International Journal of Industrial Engineering & Production Research

December 2010, Volume 21, Number 4
pp. 239-245

http://IJIEPR.iust.ac.ir/

ISSN: 2008-4889

"Technical Note"

Mining the Banking Customer Behavior Using Clustering and Association Rules Methods

Mohammad Ali Farajian & Shahriar Mohammadi *

Mohammad Ali Farajian, *K.N.Toosi University of Technology, Tehran, Iran,*
Shahriar Mohammadi, *K.N.Toosi University of Technology, Tehran, Iran*

KEYWORDS

Dataminng , data preprocessing,
K-means algorithm,
Apriori association rule inducer

ABSTRACT

The unprecedented growth of competition in the banking technology has raised the importance of retaining current customers and acquires new customers so that is important analyzing Customer behavior, which is base on bank databases. Analyzing bank databases for analyzing customer behavior is difficult since bank databases are multi-dimensional, comprised of monthly account records and daily transaction records. Few works have focused on analyzing of bank databases from the viewpoint of customer behavioral analyze. This study presents a new two-stage frame-work of customer behavior analysis that integrated a K-means algorithm and Apriori association rule inducer. The K-means algorithm was used to identify groups of customers based on recency, frequency, monetary behavioral scoring predicators; it also divides customers into three major profitable groups of customers. Apriori association rule inducer was used to characterize the groups of customers by creating customer profiles. Identifying customers by a customer behavior analysis model is helpful characteristics of customer and facilitates marketing strategy development.

1. Introduction

The important resource in contemporary marketing strategies is Customers. Therefore, it is essential to enterprises and organization to successfully acquire new customers and retain high value customers. To achieve these aims, many enterprises plan to gain their own customers' data with many of database tools which can be analyzed to achieve the customer behavioral and applied to develop new business strategies.

Economic theory has established that a business derives 80% of its income from 20% of its customers. However, instead of targeting all prospects equally or providing the same offers to all customers, enterprises select only those individuals that meet specified profitability levels based on previous behavior or individual needs. Accordingly, assuming there is same pattern between customer behaviors. Many method have been introduced to achieve better knowing of customer behaviors, the "behavioral scoring models" is one of the most successful technique that help decision makers to realize their customer behaviors. Behavioral scoring models help to analyze purchasing behavior of customers [1]. These models are highly applying data mining approaches. For a bank, most existing data mining approaches were discovered rules [2] and

* **Corresponding author. Shahriar Mohammadi**
Email*: farajian@sina.kntu.ac.ir mohammadi@kntu.ac.ir*
Paper first received Nov. 05. 2010, and in revised form Dec. 05. 2010.

predicted bankruptcy probability [3] in a bank database. Few works have focused on the analyzing of bank databases from the viewpoint of customer behavioral analyze [4]. More specifically we wanted to look at both the customers profile data and their debit cards transactions. With these data, the aim was to discover applied patterns or rules in the data that could provide information about what incentives a company could offer as better marketing strategies to its customers.

2. The Proposed Model of the Two-Stage Framework of Customer Behavior Analysis

As shown in Fig.1, this study presents a novel two-stage approach for customer behavior analysis of implicit knowledge using bank profile data of the customers and their debit cards transactions.

