A Data Mining Method for Satisfaction and Confidence of the Bank Customers

Parisa Allahverdizadeh ,Saeid Taghavi Afshord *
Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran
Emil: allahverdizadeh_p@yahoo.com,afshord@gmail.com(Corresponding auther)
Receive Date: 21 Now 2023 Accept Date:12 April 2024

Abstract

Trust is the main concern of the Bank's customers regarding electronic and Internet services. The trust of both customers is logically and experimentally important to each other, and banks need to take more steps as service providers to maintain their customers. It is necessary to increase the factors affecting the satisfaction and reliability of customers in banks using data mining. In this paper, we examine the factors affecting the increase of customers' confidence in banking and Internet banking services and the impact of any perceived credit factor by public and private banks, service providers, and infrastructure providers in electronic banking. The presented method is based on scientific data mining algorithms such as clustering and classification of the decision tree J48 and the neural network, as well as a quick and practical application of the miner. Data are analyzed using a questionnaire with the bank customers of 25 Tejart bank branches in Tehran. The experimental results demonstrate that the accuracy of the decision tree classification algorithm is 84.04 and the neural network is 72.3%.

Keywords: Trust, Bank, Data mining, Classification, Customers.

1. Introduction

In the past, organizations were only thinking about finding new customers and the improvement of customer services had no importance. Finding a new customer was a victory for the organization and there was no particular attention to existing customers. Now, the situation has changed and the maintenance and improvement of profitable customers is the first purpose of organizations. In today's society, customers are seen as a key and central factor for organizations. The orientation of all the goals, strategies, and organizational resources is attracting and keeping profitable customers.

So, the quality of service for customers in their loyalty as well as maintaining and developing bank relations and competitive position is a strategic challenge. Many expenses are spent to understand this concept and to achieve practical solutions to its strengths. Loyalty is a phenomenon that can be seen among customers who repeat purchase behavior and provide a positive attitude towards the company. Managers have successfully concluded that building consumer loyalty and keeping it can be done only in the light of their trust in organizations. Thus, changes in the direction of marketing and emphasis on customer orientation are based on trust in organizations rather than voluntary opportunity. Also, electronic customer loyalty is one of the important consequences of electronic customer relationship management. It ultimately leads to higher retention rates of these customers in organizations. An increase in customer loyalty can reduce marketing costs and increase the customer's demand.

Therefore, loyalty to business survival is one of the main priorities of organizations. The organizations emphasize creating sustainable and profitable relationships with customers

in today's challenging and competitive environment. A prerequisite for survival in saturated markets is not only attracting and satisfying customers but also establishing long-term relationships with them. This is the online world that is increasingly competitive and customers can just press a button to compare the products and services of their competitors, and its competitive advantages are achievable. This paper discusses related works in Section 2. Section 3explains our proposed method, whileSection 4 is about the implementation of the model. Section 5 contains the experimental results, andfinally, we give concluding remarks in Section 6.

2. Literature Review

Data mining is an integral part of customer relationship management. An assumption that companies can by identifying the characteristics and interests of their customers, establish successful relationships with them, has been mentioned in [8]. The impact of e-banking in increasing the bank's customer satisfaction has been investigated and concluded that the electronic services of the bank are the operational needs of customers. Also, it has been shown that the developed relationship with customer satisfaction has a direct linear relationship. Some results show that factors such as the speed of transactions, confidentiality, trusteeship, knowledge of employees, attitude, and attention to customer personality have a significant impact on customer satisfaction. The other results show that Internet banking has a very positive impact on employee attitudes and it reduces service delivery costs and improves service quality [1][4].

E-banking user's behavior was investigated by Foundation. In this study exchanges and electronic support measures and the access method to public computers for electronic banking were examined. Milne and Boza examined the possibilities of e-commerce features and customer confidence because of concerns about the use of banking services [6][2].

Kim Chang and colleagues noted in their study that generally believed security, improves confidence. And the perception of security and high trust certainly increase the use of e-commerce. This

study provides a theoretical basis for scientific and academic studies and for providers of security systems who are related to electronic payment [5][7].

