Implementing a Data Mining Solution Approach to Identify the Valuable Customers for Facilitating Electronic Banking

https://doi.org/10.3991/ijim.v14i15.16127

Mohammad Vahid Sebt (✉)
Kharazmi University, Tehran, Iran
sebt@khu.acir

Elahe Komijani
Islamic Azad University (South Tehran Branch), Tehran, Iran

Shiva S. Ghasemi
Kharazmi University, Tehran, Iran

Abstract—Nowadays, the banking system is known as one of the inherent sectors of customer relationship management systems. Its main advantage is to redesign a more responsive organization to satisfy the customers. The banking system aims to improve the structure of organizations to provide a better customer service through a set of automated and integrated processes. The final goal is to collect and reprocess the personal information of customers. To handle this dilemma, a number of new techniques in data mining provide a powerful tool to explore customers' information regarding a set of data and tools for customer relationship management. Accordingly, the customers' classification and coordination of banking system are the main challenging issues of today's world. These reasons motivate the attempts of this study to apply a composition of neural network by considering the C4.5 decision tree and the k-closest neighbor method as a variant of core boosting methodology with maximal strategy. To validate the proposed solution approach, a case study of Ansar Bank in Iran is utilized. From the results, it is observed that the proposed method provides a competitive output with the rate of 95% for the customers' classification. It also outperforms other existing methods with the rate of C4.5 decision tree, neural network, Naive Bayes and KNN with the rate of 1.04%. The main finding of this research is to propose an algorithm with the error rate of 1.9% and error squared of 0.72% as the best performance among other methods from the literature.

Keywords—Customer classification, boosting, C4.5 decision tree, neural network, KNN.

1 Introduction

Nowadays, in the banking industry, loans play a significant role, so that a large part of the assets of a bank are made up of loans paid to individuals and corporations, and

as a result, by increasing the number of loan applications from individuals and with regard to risk available in these activities, providing a way to manage these loans is necessary. One of the ways to quantify and measure credit risk and proper management is the use of credit scoring models. This research modeled based on quantitative and qualitative criteria, features and performance of past loans, to predict future performance of similar loans. A credit rating is a statistical tool used to determine degree of risk of loan payments to customers.

In CRM customer relationship management, the most important asset of most organizations is their customers. Customers, for the direct relevance of the actions of an organization, are a valuable source of opportunities, threats and operational questions related to the relevant industry. In this way, it is necessary for the organization to design and implement a system for attracting and retaining customers, a system that can manage the organization's relationships and customers well. These systems are known for customer relationship management systems and software known as CRM that can make the organization more responsive to customers, whose goal is to empower the organization to provide better services to its customers through the creation of automated and integrated processes for collecting and processing customer personal information. Customer relationship management consists of the business process, technology, and roles required managing customers in the various stages of the organization's life cycle.

Data mining is one of the new techniques for discovering patterns and trends with respect to customer data, which improves customer relationship and is one of the tools for customer relationship management. Segmenting is a way of knowing the customer and breaking the entire population of customers into smaller groups.

In this research, we intend to identify a profile for customers using the characteristics of customers and identify customers who are important and profitable for the banking system. In this way we can create more convenient facilities for them. For segmentation, a model called RFM is used, which is one of the models in the analysis of customer value, the RFM model, and presented by Hughes in 2013, and it shows the difference of customers using three variables: recency, frequency and monetary value.

The recency of the last purchase (R) represents the freshness that shows the duration between the last business engagement with the present, the longer it is, the more R is. Frequency purchase (F) represents a repetition and shows the number of transactions in a given interval, the higher the repetition, F is also larger. The monetary value of the purchase (M) shows interactions in a certain interval that the higher the monetary value, M is larger.

The RFM model variables are very efficient for customer segmentation. Other applications of the RFM model in customer segmentation based on novel variables, repetition and monetary value can be used to segment customers in order to determine optimal marketing policies for each sector. In fact, RFM is a method that examines these three features for each customer and gives them a customer benefit. The ability to identify profitable customers, build long-term loyalty with them and expand existing relationships is crucial/critical and competitive factors for a customer-centric organization. The need for these competitive factors is the existence of a customer relationship management of the organization. The evaluation of customer profitability is one of the key factors in this management. Following initial studies about the work done to determine

customer loyalty, it has become clear that a methodology based on combining popular classifications such as C4.5 decision tree and neural network and k nearest neighbor to assess customer loyalty has not been presented so far. After customer clustering, the cluster output is entered into the category, then, based on the RFM criteria, all customers model and assign each batch instance. The results of applying this method on a dataset indicate its higher accuracy than previous methods in identifying loyal and non-loyal customers. Therefore, the researcher in this study is trying to apply a method based on the clustering algorithm and classification methods such as the neural network and C4.5 decision tree or a combination of the above methods that classify valid and invalid customers in the electronic banking system with acceptable accuracy than other methods.

