International Business Management 10 (6): 1106-1119, 2016
ISSN: 1993-5250
© Medwell Journals, 2016

"Market Segmentation for Marketing of Banking Industry Products Constructing a Clustering Model for Bank Pasargad's E-Banking Customers Using RFM Technique and K-Means Algorithm"

[1]Arvin Fouladifar, [2]Elham Taghipour and [3]Arshad Hedayati
[1]Department of European and International Private Banking, Universite Nice Sophia Antipolis, France
[2]Department of Master of International Management, Universite Pierre-Mendes, France
[3]Department of Business Management, Semnan University, Iran

Abstract: Development of new business concepts like customer satisfaction, loyalty and relationship marketing obliged organizations to segment their market into smaller homogenous groups in order to have more precise marketing, advertising and product development strategies. Within present research, the importance of segmentation considering financial services sector is discussed with the emphasis on database marketing. A segmentation model for retail banking internet-based services is applied upon RFM and some demographic variables aiming to go more accurate through e-services market, in order to enable the marketers' establishing new strategies and a sample of 1478 E-banking customers is clustered via K-Means algorithm. Finally, based on the clustering results recommendations and suggestions is offered to enhance customer relations and developing services.

Key words: Market segmentation, E-banking customers segmentation, RFM, K-means, database marketing, market segmentation for internet banking services

INTRODUCTION

Classic accounting practices emphasize on assessing tangible assets on the balance sheet. Although, currently it is more typical for intangible assets such as brand, human resource and customer relationships to be the key elements of shareholders value (Sohrabi and Khanlari 2007). As Kotler (2000) mentions in his famous book Marketing Management, the change is an inevitable phenomenon which affects any business via globalization, technological advances and economic deregulations. Now a days, mass-marketing approach in any business keeps fading incessantly due to the Development of IT, distribution channels, surging globalization trend and numerous competitors showing off around the marketplace. At the current highly competitive markets of banking services where all financial institutions presenting more or less similar products and customer needs have become increasingly diverse, Mass Marketing is no longer the key strategy, while Industry owners and marketers face new terms such as customer satisfaction, loyalty, stickiness, marketing network, relationship marketing and CRM which could not be realized through a product-centric attitude implying One Size Fits All strategy to market (Kotler, 2000; Baker, 2003). Marketing has been defined by different phrases but none exclude the key concept of adding value to customers for a Profit. Kotler (2000) breaks customer's value down to the equation below:

$$\text{Value} = \frac{\text{Functional benefit} + \text{Emotional benefit}}{\text{Money cost} + \text{Time cost} + \text{Phisical cost} + \text{Energy cost}}$$

Kotler breaks benefit into human needs, Wants and demands. Needs are the most basic and vital life necessities like food, air, shelter, etc. and wants are described as the way different persons are pleased to meet their needs. Finally, Kotler mentions that with having market needs and wants marketers can do nothing unless there is an ability to pay for those needs and wants. Many need and want to possess a luxury car but few can afford one. That is why, we also need to have the third factor called demand. Now, we can better understand the marketing is truly depended on customers' demand and for using the most market potential available, the company first should be able to meet every single demand and attitude which, seems profitable to the business. Regarding many different attitudes which are spread over the market, industry owners could no longer treat all the

Corresponding Author: Arvin Fouladifar, Department of European and International Private Banking, Université Nice Sophia Antipolis, France

same, there should be market segments comprised of more or less same attitudes to be targeted with specified product(s) positioned to those segments. Following this stream, existing products also could keep being developed to maximize the customer value within a special segment. Hence, new strategies seem necessary to break the heterogeneous market into smaller homogenous groups. These fragments will allow marketers to establish a seemly strategy toward each group and generate new personalized products in order to make competitive advantage against the rivals and create value for the customers. In addition, this can help management to distinct profitable customers from loss makers (Kotler, 2000; Baker, 2003; Farquhar and Meidan, 2010).

Realizing different customers' attitudes is a key point to enhance customer relationship within a business. This is easily possible in micro-businesses like a small store but not in large ones like the banks business grows bigger. Banks are among those businesses which are benefiting a thorough database and this indeed could be the best instrument to analyze their customer behavior upon database marketing practices (Blattberg *et al*., 2008). Segmentation adds value by targeting appropriate with myriads of clients referring to hundreds of branches and automated channels. Thus, the importance of having customer's data appears more indispensable as a products to different customers and efficient usage of marketing resources and makes the organization flexible to emerging trend of markets (Kotler, 2000; Yankelovich and Meer, 2006).

Establishing a proper strategy also depends on the segmentation methods and the purpose of segmentation. Initially, we could have priori segments which are pre-defined by customer properties such as geographic attributes. In the next layer, we can divide market by demographic characteristics. Behavioral and cognitive segmentations also segment the market by psychological and behavioral attributes. Segmentation is also available upon business criteria like products or service channels. A bank easily can segment its customers into E-banking, telephone banking, mobile banking and branch customers (Farquhar and Meidan, 2010).

In spite of advantages which market segmentation brings, there are many known cases in which such approaches have been unsuccessful. Poor understanding of segmentation principles, incorrectly oriented literature and lack of real-world application guidance are only some of the possible reasons. Prior to the project's initiation, marketers should recognize the variables which will contribute to a successful result. The literature on these kinds of success factors is at an initial phase of progress. Though, proper planning, corporate level commitment and clear operation endorsements are only a few of the variables which are identified to help. During the segmentation, it is important to illuminate the qualities which segments should exhibit Dibb (1997) cited that Kotler's checklist which states that segments should be measurable, substantial, accessible and actionable and stable is probably the best known. After segmentation, further researches are needed to evaluate the segmentation success factors. Specifically, this should clarify how segmentation success can be quantified. It is essential to regard the differing orientations of academics and practitioners segmentations. For practitioners in particular, the segmentation literature remains hard to access. It is totally concerned with sophisticated questions of quantitative analysis with inadequate concentration on practical application questions.

Research questions:

- Do both behavioral (RFM) and demographic characteristics depict a good image of market segments?
- What are the characteristics of each cluster?
- What factors could also be used for more successful clustering (missing factors)?
- What are the proper strategies for each cluster?

Statement of problem: As discussed in the introduction, banking sector is a tough area of competing as a result of variety of competitors, similar products and diverse market place consisted of hundreds of attitudes.

