Customer Behavior in Banking Industry: Comparison of Data Mining Techniques

Mohammad Ali Afshar Kazemi and Negar Estemdad
Department of Economics and Management, Science and Research Branch, Islamic Azad University, Tehran, Iran
Email: {Dr.mafshar, Nestemdad} @gmail.com

Alireza Poorebrahimi
Department of Economics and Management, E-Campus, Islamic Azad University, Tehran, Iran
Email: Poorebrahimi@gmail.com

***Abstract*—Nowadays organizations have perceived the importance of managing customer relationship and its potential benefits. Customer relationship management supports organizations to deliver beneficial relations with customers. Customer satisfaction and retention are the leading objectives of any organization and this cannot be done without knowing the loyalty of the customer. Accordingly, to identify the loyalty of the customers, K-means algorithm was applied to the bank customers' data and clustering was conducted. The customer behavior is estimated by neural network and C5.0 models. Results show that C5.0 better fits the customer behavior. In addition, estimation of customer behavior leads organizations to more successful customer management strategies.**

***Index Terms*—data mining, loyalty, bank industry, neural network, C5.0 algorithm**

I. Introduction

Data mining methodology has a remarkable contribution in extraction of hidden knowledge which has been inherited in the data [1].

The main objective of the research is to provide the ultimate customer satisfaction in banking industry. To do so, customers are clustered by applying K-means algorithm and based on Davies Bouldin Index. Clustering is used to identify patterns of customer loyalty. Consequently, the loyalty degree of the customers is approximated to reach maximized win-win profits.

Following, the studies related to customer relationship management, customer loyalty, and K-means algorithm are reviewed. The Customer loyalty evaluation model is presented and findings are discussed.

A. Customer Relationship Management

Tendencies to customer relationship management (CRM) grew in 1990 [1]. Expansion of relationship with customers may lead to loyalty and retention [2]. Customer relationship management has four dimensions:

- Customer identification
- Customer attraction
- Customer retention
- Customer development [3], [4]

These four dimensions are called the closed-loop customer relationship management [5], [6]. Although this subject has been widely applied in businesses, but there has been no accepted comprehensive definition so far [2], [7]. "A comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer", is the definition by Richards and Jones [8]. CRM is the combination of processes from customer attraction to retention which utilizes data mining techniques to maximize customer value [1]. CRM ensures that a slight change in retention rate would lead to significant market share [9]. The first two dimensions of CRM, customer identification and attraction, are costly and confirm the monetary benefits of companies. Although the customer retention is of lower cost for companies, it makes great benefits as well. Reports reveal that numerous companies ignore the latter [10].

B. Customer Loyalty

Firms create loyal customer base to a) keep the number of customers overtime and b) to create the relationship with customers to encourage their future purchase and support level. Firms would manage to increase their profits if equipped with the customers' loyalty levels. Business researchers announce that a loyal customer brings a sound stream of income for a firm by remaining with the brand [11], [12]. A number of theories are proposed to link relationship and business marketing to the loyalty construct. For instance, it is proven that relationship practices have a direct impact on loyalty of the customers [13]. Other constructs such as "relationship quality, trust, involvement, satisfaction, purchase development, organizational change, and switching costs" are influencers of customer loyalty [14], [15].

Loyalty, or sometimes customer lifetime value (CLV) is estimated by variables recency, frequency, monetary (RFM). These variables are the integrating rate of each cluster, $C_{j1} = wRC_jR + wFC_jF + wMC_jM$, where wR, wF, wM are the importance of the recency, frequency and monetary variables [16]. Some researchers have proposed a model based on "computing the distance between the

Manuscript received January 20, 2014; revised April 25, 2014.

doi: 10.12720/joams.3.1.13-16

center of a cluster and zero point". In this model, highest value denotes the most customer loyalty [17].

C. K-Means Algorithm

Clustering is a technique to group homogenous data [18]. A very recognized clustering algorithm is K-means which is sensitive to the starting point (K initial clusters). The performance of clustering methods can be compared by intra-class method shown in (1) [19], [20].

$$F(x)=1/K\sum_{n=1}^{K}\sum_{ci}Dist(C_i, C^n) \quad (1)$$

II. Materials and Methods

A. Methodology

To reach the objectives of the research, four phases of the conceptual model are undertaken as: 1) data selection and preparation; 2) determining customer value and customer loyalty; 3) determining K-optimum by Davies-Bouldin Index (selecting training, testing and validation data from dataset); 4) Customer clustering based on neural networking (NN) and C5.0 techniques and 5) Evaluation of the model output.

The developed methodology has been implemented in a Bank in Iran (Refah). Fig. 1 demonstrates the conceptual model of the research. Detailed phases of the model are described in the customer loyalty evaluation model.