Liu Yang and colleagues have used classification algorithms in the field of customer trust and satisfaction in their work. OsameAbo Abbas in the same year used a clustering algorithm in the field of customer satisfaction and Abdullah has done the comparison between various classes' methods and clustering on different types of data by the use of WEKA software [8].

According to the report research of the institute (Data Monitor) which is one of the most important analysis centers of banking information in Europe, statistics of electronic banking systems in eight countries of France, Germany, Italy, Netherlands, Spain, Sweden, Switzerland, and the United Kingdom reached from 4.5 million people in 1999 to about 22, million in 2004. In 2005 more than 75 percent of companies in developed countries, at least use one of the electronic banking services [9].

Come Score Company in a study that was about the confidence-building of bank customers, has concluded that Spain after Canada, Netherlands, France, Sweden, Great Britain, New Zealand, and Belgium is the eighth country in the world that can penetrate e-banking and its reason is a trust of customers. In 2014 Jana Kriman and colleagues examined in a study the different types of data mining techniques about trust-building and satisfaction and analysis.

3. The Presented Method

In this research, we use Tejarat bank data and questionnaires, and census sampling is implemented among all key customers of Tejarat Bank in 25 branches of Tehran. The content validity is used to determine the validity of the questionnaire. The reason for choosing the key customers is that all new marketing approaches would insist on maintaining the major customers of organizations. The total number of main customers of the studied bank according to the bank statements are 1585 people of whom 91.45 percent of them with the return rate of 1450 questionnaires participated.

Then in considering some completed forms excluded due to non-compliance with reliability. Ultimately, 1300 questionnaire forms including 626 women and 674 men with different ages and educational levels are studied. Factors affecting the development of their confidence using data mining software Rapid Miner are investigated and analyzed.

4. The Presented Method

A total of 44 questions in 6 groups are selected and defined and a conceptual model is developed to identify the factors leading to trust in the answers of key customers using the questionnaire. The first 5 questions are related to customers' personal information including gender, location, occupation, age, and education. The next 39 questions in this model are divided into six main variables including:

•The first group related to the bank factors including providing information and useful advice which contains 6 questions.

• The second group is related to bank staff factors which are 3 questions.

• The third group related to the bank's reputation, including 8 questions.

• The fourth group consists of 4 questions related to the provision of complementary services.

• The fifth group related to the bank's clients, which includes 7 questions.

• The sixth group related to bank technology and its processes which include 11 questions.

To achieve the research objectives, the research method is based on the crisp method. There are different methods for the implementation and execution of data mining and one of these very strong methods is the Crisp method. The method consists of identifying the system, data understanding, data preparation, modeling, evaluation, and development of the system. The steps of this method are consistent with the phases of the project.

4.1-First phase: Business and Data Identification

This research is conducted based on data management communication with clients of Tejarat Bank. According to the study questionnaire, its validity based on controlled content is done among 25 branches of that bank. The presented model of this study is a comprehensive model including six customer-oriented and trust structures. Brand Credibility, Customer Satisfaction, Loyalty Commitment, Continued Commitment, Word of mouth Recommendation (WOM), Switching Propensity. The questionnaire is prepared according to the communication services and these six structures. The role of customer confidence in banks with customer retention and customer agreement and provided services are provided.

4.2 The second phase: Data Processing

This step of the Crisp method includes data selection, data cleaning, and preparing data for data mining. Data cleaning is the quality control phase which is done before data analysis [1].