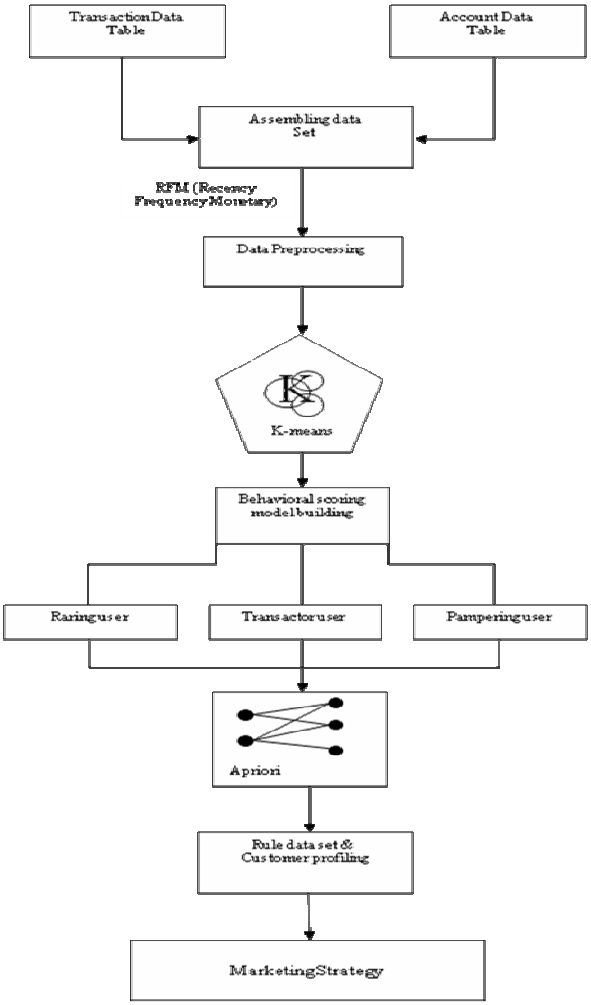

Fig. 1. the proposed model of the two-stage framework of customer behavior analysis

The key feature of the two-stage framework of customer behavior analysis is a cascade involving K-means and an Apriori association rule inducer. The *K-means* algorithm is a non-hierarchical approach to for mining good clusters.

The basic K-means algorithm has been extended in many different ways. Some of these extensions deal with additional heuristics involving the minimum cluster size and merging and splitting clusters. In the first stage of the approach presented here, K-means algorithm was used to divide customers into same groups of customers based on customer behavior and RFM[1] [5]. This K-means was employed to segmentation customers into tree major profitable groups of customer: pamper user, transactor user, and raring user.

Once the K-means identified the profitable groups of customers, And Apriori [6,7] were used to characterize the groups of customers by creating customer profiles; it is mainly used to find out the association rules and meaningful relationships between the huge numbers of items or features that occur synchronously in the database, so Apriori mechanism was used for finding relevant clustering rules. The customer profile introduced then was used to describe characters of each group of customers, and served as a tool for strategic managements to establishing better bank marketing strategies. The rest of this paper is organized as follows

3. Experimental Dataset and Preprocessing:

For this study, bank databases were provided by a major Iranians debit card issuer. The first action is an integrating data set that was intended to organize the raw data. Two data tables were obtained: a table containing effective debit card account information of 55,211 customers until October 2009, and another table storing over 11.3 million individual transaction records for these accounts from March 2007 to October 2009. Then, two data tables were joined using a customer ID to create a single assembling data Table.

The representation and quality of data is first and foremost before running an analysis so to achieve used data preprocessing. Data preprocessing was required to ensure data field consistency in customer behavior analysis model.

Obviously, not all the data are related to the chosen purposes, so knowledge extraction from the bank databases included the following two sub-actions. The first sub-action (Data cleansing) was the extraction of only that data considered useful for the analysis. Unnecessary data fields and records containing incomplete or missing data were removed from the data sets [8]. The second sub-action was the application of simple statistics to calculate an aggregate of new behavioral predicators.

The calculating of the aggregate variables was used to emphasize the customer behavior and RFM information hidden in the 12 months observation period. In this case, the values derived from the bank

[1] *Recency Frequency Monetary*

database such as minimum, maximum and average of a set of variables (e.g. Amount of transaction ,card usage cycle days, number of purchases, daily times of transaction, yearly amount of consumption, and so on) for the monthly activity over the past 12 months were considered.

So the predicators calculated used to predict which customer belongs to which profitable group. The ranges of values of numerical predictor are split into intervals so that each interval contains as many customers as possible that have a significant homogeneous behavior. Multiple predictors can be grouped together to obtain the same effect. To derive the most profitable customers, it was chosen to identify similar behavior with respect to RFM values found in the real world.

4. Analyzing the Consumer Behavior

Banks seeking newer and better ways to differentiate themselves from their competitors, customer clustering one of important way to rich this result; Customer clustering is the use of past transaction data to divide customer to the similar groups. The results produced are based on the assumptions that the customer behavior follows patterns similar to past pattern and repeats in the future.

