J. Appl. Environ. Biol. Sci., 5(5)458-464, 2015

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Classify the Data of Bank Customers Using Data Mining and Clustering Techniques (Case Study: Sepah Bank Branches Tehran)

Majid Sharahi[1], Mansoureh Aligholi[2]

[1,2] Department of Business Management, Faculty of Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

Received: November 2, 2014
Accepted: February 11, 2015

ABSTRACT

The purpose of this research is segmentation of bank customers using clustering techniques and is providing marketing strategies for each cluster of customers. Nowadays due to the large amount of customer data entities in the banking industry, the analysis of data obtained from the data bases of customers can provide useful information to detect hidden patterns in the data and can improve the level of banking services to each group of customers. In this study, we use the Excel file contains the information of 60companies from customer of Bank Sepah in database of this bank and analyze it after got approval from the relevant authorities. Data analysis is performed using the Two-steps and K-means clustering algorithms which are from data mining methods. The file contains 14 data fields and after processing of data 7 fields are selected as a final variable to enter K-means algorithm. These fields are "Type of Company", "Life time of Company", "Type of Activity", "Time of Collaboration with Bank", "Credit history", "Type of Credit", "Amount of credit". Data were categorized into two clusters based on theses even fields which 17 companies were classified into the first cluster and 43 companies were classified into the second cluster. First cluster companies often use the credit of civic participation in Bank and have the credit in the bank more than 10 years but the second cluster companies have the credit in the bank less than 5 years and use bank credit less in compared with the first cluster. We offer some suggestions about marketing strategy for each cluster after analyzing information of clusters. One of these suggestions is satisfaction strategy which could include 1- reduce services costs, 2- providing better conditions for the use of electronic service systems, 3- development of customer-oriented culture, 4-support of customer of production division as constant customers of Bank. These approaches while increasing profitability may also affect economic prosperity and employment.
KEYWORDS: Customer Relationship Management, Marketing Strategy, Market Segmentation, Data Mining, Clustering.

INTRODUCTION

Today, different organizations like banks require proper understanding of customers' behaviors and their demands in order to be successful in their businesses. In this regard, customer relationship management (CRM) is a useful instrument for acquisition of banks customers; satisfaction and increase in customers' loyalty. In fact, banks should concentrate on long-term relationship with their customers in order to increase their market share. This necessitates classification of customers based on differentiating features in different groups and preparation of appropriate marketing strategies for every class of customers. Today, considering the development of competitive environment among organizations and production and service companies, market segmentation plays an important role. In banking industry, analysis of banks customers' databases can provide us with useful information. However, it must be noted that acquisition of applied knowledge from patterns of mass data will not be possible by traditional statistical methods. Data mining is a new branch of science which can help managers with understanding customers and providing them with loans. Clustering analysis is a new technique in data mining which is used for data classification and it has been considered as one of the most popular instruments for market segmentation especially in banks (Shin and Cho, 2006). Considering the above contents, this research aims to classify bank customers using clustering analysis.

Research theoretical bases

Banks customers' classification

Iranian banking industry is one of the most effective bases of Iranian economy and plays a decisive role in economic activities. The distance between productivity of banking industry in Iran and banking industry all over the world indicates a deep gap in productivity in comparison with international standards. Considering the restrictions in banking resources for acquiring maximum productivity, relations between banking industry and

***Corresponding Author:** Mansoureh Aligholi, Assistant Professor, Department of Business Management, Faculty of Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran.
Email: man.aligholi@iauctb.ac.ir, Tel: (++98)21-66921499.