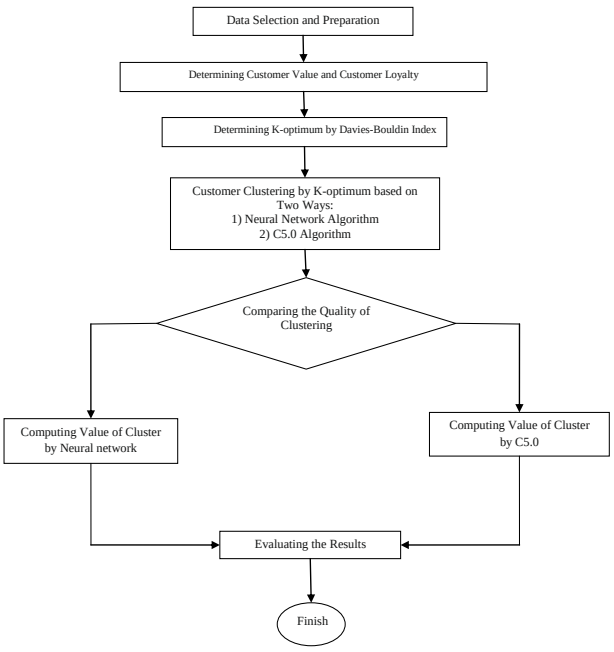

Figure 1. The method of the research

B. Data and Case Analysis

The proposed model is undertaken in an Iranian Bank. Customers' data were randomly extracted for a period of one year from March 2011 to March 2012 from short-term saving accounts. The following parameters were calculated based on the derived data:

- Recency (R)
- Frequency (F)
- Monetary (M)
- Demographic characteristics (gender, occupation and education)

Table I shows a small sample of a short-term saving account customer data. As depicted in K-mean algorithm, data are homogenous.

TABLE I. A Small Sample Data

Date	Turn Over (M)	Balance	Frequency
05/27/2011	1.000.000	150.000.000	2
06/02/2011	-10.000.000	151.000.000	1
06/03/2011	-5.000.000	141.000.000	1
06/13/2011	30.000.000	136.000.000	4
06/15/2011	800.000.000	166.000.000	2
06/18/2011	-1.000.000	966.000.000	1

III. Cutomer Loyalty Evaluation Model

Four customer loyalty factors which are generally included in loyalty models are: behavioral, demographical, environmental and cognitive factors. These factors are extensively applied to model the loyalty of the customers [21], [22]. In this research, data mining techniques were applied and K-means clustering were done to identify the loyalty of the customers (Fig. 1).

A. Data Selection and Preparation

Initially, the dataset for empirical study was selected and pre-process to be cleaned, such as deletion of the records which were missing or possess inaccurate values and elimination of the redundant attributes. This was done to reach to a data format that would shape effective processing for clustering customers.

B. Customer Value and Loyalty

To estimate the customer value, the RFM model was applied. Recency ("*How recently did the customer purchase?*"), frequency ("*How often do they purchase?*") and monetary ("*How much do they spend?*") are the variables of this model. So, the customer value is estimated based on RFM model.

C. Customer Clustering and Evaluaiton

Based on the loyalty value of the customers gender, education and occupation, clustering is done with SPSS Clementine 12.0 (K=3). Table II demonstrates the optimum Davies Bouldin Index. The results are three clusters named gold, silver and bronze. Neural network and C5.0 are independently applied to measure clustering (Tables III and IV) and results were discussed.