Therefore, there could not be a better time than now to analyze the importance of an effective new marketing strategy using the customer behavior analyze. The decisions to be made include which target groups of customers will be encouraged to use more, what terminal type to assign, how estimated probability of acceptance new products, whether to promote new products to target groups of customers, and, how to manage groups of customers to rich the customer satisfaction and direct marketing.

However, attempts to make good customer behavior analysis may be limited by the poor quality of data, poor relevant of data, or the volume of data needing to be processed. Database marketing (DM) is a systematic approach to the gathering, consolidation, and processing of customer data to help the marketers' better target their markets efforts to existing customers [9].

Additional DM analyzes customer data to look for patterns to use these patterns for a more targeted selection of the consumers [10].Over the decades, many database marketing tools were developed and used in various stages of marketing. Some of the most popular tools include: the RFM (recency, frequency, monetary), Formula, the behavior segment of the existing customers and the lifetime value of a customer [11].

Customer clustering is a process that divides customers into smaller groups; Clusters are to be homogeneous within and desirably heterogeneous in between [12]**.** Figure 2 presents the conceptual framework used to answer the questions posed in proposed the two-stage framework. This figure shows two components, customer clustering and customer profiling. In this framework, which serve as major issues to be discussed here. In this framework account and transaction tables are assumed to be input tables to customer clustering. On the other hand the Customer behavior and RFM value are assumed to be behavioral predicators affecting customer clustering.

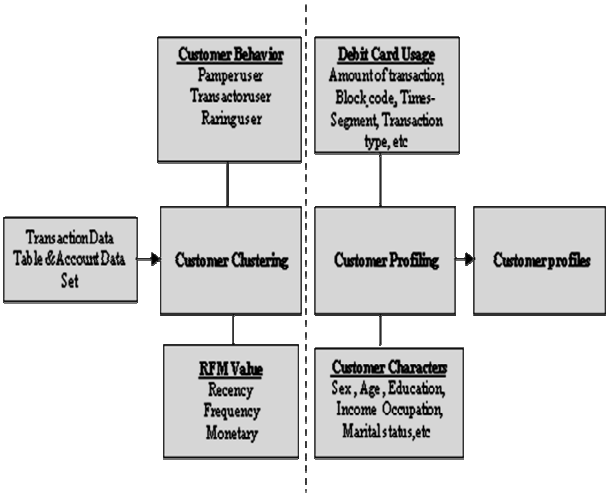

Fig. 2. A conceptual framework of customer behavior analysis.

Generally, debit card company make money from interest on account of customer balance, annual fees, and the discount collected from merchants on each transaction.

RFM analysis [13,14] has been used by direct marketers for selecting which customers to target offers. in the bank the recency (R) value measures is the date of the user's last transaction. Since the R value contributes to the RFM scoring determination, a numeric value is necessary. Therefore, a new variable, R new is defined as the number of days between the first date concerned (1/11/2009) and the date of the last active user's transaction. For example a user who has conducted his last transaction on 10/9/2009 is characterized by Rnew=51. frequency (F) value measures is defined as the count of financial transactions the user conducted within the period of interest, and monetary (M) value measures is the total value of financial transactions the user made during a yearly time period.

Next, variables such as customer characteristics and debit card usage are assumed to influence customer profiling. Finally, clusters and the associated profiles are assumed to be outputs, as well as influence of proposed frame work on strategic management decisions. In Figure 2, customer behavior is highly related to customer clustering, but is an implicit variable which cannot be retrieved directly from the data table. We need to develop a method for modeling the customer behavior.

As shown in the following equation, this study employs "Customer Power" (CP) as a customer behavior variable to model Customer behavior,

$$\text{CustomerPower} = \frac{\text{sum of the customer score on a term}}{\text{number of the months on a term}} \quad (1)$$

The default observation range is assumed to be 12 months, and CP is computed as the 'sum of the customer score on a term' divided by the 'number of the month on a term'. Score of each customer in a month calculated by flowing table:

Tab. 1.Table of Customer Score

Sum of Transaction Amount at a Month (STAM)	Customer Score(CS)
0< STAM <500	1
500<= STAM <1300	2
1300<= STAM <2500	3
2500<= STAM	4

As shown in the upper table each customer score (CS) calculated:

1. Sum of Transaction Amount on a Month (STAM)
2. score elicit according to table

For example, a customer has STAM=1500 as Customer Score (CS) elicit According to table as 3. For instance, a customer has (CS) during 12 month as (3,2,2,3,2,2,2,3,2,1,2,2) and then the degree of CP is computed as:

$$CP = \frac{3+2+2+3+2+2+2+3+2+1+2+2}{12} = 2.166$$

For each customer, if CP is between 3 and 4, then the behavior of that customer is considered a pamper user. Meanwhile, if CP is between 2 and 3 then the behavior of that customer is considered a transactor user. Finally, if the value of CP is between zero and one then the behavior of that customer is considered a raring user.

5. K-means Algorithm to the Customer Clustering

For customer clustering and segmentation, many studies have presented such as support vector regression analysis, Panel Data Clustering, linear, multiple discriminate analysis (MDA), and so on [15,16,17,18]. Furthermore Baesens[19] employed Bayesian neural networks to repeat purchase behavior modeling in direct marketing. Rather than profiling segments based on demographic or geographic characteristics, Dasgupta et al., [20] characterized potential customer segments in terms of lifestyle variables. Balakrishnan et al. [21] accomplished a six-segment classification study using coffee brand switching probabilities derived from the scanner data at a sub-household level. Setiono et al [22] utilized a rule-extraction neural network to aim at companies for the promotion of new information technology. Fish, Barnes and Aikenl [23] proposed a new methodology for industrial market segmentation by neural networks. A hierarchical clustering and a two-stage technique (average linkage then K-means) were used, and compared to delineate the meteorological conditions in Houston [24] a popular clustering method that minimizes the clustering error is the K-means algorithm, and this method performed well on the experimental data sets [25].

K-means clustering is an iterative clustering method, and divides the data into a number of clusters by minimizing an error function which can be expressed[12]. The *K-means* algorithm is a non-hierarchical approach to forming good clusters, used to group records based on similarity of values for a set of object. The basic idea is to try to discover *k* clusters and assign each object to the k cluster so as minimize a measure of dispersion within the clusters such that the records within each cluster are similar to each other. *K-means* is an iterative algorithm; an initial set of clusters is defined, and the clusters are repeatedly updated until no more improvement is possible. The K-means clustering algorithm partitions data record $X = (x_1, x_2, ..., x_n)$ into k clusters G_α ($\alpha = 1, 2, ..., k$) and the cluster G_α are associated with representatives (clustercenters) c_α ($\alpha = 1, 2, ..., k$) . K-means algorithm begins with an initial set of cluster centers and repeats this mapping process until a stopping criterion is satisfied. The Lloyd iteration for K-means clustering is given as follows [12,26].

Start: Begin with an initial group of cluster centers (centroids):

Step (t): Assign each object (data record x_β) to the group that has the closest centroid to generate the improved set of cluster representatives.

$$G_\alpha^{(t)} = \left\{ x_\beta : \left\| x_\beta - c_\alpha^{(t)} \right\| \le \left\| x_\beta - c_{\alpha^*}^{(t)} \right\| for \quad all \quad \alpha^* = 1,...,k \right\} \quad (2)$$

Update centroid: compute the positions of the centroids. If the value of the centroids didn't change then stop, else go to Step (t).

$$c_\alpha^{(t+1)} = \frac{1}{\left| G_\alpha^{(t)} \right|} \sum_{x_\beta \in G_\alpha^{(t)}} x_\beta \quad (3)$$

In this study, the K-means is built with data from existing customers, which include variables from account and transaction data tables. All of the existing customer's data are used to build the customer behavior analysis modeling order to predicate customer behavior. The utilization of K-means for customer clustering suppose that k = 3 and was used CP and RFM as predicated variables to classify each customer into clusters. As shown in Figure 3, we arranged three profitable groups of customers as 3 clusters.

