International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies

http://TuEngr.com

TuEngr Group

PAPER ID: 11A050

CUSTOMER LOYALTY EVALUATION AND PREDICTION BASED ON DECISION TREE AND ARTIFICIAL NEURAL NETWORK: CASE OF OFOGH KOOROSH STORES IN TEHRAN

Amirreza Estakhrian Haghighi [1*], Abdolreza Shirazi [1*]

[1], *Department of Computer & Information Technology, Shiraz Branch. Islamic Azad University, IRAN.*

ARTICLEINFO

Article history:
Received 30 May 2019
Received in revised form 09 December 2019
Accepted 16 December 2019
Available online 26 December 2019

Keywords:
Decision Tree method; Artificial Neural Network (ANN); Learning algorithm; Customer loyalty prediction; Data Mining; Data training; K-Means clustering; Multilayer perceptron (MLP).

ABSTRACT

The secret to staying in the business world today is having satisfied and pleased customers who buy the company services over and over again and introduce such products/services to others. These companies need to know what the customer wants and how they can adapt to the customer's needs according to customer preferences over time. In this work, the factors affecting customer satisfaction and loyalty of Ofogh Koorosh stores were studied and then analyzed using data mining techniques and methods and the extent to which each factor influenced their loyalty. The results of the decision tree and artificial neural network (ANN) with different segmentation of observed data and evaluation criteria of the obtained models show that in the decision tree model with clustering criteria and dividing data set into six clusters the burden of computation and classification accuracy have been increased, and each criterion is initially prioritized within itself. The results of these clusters are combined and the results accurately predict customer loyalty. The proportionality between accuracy and readability criteria in this algorithm indicates that the considered criteria values are well averaged and have a uniform distribution because the detection rates of the samples with low priority over the samples compared with high priority are almost equal. In other words, criteria have well recognized the high and low-priority decision tree algorithm. In neural networks both in modeling (training) and in the validation (testing) it shows the high coefficient of explanation but with the observation of error rate of both modeling and validation it can be concluded that it is possible to have unbalanced data in our dataset. The error rate of the decision tree is less than the ANN according to F-values, indicating the decision tree as a whole is more successful in estimating burnout and prioritization than ANN.

Disciplinary: Multidisciplinary (Information Sciences, Management).

*Corresponding author (A.Haghighi, A.Shirazi). Email: amirestakhrian@iau.shiraz.ac.ir, apashanishirazi@yahoo.com

1. INTRODUCTION

Today is no longer the age of loyalty of the six, neither ccustomer loyalty, employee loyalty, management loyalty, loyalty to society and principles, etc. Many studies have shown that loyalty is not the ultimate key to success and profitability. In fact, the marketing concepts that have been emphasizing it until today are no longer recognized it. Today the only happy and glad customer and a customer who has felt a sense of belonging and belonging to the organizations are assets that have long-term profitability and longevity.

Given the widespread efforts to improve the quality management tools and expand the customer-centric attitude by researchers, experts, and managers of business organizations, it seems that customer loyalty is now one of the most important factors in determining the success of organizations in business and profitability. Therefore, in order to achieve business development, assessing the behavior, attitude and perception of customers about the quality of products and services provided by business organizations has become inevitable. On the other hand, the organization's aspirations and goals have undergone major changes in recent years due to the intense expansion of global competition and the dynamics of the international economy. While it has long been the main focus of organizations to attract new customers, business strategic policies today focus on maintaining and improving loyalty and increasing customer confidence in the organization.

The most important causes of such a change are the increasing public awareness of the desired consequences of customer loyalty. Companies that have gained a larger share of loyal customers due to a variety of reasons, such as resale rates, advertising and verbal product recommendations, reduced willingness to switch and source supplier changes, have greatly increased their organization's profitability.

1.1 QUALITY OF SERVICE AND CUSTOMER LOYALTY

Today, given the fierce competition between competitors to attract new customers, creating customer loyalty is critical to the survival of organizations. To this end, many organizations try to build loyalty to their customers in a variety of ways, including pursuing relational marketing and maintaining their relationship in the long run (Jessy, 2011). Many studies of the relationship between service quality, customer loyalty, and customer loyalty have found that customer loyalty plays a mediating role in the impact of service quality on customer loyalty. In other words, in many studies believe that the quality of customer service and satisfying customer loyalty are two essential prerequisites for customer loyalty.