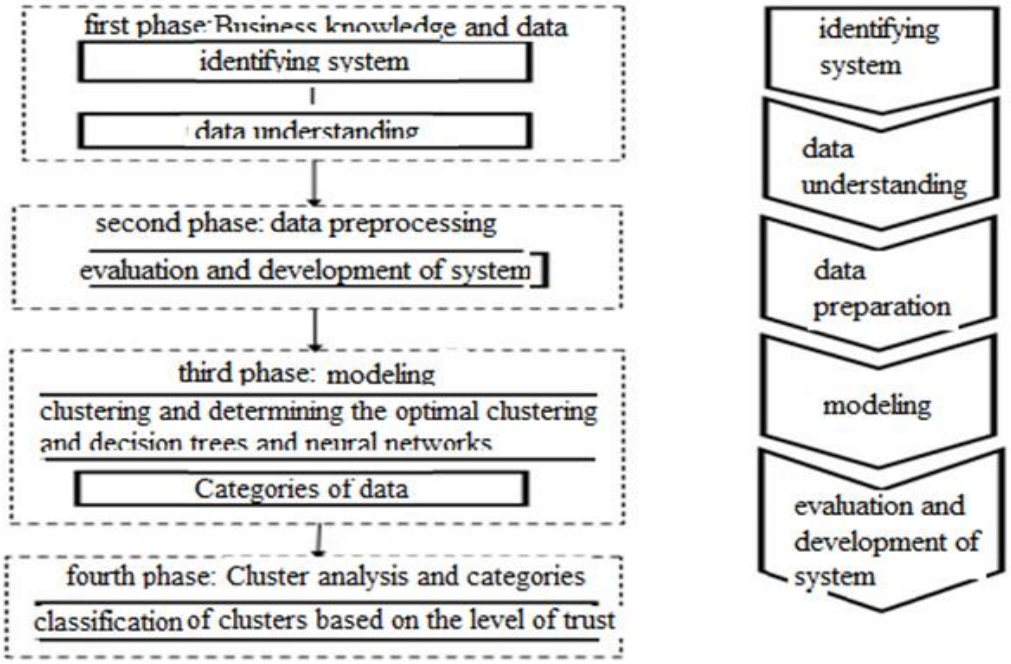

Fig.1.phases of the project implementation based on the Cresp method [3][4]

One of its responsibilities is filling or removing missed data. Missing data can be omitted if the number of non-value characteristics in a sample isn't too much, and missing values can't be filled in manually. In available data too since missing values are considered important fields. Therefore, some of them are eliminated, and from 1585 records. So, the number of investigated data is reduced to 1300 cases.

4.3- Third phase: Modeling

After studying the data and their preparation, now can go back to modeling. At this point, using data mining techniques, predictive models are discussed. Modeling is done with the help of Rapid Miner software.

✓ Modeling with KMEANS

In this stage of modeling, the KMEANS algorithm is used for clustering data. The KMEANS algorithm is a method of clustering data that because of its speed and simplicity widely used. Different method is used for clustering. However, the reason for using this method is that the optimal number of clusters is calculated using the index. This issue has output validity of clustering.

Rapid miner software is used for clustering. Bank data are divided based on trust into 6 groups and a group with personal information. Clustering is done using the KMEANS method, and then the optimal number of clusters is selected using the Davies-Bouldin index [10].

$$R_i = max_{i=1,\dots,n \;\; ci\neq j} R_{ij} \tag{1}$$

$$DB_{nc} = DB_{nc}\frac{1}{n_c}\sum_{i=1}^{nc} R_i \tag{2}$$

$$DB_{nc} = \frac{1}{C}\sum_{k=1}^{C} max\left\{\frac{s_c(Q_K)+s_c(Q_1)}{d_{ce}(Q_K,Q_1)}\right\} \tag{3}$$

DBnc:Is the mean similarity between each cluster and the most similar cluster to it. The number of clusters is tested with software. The best case is when the Davies index is less between clusters. In this study, 2-7 clusters are tested, evaluated, and implemented. The best case is for K=4 which can be seen in Table 1.

Table 1. KMEANS clustering for an optimal answer

K	Davies Bouldin
2	5.498
3	5.167
4	4.602
5	5.649
6	5.742
7	5.908
8	5.837

The number of each cluster is as follows, respectively: first cluster 150, second cluster 363, third cluster 396, and fourth cluster 391. Validation parameters are used to assess the accuracy of clustering to compare different clustering methods or to compare the results of a method with different parameters. These indicators include:

- Dunn index;
- Davies Bouldin Index;
- Root-mean-square standard deviation and R(RS) root;
- SD validity index;
- S _Dbw validity index.

