customers. In simple RFM model, the customers of "Best" segment are treated the same. But in GDRFM, company must design different and proper strategies for each segment based on the sign of change rate slope of frequency, recency and monetary.

Table 5.1 Best Ascending segment specifications

Frequency	Recency	Monetary	dM	dF	dR
High	Low	High	Positive	Positive	Negative

Retention of these best customers is critical for company. Furthermore, it is necessary for company to know why these customers prefer to use services of their company. This knowledge is useful for company in order to adopt proper and related strategies in the direction of making other customers of company to shift to this segment. On the other side, it is important to perform all of the efforts for retention of these customers.

The best strategy for Best Ascending Customers segment is to recognize that these customers are the most important customers of company and most worthy of appreciation and special treatment. These customers are required to feel appreciated. So, they must not only be rewarded by preferential discounts, but also they must be treated specially through higher quality, VIP and special services, frequent and high-appreciation communications, informing about new products or services in a timely manner, and simplifying or increasing the relations of these customers with company and other customers who share their interests by holding special events and sessions.

5.2 Best Descending Segment

Best Descending segment is the group of customers that are valuable and beneficial customers with high purchase amount and frequency similar to Best Ascending segment. But these customers have negative value of dF and dM. So, these customers are at risk of falling from beneficial segment to a non-beneficial segment. On the other word, this segment is a sub-segment of “Best” segment described in simple RFM model. The customers of this segment must be treated on the way of transferring to Best Ascending segment and also retention of these customers is so important for company. Table 5.2 shows the characteristics of this segment.

Table 5.2 Best Descending segment specifications

Frequency	Recency	Monetary	dM	dF	dR
High	Low	High	Negative	Negative	-

The best strategies for Best Descending customers segment are frequent and high-appreciation communications and informing about new products or services in a timely manner.

Recognizing the reason of decrease in purchase via communication with customer is the most helpful action that can be done for this group of customers. After that proper strategy for increasing the number of purchase and amount of purchase must be adopted. This can be done by giving information about all products and services or giving special services to these customers.

5.3 Best Frequency Descending Segment

This segment is similar to above two segments in term of R, F and M but has the positive dM and negative dF. This characteristic indicates that these customers are tending to fall into Spender Segment. So, not only we must use strategies defined for Best segments, but also we must follow the strategies specified for Spender segment.

The characteristics of this segment are shown in table 5.3.

Table 5.3 Best Frequency Descending segment specifications

Frequency	Recency	Monetary	dM	dF	dR
High	Low	High	Positive	Negative	-

5.4 Best Monetary Descending Segment

Similar to previous segment, this segment is good in term of R, F and M. Instead, it has a positive dF while has a negative dM. This means that its customers tend to fall into Frequent segment. So, it is advised to perform actions and strategies which have been designed for Frequent segment together with specified strategies for a Best segment.

Table 5.4 shows the characteristics of this segment.

Table 5.4 Best Monetary Descending segment specifications

Frequency	Recency	Monetary	dM	dF	dR
High	Low	High	Negative	Positive	-

5.5 Spenders Segment

Spenders are the customers who have a high average purchase amounts but a low average purchase frequency. So, the most appropriate strategy for this segment is to build purchase frequency. This can be done by communication. We must encourage these customers by informing them about new products and services, capabilities and unique aspects of our company in a timely fashion.

The negative sign of dM and positive sign of dR can also be used for distinguishing between customers of this segment which are at risk of falling to non-beneficial uncertain segment and those who are tending to go to Best segments. So, the retention efforts for those with negative dM value must be reinforced.

Table 5.5 shows the specification of these sub-segments.

Table 5.5Spender segment specifications

	Frequency	Recency	Monetary	dM	dF	dR
sub-segment 1	Low	Low	High	Positive	-	Negative
sub-segment 2	Low	Low	High	Negative	-	Positive

5.6 Frequent Segment

Customers of this segment are loyal customers who purchase frequently but their average purchase amount is not so considerable. So, the best strategy to follow for these customers is to increase the average purchase amount via bundling, cross-selling and up-selling.