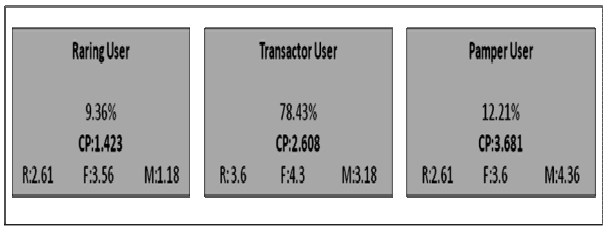

Fig. 3. Result of K-means algorithm for customer clustering

The customer behavior, ratio of number of customers relative to the overall customers, average CP and RFM values were shown for each cluster. The mass case was Transactor user , the ratio of number of customers relative to the overall customers was 78.43% and the number of customers is 43,301.nex cluster include pamper user totaling 6,741 customers. Moreover, next cluster indicated raring user totaling 5.169 customers. Table 2 lists the distribution of the relative importance for each input variable.

Tab. 2. Input variable for customer clustering

Variable Name	Comments
Education	1010, Elementary; 1020;midel school; 1030, high school; 2010Undergraduate; 2020, graduate ; 2030, Postgraduate; 3010, Dr
Days –segments	Days segment of a transaction in \month to 1, <11; 2, 10-20; 3, 20-30.
Times-Segment	Times segmentation of a transaction in day 1, <6; 2, 6-12; 3, 12-18; 4, 18-24.
Age_range	1, <27; 2, 27-50; 3, 50<.
Amount of transaction	Monthly amount of transaction <900,000
Transaction type	Encoded field
Sex	1, men; 2, women.
Terminal type	Code of terminal type
Occupation	Encoded field
Block code	Card usage limit or not

6. Create Customer Profiles by Association Rules Inducer:

The study's aim is to try to find out most important and Practical patterns in bank databases so that it could better understand different behaviors about different customers and develop new strategies to provide better service and satisfying their needs better than the competition.

In the previous sections, we used K-means clustering model to classify customers into clusters with shared characteristics. The employment of mining association rules was used to create customer profile in each cluster. The purpose of association rule extraction is to discover significant relationships between items or features that occur frequently in a transaction database.

Let I= {i1, i2,..., im} be a set of items[6,7]. Let DB be a database of transactions, where each transaction T consists of a set of items such that $T \subseteq I$. Given a set of items $X \subseteq I$, a transaction T contains X only and only if $X \subseteq T$. Support (X,DB) denotes the rate of X in DB.'X Y(s%, c%, l)' denotes An association rule , where $X \subseteq I$, $Y \subseteq I$ and $X \cap Y = \varnothing$. The association rule X Y has support s in DB if the probability of the transaction in DB contains $X \cup Y$ is s (i.e.Supp (X Y)= Support ($X \cup Y$, DB)). The association rule X Y with *confidence c* in DB if the probability of the transaction in DB which contains X also contains Y is c (i.e. Conf(X Y)= Supp (X Y)/Support (X,DB)). Apriori algorithm [6,7] is one of most successful algorithm has been proposed for mining association rules in a database. Rule candidates are considered useful and become association rules only if whose support is larger than a minimum support (minsupp) threshold and whose confidence is larger than a minimum confidence (minconf) threshold. The association rules must have two conditions: Supp (X Y) ≥ minsupp, Conf (X Y) ≥ minconf.

Once the clusters and the associated statistical summarized data are made by K-means algorithm. The customers are fall into three major profitable groups of customers. The association rule inducer is assisted to create and verify the customer profiles. The variables deriving from the sensitivity analysis were chosen as predicate variables for association rule analysis. For explanation, we chose only cluster-2 for mining association rules.

Tab. 3. Result of association rules

Rule ID	Association rules	Support	Confidence
1.	**Terminal type =3010← Days –segments=1&sex=1& Times-Segment=3**	6.1%	83%
2.	**Terminal type =3010← Days –segments=1&sex=1& Times-Segment=4**	7.2%	87%
3.	**Terminal type =3010← Days –segments=1&sex=1**	34.3%	93%
4.	**Terminal type =3010← Days –segments=1&sex=1& Transaction type=1017**	12.3%	84%
5.	**Terminal type =3010← Days –segments=1&sex=1& Transaction type=1011**	9.8%	85%
6.	**Terminal type =3010← Days –segments=1&sex=1& Transaction type=1031**	6.3%	87%
7.	**Terminal type =3010← Days –segments=1&sex=1& Transaction type=1032**	11.8%	83%