In this paper, the Davies-Bouldin Index is used because this index calculates the mean similarity between each cluster with the most similar clusters to it. It can be concluded that, whatever the amount of this index is higher, better clusters have been produced. But, in the Rapid Miner software, as the amount of this index is negative, its minimum case will be the best amount.

✓ Modeling with KMEDOIDS

This algorithm is shown with the Partitioning Around Medoids (PAM) performance [10], and its property is that K is the first representative object (Medoid), and is selected arbitrarily from n object of the database. Each remaining object is clustering with a medoid that has the most closely resembles it. Then this strategy frequently replaces one of the medoid objects with the non-medoid object So that the quality of clustering results is improved. This quality is estimated using a cost function that the average of dissimilarity between an object and its medoid measures the cluster. In other words, this method is very similar to the previous method. The difference is that in the prior, each cluster is shown with an average of that cluster's objects (cluster center). But in this method, each cluster displays one of the objects located near the center of the cluster. The basis of this method is showing each cluster by one of the data which is located near the center of the cluster, and instead of using of cluster center as a reference, can be used medoids, which means the data which located in the most central part of the cluster.

In this method, a representative of each cluster is selected first, then the distance between the individual points is calculated and finally, the total cost function is counted. In the next step, we change one of the preferred points and then follow the same steps again. If the difference between the current total cost and the previous total cost is negative, a transfer is done and the representative of the cluster is changed. These steps must be tested for every point. The cost of transfer (S) can be calculated after changing the medoid of the cluster as follows: if it wasn't positive, the transmission of the intended medoid is ideally suited.

S = Current Total Cost – Past Total Cost (4)

To find the cost function or total cost, the cost per data is calculated from its cluster center, and finally, these numbers accumulate together. When the aforementioned issue is clustering with the Kmedoid cluster method, more time is consumed compared to the previous Kmeans. Also, the results of the Davies Bouldin Index of this part compared to the previous one are high. When comparing the clustering algorithms, simplicity, accuracy, speed, and better results are taken into consideration. Therefore, the results of both clustering can be seen in Table 2.

✓ Modeling with neural network

Table 2. Comparisons of two types of clustering

Measurement	Parameter	KMeans	KMedoids
Davies Bouldin	K=4	4.602	5.867
Avg.within centroid distance	All clusters	30.030	27.978
Execution Time	Processing speed	4min	30min

At this stage, the predictive models are provided using data mining techniques. Modeling is performed using the Rapid Miner software. The neural network classification algorithm is created using input variables and the determination of the target variable. To build a neural network model some variables such as gender, age, education, occupation, speed of service, bank employees, bank brand, technology, and staff dealing with customers are determined as predictor variables, and internet banking variables, quality and speed of service provided are determined as a target variable. Then, data is divided into two parts: training and testing, 90 percent for education and 10 percent for the test.

The neural network has an input layer, some hidden layers, and an output layer. Each node receives input, processes it, and produces output. To determine whether each input will receive to output node or not, depends on its weight. To predict field validation the neural network software is used which contains three groups: Perception, Neural Net, and AutoMIp. Among the three types of networks, Neural Net shows better accuracy in output. For this network, some parameters such as training cycles with 500, learning rate with 0.2, and Error Epsilon equal to 1.0E-5 are used. At this stage, the learning accuracy is 72.3 percent.

✓ Modeling with decision tree

Predictive models are discussed using data mining techniques. Modeling is performed using the Rapid Miner software. At this point, decision tree algorithms, including J48, and C5.0 are created using input variables and determination of target variable. To build decision tree models some variables such as gender, age, education, occupation, speed of service, bank employees, bank brand, technology, and staff dealing with customers are determined as predictor variables, and internet banking variables, quality and speed of service provided are determined as a target variable. data is divided into two parts: training and testing, 90 percent for education and 10 percent for the test. A decision tree is a combination of some logical requirements (if-then rule). Decision trees aren't just a demonstration of the decision-making process but can also be them to solve classification problems. Usually, a set of rules extracted from a decision tree, are the most important studies that come from them. In the created model in this application for dividing branches, the Gini index is used. In this study, the confidence standard for created rules is determined 80 percent. The reason for choosing this model is that the calculation shows the intended indices have the highest accuracy among the models that are implemented. Index calculation is presented in the assessment section.