These customers are valuable for company because of their proven pattern of repeat purchases but they have low level of revenue for company. They always buy cheap services or use only small number of company services or products.

The cross-selling of other services or products can help to increase the money they spend in each purchase. Providing online shopping channel for these customers can help to increase their purchase each time they visit online store. By this tactic, this group of customers will face with various products and services that they may not visit in formal and traditional shopping. The probability of purchasing more and more will be increased. By adopting these strategies, the customers of this segment will be migrated to best segment.

The negative sign of dF and positive sign of dM can also be used for distinguishing between customers of this segment which are at risk of falling to non-beneficial Uncertain segment and those who are tending to go to Best segments. So, the retention efforts for those with negative dF value must be reinforced.

Table 5.6 shows the specification of these sub-segments.

Table 5.6Frequent segment specifications

	Frequency	Recency	Monetary	dM	dF	dR
sub-segment 1	High	Low	Low	Positive	-	-
sub-segment 2	High	Low	Low	-	Negative	-

5.7 Uncertain Segment

These customers spend very little and rarely. It is so important to investigate about why these customers do not shop frequently and in large amount.

Customers with negative dF and dMare at risk of leaving company services, so we must adopt proper strategies for them. One of the actions that we can do is promotion plans and some incentives or offers in order to get these customers to become more engaged. These offers must be adequate and profitable for company. If this action only lead to one more visit, it will not useful for company. So, we must define a set of best and most appropriate offers for distinct groups of this segment. On the other side, we must consider that offers, special discounts and promotional plans have some cost for company. So, there must be some trade-off between costs and incomes of these plans or it is better to optimize our offer plans by using predictive models and more adequate analysis.

We can define two sub-segments for this group of customers. The first group includes uncertain customers with large negative value of dM, dF and positive value of dR. So these customers are tending incrementally to cancel their services from the company. These customers are not so beneficial customers for company. So, the strategies for these customers must be adopted carefully considering a trade-off between retention costs and their revenue.
Promotional plans and special discounts are useful for this group of customers.

The next sub-segment consists of customers with positive value of dF or dM. For these customers, proper strategies can be cross-selling, special discount and shifting to online shopping channel. These customers must be treated on the way that transferring from non-beneficial segment to other beneficial segments.

Finally, it must be noted that company can focus its efforts only on those Uncertain Customers who are new or have a great affinity to a specific type of service or have a positive value of dM or dF.

Table 5.7Uncertain segment specifications

	Frequency	Recency	Monetary	dM	dF	dR
sub-segment 1	Low	High	Low	Negative	Negative	-
sub-segment 2	Low	High	Low	Positive	Positive	-

5.8 Chapter Summary

In this chapter, the detail description and specification of all segments found in our case study were presented. Based on these specifications, some useful strategies were proposed. Table 5.8 summarizes these characteristics and strategies.

It must be noted that effectiveness of these strategies must be studied by a separate analysis.

Table 5.8 characteristics and strategies for all customer segments -*continue*

Segment	Sub-segment	Attribute	Attribute value	Strategies
Best Ascending Segment		R F M dF dM dR	Low High High Positive Positive Negative	• Recognizing the importance of customer • Communication • VIP and special Services • Preferential discounts • Informing about new products or services • Simplifying or increasing the relations • Increasing the relations of these customers with company and other customers who share their interests by holding special events and sessions.
Best Descending Segment		R F M dF dM dR	Low High high Low Low -	• Frequent and high-appreciation communications • Informing about new products or services • Recognizing the reason of decrease in purchase • Giving information about all products and services • Giving special services to these customers
Best Frequency Descending Segment		R F M dF dM dR	Low High High Low High -	• Frequent and high-appreciation communications • Informing about new products or services • Recognizing the reason of decrease in purchase • Giving information about all products and services • Giving special services to these customers
Best Monetary Descending Segment		R F M dF dM dR	Low High High High Low -	• Frequent and high-appreciation communications • Informing about new products or services • Recognizing the reason of decrease in purchase • Giving information about all products and services • Giving special services to these customers