customers as the main resource of income and organizational success should be revised basically. Performance of banking system is productive when resources are spent on the best customers and customers who receive services can have long-term satisfaction with the services and remain loyal to the bank. In other words, if banking services are satisfactory, maximum productivity is reached (Asghari and Amin, 2005). CRM and its processes and technologies can play an effective role in promotion of Iranian banking system services. Within the past few years, CRM has converted from operational form to analytical form. Operational CRM is a kind of system in which all customers' data are gathered via contact points like contact centers, contact management systems, post, fax, salespeople and web. The data are stored and organized in a middle database. This database is accessible for all employees and users related to customers. This system can provide a complete and comprehensive investigation of customers' data. This kind of CRM has the advantage of privatization of relationship with customers and expansion of organizational responses required by customers. On the other hand, CRM is a systematic analysis by which data stored in database are analyzed by some instruments and customers' profiles are produced. Moreover, behavioral models of customers are identified, their satisfaction levels are specified and customers are classified in similar classes. Knowledge obtained from analytical CRM is effective in promotion of marketing strategies. Technologies which support analytical CRM (ACRM) include databases, analysis and prediction motors, pattern discovery association rules, sequential patterns. These led to classification of customers and offering better services and products to every group of customers (Elahi and Heidari, 2005). Today, in banking information systems, ACRM makes use of methods for analyzing data acquired from operational CRM and prepares its results for performance management (Albadavi et al, 2005). Therefore, banks can obtain useful information about their customers using ACRM and they can concentrate on discovered patterns and analyze them in order to make market progress. For instance, if a bank knows that which customer will respond positively to a new product can offer the product effectively or if an organization can predict goodness or badness of a new customer, it can avoid many costs spent on bad customers and spend them on good customers (Moradi, 2001). In general, a bank or national institute should rank and classify its customers in order to maximize its productivity via offering services and products to those customers which have higher added values. Today, classification is a key method which is used by banks in order to understand customers in competitive environments (Durkin, 2004).

Clustering

Data mining means extraction of knowledge and discovery of latent patterns in large databases. Data mining and discovery of valuable information from large databases is an attractive field of study which has received a lot of attention within the past two decades. In fact, data mining aims to create models for decision-making. Different data mining techniques including clustering, classification, decision trees, regression, association rules, succession models and artificial neural networks allow analysts to uncover latent knowledge in raw data and predict future trends based on past trends (Shin and Chu, 2006). Clustering means division of a large group of observations into smaller groups so that observations in each group are relatively similar to each other and different groups observations are relatively different. Clustering is a data mining technique which has many applications in scientific areas especially in management. Data mining techniques are divided into two groups: learning with supervision and learning without supervision. Contrary to classification, clustering is one of the without-supervision learning techniques. In classification, the number and features of classes are specified in advance and we aim to put data in pre-determined classes. In clustering, however, we do not have any knowledge about structure and features of clusters and we want to put similar data in clusters. Therefore, classification is learning with supervision and clustering is learning without supervision (Levang Jing, 2002). Clustering algorithms are classified into two groups: hierarchical and non-hierarchical. Hierarchical clustering methods are also classified into two groups: top-down methods and bottom-up methods. Top-down hierarchical clustering starts with a cluster including all data and appropriate clusters are separated from this large cluster. In bottom-up hierarchical clustering, all data are considered as a cluster at first, then we combine similar data and clusters become larger at each stage. The output of hierarchical clustering on a data set is displayed with a tree-like structure called dendogram. Two Step clustering algorithm is a hierarchical clustering technique which is used for elimination of odd variables in large volumes of data before implementation of non-hierarchical clustering algorithms. The most famous non-hierarchical clustering algorithms include K-means method and fuzzy C-means method (Tior et al, 2005). K-means algorithm is a common non-hierarchical clustering method in classification of data. This algorithm has a parameter called K which indicates the number of clusters. In this method, the researcher should specify the number of clusters contrary to hierarchical clustering methods. This algorithm procedure involves selection of points randomly as many as required clusters. Then, the problem data are corresponded to one of these clusters considering the degree of proximity to these points. Therefore, new clusters are obtained. In new clusters, we specify the centers of clusters by taking average from data in each cluster. By repetition of this procedure, we can find new cluster centers each time and then we can make data correspond to new clusters. This trend continues until when no other change is resulted in data. K-means clustering algorithm covers an evolutionary procedure and clusters centers move towards optimization in each algorithm repetition (Chiang, 2014).