Considering the importance of the classification of credible customers in the electronic banking system, today, classification methods have become one of the most accurate methods of credit analysis among other tools. Therefore, the researcher in this study is trying to apply a method based on clustering algorithm and combining classification methods that provide valid and invalid customers in the electronic banking system with acceptable accuracy compared to other methods. It should be noted that in order to provide a new method, we can use the combination of classification methods in the current research and classify the customers with the appropriate precision.

In this research, using the user profile and RFM technique, we introduce a combination of popular categorization algorithms and then identify valid users in a banking system with a degree of affiliation. Therefore, the main objective of the current research is to increase the accuracy of customer ratings in the electronic banking system, so that the proposed method is more appropriate than other proposed methods.

The objectives of this study are to provide a template or comprehensive model for providing or not providing facilities to future clients of the facility, and the classifications and classifications of current customers of Ansar Bank.

2 Literature Review

It is possible to classify customers by using a data set of depositors of a loan, taking into account information on the characteristics of customers of different banks, such as age, sex, occupation, annual income, total assets, other earnings, etc. [1].

Customer revenue performance plays an important role in loan contracts. In a study using a sample of 3725 loan facilities from banking customers during the period from 1995 to 2011, the impact of customer earnings performance on the prevailing and non-prevailing conditions of loans was examined and considered three main hypotheses and obtained results using regression [2].

The clustering and data mining methods are widely used in different regions. The prediction of oil performance in oil fields [3], the forecast of non-performing loans in Post Bank branches [4], the identification of rules and relations between grades Input tests and other personal and occupational variables and the status of employees with job performance [5], maintaining customers in a telecom company [6], assessing customer needs [7] and the use of direct marketing methods by corporations and banks to

reach customers [8] are some examples in different fields that researchers use using data techniques K They have done it.

In an article, a framework is proposed in which all communication rules are clustered using a new similarity criterion, and the levels of real satisfaction are embedded in the communication rules to enrich them in an innovative manner [9].

Particularly, in customer relationship management (CRM) and customer needs assessment, the clustering and classification of the dataset using customer attributes are used in order to increase profits and increase investment (ROI). [10] and [11]. customer needs assessment by the VARCLUS algorithm [12], customer needs assessment by self-organized mapping algorithm and WRFM model [13], customer needs assessment by the data-mining prediction approach [14] customer loyalty assessment through k-mean clustering and WRFM model [15], and consequently customer loyalty predictors by PLS-SEM [16] are provided.

[17], [18] and [19] investigated customer profiles to improve customer relationship management performance. [20] has criticized the concept of data mining and customer relationship management in banking and retail organizations. They also discussed standard data mining tasks and evaluates various data mining applications in different sectors. [21] wrote about the responsibility of companies to ensure the accuracy of information stored and to remove inappropriate data in their research.

By integrating two methods of clustering and data envelopment analysis, it is possible to identify bank cluster management branches and to evaluate the efficiency of payment efficiency [22] or with the aim of customer clustering based on hierarchical structure with optimization of logistic network [23].

Using the data mining approaches, the Stock Profit Framework (StockProF) has been redefined to create stock portfolios [24]. And, data mining-based performance appraisal framework, to conduct an *automatic* and *comprehensive* assessment of the employees on their working ability and job competency [25].

The decision tree model is one of the most effective data mining techniques used in many sources: The decision tree algorithm in the Commercial Bank Recruitment testing database to examine the factors affecting performance and human resource development [26], analyzing both greedy and random decision trees, and the conflicts that arise when trying to balance privacy requirements with the accuracy of the model [27], the Kornel decision tree classification approach for model development for future sectors at the Retail Commercial Bank in the city of Tangayil, Bangladesh [28], C5.0 decision tree, Multi-Layer Understanding Neural Network (MLPNN), Naive Bayes expanding tree and logistic regression of the Portuguese bank data collection [29] have been used and their results have been analyzed.

In [29], the results show that the decision tree C5.0 has the best proportion of accuracy than other methods. Another work in this field and similar data sets indicate that MLPNN accuracy is better than Naive Bayes [30]. Decision tree is a useful tool for discovering data for predictions in the form of rules. The classification and regression tree (CART) algorithm is the most popular method used in HRM to decide on recruitment [31].