Table 5.8 characteristics and strategies for all customer segments

Segment	Sub-segment	Attribute	Attribute value	Strategies
Spenders Segment	Sub-segment 1	R	Low	• Communication. • Informing them about new products and services, capabilities and unique aspects of our company in a timely fashion.
		F	Low	
		M	High	
		dF	-	
		dM	High	
		dR	Low	
Spenders Segment	Sub-segment 2	R	Low	• Communication. • Informing them about new products and services, capabilities and unique aspects of our company in a timely fashion.
		F	Low	
		M	High	
		dF	-	
		dM	Low	
		dR	High	
Frequent Segment	Sub-segment 1	R	Low	• Bundling, • Cross-selling • Up-selling. • Providing online shopping channel
		F	High	
		M	Low	
		dF	-	
		dM	High	
		dR	-	
Frequent Segment	Sub-segment 2	R	Low	• Bundling, • Cross-selling • Up-selling. • Providing online shopping channel
		F	High	
		M	Low	
		dF	Low	
		dM	-	
		dR	-	

Table 5.8 characteristics and strategies for all customer segments

Segment	Sub-segment	Attribute	Attribute value	Strategies
Uncertain Segment	Sub-segment 1	R	Hugh	• Frequently promotion plans • Frequently incentives or offers
		F	Low	
		M	Low	
		dF	Low	
		dM	Low	
		dR	-	
Uncertain Segment	Sub-segment 2	R	High	• Frequently promotion plans • Frequently incentives or offers • Cross-selling • Special discount • Shifting to online shopping channel.
		F	Low	
		M	Low	
		dF	High	
		dM	High	
		dR	-	

Chapter6: Conclusion and Further Research

Conclusion

Contributions

Limitations

Future Works

6.1 Conclusion

Customer segmentation is a method for grouping customers based upon similarities they share with respect to any dimension, whether it is customer needs, channel preferences, interest in certain product features, customer profitability, etc.

Common customer segmentation objectives are developing new products and services, creating different marketing communications for different customer groups, developing different customer servicing and retention strategies, targeting company efforts to segments with the greatest profit potential and developing any strategy that may help the company in increasing its profits and customer retention.

Customer segmentation and definition of proper strategies for each segment can provide tremendous returns for companies. In this way, there are various models of implementing customer segmentation. Some of these methods are RFM, customer value matrix, CLV and data mining methods. But it must be considered that there is great value to keeping things simple, especially for small and medium sized businesses. Methods that are derived from complex statistical modeling techniques can provide useful information for experts but are hard to implement for these businesses and are likely to present a challenge to the development and implementation of strategies.

In this study Recency, Frequency and Monetary method which also known as RFM method has been used for customer segmentation in an Iranian internet service provider.

Customer data and their attitudes were mined in order to perform customer segmentation and consequently defining proper and useful strategies for having a better view of company customers and their behaviors and also increasing its profitability. Also company can recognize and classify an important or less important potential customer to set up proper marketing plan for those particular customers.

By definition of some new variables in RFM method, two new RFM variant methods have been proposed which have some advantages with respect to simple RFM model. The results of applying these new methods show their effectiveness for customer segmentation and also their ability in identification of customer behaviors especially the risk of cancelling company services.

Customers with different reflected purchase behavior must be treated differently. In order to convert this idea in to a computable parameter, all of the purchase amounts of customers in each period of analysis collected and then a parameter named change rate of purchase amount in each time section was defined. The computation of the slope of purchase amount in time can be based on a best-fit regression line plotted through the known x-values (which are time of purchase) and known y-values (which are purchase amount in each time section) or a new approach proposed in this project. This approach is based on computation of a new parameter which we named it discounted purchase amount slope (DPS). According to this approach, the purchase amount slope of each customer is computed by the sum of the discounted slopes of purchase amount in all time sequences. The discount rate determines the present value of past slopes. By defining this parameter, we reinforce the effect of recent purchase behaviors of customer in computation of total purchase amount slope while mitigating the importance of previous purchase slopes by inserting a discount factor.