1.2 SERVICE QUALITY MODELS

Quality is not a one-dimensional phenomenon, but a multi-dimensional phenomenon. Therefore, it is not possible to achieve service quality without identifying important aspects of service quality. In the following, the factors and models that determine the quality of service will be discussed.

1.3 GRONROOS MODEL

Gronroos introduces three dimensions of quality service:

Technical quality: The technical or outcome quality that refers to the actual outcome of the service being evaluated after the service has been provided. Outcome is what the customer receives from the organization. Service outcome is often evaluated by the consumer in an objective way.

Functional Quality: Process or operational quality refers to the quality of processes and procedures in producing and delivering customer services. Due to the synchronization of production and service consumption, the quality of the process is usually evaluated by the customer when performing the service. This element of quality refers to the interaction between the provider and the recipient of service and is often perceived in a mental way.

Company image: This dimension relates to customer perceptions of the service organization. The subjective image depends on the outcome and functional quality, the price, the extracurricular activities, the physical position, the fitness of the branch, the competence and behavior of the organization's employees (Bayraktar et al., 2012).

2. RESEARCH METHOD

Three main phases were implemented to apply a data mining algorithm to research data to identify and explain hidden patterns between data and obtain applied and functional results:

1) Data mining of research data;

2) Analyzing the results of data mining;

3) Explaining the hidden patterns between the research data and the presentation of applied results.

It is important to note that the sequence of these phases greatly enhances the validity and accuracy of the results and can, therefore, serve as a basis for strategic management decisions of organizations in the field of customer interaction and communication, and how customer loyalty systems are adopted, implemented and applied by organizations wanting to use this management tool. Figure 1 shows the research process.

2.1 STATISTICAL POPULATION OF RESEARCH

According to the research done in Ofogh Kourosh stores in Tehran, the selected statistical population is also from this collection, consisting of 800 people. Experts and experienced experts in each department who are familiar with the concepts of knowledge management, its components and factors affecting its adoption and application in organizations, have been utilized. Therefore, in each store, 10 people were selected as the statistical sample, in which for the most part, the selected individuals included fixed customers.

2.2 RESEARCH RELIABILITY

Validity means the extent to which the method or tool used can accurately measure the desired properties. The three main types of validity are as follows:

- Content validity
- Criterion-based validity
- Construct validity

In this study, after distributing the questionnaire to 20 organizational experts in customer loyalty, Cronbach's alpha coefficient was calculated for it, see Table 1.

Table 1: Cronbach's alpha coefficient.

N of Items	N	Cronbach's Alpha
20	26	0.949

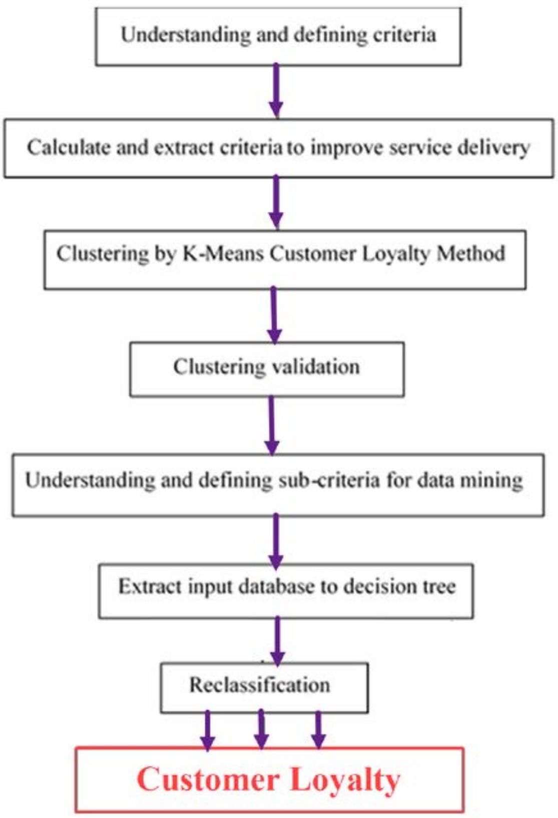

Figure 1: Research process.