Research background

Taghva et al (2009) analyzed bank customers' data by means of data mining techniques and used it as a means for improvement of CRM management in banks. They used data mining instruments like clustering and regression analysis as means for applications like marketing, risk management, deception detection and attraction and maintenance of bank customers. Khan Babayee (2010) conducted a research titled: "application of clustering techniques and genetic algorithm in making decision trees for classification of banks customers". They used banks customers clustering for validation of banks customers and granting credit loans to them. He analyzed data obtained from one thousand bank transactions made by bank customers including checks status, time period, credit background, credit value, saving status, work experience, number of installments, personal status, age, gender, residence area, assets and properties, present credit status, job, number of children and so on by means of K-means and Two-steps clustering techniques and genetic algorithm and specified a routine for identification of valuable customers for granting bank loans (Khan Babayee, 2010).Moradi (2011) conducted a research titled: "classification of Mellat Bank customers in Arak City and determination of CRM strategies in each section" and clustered bank customers data by means of RFM and K-means methods. Alfansi&Sargeant (2000) conducted a research in Indonesia. In this research, they used principal components analysis and identified 8 factors for clustering of customers and put customers in three clusters by using cluster analysis. Previous studies conducted on clustering bank customers' show that managers and researchers use different criteria for classification of banks customers market. One of these methods is classification based on demographic variables. This kind of classification involves segmentation of market into groups based on demographic variables like age, gender, family lifetime period, job, education level, religion, culture and nationality. Use of these variables is very common in market segmentation because its measurement and control is relatively easy (Kotler and Armstrong, 2004). Bobinski, S. & Assar (1994), Mathur&Moschis (1994) were researchers who found significant differences in behaviors and demands of bank customers with different demographic features. Another method in classification of banks customers is use of psychological variables. in psychological segmentation, customers are classified into different groups based on lifestyle or personality. A qualitative example of psychological segmentation is Harrison's study (1994). In his research, four different segments of banks customers were identified based on knowledge level and financial maturity degree. In another research, Lawson and Todd (2010) classified customers based on their payment methods preferences. Kruke et al conducted a research in 2001 and investigated application of data mining in banking industry and after implementation data mining techniques; they used them for identification of valid customers and presentation of credit cards to them and verification of their loans. Verhouf et al (2002) used data mining technique for classification and prediction of customers' information in Netherlands. Ture et al (2005) compared useful data mining techniques in marketing and investigated the roles of these techniques in prediction of customers behaviors. Considering the contents and studies conducted in this field, the present research tries to gather and prepare banks customers' data and implements clustering methods for them and the results are offered in the form of useful models for classification of banks customers and promotion of bank services.

RESEARCH METHODOLOGY

The present research aims to respond to a quantitative and applied scientific problem and it is of survey nature. It deals with topics like market segmentation and data mining techniques and it also investigates previous studies regarding research subject. it is also a descriptive study and since customers' raw data file was used for analysis, it is a descriptive survey. SPSS statistics (version 20) was used for data analysis. After descriptive investigation of data, Two Steps clustering method was used for determination of variables which are effective in clustering and the number of clusters for implementation of K-means clustering algorithm. Statistical population of the research included 60 companies which were legal customers of Sepah Bank. They were selected from branches of Sepah Bank in Tehran City randomly. Data obtained from them were analyzed from 2009 till the end of 2013. Data base of legal customers of Sepah Bank was used for data gathering. The data were in the form of an Excel File and were prepared for implementation of data mining after elimination of incomplete data. the Excel File had 14 data fields which included company type, activity type, company lifetime, the time of cooperation with bank, different types of bank accounts of company, credit background, loans types, collateral type, the number of rebounded checks, use of POS device, the number of letters of guarantee received, the number of letters of Credit, amount of loans received and amount of receivables.

Data analysis

Descriptive statistics

Regarding "company type" variable, 50 companies out of the 60 companies (73.3%) were private stock companies. 5 companies were limited liabilities companies, 4 companies were public stock companies and 1 company was of cooperative type. Regarding "activity type" variable, 32 companies out of 60 companies (53.3%) were production companies. 19 companies had commercial activities and 9 companies were service companies. Regarding "company lifetime" variable, 31 companies out of 60 companies (51.6%) had lifetime above 15 years. 22 companies had lifetime between 7 to 15 years and 7 companies had lifetime below 7 years. Regarding "cooperation with bank" variable, 24 companies (40%) had a life time between 5 to 10 years and 24 companies had a background of cooperation with Sepahbank over 10 years. The remaining 12 companies had lower than 5 years of cooperation with bank. Regarding "different types of bank accounts" variable, every company can select one of the following options: "interest-free account", "traditional current account", "golden current account", "ATM account", "short-term account" and "long-term account", or simultaneously several options.