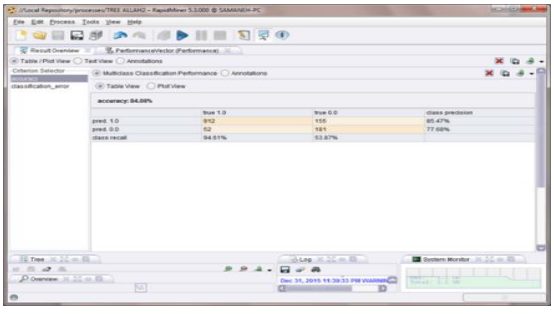

Fig.2. the accuracy made by decision tree J48.

5. The experimental results

The highest and the best clusters are related to the third and fourth clusters, each of them has 396 and 391 records. The reason is that data distance and their similarity are much less than others. Then, the second and fourth clusters have 363 and 391 records and finally, the first cluster has 150 ones.

✓ Analysis of the first cluster

This cluster with 150 records has the lowest population. Members of this group are women, 70 percent of them are living in Tehran and 75 percent of them are employees. Indeed, 50 percent of them are young and they have bachelor's degrees. They are satisfied on average with the provided information and bank consultations. Employee behavior with customers in this cluster is at a high level. They have medium satisfaction with the bank technology and SMS notification of the bank is at a high level.

✓ The second cluster analysis

The population of this cluster is 363 people that have the third largest population. The members of this group are women and they are self-employed. Indeed, 14 percent of customers in this group are elderly and 32 percent are middle-aged, 18 percent of customers of this group have a master's or higher educational degree. They are satisfied with the provided information and bank consultation on average. The employee's behavior toward customers is at a high level in this cluster. They are moderately satisfied with the bank's brand, and their satisfaction with the SMS notification and technology of the bank is at an average level.

✓ The third cluster analysis

This cluster with 396 people is in the first place. The people of this group are men, 95 percent of them are living in Tehran, 20 percent of them are self-employed and 80 percent are employees. Also, 46 percent of customers in this group are young and 35 percent have a high school diploma. They are on average satisfied with the provided information and bank consultation. The employee's behavior with customers in this cluster is at a medium level. They are satisfied with the bank's brand moderately. Their satisfaction with technology and SMS notifications is at a medium level.

✓ The fourth cluster analysis

This cluster with 391 populations is in second place, 71 percent of this group members are men and 29 percent are women. Indeed, 68 percent of them live in Tehran and they are self-employed, 44 percent of customers of this group are young and 49 percent of them have an associate degree or lower. They are satisfied with the provided information and bank consultation on average. The employees' behavior with customers at this cluster is at a medium level. They are moderately satisfied with the bank's brand. Their satisfaction with technology and SMS notifications of banks is at a medium level.

Conclusion

Nowadays, customers have many options to select the type of service and its provider. Managers in banks should be aware of the fact that if they don't pay attention to each customer specifically, the customer will be able to choose another bank for getting services. Therefore, many banks are trying to be customer-oriented institutions, unlike product-based companies. Marketing managers if can explore customer behavior and predict it, then will be able to have long-term and better relationships with customers. Flexibility, intimate relationships, and hostelry access are affecting parameters in gaining the trust of customers. Customers, mostly due to the basic

rights and lack of handling the complaints aren't satisfied. Therefore, in this case, the bank can improve its weakness to gain the trust of customers

References

[1]Allahyary Fard, Manhood, (1384), "Electronic service administrative": a series of lectures in the fifteenth conference of monetary policy and currency, central monetary institute of Iran.

[2] Khan Babaei, Mohammad, Nadali, Ahmad, (1387), "The use of decision trees techniques and genetic algorithms to validate the bank's clients in a decision support system": second national conference on data mining. Tehran: Amirkabir University of Technology.

[3] SHahrabi, Jamal, Shakurnia, Venus, (1387), "Data mining concepts in Oracle": Amirkabir research publications, first edition.