Table continued. 4. Result of association rules

Rule ID	Association rules	Support	Confidence
8.	**Terminal type =3010← Days –segments=1&sex=1& Transaction type=1012& Times-Segment=3**	12.6%	87%
9.	**Terminal type =3010← Days –segments=1&sex=1& Transaction type=1012**	23.4%	91%
10.	**Terminal type =3010← Days –segments=1&sex=1& Transaction type=1012& Times-Segment=4**	14.0%	93%
11.	**Terminal type =3010← Days –segments=1&sex=0& Transaction type=1031**	15.1%	85%
12.	**Terminal type =3010← Days –segments=1&sex=0& Transaction type=1030**	7.3%	87%
13.	**Terminal type =3020← Age-Segments=2&sex=1& Times-Segment=3**	8.7%	83%
14.	**Terminal type =3020← Age-Segments=2&sex=1& Times-Segment=4**	9.8%	87%
15.	**Terminal type =3020← Age-Segments=2&sex=1&Transaction type=1011& Marital_Status=1**	6.3%	83%
16.	**Terminal type =3020← Age-Segments=2&sex=1&Transaction type=1031& Marital_Status=1**	11.8%	87%
17.	**Terminal type =3020← Age-Segments=2&sex=1&Transaction type=1032&** Marital_Status=1	12.6%	89%
18.	:		

Parameters were set up to identify association rules that had at least 80% confidence and 5% support imposed on the Apriori association rule inducer. Table 3 lists the cluster profile of cluster-2 in the form of association rules, where each rule represents a customer profile that was dominant or most strongly associated with the customers matching that cluster. For discriminating purposes, we have grouped customers with shared behavioral characteristics. From this, marketers can create more accurate campaigns towards each target group of customers for cross-selling and encouraging consumption. After briefly reviewing the 3 clusters using cluster profiles, the customers with values tend to R↓F↑M↑ can be targeted with greater accuracy.

7. Conclusion

This study proposes a new two-stage framework of customer behavior analysis using K-means clustering algorithm and an association rule inducer for analyzing bank databases. For differentiation purposes, we grouped customers with shared customer behavior and RFM value.

After briefly reviewing the customer profiles using the association rule inducer, the customers with a higher CP or RFM might be the target customer groups of precedence. The existing customers were divided into three profitable groups of customers according to their shared behavior and characteristics. Marketers then can infer the profiles of customers in each group and propose management strategies appropriate to the each group. This study provides a new method of analyzing bank databases. Beyond simply understanding customer value, the bank gains the opportunities to establish better customer relationships.

8. Further Research

Further research may aim at time-series behavioral analyzing models that could include the change of behavioral status in every period. This can happen by establishment of a simulation to have a more precise analysis by changing parameters.

9. Appreciation

The author sincerely appreciates the Iran Telecommunication Research Center due to their kind financial as well as moral support of this research.

References

[1] Setiono, R., Thong, J.Y.L., Yap, C.S., "*Symbolic Rule Extraction from Neural Networks-an Application to Identifying Organizations Adopting IT*". Information and Management, 34(2), 1998, pp 91–101.

[2] Au, W.H., Chan, K.C.C., *Mining Fuzzy Association Rules in a Bank-Account Database*, IEEE Transactions on Fuzzy Systems, 2003, Vol. 11.

[3] Donato, J.C., Schryver, G.C., Hinkel, R.L., Schmoyer, J., Leuze, M.R., Grandy, N.W., *Mining Multi-Dimensional Data for Decision Support,* Future Generation Computer Systems, Vol. 15, 1999, pp.433-441.

[4] Sharda, R., Wilson, R., *Neural Network Experiments in Business Failures Predication: a Review of Predictive Performance Issues.* International Journal of Computational Intelligence and Organizations,1(2), 1996, pp 107–117.

[5] Bult, J.R., Wansbeek, T., *Optimal Selection for Direct Mail.* Marketing Science, 14(4), 1995, pp 378–381.

[6] Agrawal, R., Imielinski, T., Swami, A., *Mining Association Rules Between Sets of Items in Large Databases*. Proceedings of the SIGMOD'93, Washington, DC, 1993, pp 207–216.