Other research is expected through studying the key factors in customer relationship management and the use of data mining in the bank. Bank customers classify using

decision tree algorithm. Three decision tree models, including ID3, C4.5 and CART, are used for categorization and prediction [32].

When an unbalanced data set exists, applied data mining techniques produce misleading results, since it provides for a complete negative result. An approach to overcome the incompatibility of the data set is to use hybrid data extraction solutions to provide more accurate results using more than one classification in the data set [33]. Collected data based on customer characteristics may be complex, especially when they are generally large and unbalanced information collections. So, if the techniques applied incorrectly, the cost of the campaign can be dramatically increased [34].

Author in [35] suggested the use of genetic algorithm for customer classification. In general, RFM represents the dynamic behavior of the customer and evaluates LTV models of value or customer participation. [36] offers the use of neural networks in the classification of customers. In this way, they aim to enhance customer satisfaction with the color and model of goods with the help of the intelligent system created using artificial neural networks.

The priori algorithm was used to improve the bank segmentation of customers [37]. Experimental results showed that this algorithm can effectively overcome the traditional algorithm, resulting in precision in customer segmentation, more reasonable results, effective decision-making and more benefits to the bank.

In another study, analysis of data mining techniques and its applications in the banking sector, such as the prevention and classification of customers, customer retention, marketing, and risk management, have been addressed [38].

Due to security concerns and lack of access to real financial information from banks, [39] applies credit scoring methods based on the payment dates for members of the club and shows the use of data mining to improve credit assessment using credit estimation models. Credit card model classification, logistic regression model and decision tree model were compared. Using fuzzy neural hybrid model, the credit ratings of customers were presented and used the ant colony algorithm to optimize the model. In the research, [40] ranked customers and identified their superior segments using the neural network, which resulted in the brokerage company being mechanized for the allocation of credits. The analysis of these results was intended to make decisions and strategies suitable for determining the facility to customers. [41] evaluated data for finding models based on customer specifications regarding their use of Mobile Bank by applying two techniques of artificial neural networks and simple bans. In this research, the results have led to customer attraction, current client maintenance, and increased customer satisfaction. Finally, it was determined that the artificial neural network technique has a higher accuracy than the simple binary technique.

Another study [42] intended to optimize particle parameters (PSO) to obtain appropriate parameters for the vector support vector (SVM) and decision tree (DT), and to select a subset of useful features, without reducing the degree of classification accuracy, to its work. To evaluate proposed approaches, data collected from Taiwanese commercial banks are used as sources of data. The experimental results showed that the proposed methods could achieve better parameterization, reduce unnecessary characteristics and significantly improve the classification accuracy.

Author in [43] examine various data mining techniques that can be used in banking areas. This provides an overview of the techniques and methods of data mining and sees how these techniques can be used in banking areas to make decision making easier and more productive.

The customer status of the bank was analyzed by using a model called CLV [44]. Their main goal in this study was to calculate real values or customers in order to discover and analyze their position in markets or the banking system. [45] analyzes the criteria and relationships for analyzing how customers use banks. In his research, he examined customer status in 2010-2014 and used the CPA model. Authors in [46] analyses data mining techniques in business especially in healthcare.

In the following sections, we will discuss and investigate the following: In the third section, the data sources used, the algorithm and flowchart of the proposed method, and the full description of the implementation steps of the method proposed in this study are examined. In the fourth section, the simulation results of the proposed method, which includes the initial clustering of data, customer classifications using neural network algorithms, decision trees, and so on are mentioned. The research findings have been compared with the results of other methods. In the fifth section, which is the final chapter, the final result and suggestions for developing current research are presented.

3 Modeling the Proposed Method

Data is collected from different sources and dispersed centrally in a collection area and a concentrated data warehouse. Ansar Bank's customers' transactions and information are used here to form a data bank. The data collected from the Ansar Bank was during the years 2011-2016 and the total number of records was 5,000.

A preprocessing takes place on data and eliminates data problems (cleaning, selecting a subset of features, sample filtering, sampling, data conversion, discretization, dimming, data accumulation, and feature creation). With preprocessing of the data, the volume of data is reduced, useful data is provided, and therefore suggestions can be made with better speed and accuracy. In this research, due to the nature of the data required, it eliminates some of the features so that less processing is needed to be done in the class operations and other operations, and at a more appropriate time for the discovery of credible customers. In this study, the variables are examined in three different characteristics: Names of Features, Data type, Variable type respectively.