2.3 CRITERIA ANALYSIS

In this section, we first identify the criteria and consider the impact of different criteria on the impact of customer loyalty. Subsequently, according to the comparative matrix of quantitative and qualitative valuation of the sub-criteria, the value of the sub-criteria is determined. It is noteworthy that the final customer loyalty maps are derived from the combination and overlapping criteria and sub-criteria maps.

2.3.1 EVALUATION CRITERIA

In order to evaluate the performance of the prioritization algorithms, the total set of data collected for this research is randomly divided into two parts T (training set) and P (test set). The training set contains the data used to train the classifier and the output of this training is the classification that is measured by receiving the test dataset of classification accuracy. Performance evaluation criteria for the proposed method for decision tree will be accuracy, precision, Recall, and F-measure. These criteria have been chosen because of their widespread use in data mining researches. Accuracy and recall are widely used to measure the efficiency of the selection of worn-out areas by the proposed algorithm. Also, performance evaluation criteria are for the neural network method, coefficient of determination, mean square error, and root means square error percentage. Here, for classification method with a decision tree, the customer loyalty problem is turning into a two-class problem and positive class samples represent a high priority for customer loyalty allocation, and negative class samples represent a low priority for customer loyalty. One of the solutions for checking the accuracy of the classification algorithm is the use of a confusion matrix. Figure 2 shows the two-class confusion matrix. The positive class represents the high priority and the negative class indicates the low priority.

Table 2: Confusion matrix

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

According to the confusion matrix, TP is the number of samples of high priority and the proposed method has correctly identified the highest priority. The term FP indicates that the classifier has mistakenly placed a high priority on customer loyalty. The term TN is the number of samples that have low priority for loyalty and the proposed method correctly identifies them as a low priority for loyalty. The term FN indicates that the classification has been given a low priority by mistake. The most commonly used scale to evaluate the performance of predictive models is predictive accuracy, which is the ratio of the number of correctly predicted classes (from free and error-free classes) is defined as the total number of classes. The criterion of accuracy (or success rate) for measuring the overall accuracy of predictive precision that from the confusion matrix, the accuracy of the estimation is calculated according to the following relation (Pollard, 1981).

$$AC = (TP + TN) / (TP + TN + FP + FN) \quad (1).$$

2.3.2 PRECISION

Assuming high priority allocation, the prediction accuracy is obtained by proportionally sampling correctly classified to the total number of available samples. The higher the Precision value, the higher the quality of the algorithm (Fahad et al., 2014).

$$Precision = TP / (TP + FP) \quad (2).$$

Recall: Assuming low priority allocation, the prediction accuracy is in the ratio of correctly classified samples to the total number of correctly classified samples. The best value is one. The high recall rate means less FN (Singh & Reddy, 2015).

$$Recall = TP / (TP + FN) \quad (3).$$

2.3.3 COMBINED MEASURE F

F-measure considers both accuracy and recall for accuracy calculation, which can be interpreted as a weighted measure of accuracy and recall. This weight is denoted by α and is usually considered one. The F-value is between 0 and 1, and the closer it is to 1, the better the performance of the classification results (Brijain et al., 2014).

$$F = 2 * (precision. recall) / (precision + recall) \quad (5).$$

The main advantage of the root means square error percentage is that the predictions made by the different models allow comparisons to be made and the appropriate models are chosen using the above models (Teli & Kanikar, 2015). In the above relationship, n, yt, $\hat{y}_t$, $\bar{y}_t$ are respectively the