Therefore, using the highest market potential, gaining competitive advantage and evading waste of resources and reducing costs, there should be a method of classifying and categorizing the assorted vast market into different standardized smaller segments for establishing a comprehensive bilateral relation between the business and clients. To do so, business owners and marketers have to find proper variables and segmentation methods to find hidden patterns buried in database to use them for making an efficient marketing strategy.

Literary review: Financial services seem less tangible in comparison to services whichcustomers typically use. The vast spectrum of such services offered by various financial institutes, the ability of being tailored to each customer's needs, close ties with ever-changing regulators' legislations, inclusion of many sophisticated calculations, etc., made them tough enough to be comprehended by clients. Hence, aiming to achieve competitive advantage, banks and financial institutes have tried to work on the cutting edge of financial services. Despite such investments, financial institutions are also need to establish special marketing strategies to

meet their goals. Although, macro-segmentation methods are also exist and many institutions still are working upon them but the advent of information technology in possessing complicated databases enable the banks to perform more sophisticated segmentations and sub-segmentations. Additionally, attempting to imply a single basis for segmentation (as psychographics, demographics, behavioral, brand preference or product usage) for all marketing decisions may result in incorrect marketing strategies as well as a waste of resources.

Many banks initially have macro-segmentations as personal, business and corporations which can be developed considering demographic and behavioral criteria to make sub-segments (Yankelovich and Meer, 2006; Farquhar and Meidan, 2010; Kotler 2000; Dibb, 1997; Durkin, 2004). Some researchers suggest that a priori segmentation which charges the researcher with defining the size and character of segments is the most widely used. This approach involves implication of demographic data. On the contrary, post-hoc approaches are less widely used and cover the grouping of respondents based on their responses to particular variables.

Multivariate methods such as cluster analysis and factor analysis can then be applied to these responses. Increasingly there is a focus on behavioral segmentation. The behavioral approach contrasts with the process of segmentation based on customer characteristics in that the focus is more driven by customer needs. It is argued that such a need identification approach is more robust than a classification of characteristics and that it is more probable that the segments that are consequently identified will be ultimately more predictive of purchase behavior. In support of this approach, it is argued that any approach to segmentation that is not focused on clustering customers according to their motivations is simply an approximation based on the assumption that descriptors and motivations are closely aligned usually they are not. A priori and post-hoc segmentation methods currently employed reveal little of predictive use to bank marketers. In support of this, qualitative research conducted with an international sample of senior bank executives found a lack of clarity in terms of what market segments these bank executives felt were best served by their internet banking proposition and what motivations of different customer groups were in adoption. Adopters of consumer innovations differ in their usage of technological products because of their behavior and motivations for usage. Thus, similar usage patterns may actually hide different motives for use. Their study based on financial service consumers identified distinct motivational clusters that were independent of the more established demographic segmentation variables banks used in targeting and communicating. The research suggests that customer motivations may be useful in predicting their response to new products as well as persuading them to use existing services for the specific benefits they value. The researchers conclude that all of the clusters identified needed to be informed about the benefits given their own specific personal motivations for managing money as many of the generic advertising and merchandising messages undertaken by the bank were not picked up on by these distinct clusters. Similarly, Machauer and Morgner (2001) adopted a cluster analysis methodology as they attempted to better understand customer perceived benefits of the bank relationship and in particular the E-banking channel. In summary, therefore, it is evident that there is a general lack of clarity as to what segments of banks' customer bases are adopting the Internet innovation for their banking and that the important issues of motivation and how decisions regarding adoption of the E-banking platform are not well served through traditional segmentation and profiling tool (Durkin 2004; Machauer and Morgner, 2001).

Key concepts description

Data mining: Data Mining (DM) is the procedure of extracting knowledge hidden in the large quantities of raw data by means of automated or semi-automated tools, in order to explore new patterns or rules. Data mining usually is conducted to reach two main objectives: first, predictions upon hidden patterns or relations between the quantities within the database and second, description which is actually the interpretation of database (Bhambri, 2011; Garg *et al*., 2008). Data mining methods are classified in two categories: descriptive and predictive. Classification is predictive method and clustering is descriptive method. Classification is the process of finding a model that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of items whose class label is unknown. Unlike classification and prediction which analyze class-labeled data objects, clustering analyzes data objects without consulting a known class label (Khajvand and Tarokh, 2011).

Clustering: One of the most common DM methods is clustering upon which, the similar quantities are collected into same clusters (Garg *et al*., 2008). In fact, it is about finding the sets in heterogeneous data by minimizing some measure of dissimilarity (Amiri and Fathian, 2007).Clustering is based on measuring Euclidean distance between two data observations and performs disjoint analysis computed from one or more variables which market researchers define. It is assumed that the

samples in each cluster have more similar characteristics rather than features in other clusters and clusters are dissimilar (Garg *et al*., 2008). Many researchers have adopted different approaches to solve the data clustering problem, including K-means, Self-Organizing Map (SOM), Genetic Algorithm (GA), Fuzzy System, Tabu Search (TS), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO, Support Vector Machine (SVM) and Artificial Immune System (AIS). Of them, K-means is an extensively adopted (Chiang *et al*., 2011).

K-Means algorithm: K-Means algorithms, originally known as Forgy's Method (Chenge and Chen 2009; Forgy, 1965) has been applied widely in different areas comprising of data mining, statistical data analysis and other business applications. Thus, this research offers the K-Means algorithm to shape the clusters by attributes. K-Means algorithms allow marketers to find out which segment could exist and so they are able to make suitable strategies to touch the optimal point (Amiri and Fathian, 2007; Sohrabi and Khanlari, 2007). To find k centers, the problem is defined as an optimization of a performance function, Perf (X, C), defined on both data items and center locations. A popular performance function for measuring goodness of the K clustering is the total within-cluster variance or the total Mean-Square quantization Error (MSE), equation below (Amiri and Fathian, 2007; Zulal and Alper, 2006):

$$\text{Perf}(x \times c) = \sum_{i=1}^{N} \text{Min}\left\{\left\|x_i - c_l\right\|^2 \mid l = 1,\ldots K\right\}$$

In K-means clustering technique, the number of clusters should be determined by decision maker (Khajvand *et al*., 2010).