The next variant of RFM method proposed in this project is based on the idea of value change rate stated in the above method and we named it Generalized Differential RFM or GDRFM. If we generalize the computation of purchase amount change rate to R, F and M parameters we can distinguish at risk customers and customer segments more adequately. It is based on the fact that customer behaviors such as decrease in the purchasing amount, decrease in the number of purchases, decrease in number of product categories purchased by customer and also increase in the length of time between shopping can be useful in predicting a potential decline in retention of customers. These indicators addressed by computing the change rates of RFM variables. In another word, not only considering RFM values is necessary for segmenting customers but also computation of derivatives of recency, frequency and monetary amount of customers with respect to time can be useful for obtaining better and more adequate results in segmentation.

The advantage of this method with respect to simple RFM method is on the fact that in GDRFM method, changes in behaviors of customer during the time is considered. Therefore changes in frequency and recency of purchase for a customer are taken in to account with change slope of purchase amount simultaneously. So, customers with positive monetary change slope and positive frequency change slope will be treated differently from customers with negative or different frequency and monetary change slopes.

The clustering algorithms used for segmentation of our data were k-means and EM algorithm. Finally, the detail description and specification of all segments found in our case study were presented and based on their specifications, some useful strategies were proposed.

The results of applying RFM method and its variants (GDRFM and FMC) for customer segmentation show that GDRFM and FMC methods can be so useful not only for segmenting customers with different values of recency, frequency and monetary but also they can indicate easily which customers are at high risk of cancelling services of company or falling from beneficial segments to other non-beneficial segments.

There are some points that must be considered here.

Since, there are too many methods for customer segmentation, and it is difficult to compare all of them it can be useful to develop an experiment for comparing the advantages and disadvantages between existing segmentation methodologies in the future.

The other point is that effectiveness of strategies defined for each segment must be studied and investigated by a separate analysis. This will guide us and conduct us to have a better understanding on usefulness or weakness of our proposed methods and strategies.

6.2 Contribution

In this project we proposed two new variants of RFM method which are GDRFM and FMC method. These methods use new variables that indicate the purchase behavior changes of customers. We also proposed a novel approach for formulating these changes by proposing a discount parameter. These discount parameters reinforce the effect of newly visited purchase change in comparison with old ones.

Customer segmentation using the GDRFM method provides a particularly viable alternative, simple to implement relative to the amount of effort involved and easy to understand method for companies. The ability in making difference among customers based on their behavioral purchase changes and identifying customers who are at risk of cancelling the company services are the main features of these newly proposed methods.

6.3 Limitations

However, this study suffers from some limitations.

The proposed methods require numerous customer data in order to be validated adequately. In this study, we had limited and incomplete information about customers, their purchase history and especially what services or product they had purchased from company.

Because of our incomplete and improper database we couldn't analyze the other customer segmentation methods and consequently it was impossible to compare our proposed methods with other methods.

6.4 Future Works

The future works proposed to be followed after this study, are as follow:

- Comparison of GDRFM method with other customer segmentation methods such as CLTV must be investigated.
- Checking the effects of changing γ (discount factor) on results of segmentation for monetary, frequency and recency can be investigated.
- Checking the effects of number of time steps on results and definition of optimum number of time segments is the other work that can be done in future.
- Comparison and analyzing the effectiveness of GDRFM method for large, medium and small size businesses must be studied. The advantages and disadvantages of the proposed methods in the case of different size companies can be analyzed and investigated.
- In the case of strategy definition, the recognition of effectiveness and profitability of these strategies and optimization of decisions based on their costs and revenue must be investigated in the future and in a specific study.
- …

References

Claudio Marcus, (1998), A practical yet meaningful approach to customer segmentation, Journal of Consumer Marketing, Volume 15, issue 5, pp.494–504.

Samira Madani, (2009), Mining Changes in Customer Purchasing Behavior - a Data Mining Approach. Master Thesis, Lulea University, Department of Business Administration and Social Sciences Division of Industrial marketing and e-commerce.

Brent A. Gloy Jay T. Akridge Paul V. Preckel, (1997), Customer Lifetime Value: An Application in the Rural Petroleum Market, Wiley & Sons, Inc. Agribusiness, Volume 13, No. 3, pp.335–347.