Age*/ Numeric/ Input*
Education*/ Discipline/ Input*
Year emp*/ Numeric/ Input*
Income/ Numeric/ Input
Debt income/ Numeric/ Input
Other debt/ Numeric/ Input
Loan/ Boolean/ Target variable
Sex/ Boolean – Discipline/ Input
Married/ Boolean/ Input
Level/ Discipline/ Input

Balance/ Numeric/ Input
Day Count Fix Balance/ Numeric/ Input
Account Strat/ Numeric/ Input
Transaction Count/ Numeric/ Input
End Transaction Day/ Numeric/ Input
Count/ Numeric/ Input
Loan Price/ Numeric/ Input
Arrive Count Loan/ Numeric/ Input
Q Count/ Numeric/ Input
Q Count Pending/ Numeric/ Input
Q Avg Pay/ Numeric/ Input

Due to the fact that the data is extracted from a valid and standard source, there is little or no outlier between them. This means that some of the data has null values, which is excluded from the database.

Subsequently, the data are processed and converted into data mining tools such as Rapid Miner, Weka and SPSS.

In the proposed method, the most popular classification method is used to classify valid customers in an electronic banking system to provide facilities. Finally, the above methods are combined and at each stage, the best answer is selected from the answers given and is determined as the final result. Fig. 1 illustrates the flowchart and hybrid system architectures.

Considering the nature and sensitivity of customer classification environments, the proposed method is considered to be the combination of the neural network, the C4.5 decision tree and the k closest neighbor method. The combination of these two methods is presented in the form of a boosting system.

The routine of doing this is that at first financial information of all customers of the bank is clustered using the k-means algorithm. The purpose of this step is to initialize the data. By determining the initial data group, we can validate other new samples using classifier algorithms and predict the categories of customers. In fact, by clustering data, samples are made for applying classification algorithms.

The output of the clustering phase is separated and classified into two categories of training and experimental data for the classification of customers. 70% of the data is used as training data to train classification models and 30% of data is used to evaluate and determine the category and credibility of customers. The procedure for classifying training and the experiment data is in accordance with a balanced technique. Using a balanced technique, data can be extracted and separated from each group or batch in the data. Data split operations are done in the Rapid Miner data mining software using a balanced sampling.

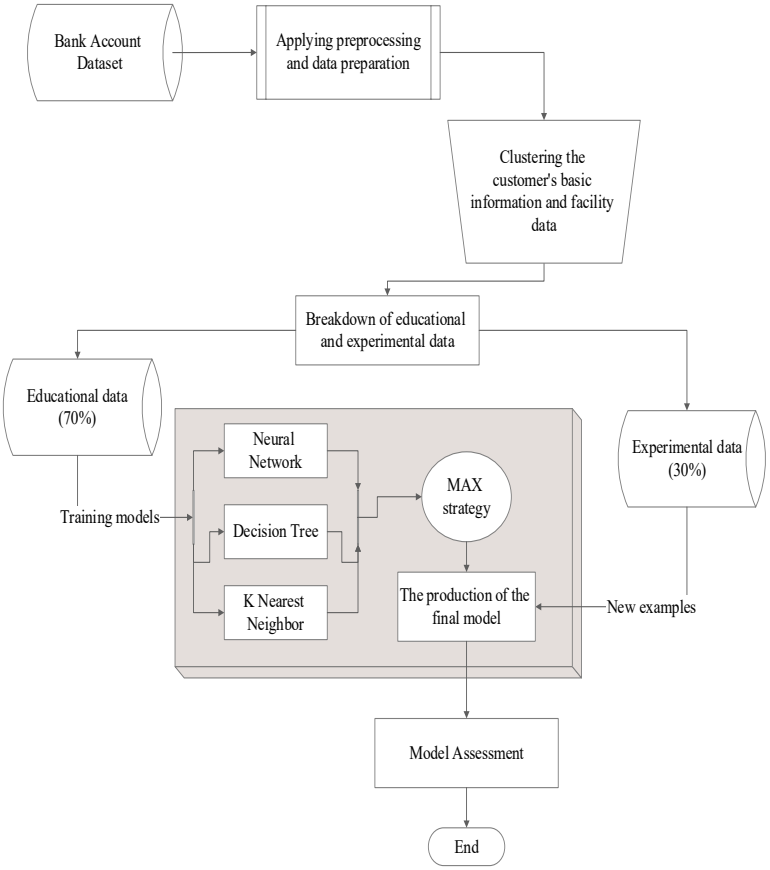

Fig. 1. Flowchart of the Proposed Method

Then, the training data is entered into the model of the neural network with the number of two hidden layers, the C4.5 decision tree and the KNN classification algorithm. The number of data or nodes for teaching 3500 model samples. The type of neural network algorithm used is the neural network standard algorithm.