RFM: RFM models have been being applying in direct marketing since 30 years, ago (Sohrabi and Khanlari, 2007). The model is the most frequently adopted segmentation technique that comprises three main factors: recency, frequency and monetary which can address different concepts in each case and research depending on industry, sector and nature of product (Wei *et al*., 2012). This analytical technique grew out of an informal recognition by catalog marketers that three factors seem mostly related to the likelihood that customers in their house data files would respond to specific offers. Customers who recently purchased from a marketer (recency), those who purchase many times from a marketer (frequency) and those who spend more money (monetary value) usually represent the best landscape for new offerings (McCarty and Hastak, 2007). RFM is a method to segment the market upon customers' behavioral characteristics usually related with how they react toward a single product or a group of them. With the advent of information technology, database marketing allows marketers to use such models usually with a large number of customers and their transactional information (Wei *et al*., 2010; Wu and Pan, 2009; Kotler, 2000).

The relative benefits and shortcomings of RFM and other models: Recently, new models have emerged into the RFM Model to increase predictability. For instance, Liu and Shih (2004, 2005) suggested two hybrid methods that exploit the benefit of a weighted RFM-Based Method which is called WRFM-Based Method or the preference-based Collaborative Filtering (CF) Method in improving the quality of recommendations of products (Wei *et al*., 2010). Their conclusions showed that the proposed hybrid methods are superior to the other methods.

Wei *et al*. (2010) quoted Rust and Verhoef (2005) which provided a fully modified model for optimizing multiple marketing interventions in intermediate-term (CRM) by conducting a longitudinal validation test to compare the performance of the model with that of the commonly used segmentation models in predicting the intermediate-term and customer-specific gross profit change, comprising demographic model, RFM Model and finite mixture models. Their consequences indicate that the proposed model outperformed classic segmentation models in forecasting the efficiency of the intermediate-term (CRM). McCarty and Hastak (2007) inspected different approaches for direct marketing segmentation, namely RFM, CHAID (Chi-square Automatic Interaction Detection) and logistic regression. Their results concluded that CHAID outperforms RFM in the situation that the response rate to mailing is low and the mailing would be limited to a small portion of the database.

Though, RFM is an acceptable technique in other circumstances. Wei *et al*. (2010) also mentioned Wang (2010) which adopted a hybrid method that incorporates kernel induced fuzzy clustering techniques to detect outliers efficiently and to segment customers more effectively, including robust possibility clustering method and robust fuzzy clustering method by using two real dataset, regarding the WINE and the RFM dataset to validate the hybrid method. The consequences exposed that the mentioned method can fulfill both robust classification and robust segmentation in the use of the noisy dataset (Wei *et al*., 2010).

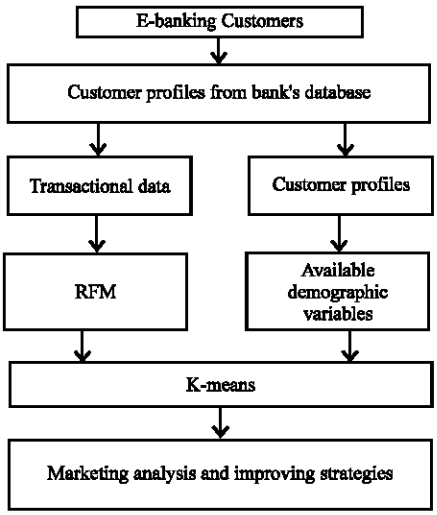

Fig. 1: Model architecture in general

Underlining the central model: Varun *et al.* (2012) developed a model to segment the market based on both demographic and behavioural characteristics of customers. They use K-means and SOM to cluster the database into homogenous segments. K-means is used for both RFM parameters and demographic attributes.

However, a SOM Model is used to classify all customers into clusters. Finally after clustering all data by K-Means algorithm and classifying them through a neural network, LTV prediction is conducted. The model is used within this research is the same nevertheless it uses judgemental sampling for selecting customers from the bank's databanks instead of SOM approach. Also the clusters, result after K-Means algorithm is analysed to give recommendations upon their behavioural and demographic characteristics in case of suggesting suitable products and services and advertising channels. Thus the model excludes CLV (LTV) predictions. The model architecture can be demonstrated in Fig. 1.

MATERIALS AND METHODS

Research statistical population and sample: Population for this study comprises of a sample of all Bank Pasargad'sE-banking customers who use electronic services through the bank's website during the Persian year 1390 (21 March, 2011-2012). RFM variables and transactional data could be easily traced via bank's core-software and also some customers' demographic characteristics are stored in the bank's databanks. This research used judgmental sampling because of the numerous difficulties with analyzing the massive volume of entire E-banking customer's data at this research level. Nonetheless, the sample size of this paper is seemly suitable in compare to similar researches with having >1,478 profile data and that is actually one of the research's strengths.

Data collecting method: In order to analyze Bank Pasargad's E-banking customers and cluster them via RFM model, data collecting procedure was conducted in following stages:

- Using literature review and interviewing with the banking experts: to develop the model framework in order to implement the RFM, the research uses existing models previously applied in other similar works, the components of the model are recognized
- Research Conceptual Framework: considering the research goals and questions, the relevant framework is carefully chosen and other aspects and their applications in management and strategic market planning are presented
- Using the Bank Pasargad's data: the next step in the research procedure includes extracting applicable data from bank's database in order to examine mentioned framework in real world. Regarding to the common RFM frameworks, transactional data comprising transactions amount, last login intervals and the times each customer visited the bank's website through the year are extracted by the bank's IT Department. Some demographic data also was requested like gender, age, occupation, customers' residence, etc. among which only age and gender were available to be analyzed within the framework beside RFM variables
- Final step: within this step, collected data from the bank's database is set as the foundation of research model to be clustered into different customer segments upon RFM variables

Resear ch procedure: To be successful in DM studies, a systematic approach is required. Research process is based on CRIPS-DM which is accepted as one of the most complete methodologies in Data Mining (DM) and includes six main phases.