*Corresponding author (A.Haghighi, A.Shirazi). Email: amirestakhrian@iau.shiraz.ac.ir, apashanishirazi@yahoo.com

number of observations, the observed customer loyalty, the predicted customer loyalty, the average customer loyalty observed. Many classifiers such as decision tree-based or law-based methods are designed so that they generate only one binary output (based on belonging to one of two possible classes). This type of classifier, which produces only one specific output for each input, is called a discrete classifier. Similarly, there are other classifiers such as Bayesian-based classifier or neural networks that generate a probability or score for each input, which represents the degree of input belonging to one of the two available classes. These classes are called continuous classes, and because of the specific output of these classes, a threshold is used to determine the final output. Among classification algorithms evaluation methods (in this algorithm, working routine is such that classification model is built by training dataset and evaluated by the experimental dataset.), the Holdout method can be used to determine how the ratio of data sets (to two training datasets and experimental datasets) depends on the detection of the analyzer, usually two-thirds for training and one-third for evaluating. The main advantage of this method is the simplicity and high speed of the evaluation process, but the Holdout method has many disadvantages, including the training and experimental data sets that will become dependent on each other. In fact, part of the initial dataset that is separated for testing does not have a chance of being in the training phase, and similarly, if one chooses a record for training, there will be no chance of using this record to evaluate the model built. Also, the model built depends largely on how the initial dataset is divided into training and experimental datasets. If we run the Holdout method multiple times and average the results, we have used a method called Random Sub-sampling. The most important disadvantage of this method is the lack of control over the number of times a record is used as a training sample or a testing sample. In other words, some records may be used for learning or evaluation more than others. Also in the comprehensive k-Fold Cross Validation method, the entire data set is divided into k equal parts. The k-1 part is used as the training dataset and the model is built accordingly and the evaluation is performed with the remaining part. The process will be repeated k times so that each k part is used only once for evaluation and each time accuracy is calculated for the built model. In this method, the estimation of the final accuracy of the classifier will be equal to the average of the k calculated accuracy. The most common value for k in scientific literature is 10. Obviously, the larger the value of k, the more accurate the computed accuracy for the classifier will be, and the resulting knowledge will be more comprehensive. The most important problem is increasing the evaluation time. The maximum value of k is equal to the number of records in the original dataset, known as Leaving One Out.

3. SIMULATION ENVIRONMENT

To implement the proposed method, the program is written in Java. For the data pre-processing step, Java is coded using the Eclipse environment and the Weka software APIs in Eclipse are called for clustering, classifying and neural network applications. The results are saved as a file and output has been designed in the form of charts and tables using Excel software. For each version, packages are implemented for grouping and applying the learner and computing the results. This program is executed using the SDK compiler in the Windows operating system and the evaluation charts are drawn using Excel software. The hardware specifications of the proposed operating system are 16 GB main memory and Intel Core i7 CPU and Windows 8.1 operating system. Because of the proposed method on real datasets, we need to implement a scalable and efficient algorithm. For this reason, in all functions, structures such as a vector or one-dimensional arrays are used for each block and neighborhood information. For classifier applications, classification algorithms are implemented in a Java environment and evaluated with different parameters and using Weka software for decision tree and neural networks to obtain the best classifier. Finally, the experimental data are tested and the evaluation functions are calculated.

3.1 EVALUATION OF METHODS

In this section, the results of applying the methods used to prioritize are presented and the results are analyzed below.

3.2 INITIAL PREPROCESSING

Iran's office system is constantly changing and it is difficult to gather and record this information properly which this process requires a lot of precision. One of the challenges in the field of data collection is the recording and storing of all information without losing data because at any stage it is possible to lose or lose data so in case studies given the amount of data lost, it is possible that the results can be mistaken. In this regard, we tried to examine the preliminary data first and replace the fields whose information is missing or lacking in ways such as averaging or reasoning with new data to achieve acceptable results. For the preprocessing of the collected data, alternate valid data is selected for each field where valid data is not available.

4. RESULTS OF THE DECISION TREE METHOD

To apply the decision tree method to the existing dataset, first, the clustering is done and then we need for a field labeled as described in Table 3.

Table 3: K-Means clustering results.