TABLE II. Desired K in Davies Bouldin Index

K	Davies Bouldin Index
3	0.07
4	0.09
5	0.08
6	0.08
7	0.09
8	0.15

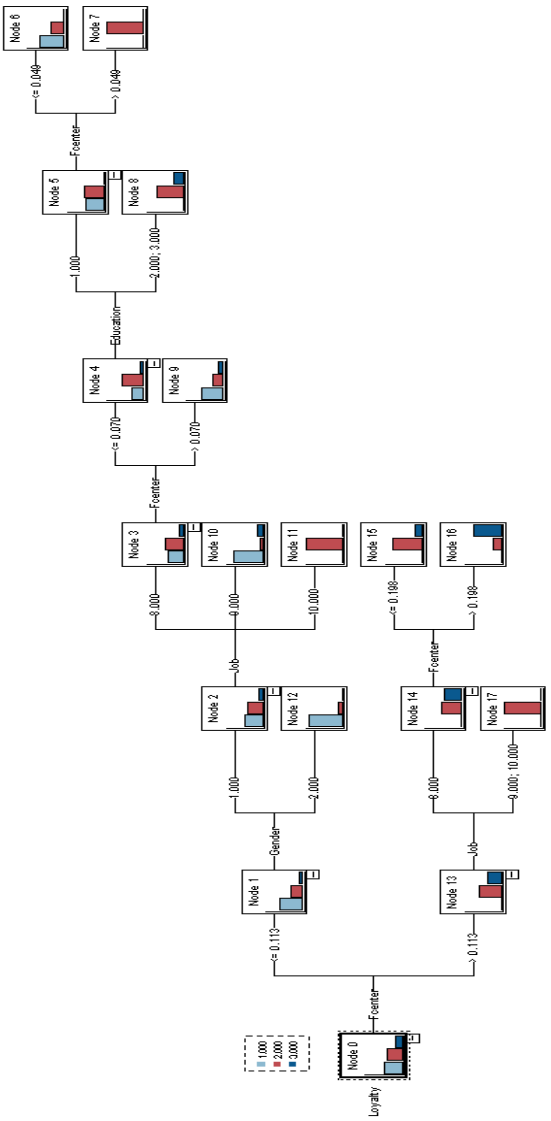

Figure 2. C5.0 tree view

TABLE III. RESULTS OF NEURAL NETWORK

Results for output field Loyalty
Comparing $N-Loyalty with Loyalty

'Partition'	1_Training		2_Testing		3_Validation	
Correct	46	61.33%	15	68.18%	20	60.61%
Wrong	29	38.67%	7	31.82%	13	39.39%
Total	75		22		33	

Coincidence Matrix for $N-Loyalty (rows show actuals)

'Partition' = 1_Training	1	2
1	23	12
2	5	23
3	0	12

'Partition' = 2_Testing	1	2
1	7	3
2	3	8
3	0	1

'Partition' = 3_Validation	1	2
1	9	6
2	0	11
3	0	7

Performance Evaluation

'Partition' = 1_Training	
1	0.565
2	0.271

'Partition' = 2_Testing	
1	0.432
2	0.288

'Partition' = 3_Validation	
1	0.788
2	0.318

Fig. 2 demonstrates the generated tree of the C5.0 algorithm. Accordingly, it shows the conditional rules of the loyalty. For instance, if the distance from the center of the cluster is equal or greater than 0.113 and the customer is female, they are clustered as disloyal. Or if the distance is less than 0.19 and customers are in the first category of the education, then they are clustered as loyal.

In general, the tree notes that if the distance is less than 0.113, regardless of customers' gender, occupation and education, customers are disloyal.

TABLE IV. RESULTS OF C5.0 ALGORITHM

Results for output field Loyalty
Comparing $C-Loyalty with Loyalty

'Partition'	1_Training		2_Testing		3_Validation	
Correct	61	81.33%	14	63.64%	25	75.76%
Wrong	14	18.67%	8	36.36%	8	24.24%
Total	75		22		33	

Coincidence Matrix for $C-Loyalty (rows show actuals)

'Partition' = 1_Training	1	2	3
1	34	1	0
2	5	21	2
3	3	3	6

'Partition' = 2_Testing	1	2
1	10	0
2	7	4
3	0	1

'Partition' = 3_Validation	1	2	3
1	14	1	0
2	3	7	1
3	1	2	4

Performance Evaluation

'Partition' = 1_Training	
1	0.551
2	0.811
3	1.545

'Partition' = 2_Testing	
1	0.258
2	0.47

'Partition' = 3_Validation	
1	0.537
2	0.742
3	1.327

IV. RESULTS AND DISCUSSION

The objective of this research was to cluster the customers of a bank and to evaluate the clustering by neural network and C5.0 algorithms. Clustering was based of the loyalty level, which is homogenous data [18]. According to Table V which summarizes the results of Tables III and IV, it is revealed that validity of train, test and validation data by neural network are 61.33%, 68.18% and 60.61% respectively. Besides, C5.0 delivers 81.33%, 63.64% and 75.76% validity of data.

In overall, C5.0 delivers more precise evaluation than neural network. This might be according to the tree-based nature of C5.0 algorithm [23].

TABLE V. COMPARISON OF C5.0 AND NEURAL NETWORK

Algorithm	Data Validity Test		
	Train	Test	Validity
C5.0	81.33	63.64	75.76
NN	61.33	68.18	60.61

Results revealed that when RFM model is combined with demographic characteristics and applied by K-means algorithm, remarkable improvement in the accuracy of classification is maintained. This would reach to an admirable customer relationship management.

REFERENCES

[1] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: Aliterature review and classification," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2592–2602, 2009.

[2] E. W. T. Ngai, "Customer relationship management research (1992-2002): An academic literature review and classification," *Marketing Intelligence & Planning*, vol. 23, no. 6, pp. 582–605, 2005.

[3] A. Parvatiyar and J. N. Sheth, "Customer relationship management: Emerging practice, process, and discipline," *Journal of Economic and Social Research*, vol. 3, no. 2, pp. 1–34, 2001.