CRISP-DM methodology: There are different methodologies presented for implementation of data mining projects among which one of the strongest methodologies is Cross-Industry Standard Process for Data Mining (CRISP–DM). Based on CRISP-DM, a given

data mining project has a life cycle involving of six phases. Note that the phase sequence is adaptive. That is, the succeeding phase in the sequence regularly depends on the consequences related to the preceding phase. The most significant dependencies between phases are indicated by the arrows. For instance, suppose that we are in the modeling phase. Depending on the behavior and characteristics of the model, we may have to return to the data preparation phase for further refinement before moving forward to the model evaluation phase. The six phases are as follows (Larose, 2006).

Business (Research) understanding: The first phase emphases on understanding the project objectives and requirements from a business viewpoint and then converting this knowledge into a data mining problem definition and an initial project plan designed to attain the objectives (Fernandez *et al*., 2002; Wirth and Hipp, 2000) as the following steps:

- Enunciate the project objectives and requirements clearly in terms of the business or research unit as a whole
- Translate these goals and limitations into the construction of a data mining problem definition
- Prepare a preliminary strategy for realizing these objectives (Larose, 2006)

Data understanding: Second stage is called as the understanding of the data content. First thing to do is getting the expressive data from database or data warehouses for the selected application. In the meantime, in this phase data quality and discovering first insights into the data are seen as two important aspects. The procedure can be summarized as follows:

- Collect the data
- Use exploratory data analysis to familiarize yourself with the data and discover initial insights
- Evaluate the quality of the data
- If desired, select interesting subsets that may contain actionable patterns (Larose, 2006)

Data preparation: The data preparation phase covers all actions to construct the final data set or the data that will be fed into the modeling tool from the early raw data. Actions include table, record and attribute selection as well as transformation and cleaning of data for modeling tools. The five steps in data preparation are data selection, data cleaning, data construction, data integration and data formatting (Shearar, 2000). The purpose of the cleaning step is picking unfitting or incorrectly entered data. Data conversion is necessary for recording data in different formats or values as some data mining algorithms work only with data in digital format. In this case, it needs to convert data in text format to the digital one. Purpose of the feature selection is determination of the most main parameters in predicting a value. It might be assigned many features to estimate a value. Though, it is not always simple to collect the determined data. For this case, by finding the effective properties data acquisition can be fast and simple. Additionally, data are divided into two groups as training and testing data (Saltan *et al*., 2011).

Modeling: During this phase, several modeling techniques are selected and applied and their parameters are regulated to the optimal values. Normally, a number of techniques exist for the same data mining problem class. Some techniques have specific requirements on the form of data. Thus, stepping back to the data preparation (phase 3) might be needed. Modeling steps comprise the selection of the modeling technique, the generation of test design, the creation of models and the assessment of models (Shearar, 2000). If the task is completely accomplished, in this case, selection of the right algorithm is much easier. Each task entails different algorithms and it is not known, which one provides the best outcome without constructing the model. It may be only possible to guess according to the condition of the data in hand. If there is a linear relationship between whole input and estimation variables, choosing the decision tree algorithm can be good choice. If there is a complex relation among the variables, in this situation neural network algorithm can be selected.

Evaluation: Before continuing to final deployment of the model built, it is essential to more comprehensively evaluate the model and review the model's construction to be certain it correctly attains the business objectives. It is critical to determine if some important business subjects have not been adequately reflected. At the end of this phase, the project leader then should decide precisely how to use the data mining outcomes. The key steps here are the evaluation of outcomes, the process review and the determination of the next steps (Shearar, 2000).

Deployment: Model creation is mostly not the termination of the project. The knowledge achieved should be organized and presented in a way that the customer can use it which often involves applying live models within an

organization's decision-making processes. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. Even though, it is often the customer, not the data analyst, who carries out the deployment steps, it is important for the customer to understand up front what actions must be taken to make use of the created models. The key steps are plan deployment, plan monitoring and maintenance, the production of the final report and review of the project (Shearar, 2000).Knowledge discovery uses data mining and machine learning techniques that have developed over a synergy in artificial intelligence, Computer Science, statistics and other related fields. CRISP-DM is a Flexible Model which can adopt with different data mining applications.

Research framework

Phase one

Describing the investigated company: Bank Pasargad was established in 2005. Bank Pasargad's initial capital was IRR3500 billion 2005 thereafter increased to IRR7700 billion in different phases. Subsequently, capital increased to IRR23100 billion with a 200% growth in 2010-2011 year which was the largest public subscription issue, ever occurred in Iranian Economical History. In year 2015, initial capital increasedto IRR50,400 billion.

The bank field of activity is retail banking services which are available through electronic ports and over 330 branches around Iran such as deposits, LCs, certificates, cards, exchanges, facilities, etc. and as well represents services like investment banking, insurance, stock exchange and so on, through its affiliates. In 2011, the bank entered to TSE (Tehran Stock Exchange) as a listed company. Bank Pasargad was the first in Iran in terms of Virtual Banking and innovative E-banking services as it is confirmed and registered by international union of invention and industrial innovation IUI5002(2012, Switzerland). Bank's E-banking services cover a range of services such as: inter-bank and intra-bank transactions, account balance reporting, bill payments, facilities reporting, installment payment, account signup, checkbook enquiry, checkbook reports, gift cards, modifying accounts (e.g., blocking, changing supportive account, etc.), secured payment, customers' club, etc. in the research date, the bank has over 500,000 internet users who use the mentioned services through the net.

The bank uses its own self-developed corE-banking software which is coded by an affiliate. The software made the bank able to store a thorough database of the customers as individuals are asked to fill out a form at signup time including personal identity (name, surname, date and place of birth and ID number), gender, marital status, home and workplace address, education, religion, prefix and contact details. However some fields are optional and the customers typically do not incline to update some of them (e.g., marital status), the database is quite detailed to be used as a database marketing source. Additionally, all transactional data, regarding the nature of banking services are stored to be exploited.

Business problem: Customer segmentation in order to improve CRM, enhance loyalty and satisfaction and product development, is considered as the research's main problem. As discussed before, now a days companies should go through more market details and present what the customers need to maintain their competitive advantage. To do so, the business strengths and weaknesses should be recognized. Here in this study some of them for the bank are listed as: customer segmentation could help the bank's marketers to establish a superior CRM system and also helps R&D to have a brighter view on market to develop current services or design new ones to be presented to the target groups. Furthermore, recognizing profitable customers can help directing resources more efficiently and enhances the mentioned group satisfaction and loyalty (Table 1).