[7] Agrawal, R., Srikant, R., *Fast Algorithms for Mining Association rules*. In Proc. 20th Int. Conference on Very Large Data Bases, 1994, pp. 487-499.

[8] Fish, K.E., Barnes, J.H., Aiken, M.W., *Artificial Neural Networks—a New Methodology for Industrial Market Segmentation*. Industrial Marketing Management, 24, 1995, pp. 431–438.

[9] Tao, Y.H. Yeh, C.C.R., *Simple Database Marketing Tools in Customer Analysis and Retention*, International Journal of Information Management, Vol. 23, 2003, pp.291-301.

[10] Qizhong, Zhang, An Approach to Rough Set Decomposition of Incomplete Information Systems. 2nd IEEE Conference on Industrial Electronics and Applications, ICIEA, 2007, pp. 2455-2460.

[11] Feelders, A.J., *Credit Scoring and Reject Inference with Mixture Models*, International Journal of Intelligent Systems in Accounting, Finance and Management, Vol. 9, 2000, pp.1-8.

[12] Anil, K.J., *Data Clustering: 50 Years Beyond K-Means*, International Journal of Pattern Recognition Letters 2009.

[13] Madeira, S.A., "*Comparison of Target Selection Methods in Direct Marketing*", MSc Thesis, Technical University of Lisbon, 2002.

[14] Ching-Hsue Cheng, You-Shyang Chen: Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications. 36 (3): 2009, pp. 4176-4184.

[15] Levis and Papageorgiou. Customer demand forecasting via support vector regression analysis. Chemical Engineering Research and Design. v83. 2005, pp 1009-1018.

[16] T. Zheng, et al. Panel Data Clustering and its Application to Discount Rate of B Stock in China, in 2009 Second International Conference on Information and Computing Science. 2009.

[17] Guangli Niea, Yibing Chen, Lingling Zhanga, and Yuhong Guo, Credit card customer analysis based on panel data clustering, International Conference on Computational Science, ICCS 2010.

[18] Malhotra, R., Malhotra, D.K., *Evaluating Consumer Loans using Neural Networks*, Omega, Vol.31, 2003, pp.83-96.

[19] Baesens, B., Viaene, S., Poel, D., Vanthienen, J., Dedene, G., *Bayesian Neural Network for Repeat Purchase Modelling in Direct Marketing*. European Journal of Operational Research, 2002, 138, 191–211.

[20] Dasgupta, C.G., Dispensa, G.S., Ghose, S., *Comparing the Predictive Pperformance of a Neural Network Model with Some Traditional Market Response Models*, International Journal of Forecasting,Vol. 10, 1994, pp.235-244.

[21] Balakrishnan, P.V.S., Cooper, M.C., Jacob, V.S., Lewis, P.A., *Comparative Performance of the FSCL Neural Net and K-Means Algorithm for Market Segmentation*, European Journal of Operational Research, Vol. 93, 1996, pp.346-357.

[22] Setiono, R., Thong, J.Y.L., Yap, C.S., *Symbolic Rule Extraction from Neural Networks – An Application to Identifying Organizations Adopting IT,* Information & Management, Vol. 34, 1998, pp.91-101.

[23] Fish, K.E., Barnes, J.H., Aiken, M.W., *Artificial Neural Networks - A New Methodology for Industrial Market Segmentation*, Industrial Marketing Management, Vol. 24, 1995, pp.431-438.

[24] Davis, J.M., Eder, B.K., Nychka, D., Yang, Q., *Modeling the Effects of Meteorology on Ozone in Houston using Cluster Analysis and Generalized additive Models*. Atmospheric Environment 32, 1998, pp. 2505–2520.

[25] Balakrishnan, P.V., Cooper, M.C., Jacob, V.S., Lewis, P.A., *A Study of the Classification Capabilities of Neural Networks using Unsupervised Learning: a Comparison with k-Means Clustering*, Psychometrika 59(4), 1994, pp 509-525.

[26] Jim, Z.C., Laia, Tsung-JenHuanga, Yi-ChingLiawb. *A Fast k-Means Clustering Algorithm using Cluster Center Displacement*, Pattern Recognition 42. 2009, pp. 2551-2556.