Cluster	Member of cluster
Cluster one	•level of closed opportunities •customer return rate •number of sale calls •number of sale calls for each opportunity
Cluster two	•Average number of written servicing requests •Average time to fix the problem •Average number of calls per service day •Percentage of Compliance with Service Level Agreements (SLAs)
Cluster three	•Percentage of service renewal and support •Level of customer loyalty •Time from complaint to problem resolution •Create interest for customers away from the collection
Cluster four	•Files that are closed in one day •Number of cases handled by the employee •Number of service calls
Cluster five	•Amount of new revenue •The amount of income in a given time period •Channel closing time •The difference between the selling price of a product or service and the cost of producing it •Duration of the sales phase
Cluster six	•Number of potential customers •Number of new customers •Number of customers retained •Number of open opportunities

*Corresponding author (A.Haghighi, A.Shirazi). Email: amirestakhrian@iau.shiraz.ac.ir, apashanishirazi@yahoo.com

4.1 K-MEANS CLUSTERING

Cluster formation is a process by which a set of localities or neighborhoods are divided into clusters or groups in such a way that the criteria within each cluster are as similar as possible while the criteria in different clusters are as distinct as possible. The clustering algorithm used in this study is the K-Means algorithm. In this step, clustering is performed using the finalized data set of the previous step. By applying the K-Means algorithm to the dataset, all criteria are finally divided into six clusters.

4.2 LABELING EXISTING SAMPLES

In data mining problems, each row of data is called a sample. Here a row of data about a sub-criterion is called a sample. To classify existing sub-criteria, a label field is required, indicating whether the existing sample has a high priority or a low priority. In other words, if the label field is equal to one, it indicates a high priority, and if the label value is zero, this sample has a low priority. In this study, by combining several fields and their arguments for each field record label is added. To check the value of the label field, to determine the importance level, the impact of the considered criteria, the six criteria; the number of sales calls - the number of sales calls for each opportunity - the amount of new revenue - the amount of revenue within a given time period- the closing time through the channel - the difference between the selling price of a product or service and its cost of production are defined and for the first two criteria are specified sub-criteria.

4.3 INVESTIGATION OF TREE STRUCTURE AND CLASSIFICATION BY DECISION TREE METHOD

The decision tree is a decision support tool that uses trees to model. Regarding data, decision trees create a tree-like structure that acts like the IF and ELSE rules, and ultimately reaches our preferred tags learned from the training data. In fact, the decision tree learning operation is to build the elements and leaves of a tree. Here, the decision tree is extracted by examining the values of existing criteria using the Weka software.

4.4 EVALUATING THE RESULTS OF THE DECISION TREE METHOD

To this point, samples have been prepared to apply the classifier. At this stage, the decision tree classification is applied to the samples, see results Table 4.

Table 4: Decision tree results in selective datasets by clustering.

Cluster	Accuracy	F-measure	Precision	Recall
Cluster one	0.95	0.96	0.95	**0.95**
Cluster two	0.97	0.97	0.96	**0.97**
Cluster three	0.96	0.96	0.97	**0.96**
Cluster four	0.95	0.95	0.96	**0.95**
Cluster five	0.97	0.97	0.97	**0.98**
Cluster six	0.98	0.98	0.98	**0.98**

The values in Table 4 show that you can rely more on the criteria stated by the sixth cluster. Because the accuracy of the algorithm for the sixth cluster is 98%.

4.5 NEURAL NETWORK METHOD RESULTS

In order to apply the neural network method to the existing dataset, the initial data must first be standardized, which in the following the steps of the neural network until the results are described:

4.6 PREPARING DATA FOR NEURAL NETWORK MODELING

The data used before entering the neural network were standardized to increase the accuracy and speed of neural network processes. To standardize, these relationships are used such that the data are in the range of 0 and 1.

$$\overline{\mathrm{X}} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{6},$$

where x is standardized data, x_i is used data, x_{min} and x_{max} are the lowest and highest data sets in the data set. After data standardization, in order to model data, the ten-segment method is used. In order to model with neural networks, the whole dataset is divided into ten equal parts. The nine parts are used as the training dataset and the model is built on it and the remaining part is evaluated. The process will be repeated 10 times so that each of the 10 parts is used only once for evaluation and each time precision is calculated for the built model. Finally, from the coefficient of determination, the training phase error, the accuracy is averaged. Next, all the data fields are entered to the neural network as input.