The results of descriptive statistics showed that 26 companies out of 60 companies (43.3%) had traditional current accounts, 4 companies had golden current accounts and 2 companies had only short-term accounts in Sepah Bank. Furthermore, none of the companies had only interest-free account or only ATM account or only long-term account. The other 28 companies had combinations of several different accounts I Sepah Bank. The maximum number belonged to owners of both traditional current accounts and short-term account (12 companies out of 28 companies). Regarding "credit background" variable, 27 companies out of 60 companies (45%) had credit backgrounds below five years. 18 companies had credit backgrounds between 5 to 10 years and 15 companies had credit backgrounds above 10 years. Regarding "loans type" variable, every company can have one or several options: "interest-free loan", "civil partnership", "installment sales", "dormant partnership", "hire-purchase", "futures", "compensation", "debt purchase", "Murabaha". In this research, 11 companies out of 60 companies (18.3%) used only civil partnership loans. 7 companies (11.6%) used only dormant partnership loans, 3 companies used only Murabaha loans and 3 companies used only one of these three loan types: installment sales, compensation and debt purchase. Of combination of loans, the maximum value belonged to simultaneous use of civil partnership and Murabaha. 9 companies used these two options. 11 companies used other combinations of loans and 16 companies out of the 60 companies (26.6%) did not use any of the loans of Sepah Bank. It is clear that most companies which used Sepah Bank loans used partnership contracts like civil partnership and dormant partnership. Regarding "collateral type variable, every company can select one of the following options: "credit insurance policy", "deposit and securities", "promissory note", "letter of guarantee", "immovable", and "contract". 23 companies (38.33%) had used only immovable collateral. 3 companies had used only contract collateral and two companies had used only one of the two collateral deposit or securities and promissory note. Of companies which had used combinations of several collateral, the maximum value belonged to 11 companies who had used collaterals immovable and contract. Furthermore, it must be mentioned that 13 companies (26.6%) had not used any kind of collateral. Regarding "the number of rebounded check" variable which indicated the number of rebounded check for every company since beginning of 2009 till the end of 2013, 54 companies (90%) did not have any rebounded check. 3 companies had 1 to 3 rebounded check, 1 company had 10 to 20 rebounded check and two companies had more than 20 rebounded check. Regarding "use of Sepah Bank POS device" variable, 52 companies (86.6%) had not used POS device and only 8 companies had used it. Regarding "the number of letter of guarantee received", 48 companies (80%) had received letter of guarantee lower than 10 times. 7 companies had received between 10 to 20 times and 5 companies had received letter of guarantee more than 20 times. In terms of "the number of LCs opened", 56 companies (93.3%) had not received any LCs. Two companies had received LC between 1 to 5 times and 2 companies had more than 5 LCs. Regarding "amount of loans received", 17 companies (28.3%) had not received any loans, 19 companies (31.7%) had received loans below 20000 million Rials. 20 companies had received loans between 20000 to 100000 million rials. 4 companies had received loans more than 140000 million Rials. Finally, in terms of "amount of receivables" variable, 57 companies (95%) did not have any receivables and only three companies had receivables equal to 5500, 16000 and 46000 million Rials.

Clustering analysis

Insertion of all 14 variables into clustering algorithm reduces quality of clusters and reduces profile index (Silhoutte). It is common that using Two Steps method, we first identify variables which were effective in increasing quality of clustering and the number of clusters and then we use K-means method -which was a powerful algorithm in this field- to classify data (Liu and Ang, 2008). After elimination of variables which had prediction importance indices smaller than 0.1, 7 variables out of 14 variables were selected for clustering. These variables included: company type, activity type, company lifetime, time period of cooperation with bank, credit background, loans type and amount of loans received.

For analysis, it can be said that when data are similar to each other in terms of several variables, those variables have weaker impacts on separation of data from each other and insertion of these variables into clustering algorithm results in odd data (data with large scattering). After identification of the seven factors which affect clustering of data, the companies were put into two groups based on these seven variables and using K-means algorithm. Of the 60 companies, 17 companies (28.3%) were put into the first cluster and the remaining 43 companies (71.7%) were put into the second cluster. The ratio of the number of the second cluster companies over the first cluster was equal to 2.53. after specification of the appropriate number of clusters (k=2) and determination of variables which are effective in formation of clusters, we use K-means algorithm for segmentation of data. in the software, we specified the number of iterations to be equal to 10but the algorithm ended after 4 iterations because centers of clusters had no change in the fourth iteration. Table 1 indicates final values of centers of two clusters with respect to 7 main variables.