Determining the research's objectives: The research objective is to suggest a segmentation model for Bank Pasargad's E-banking customers based on data mining techniques. There are a lot of researches on RFM and market segmentation and many of them study RFM beside other models like CLV to calculate the future life time value of each segment.

RFM can be easily applied in the banks as most of them benefits from a thorough database comprise transactional data and many demographic ones. As mentioned, before many research only use RFM to segment the market and other mix the parameters with other factors like RFMCI (RFM+Count Item) or use Weighted RFM (WRFM) and so. This study is use RFM combined with two demographic parameters including age and gender to see if there will be meaningful clusters resulted from the calculation. There are no coefficients as weight to any parameters.

Phase two; data preparing and preprocessing: This phase includes 4 stages: Eliminating defective and confusing data. Data extraction and creating data warehouse: The goal of this stage is creating an integrated data attained from different sources. Creating a data warehouse usually take place during the first stage

Table 1: Business strengths and weaknesses

Strengths	Weaknesses
Bank's self-developed powerful core-banking software and database	Absence of a specialized and discrete market research department within the bank
Ability of making changes and adding new features to the software anytime when needed	Some customer important data are not recorded or updated in order to go more precisely in segmentation,targeting appropriate products and cross-selling (e.g., customers' daily shopping records, occupation, hobbies, number of kids, vehicle or real estate ownership, etc.)
A powerful E-banking system and related applications like mobile banking software	
Benefiting from educated young colleagues and a well-institutionalized organizational culture within the bank	A very large group of customers are not familiar with internet based banking products and do not use them regularly
Branch's suitable geographical distribution	
Top management's (corporate level)positive attitude toward new techniques and modern channels	Internet infrastructural problems within the country

(above). According to Inmon (2005), a data warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of management's decision making process. This short definition presents the main structures of a data warehouse. The four keywords, subject-oriented, integrated, time-variant and nonvolatile, distinguish data warehouses from other data repository systems such as relational database systems, transaction processing systems and file systems.

Subject-oriented: a data warehouse is organized around major subjects, such as customer, vendor, product and sales. Rather than concentrating on the day-to-day operations and transaction processing of an organization, a data warehouse focuses on the modeling and analysis of data for decision makers. Hence, data warehouses typically provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.

Integrated: a data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files and on-line transaction records. Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures and so on.

Time-variant: data are stored to provide information from a historical perspective (e.g., the past 5-10 year). Every key structure in the data warehouse contains, either implicitly or explicitly, an element of time.

Nonvolatile: a data warehouse is always a physically separate store of data transformed from the application data found in the operational environment. Due to this separation, a data warehouse does not require transaction processing, recovery and concurrency control mechanisms. It usually requires only two operations in data accessing, initial loading of data and access of data (Inmon, 2005).

Transactional data scaling: within this stage, we scale the transactional data including the interval between last login to the bank's website, times each customers logged into the website during the year and the amount of all transactions he/she made during the year. The results and more details will be described in the following chapter.

Data normalization: this stage is not along with the others and indeed is taken place after the modeling process. Normalization of data means adjusting values measured on different scales to a notionally common scale, often prior to averaging (Dodge, 2003).

Phase three; modeling: Modeling embraces following four steps:

- Optimized cluster numbers is determined upon experts' point of view. Also other numbers are tested with the software (SPSS) to see whether the numbers of clusters are optimized or not
- Customers are clustered through K-Means algorithm
- If the result from K-means is compatible with which experts determined, we go through the next step. Nevertheless, the process will be repeated from the beginning
- After clustering, the characteristics of each cluster is determined and it would be prepared for making recommendations and planning (marketing strategies, CLV, etc.)

Phase four; evaluation: Finally and based on the guidelines given by the banking experts and also each cluster's characteristics, the results will be analyzed and recommendations will be suggested. This study is not going through other models which could use the results as input materials like CLV. However, further studies could benefit the research findings for implementing of other models.

Descriptive statistics: In order to make data analysis, descriptive statistical techniques such as frequency tables and statistical diagrams is applied and sample distribution in terms of variables like age and gender is investigated.

Population gender attributes: As demonstrated in Table 2, 1478 valid observations gathered from a 7000 person sample of over than 500,000 Bank Pasargad's

Table 2: Gender distribution

Male	Female	Total
1045	433	1478
71%	29%	100%

Table 3: Age distribution

Age	Occurrence	Percentage
18-30	681	46
31-40	411	28
41-50	233	16
Over 50	153	10
Total	1478	100

E-banking customers which were selected upon a judgmental sampling process.about 1045 persons are male (71%) and 434 others are female (29%).

Population's age attributes: Age groups within this research start at 18 and the eldest customer aged 89. The age distribution shown as follows (Table 3).

Applied software and analysis environment: This project carried out through two main software including Microsoft Excel 2010 and SPSS clementine.

Model implementation with data: Initially a data set of about 7000 Bank Pasargad's E-banking customers were dig out from the bank's database from 21 March 2011-2012 (Persian year 1390). About 80% of the data were recognized as invalid or confusing as most of the customers have not had any visits during the year and consequently the RFM parameters during that year had no value to be considered as a valid observation, about 1% of the whole data was also had missing parameters. A few customers had closed their account during the previous year and so are not counted as bank's customers to be analyzed.

Data preparation: The data came in different formats from different databanks and there were unnecessary fields like customer ID or branch code which in this study were impractical. At this stage, a data warehouse is created through Excel 2010 in an acceptable way for importing to SPSS. Thus, all fields were sorted and confusing of defective data was eliminated form data warehouse. At the end 1,478 valid rows created and organized to be imported to SPSS for later analysis.

Adopted parameters F, M and R: The R parameter which presents recency is defined based on the interval between the customer's last login to website before 21 March 2012. This is presented in days (e.g., 170 days) and the lower an interval is the higher score can be allocated to that customer. So, a customer with a 5 days interval gain a higher score that one with 50 days and therefore, he/she seems to be more loyal, F parameter presents recency which is assumed as the times a customer logged in to the bank's website during the research time span. It is given that more logins will cause more score in calculation the F. Finally M presents monetary which is considered as the total amount of transaction a customer made via internet banking system during March 2011-2012.