Jinghua Zhao, Wenbo Zhang and Yanwei Liu, (2010), Improved K-Means Cluster Algorithm in Telecommunications Enterprises Customer Segmentation. Information Theory and Information Security (ICITIS), IEEE International Conference on, pp.167 – 169

Xiaoping Qin, ShijueZheng, Ying Huang and Guangsheng Deng, (2010), Improved K-Means algorithm and application in customer segmentation. IEEE International Conference on Web Information Systems and Mining, pp. 13 – 16

Yuerong Chen and Xueping Li, (2009), The Effect of Customer Segmentation on an Inventory System in the Presence of Supply Distributions. IEEE International Conference on Winter Simulation Conference (WSC), pp.2343 – 2352.

Zhang xiao-bin and Gaofeng, Huang hui, (2009), Customer-churn Research Based on Customer Segmentation. IEEE International Conference on Electronic Commerce and Business Intelligence, pp.443 – 446.

Burcu Sagˇlam, F. Sibel Salman, SerpilSayın and MetinTu¨rkay, (2006), A mixed-integer programming approach to the clustering problem with an application in customer segmentation. Elsevier Conference on European Journal of Operational Research 173, pp. 866–879.

Zhou Zhongding, Miao Xuemei and Liu Guangcan, (2009), Customer Segmentation Algorithm of Wireless Content Service Based on Ant K-Means. IEEE Conference on

International Forum on Computer Science Technology and Applications, Volume 1, pp.267 – 269.

Izak Benbasat, David K. Goldstein and Melissa Mead (1987) The case research strategy in studies of information systems. MIS Quarterly, 11(3), pp. 369-386.

Mike McGuirk, (2007), Customer Segmentation and Predictive Modeling. Behavioral Sciences, marketing and analytic consulting firm located in Burlington, Massachusetts. Available at : www.emarketingpapers.com/content10131

Osmar R. Zaïane, (1999), CMPUT690 Principles of Knowledge Discovery in Databases, Available at: www.di.ubi.pt/~ddg/aulas/licenciatura/dwdm/dw/ch1s.pdf.

Xi Zhang and Yu Tang, (2006), Customer Perceived E-service Quality in Online Shopping. Master Thesis, Lulea University, Department of Business Administration and Social Sciences Division of Industrial marketing and e-commerce.

LalehNosrati, (2008), The Impact of Website Quality on Customer Satisfaction. A Research on Iranian Online Bookstores, Master Thesis, Lulea University, Department of Business Administration and Social Sciences Division of Industrial marketing and e-commerce.

Chun Wang Zheng Wang, (2006), The Impact of Internet on Service Quality in the Banking Sector. Master Thesis, Lulea University, Department of Business Administration and Social Sciences Division of Industrial marketing and e-commerce.

Yin, R. K. (2003), Case study Research Design and Methods (3rd ed.) California: Sage Publications.

Vasilis Aggelis and Dimitris Christodoulakis, (2005), Customer Clustering using RFM analysis, 9th WSEAS International Conference on Computers, Special Session Data Mining, Techniques and Application.

Chu Chai Henry Chan, (2008), Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. Elsevier Conference on Expert Systems with Applications Volume 34, pp. 2754–2762.

Su-Yeon Kim, Tae-Soo Jung, Eui-Ho Suh and Hyun-Seok Hwang (2006), Customer segmentation and strategy development based on customer lifetime value: A case study, Elsevier Conference on Expert Systems with Applications Volume 31, pp. 101–107.

Cheng Li (2008) Research on Segmentation implementation process of air cargo customer based on Data Mining, IEEE 4th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1-4.

Huaping Gong, Qiong Xia (2009), Study on Application of Customer Segmentation Based on Data Mining Technology, IEEE Conference on ETP International Conference on Future Computer and Communication, pp. 167 – 170.

Xin-an Lai (2009), Segmentation Study on Enterprise Customers Based on Data Mining Technology, IEEE First International Workshop on Database Technology and Applications, pp. 247 – 250.