[4] R. S. Swift, *Accelerating Customer Relationships: Using CRM and Relationship Technologies*, Prentice Hall, 2001.
[5] W. H. Au, K. C. Chan, and X. Yao, "A novel evolutionary data mining algorithm with applications to churn prediction," *IEEE Trans Evolutionary Computation*, vol. 7, no. 6, pp. 532–45, 2003.
[6] A. H. Kracklauer, D. Q. Mills, and D. Seifert, *(Hrsg): Collaborative Customer Relationship Management-Taking CRM to the Next Level*, New York, 2004.
[7] R. Ling and D. C. Yen, "Customer relationship management: An analysis framework and implementation strategies," *Journal of Computer Information Systems*, vol. 41, no. 3, pp. 82–97, 2001.
[8] K. A. Richards and E. Jones, "Customer relationship management: Finding value drivers," *Industrial Marketing Management*, vol. 37, no. 2, pp. 120–30, 2008.
[9] D. V. D. Poel and B. Lariviere, "Customer attrition analysis for financial services using proportional hazard models," *European Journal of Operational Research*, vol. 157, no. 1, pp. 196–217, 2004.
[10] G. Cui, "A methodologic application of customer retention based on back propagation Neural Network prediction," in *Proc. 2nd International IEEE Conference on Computer Engineering and Technology*, China, 2010, pp. 418–422.
[11] S. Y. Lam, V. Shankar, M. K. Erramilli, and B. Murthy, "Customer value, satisfaction, loyalty, and switching costs: An illustration from a business-to business service context," *Journal of the Academy of Marketing Science*, vol. 32, no. 3, pp. 293-311, 2004.
[12] F. F. Reichheld and T. Teal, *The Loyalty Effect: The Hidden Force behind Growth, Profits, and Lasting Value*, Trans. Boston, Mass: Harvard Business School Press 1996.
[13] L. Ricard and J. Perrien, "Explaining and evaluating the implementation of organizational relationship marketing in the banking industry: Clients' perception," *Journal of Business Research*, vol. 45, pp. 199–209, 1999.
[14] S. Chow and R. Holden, "Toward an understanding of loyalty: The moderating role of trust," *Journal of Management Issue*, vol. 9, no. 3, pp. 275-298, 1997.
[15] K. Eriksson and A. L. Vaghult, "Customer retention, purchasing behavior and relationship substance in professional services," *Industrial Marketing Management*, vol. 29, no. 4, pp. 363-372, 2000.
[16] D. R. Liu and Y. Y. Shih, "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information & Management*, vol. 42, no. 3, pp. 387–400, 2005.
[17] C. H. Cheng and Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications: An International Journal*, vol. 36, no. 3, pp. 4176-4184, 2009.
[18] S. R. Ahmed, "Applications of data mining in retail business," in *Proc. International Conference on Information Technology: Coding and Computing*, North Georgia, USA, 2004, pp. 455–459.
[19] P. Michaud, "Clustering techniques," *Future Generation Computer System*, pp. 135–147, 1997.
[20] H. W. Shina and S. Y. Sohnb, "Segmentation of stock trading customers according to potential value," *Expert Systems with Applications*, vol. 27, pp. 27–33, 2004.
[21] D. C. Schmittlein and R. A. Peterson, "Customer base analysis: An industrial purchase process application," *Marketing Science*, vol. 13, no. 1, pp. 41–67, 1994.
[22] W. Buckinx, E. Moons, D. V. D. Poel, and G. Wets, "Customer-adapted coupon targeting using feature selection," *Expert Systems with Applications*, vol. 26 , no. 4, pp. 509–518, 2004.
[23] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

Mohammad Ali Afshar Kazemi holds a Ph.D. and is the associate professor of Islamic Azad University in the Central Branch. He is with the department of industrial management. His research interests are data mining, knowledge management, system dynamics, simulation, and product life cycle.

Alireza Poorebrahimi holds a Ph.D. in Industrial Management and is the assistant professor of the Islamic Azad University. He is now a member of the scientific board with the department of industrial management in the Islamic Azad University Electronic Campus. Dr. Poorebrahimi has led over 50 dissertations in masters' level. He is the official manager of the International journal of Information, Security and Systems Management. His research interests are information security, data analysis and data mining, customer service, Electronic Support Services, and business analysis.

Negar Estemdad holds a Master's Degree in information technology management from Islamic Azad University, Science and Research Branch (Iran). She obtained her Bachelor's in basic science with specialization in applied mathematics from Mashhad University in Iran. Her research interests are customer relationship management and data mining. She is currently focusing on a research topic about integration of customer relationship management and Quality function deployment (QFD).