Demographic data: Fields in demographic data included gender, age, occupation, residence, marital status and education but as many data in some parameters seemed missing or confusing were eliminated and the research is conducted just on gender and age.

Modeling and software: As mentioned in literature review, segmentation goes through two different approaches: first segmentation based on customers' value and second, customer segmentation based on customer needs. First approach leads to a complete CRM system and more customer loyalty and satisfaction and the second one can be an appropriate starting-point for emerging strategies and product development plans.

There is a variety of algorithms to cluster different data. K-means is one of the most common algorithms to be used in clustering problems. It is categorized as an unsupervised classification through which we are able to cluster customers. A major weakness of the mentioned algorithms is that by which the hidden relations and patterns within the data warehouse cannot be revealed. As this study is not going to recognize those hidden patterns within the clustering process, K-means could be considered as the right tool to be used in clustering our data into homogenous groups.

Results attained by K-Means algorithm: Table 4 shows the initial cluster center of each parameter. Table 5 demonstrate the clustering progress in each stage. During the first iterations cluster centers will change slightly.

The final cluster centers in fact are the mean of each parameter within each cluster. These final centers actually reflect the customer characteristics in each cluster in terms of the variable which its final center is considered. Table 6 demonstrates that RFM parameters means are very close in each of three clusters but age mean in cluster 1 is more than two other.

Table 7 below shows the Euclidian distances between clusters. The more distance between the cluster exist the more we can have heterogeneous clusters and contrary less distance between the clusters creates less heterogeneous clusters. For instance the distance between cluster 1 and two is 9.936.

Table 4: Initial cluster centers

	Clusters		
Variables	1	2	3
Gender	1.00	1.00	2.00
Age	89.00	18.00	54.00
Monetary	0.00	0.00	0.00
Frequency	0.00	0.00	0.00
Recency	0.10	0.22	0.37

Table 5: Iteration history

	Clusters		
Iterations	1	2	3
1	8.629	9.873	7.116
2	8.967	0.404	0.195
3	5.010	0.000	1.016
4	2.780	.000	0.873
5	1.064	0.239	0.910
6	0.780	0.000	0.256
7	0.800	0.000	0.267
8	0.000	0.224	0.529
9	0.000	0.000	0.000

Table 6: Final clusters centers

	Clusters		
Iterations	1	2	3
Gender	1.68	1.71	1.72
Age	61.00	27.79	43.23
Monetary	0.00	0.00	0.01
Frequency	0.02	0.02	0.02
Recency	0.21	0.18	0.18

Table 7: Distances between final clusters center

Clusters	1	2	3
1	-	33.211	17.765
2	33.211	-	15.446
3	17.765	15.446	-

Table 8: ANOVA

	Cluster		Error			
Variables	Mean square	df	Mean square	df	F	Sig.
Gender	0.073	2	0.207	1475	0.350	0.705
Age	79219.328	2	21.730	1475	3.646E3	0.000
Monetary	0.001	2	0.001	1475	0.550	0.577
Frequency	0.008	2	0.005	1475	1.750	0.174
Recency	0.037	2	0.015	1475	2.443	0.087

Table 9: Number of cases in each cluster

Parameters	Values
Cluster	
1	121
2	961
3	396
Valid	1478
Missing	1

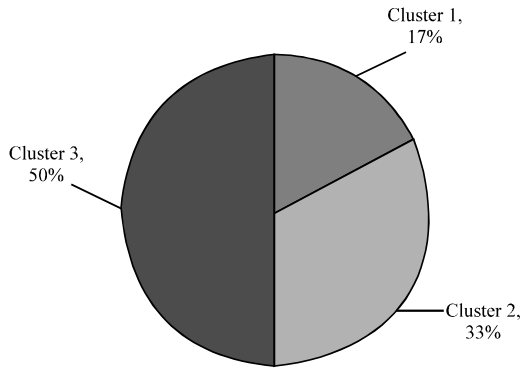

Fig. 2: Percentage of customers in each cluster

Table 8 shows that which variable has the most significant role within the clusters. That much F comes higher for a parameter that variable plays a more important role in splitting the clusters from each other. Within this research, age variable has the biggest role (F = 3.64 with sig. <0.01) and dissimilarly gender has the lowest effect on clustering procedure.

Finally, Table 9, illustrates how many customers located in each cluster. From totally 1.478 customers, 121 customers in fist cluster 961 in second cluster and 396 customers are categorized in cluster 3.

As in calculation of RFM parameters the nature of all factors are not the same and a huge amount or small amount of a single parameter could affect the calculation, all fields are normalized and converted to a value between 1 and 0. Normalized values are calculated through the formula below:

$$\upsilon = \frac{\upsilon - \min_A}{\max_A - \min_A}$$

After the results shows how many customers belong to each cluster and effective parameters which play the most significant roles in each one, analysis could be conducted over the table of final cluster centers (Table 6, Fig. 2). As it is demonstrated in cluster 1 includes customers with older ages and higher recency score in compare to the two other clusters. With this in hand we can construct proper strategies upon ethology studies, research realities and literature review.

CONCLUSION

Within present research a framework for segmentation Bank Pasargad's E-banking customers is established and some demographic parameters such as age and gender are also considered within the applied model. Extracting other important demographic variables was unavailable due to lot of missing data and confusing values. Though, the sample is considered rich for conclusion in compare to similar papers but the huge amount of eliminated invalid data (about 5520 customers) is considered as a remarkable problem. The data is related to the customers' activities during the Persian Year 1390 (21 March 2011-2012) and none of the customers within

the database were signed up to the bank during the year. In the next stage, the RFM parameters were defined for the sample and all data was normalized to be used with SPSS software.

F stands for frequency which means the times each customer entered to the bank's website to use E-banking services and it includes every kind of services like balance check, financial transactions, paying bills, etc.

Finally, M is for monetary parameter by which the total amount of a customer's financial transactions is defined.

Organized and normalized data from data warehouse is imported to the SPSS software in the next stage and K-Means algorithms implanted to create the supposed clusters.

According to the results, recommendations and suggestions is given to be used by marketers or applied for further works like calculation CLV for each cluster.

RECOMMENDATIONS

Recommendations upon clustering results: This study clusters E-banking customers who have visited the Bank Pasargad's E-banking system during the year. The variables extracted from bank's database can show meaningful relations but for a thoroughgoing study, more variables and parameters are needed.