[4] Ghazanfarri, Mehdi, Alizadeh, Somayeh, Teimorpoor, Babak, (1387), "Data mining and knowledge discovery", the Center for Science and Technology University Press, first and second edition (1390).

[5] Chen, M. S., Han, J., Yu, P. S. (2012); "Data Mining: An Overview from A Database Perspective", IEEE Trans. Knowledge Data Engineering.no. 8 (6), pp. 866-883. doi:10.1109/69.553155.

[6] Gill A., Flaschner A., Shachar M.;(2006)." Factors that affect the trust of business clients in their banks; International Journal of Bank Marketing", Vol. 24 No.6, pp. 384-405.

[7] Hussain, F. K., Chang, E. and Dillon, T. (2006). "Trust and reputation for service-oriented environments."Vol.1: Technologies for building business intelligence and consumer confidence. West Sussex, England: John Wiley and Sons.

[8] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). "Application of data mining techniques in customer relationship management": A literature review and classification. Expert Systems With Applications, 36, 2592–2602.

[9] Sabzevari H., Soleymani M., Noorbakhsh E. (n.d.)(2013) ." A Comparison between Statistical and Data Mining Methods for Credit Scoring in Case of Limited Available Data."

[10] Witten, I.H., Frank, E. (2005). "Data Mining: Practical Machine Learning Tools and Techniques", 2nd Edition, San Francisco: Morgan Kaufmann.

[11]Akhter, F., Hobbs, D., & Maamar, Z. (2005). Using fuzzy cognitive time maps for modeling and evaluating trust dynamics in the virtual enterprises. Expert Systems with Applications, 28, 623–628.

[12] Alberto, Guillén, Héctor, Pomares, Jesús, González, Ignacio, Rojas, Olga, Valenzuela, & Beatriz, Prieto (2009). Parallel multiobjective memetic RBFNNs design and feature selection for function approximation problems. Neurocomputing, 72(16– 18), 3541–3555.

[13]Álvarez, J. M. (2008). La banca española ante la actual crisis financiera. Estabilidad Financiera, 15(November), 23–38.

[14]Azcorra, A., Bernardos, C. J., Gallego, O., & Soto, I. (2001). Informe sobre el estado de la teleeducación en España, Asociación de Usuarios de Internet.

[15]Belanger, F., Hiller, J. S., & Smith, W. J. (2002). Trustworthiness in electronic commerce: The role of privacy, security, and site attributes. Journal of Strategic Information Systems, 11, 245–270.

[16] Bellman, S. G. L. (1999). Predictors of online buying behavior. Communications of the ACM, 42(12), 32–38.

[17]Flavián, C., & Guinalíu, M. (2007). Un análisis de la influencia de la confianza y del riesgo percibido sobre la lealtad a un sitio web: El caso de la distribución de servicios gratuitos. Revista Europea de Dirección y Economía de la Empresa, 16(1), 159–178.

[18] García, N., Santos, M. L., Sanzo, M. J., & Trespalacios, J. A. (2008a). El papel del marketing interno como antecedente de la capacidad de innovación de la PYME. Efecto sobre los resultados empresariales. XXII Congreso anual AEDEM. Salamanca, 18, 19 y 20 de Junio.

[19] García, N., Sanzo, M. J., & Trespalacios, J. A. (2008b). Can a good organizational climate compensate for a lack of top management commitment to new product development? Journal of Business Research, 61, 118–131.

[20]Gefen, D. (2000). E-commerce: The role of familiarity and trust. The International Journal of Management Science, 28, 725–737.

[21]González, E., López, M. J., & Lampón, J. F. (2011). La confianza del consumidor como factor clave en la construcción de lealtad en ambientes electrónicos, XXI Jornadas Hispano Lusas de Gestión Científica, 2–4 de Febrero, Córdoba.

[22]Guillén, A., Sovilj, D., Lendasse, A., Mateo, F., & Rojas, I. (2008). Minimizing the delta test for variable selection in regression problems. International Journal High-Performance Systems Architecture, 1(4).