Tian Yuanli and shaoliangshan (2010), Customer segmentation based on Ant clustering Algorithm, IEEE Second Conference On Computational Intelligence and Neural computing (CINC) Volume 1: pp. 133 – 136.

Babak Sohrabi, Amir Khanlari (2007), Iranian Accounting & Auditing Review, Volume 14 No. 47, pp. 7- 20.

Hee Seok Song, Jae kyeong Kim and Soung Hie Kim (2007), Mining the change of customer behavior in an internet shopping mall, Elsevier Conference on Expert Systems with Applications Volume 21 pp. 157±168.

Berger, P. D. and Nasr, N. I. (1998), Customer lifetime value: marketing models and applications, Journal of Interactive Marketing, volume 12(1), pp.17- 30.

Mirko Bo¨ttcher , Martin Spott , Detlef Nauck and Rudolf Kruse(2009), Mining changing customer segments in dynamic markets, Elsevier conference on Expert Systems with Applications Volume 36 pp. 155–164.

Gupta, S. and Lehmann, D. R. (2003), Customers as assets, Journal of Interactive Marketing, volume 17(1), pp.9- 24.

Jain, D. and Singh, S. S. (2002), Customer lifetime value research in marketing: a review and future directions, Journal of Interactive Marketing, Volume 16 , pp. 34- 45.

Jean-Paul Ruiz, Jean-Charles Chebat and Pierre Hansen (2004) Another trip to the mall: a segmentation study of customers based on their activities, Elsevier Conference on Journal of Retailing and Consumer Services volume 11 pp. 333–350.

Hwang, Hyunseok, Jung, Taesoo, & Suh, Euiho (2004), An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry, Expert Systems with Applications, Volume26, pp.181–188.

Tianyi Jiang and Alexander Tuzhilin(2006) Improving Personalization Solutions through Optimal Segmentation of Customer Bases, IEEE Sixth International Conference on Data Mining, ICDM, pp. 307 – 318

Reichheld, F.F. (1996), The Loyalty Effect, Harvard Business School Press.

Massnick, F. (1997), The Customer is CEO: How to Measure what Your Customers Want and Make Sure They Get It, Amacom.

Suqun Cao, Quanyin Zhu and Zhiwei Hou(2009), Customer Segmentation based on a Novel Hierarchical Clustering Algorithm, IEEE Chinese Conference on Pattern Recognition, pp. 1 – 5.

Minghua Han (2008),Customer Segmentation Model Based on Retail Consumer Behavior Analysis, IEEE International Symposium on Intelligent Information Technology Application Workshops, IITAW, pp.914 – 917.

Zhang xiao-bin and Gaofeng, Huang hui (2009), Customer-churn Research Based on Customer Segmentation, IEEE International Conference on Electronic Commerce and Business Intelligence, pp. 443 – 446.

HAO Su-li (2010),The Customer Segmentation of Commercial Banks Based on Unascertained Clustering, IEEE International Conference on Logistics Systems and Intelligent Management, Volume 1, pp.297 – 300.

Tan Haining, XuJuanjuan and Zhao Bian (2009), Research on Index System of Dynamic Customer Segmentation: Based on the Case Study of China Telecom, IEEE International Conference on Information Management and Engineering, ICIME, pp.441 – 445.

S. Sumathi, S. N. Sivanandam (2006), Introduction to data mining and its applications, Springer Book, Volume 24, pp. 204-207

Bezdek, James C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms

Paulo Batista, Mário J. Silva (2004), Mining On-line Newspaper Web Access Logs, IEEE International Conference on Information Technology: Coding and Computing, Proceedings. ITCC 2004, Volume: 1 pp.392 - 397

Hill, T. & Lewicki, P. (2007), Statistic Methods and Applications. Stat Soft, Tulsa, OK. Available on http://www.statsoft.com/textbook/cluster-analysis/#k

Hai Wang, and Shouhong Wang (2006), A Purchasing Sequences Data Mining Method for Customer Segmentation, IEEE Conference on Service Operations and Logistics, and Informatics, SOLI '06, pp.883 - 886

Software
Weka
SQL server 2000