4.7 TOPOLOGY OF THE PROBLEM

At this stage, the network type, the number of hidden layers, the number of neurons in each layer, and the base and stimulus functions are selected. Since feedforward networks with backpropagation algorithm are one of the most commonly used neural networks, therefore, the present study used feed-forward neural network with backpropagation algorithm (multilayer perceptron networks). In the present study, 76 variables were introduced as input layer neurons and burnout rate as the priority of facility allocation as output layer neurons to the neural network. Determining the number of neurons in the hidden layer is not easy, one of the most appropriate methods to determine the number of hidden layer neurons for the least amount of error is to use trial and error method [8]. In the present study, the trial and error method was used to determine the number of neurons in the hidden layer. In this way, different neural networks with a different numbers of neurons were constructed in the hidden layer (from eight neurons upwards). This continued until the best model was selected. Among these models with different numbers of neurons was selected the best model with the least error and the most accuracy (in training and experiment stage). Also in feed-forward neural network, sigmoid logarithm function as transfer function and The Levenberg-Marquardt Learning Machine (LM) algorithm has been used as a learning algorithm.

4.8 MULTILAYER PERCEPTRON (MLP) NEURAL NETWORK RESULTS FOR THE DATASET

One of the important indicators in a neural network is the calculation of the root mean square error (RMSE). The R^2 criterion also indicates the correlation coefficient for the test data. The values of these two for the selected dataset are reported in Table 5.

The results of the above table show that in feed-forward network modeling the coefficient of determination in the training dataset is higher than the test dataset and the average of these two datasets is 0.98. Despite the high coefficient of determination, the root mean square error in both datasets was high, indicating that some data values attempted to mislead the learner, in other words, the data were unbalanced data. Therefore, among the models, in terms of the coefficient of determination and error of training and testing, it can be concluded that the feed-forward neural network with 12 neurons in the hidden layer and the topology variables as their inputs can provide better results.

Table 5: Simulated model results with different error measurement statistical criteria

Type of network	Number of fold	Topology Factors	
Feedforward network		Activation function	Logsig
		Learning algorithm	LM
	R^2	training	0.99
		testing	0.97
		total	0.98
	RMSE	training	0.7839
		testing	0.8123
		total	0.79