Table 1.final centers of clusters in K-means algorithm

	Cluster	
	1	2
Time of cooperation with bank	3	2
Credit background	3	1
Amount of loans received	56,000	15,000
Loans type	1	0
Company lifetime	3	2
Company time	2	2
Activity type	1	1

The variable "time period of cooperation with bank" can have one of the following values: "below 5 years" (equal to 1), "5 to 10 years" (equal to 2 and "more than 10 years" (equal to 3). According to the results of K-means clustering, center of the first cluster for "cooperation time" is equal to 3 (more than 10 years) and center of the second cluster for this variable was equal to 2 (between 5 to 10 years). The variable "credit background" can also have one of the following values: "below 5 years" (equal to 1), "5 to 10 years" (equal to 2) and more than 10 years" (equal to 3). According to the results of K-means clustering, center of the first cluster for "credit background" variable is equal to 3 (more than 10 years) and center of the second cluster for this variable is equal to 1 (below 5 years). The variable "amount of loans received" (in million Rials) which is a ratio variable contrary to the other six variables can have any quantitative value. According to the results of K-means clustering, the average of this variable for center of the first cluster was equal to 56000 million rials and for the center of the second cluster was equal to 15000 million rials. Regarding "loan type" variable, since "civil partnership" had the maximum frequency in comparison with other options, two values 1 (meaning use of civil partnership loan) and zero (use of other loans other than civil partnership) were used. According to the results of K-means clustering, center of the first cluster for variable "loans type" has the value of 1 (use of civil partnership loan) and center of the second cluster for this variable has the value zero (use of loans other than civil partnership loan). The variable "company lifetime" can have one of the following values (below 7 years"(equal to 1), "7 to 15 years" (equal to 2) and "more than 15 years" (equal to 3). According to the results of K-means clustering, center of the first cluster for variable "company lifetime" is equal to 3 (more than 15 years) and center of the second cluster for this variable is equal to 2 (between 7 to 15 years). The variable "company life" can have one of the following four values: "public stock (equal to 1), private stock (equal to 2), cooperative (equal to 3) and limited liabilities (equal to 4). According to the results of K-means clustering, centers for both first and second clusters for variable "company type" is equal to 2 (private stock). The variable "activity type" can have one of the following values: production (equal to 1), commercial (equal to 2) or service (equal to 3). According to the results of K-means clustering, centers of both first and second clusters for variable "activity type" had the value 1 (production). After investigation of the results of clustering, the companies of the two clusters can be compared with each other as follows: the companies of the first cluster which have all backgrounds above 10 years have mostly more than 15 years of establishment while companies of the second cluster have lower than 10 years of cooperation background and most of them have lower than 15 years of establishment. Companies of the first cluster have used mainly civil partnership loan and dormant partnership loan and have more than 10 years of credit background in Sepah Bank while companies of the second cluster have lower than 5 years of credit background and have used Sepah Bank loans lesser than the first cluster. Average of the amount of loans received by companies of the first cluster is more than amount of loans received by companies of the second cluster. In fact, average of the amount of loans received by the companies of the first cluster which are half of the companies of the second cluster in number is four times as much as the amount of loans received by companies of the second cluster. In fact, companies of the first cluster are more loyal customers because of long background of cooperation with Sepah Bank and have received more loans. It must be mentioned that most companies in both clusters are production and private stock companies.

Conclusion

Kotler believes that customers of different markets can be classified based on demographic, geographic, psychological and behavioral variables. Segmentation of 60 companies which were customers of Sepah Bank was a kind of demographic and behavioral segmentation and it can help identify loyal customers. Attention to customers segmentation based on loyalty allows an organization to link their communications and marketing strategies to the most loyal customers so that it can maximize its profitability (Elahi and Heidari, 2005). Salmen and Moyer presented some strategies for creation of loyalty in customers in a bank. They introduced two strategies: "locking" and "customer satisfaction". Customer satisfaction strategy is characterized by his or her appropriate behavior and trust in future which causes customers not to have any tendency to change organization. On the other hand, "locking" strategy refers to factors which can help prevent from organizational change by customers. Locking strategy is usually implemented on newer customers (Moradi, 2011). In this research, we used clustering technique and seven effective variables to classify legal customers of Sepah Bank in two clusters. 17 companies which were put in the first cluster had longer backgrounds and had taken more loans from Sepah Bank and were more loyal than the companies of the second cluster. Therefore, "customer satisfaction" strategy should be used for these companies. Salman and Muir believe that loyalty improvement programs and service purchase repetition should be proposed to banks in order to reach the outlooks of this strategy. Loyalty improvement programs are subsets of "customer satisfaction" strategy which are mainly used regarding real and legal customers with long background:

1. reduction in service provision costs, 2. Creation of better conditions for use of e-service systems, 3.Expansion of customer orientation culture, 4. Support for production sector customers as bank loyal customers

REFERENCES

Albadvi, Amir; Fesengheri, Mahdi; and Ghopanchi, Amir Hosein (2005). Use of data mining techniques for implementation of CRM in banking industry, commercial studies monthly, number 24.

Asgharizadeh, Ezzatollah, Amin, Fereshteh. (2005). Increase in productivity of banking services by ranking of customers using quantitative techniques, quarterly of economic policies and studies, number 36.

Elahi, SHaban; Heidari, Bahman (2005), CRM, Tehran: commercial publication and printing company.

Alfansi, L., Sergeant, A. (2000), Market Segmentation in the Indonesian banking Sector, International journal of Bank Marketing, 18 (2), 64-74.

Bobinski, G.S. and Assar, A. (1994), "Division of financial responsibility in baby boomer couples: routine tasks versus Investment:, In Costa, J.A. (Ed.)., Gender Issues and Consumer Behavior, Sage Publications, London.

Chiang, W.(2014), Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan, Transportation Letters 6 (2), 89-97.a

Durkin M.g. (2004), In search of the Internet – banking customers, International journal of Bank Marketing, 22:7, 484-503.

Harrison, T. (1994).Mapping Customer Segments for Personal Financial. International Journal of Bank Marketing, 12(8), 17-25.

Khan Babayee, Muhamamd (2009), use of clustering technique and genetic algorithm in making decision trees for optimal classification of banks customers, the third conference on data mining in Iran, Amir KabirUniversity, Tehran.

Kotler, Philip and Armstrong, Gary; principles of marketing, Ali Parsaeeyan, Tehran, Adabestan (2004), third printing, first volume.

Lawson, R., & Todd, S. (2010). Consumer preferences for payment methods: a segmentation analysis. International Journal of Bank Marketing, 21(2), 72-79.

Liu, Hsiang-His, Ong, Chorng-Shyong (2008), Variable selection in clustering for marketing segmentation using genetic algorithms, Expert Systems with Applications 34 (2008) 502–510.

Luan Jing. (2002). Data Mining and Knowledge Management in Higher Education.Knowledge and Data Management white papers, Cabrillo College, Toronto, Canada, 6-16.

Mathur, A., &Moschis, G. P. (1994).Use of Credit Cards by Older Americans. Journal of Services Marketing, 8, 27-36.

Moradi, Muhamamd (2010). Classification of customers of Mellat Bank in Arak City and determination of CRM strategies for each segment, Master Degree thesis, Sanandaj Islamic Azad University

Ngai, E.W.T, Xiu, Li, Chau, D.C.K. (2009), Application of data mining techniques in customer relationship management:A literature review and classification, Expert Systems with Applications 36 ,2592–2602.

Shin, H., Cho, S. (2006). Response Modeling with Support vector Machines. Expert Systems with Applications 30: 746-760.

Taghva, Muhamamd Reza; Olfat, Laya; HoseiniBamakan, SeyyedMojtaba (2010). Application of data mining techniques for improvement of CRM in banking industry, the fourth data mining conference in Iran, Amir Kabir University, Tehran.

Ture, M., Kurt, I., Turhan, K. A., &Ozdamar, K. (2005).Comparing classification techniques for predicting essential hypertension. Expert Systems with Applications, 16(4), 379–384.

Verhoef, p., Spring, p., Hoekstra, J., Leeflang, P. (2002), The commercial use of segmentation and predictive modeling techniques for data base marketing in the Netherlands, Journal of Decision support systems, 34, 471-481.