As it is demonstrated in Table 3-9, K-Means algorithm and experts' point of view split the customers into three different clusters. Cluster 1 includes17% of the database and contains people with average age of 61 years old who merely visit the banks website and make few transactions(amount and number). Cluster 2 comprise of young customers who are 33% of whole data with a higher visit times but also make few transaction (amount and number) and finally cluster 3 contains 50% of customers with average age of 43 years old, higher transactions' amount and higher visit rate.

Frequency and gender have not affected clustering and it can be claimed that the data set is homogenous in terms of this variable.

Cluster 1 comprises 17% of customers with the average age of 61 years old with a long recency which indicates that this group of customers seldom uses E-banking services which could be due to their age and unfamiliarly with new channels like internet. Also from the table, we can realize that the amount of their transactions in the cluster is not remarkable and thus they cannot be considered as profitable customers in short and long term due to their age. As debated in previous chapters, non-profitable customers should be recognized to avoid waste of company's resources and therefore investing on this segment does not make sense. Though, the bank can conduct programs to make these customers familiar with E-banking services to enhance their involvement and stickiness with e-services and increase their loyalty. This will significantly depend on the cost-and-benefit analysis the marketers make on this segment and this highlights the role of further studies like CLV to calculate customer values to evaluate whether investing on mentioned group is beneficiary or not. At this stage, cluster 1 is just considered as the least profitable segment with no bright prospect due to the age attributes and subsequently investing on this cluster takes the last priority. Another issue is the way company can inform mentioned customers about the bank's services (advertisements). It is obvious that the internet channels are not the best choice as this cluster has not regular visits to the website. Thus, alternative media are suggested to target the customers within the cluster. Finally, we should consider that the most important factor for this cluster in terms of internet services is simplicity of working with E-banking menus. Thus, a suitable product the bank can offer to this cluster could be a simplified versionof E-banking website which presents e-services in a more user-friendly environment. Products such as pensions and health insurance policies are also could be presented through E-banking systems to encourage this cluster to use e-services.

Cluster 2 embraces 33% of total customers with average age of 27. It is obvious the intention of using e-based services is higher in compare to the first cluster ($R_1 = 0.21$ in compare to $R_2 = 0.18$). However, these customers do not make many transactions but they regular visit from bank's website is higher. Maybe at this time, they are not considered as the most profitable customers but as they are young they can have a bright prospect to be potential profitable users. Therefore, investing on this cluster seems logical to get them involved with the bank's services. The skill of using online services and their regular visits which this cluster benefits from, paved the way for internet base informing channels to make them familiar with the bank's features.

Note that the level of investment on each group should be precisely investigated with models like CLV. Many incentives could be offered to this group aiming to increase their participation. Designing gifts in proper with their internet usage seems reasonable to enhance their involvement and loyalty. Some new products also can be designed targeting the youth needs like sport and travel insurances, mortgages or more sophisticated e-services such as virtual debit card which is now available by the bank.

Clustering results show that the most profitable customers can be found within cluster 3 as they more regularly visit the bank's E-banking website. They use the e-services more and make more transactions with remarkable amounts. As they make up a half of the customers and are middle-aged, they can be considered as the long term loyal partners who are involved with business activities (in proper to their age) and gain higher income to put into bank deposits. This is a very bright sign indicates that at least half of the E-banking users are profitable to the bank. As one of the main clustering objectives are distinguishing profitable customers from non-profitable ones to allocate the company's resources in an efficient manner, the results should be used to recognize the customers within the cluster 3 in order to allocate more attention and care. Resources should be conducted to this group of customers in terms of internet services. Increasing their maximum amount allowed for daily transactions and withdraws or providing them with incentives through other channels and services like an insurance premium discount or higher credit card amount can be considered as some of recommended services.

Similarly, the results can be used to develop CRM programs to define customer care services toward mentioned group. Their information can be imported into CRM software to define services such as calling priority when they call bank's call center.

Recommendations upon research process: One of the most involving difficulties during the research process was the lack of key factors such as customers' occupation, residence, education level, marital status, shopping records, etc. Unfortunately, some of them like shopping records are not systematically recorded in bank's database due to the variety of payment channels, cultural issues (Generality of cash payment instead of e-payment) and telecommunications infrastructures. Other variables like occupation, marital status and residence are not accurate as people might state confusing terms or the changes might not be updated. So, no market research could achieve its perceived objectives without having enough variables and the results do not have the perfect quality. So, it is highly recommended that such parameters be stored with more accuracy to feed market researches.

Establishing a discrete specialized unit for market studies may help speeding up the process. Moreover, it is suggested that marketer have access to the bank's database with the ability to investigate any part of market by any desired variable(s) to conduct more effective marker researches.

Also, it is suggested that variables like customers deposit amounts be modified to be used beside the other parameters in clustering models. Using the customers' deposit amounts (e.g., IRR 4,560,000) beside other parameters causes irrelevant results in clustering. Amount should be standardized in terms of scores given to each customer in proper to his/her deposit amount and by that mentioned parameter could play an important role in clustering process.

Further studies: Within this part several suggestions for further researches is given based on research results. Market segmentation is the base action of other market studies and strategies. Processes like targeting and positioning are remarkably depended on the results of market segmentation. This study focuses on RFM variable but other similar researches can be conducted to gain more market share or designing more competing products. Some recommended projects can be listed as follow.

It is suggested that for further studies market segmentation be conducted based on other variables which are missing in present paper (e.g., occupation). Also, the significance of each parameter be determined by experts do conduct a weighted RFM model for more accuracy. Also, it is recommended that segmentation be performed with other Models like neural networks and other clustering algorithms. Then, the marketers can compare the results and decide which model works more efficiently for the bank. Additionally, it is suggested that same researches can be performed to segment other markets (e.g., loans, deposits, etc.).

Further, researches can be fed by the entire database instead of a part of it in order to gain more accuracy in findings and conclusions.

Related researches can be executed to calculate customers' CLV for each cluster to have a more clear view of the costs or investments that is made for each group aiming to see whether the costs or investments on a special group of customers is efficient or not. Additionally, calculating CLV can be used to predict the bank's revenues and company valuation. CLV similarly can be used in some researches which are aimed to predict paradigm shifts in marketing and revenue management projects.