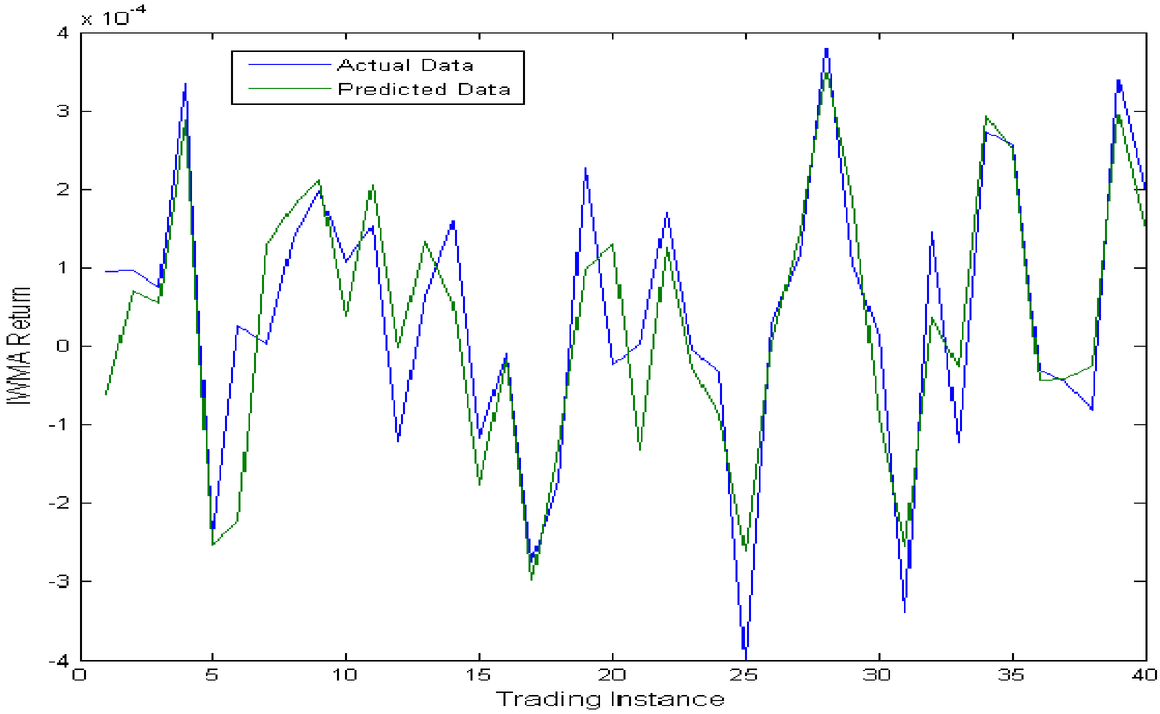

Figure 2: Comparison of actually predicted outputs by neural networks in estimating customer loyalty.

5. CONCLUSION

In this research, by first evaluating the criteria and sub-criteria, an initial pre-processing of the data was performed and then the database needed to enter the decision-tree learning algorithm has been extracted. Before applying this algorithm to the dataset, the criteria were clustered using the K-Means clustering algorithm. Since the most important task of a clustering algorithm is to optimally minimize the intra-cluster distance and maximize inter-cluster distance, we have implemented K-Means clustering algorithms and obtained six clusters with similar data within each cluster. Then, the decision tree algorithm was applied for each cluster separately and the priority of different criteria was evaluated. According to the results of the decision tree algorithm, it can be said that the classification pattern of the relevant data was well analyzed and the results were very good. Because each cluster has a high accuracy rate of 96%. In this research, for modeling the problem using MLP grids, a hidden layer starting with eight different neurons and sigmoid function boundary stimulus functions, and among various methods of training, error back-propagation method by Levenberg-Marquardt algorithm for faster convergence in network education have been used. The reason for choosing a hidden layer is because the middle layers are not directly related to the output, so the changes in these layers do not have much effect on prioritizing adjustment. This layer is also selected to reduce the error. The basis of the error backpropagation method is based on the error correction learning law, which consists of two main paths. Along the way, the input vector is applied to the network and its effects propagate through the middle layer to the output layer, and the output vector produces the real response of the network. One of the challenges of this method is the unbalance of the data that increases the neural network error.

6. DATA AND MATERIAL AVAILABILITY

Information regarding this study is available by contacting the corresponding author.

7. REFERENCES

Bayraktar, E., Tatoglu, E., Turkyilmaz, A., Delen, D., & Zaim, S. (2012). Measuring the efficiency of customer satisfaction and loyalty for mobile phone brands with DEA. *Expert Systems with Applications*, *39*(1), 99-106.

Brijain, M., Patel, R., Kushik, M., & Rana, K. (2014). A survey on decision tree algorithm for classification.

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., ... & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, *2*(3), 267-279.

Jessy, J. (2011). An analysis on the customer loyalty in telecom sector: Special reference to Bharath Sanchar Nigam limited. *African Journal of Marketing Management*, *3*(1), 1-5.

Pollard, D. (1981). Strong consistency of k-means clustering. *The Annals of Statistics*, *9*(1), 135-140.

Raza, H., Burney, A.I., Ahsanullah. (2020). Impacts of Service Quality on Customer Satisfaction: A Comparative Study on Banking Sector of Pakistan Through Weighted Servperf Model. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*. 11(2), 11A02I: 1-15.

Singh, D., & Reddy, C. K. (2015). A survey on platforms for big data analytics. *Journal of big data*, *2*(1), 8.

Teli, S., & Kanikar, P. (2015). A Decision Tree Based Survey in Data Mining Approaches. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, *5*(4), 613–617.

Dr.Amirreza Estakhrian Haghighi is an Assistant Professor at Computer & Information Technology Department, Shiraz branch, Islamic Azad University, Shiraz, Iran.

Abdolreza Shirazi is a Master's degree student in Information Technology (IT) in the e-commerce branch. He has a BA in Software engineering from Islamic Azad University, Shiraz branch. He is interested in applying IT to Management.