Further researches also can be executed to be used in Customer Lifecycle Management or CLM which is the measurement of multiple customer related metrics which when analyzed for a period of time, indicate performance of a business. The purpose of the customer life cycle is to

define and communicate the stages through which a customer progresses when considering, purchasing and using products and the associated business processes a company uses to move the customer through the customer life cycle. Additional research could be conducted to perform customers' risk assessment that can be used in credit card services, mortgages and loans.

REFERENCES

Amiri, B. and M. Fathian, 2007. Integration of Self-organizing feature maps and Honey Bee Mating Optimization Algorithm for Market Segmentation. J. Theoreti. Appli. Inform. Technolo.

Baker, M.J., 2003. The Marketing Book. Linacre House, Jordan Hill, Oxford OX2 8DP.

Bhambri, V., 2011. Application of Data Mining in Banking Sector. IJCST, Mandi Gobindgarh, Punjab, India, Vol. 2, Issue 2.

Blattberg, R.C., B. Kim and S.A. Neslin, 2008. Database Marketing. New York: Springer Science and Business Med., LLC.

Cheng, C. and Y. Chen, 2009. Classifying The Segmentation pf Customer Value via RFM Model and RS Theory. Expert Systems with Applications, Elsevier, 36: 4176-4184.

Chiang, M.C., C.W. Tsai and C.S. Yang, 2011. A time-efficient pattern reduction algorithm for k-means clustering. Elsevier. J. Information Sci., 181: 716-731.

Dibb, S., 1997. Market Segmentation: Strategies for Success. MCB University Press: UK.

Dodge, Y., 2003. The Oxford Dictionary of Statistical Terms. OUP. ISBN 0-19-920613-9 (entry for normalization of scores).

Durkin, M., 2004. In Search of the Internet-Banking Customer, Exploring the use of Decision Styles. The Intl. J. Bank Market., 22 (7): 484-503.

Farquhar, J.D. and A. Meidan, 2010. Marketing Financial Services. London: Palgrave.

Fernandez, I.B., S.H. Zanakis and S. Walczak, 2002. Knowledge discovery techniques for predicting country investment risk. Computers and Industrial Eng., 43: 787-800.

Forgy, E., 1965. Cluster analysis of multivariate data: efficiency versus interpret ability of classifications. Biomertrics, pp: 21, 768.

Garg, K., D. Kumar and M.C. Garg, 2008. Data Mining Techniques for Identifying the Customer Behavior of Investment in Life Insurance Sector in India. Internat. J. Informat. Technolo. Knowled. Manage., 1 (1): 51-56.

Inmon, W.H., 2005. Building the Data Warehouse. Wiley Technology, Original from the University of Michigan.

Khajvand, M., K. Zolfaghar, S. Ashoori and S. Alizadeh, 2010. Estimating Customer Lifetime Value Based On RFM Analysis of Customer Purchase Behavior: Case Study. Procedia Comput. Sci., Elsevier, 3: 57-63.

Khajvand, M. and M.J. Tarokh, 2011. Estimating Customer Future Value of Different Customer Segments Based on Adapted RFM Model in Retail Banking Context. Elsevier Ltd., Procedia Computer Sci., 3: 1327-1332.

Kotler, P., 2000. Marketing Management Millennium Edition. 10th Edn. Pearson Custom Publishing: USA.

Larose, D.T., 2006. Data Mining Methods and Models. New Jersey: John Wiley and Sons, Inc.

Liu, D.R. and Y.Y. Shih, 2004. Integrating AHP and data mining for product recommendation based on customer lifetime value. Informat. Manage. Elsevier, 42: 387-400.

Liu, D.R. and Y.Y. Shih, 2005. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. J. Syst. Software Elsevier, 77: 181-191.

Machauer, A. and S. Morgner, 2001. Segmentation of bank customers by expected benefits and attitudes. Intl. J. Bank Marke., 19 (1): 6-17.

McCarty, J.A. and M. Hastak, 2007. Segmentation approaches in data-mining: A comparison of RFM, CHAID and logistic regression. J. Busin. Res., Elsevier, 60: 656-662.

Rust, R.T. and P.C. Verhoef, 2005. Optimizing the Marketing interventions mix in intermediate-term CRM. Mark. Sci., 24 (3): 477-489.

Saltan, M., S. Terzi and E.U. Kucuksille, 2011. Back-calculation of pavement layer moduli and poisson's ratio using data mining. Expert Syst. Application, 38: 2600-2608.

Shearar, C., 2000. The CRISP-DM Model: The new blueprint for data mining. J. Data Warehous., 5 (4): 13.

Sohrabi, B. and A. Khanlari, 2007. Customer Lifetime Value (CLV) measurement based on RFM Model. Iranian Account. Audit. Rev., 14 (47): 7-20.

Varun, K.M., C.M. Vishnu and M. Madhavan, 2012. Segmenting the Banking Market Strategy by Clustering. Intl. J. Compu. Applicat., 45 (17).

Wang, C.H., 2010. Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques. Expert Syst. Appl., 37: 8395-8400.

Wei, J.T., S.Y. Lin and H.H. Wu, 2010. A Review of the Application of RFM Model. African J. Busin. Manage., 4 (19): 4199-4206.

Wei, J.T., S.Y. Lin, C.C. Weng and H.H. Wu, 2012. A case study of applying lrfm model in market segmentation of a children's dental clinic. Expert Systems with Applicati., Elsevier, 39: 5529-5533.

Wirth, R. and J. Hipp, 2000. CRIPS-DM: Towards A Standard Process Model for Data Mining. In Proceedings of The 4th International Conference on The Practical Applications of Knowledge Discovery And Data mining, Manchester, UK, pp: 29-39.

Wu, H.H. and W.R. Pan, 2009. An Integrated Approach of Kano Model and ANOVA Technique in Market Segmentation: A case of a coach company. J. Stat. Manage. Syst., 12 (4): 679-691.

Yankelovich, D. and D. Meer, 2006. Rediscovering Market Segmentation, Harvard Business Review.

Zulal, G. and U. Alper, 2006. Kharmonic Means Data Clustering with Simulated Annealing Heuristic. Appli. Mathemat. and Computation.