All three methods are taught to teach their models and are ready to receive data to identify valid and invalid customers.

After the relevant models are taught, experimental data, which is 30% of the total data, is introduced to evaluate the models. Each model returns its output and its prediction and identification as output.

The output of these methods is connected to the core input of the boost system and according to the parameter in the boosting system, min, max and avg, the best prediction and classification is selected and used as output.

By providing the above-mentioned method in each run and predicting a new case, the best response to the output is sent, and finally, the optimal response will be high precision and the least error. Fig. 2 provides an overview of the hybrid boosting system.

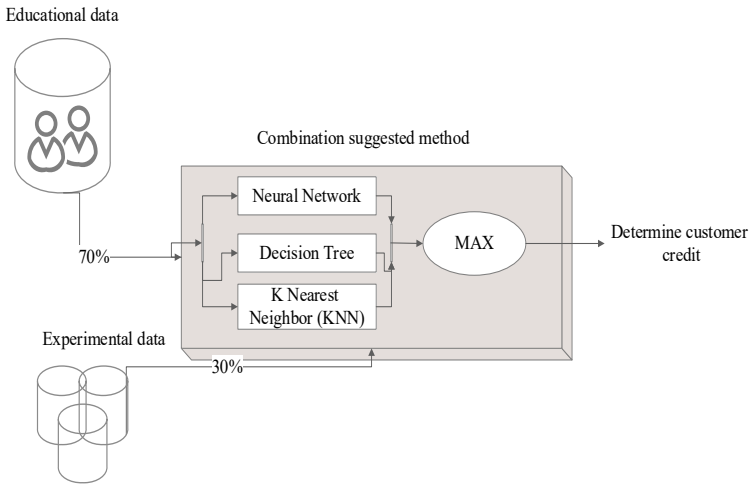

Fig. 2. Flowchart of Combined Boosting Algorithm from customers classification

So, according to the view presented in this section, we simulate the proposed method in the next section and evaluate the results with other methods.
In order to evaluate the classification methods used to classify customers, there are criteria that calculate the accuracy, tree accuracy, and accuracy of the methods. Among the most important of these criteria are: 1) accuracy, 2) precision, 3) recall, 4) error rate. In the following relationships, methods for calculating accuracy, precision, recall, and classification error rate are shown.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

In formula 1, True Positive (TP) represents instances where the values of their type are correct and are also correctly recognized. The True Negative (TN) represents the number of instances that are false and incorrectly detected. The False Positive (FP) also represents the number of instances that were false and correctly recognized. Finally, False Negative (FN) also shows samples that are false and precisely incorrect.

The formula for verifying the correctness and call is as follows.

$$\text{Pr}ecision = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Re}Call = \frac{TP}{TP + FN} \tag{3}$$

Finally, the error rate is calculated as follows:

$$ErrorRate = 100 - \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \tag{4}$$

4 Simulate the Proposed Method and Evaluate the Results

The proposed method was implemented using the Rapid Miner simulator. The tests were performed on a desktop with win 7 and Intel's Core 7, Core ™ i7 CPU, Q720 @ 1.60GHz 1.60 GHz processor with 3.06 gigabytes of RAM, and 32-bit operating system. One of the reasons for selecting the Rapid Miner data mining software to simulate the current research is the ability to perform pre-processing on the data, standard and acceptable sampling, modeling the algorithms used, and ultimately the ease of implementation of the algorithms.

4.1 Data clustering by using K-Means algorithm

In a simple kind of k-means clustering algorithm, first, the number of clusters needed for points is randomly selected. Then, in the data, they are attributed to one of these clusters according to the degree of similarity, and thus new clusters are obtained. By repeating the same procedure, it is possible to calculate new centers for each of them by averaging the data, and again attributing the data to new clusters. This process continues until there is no further change in the data. The following function is considered as the objective function, which is $\| \|$ the standard of distance between points, $x_i^{(j)}$ is the observation i of the cluster j and c_j is the center of the cluster j.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{5}$$

The algorithm below is the basic algorithm for this method:

1. At the beginning, the point K is selected as the points of the cluster centers.
2. Each data sample is attributed to the cluster whose center of the cluster has the smallest distance to that data.
3. After all data is added to one of the clusters for each cluster, a new point is calculated as the center. (Average points belonging to each cluster)
4. Steps 2 and 3 are repeated until there is no change in the cluster centers.

The routine of work is that the data are first normalized and then given as input to the k-means clustering algorithm. Given that nodes data does not have a specific category and group, it is necessary for the first reason to determine for each batch instance.

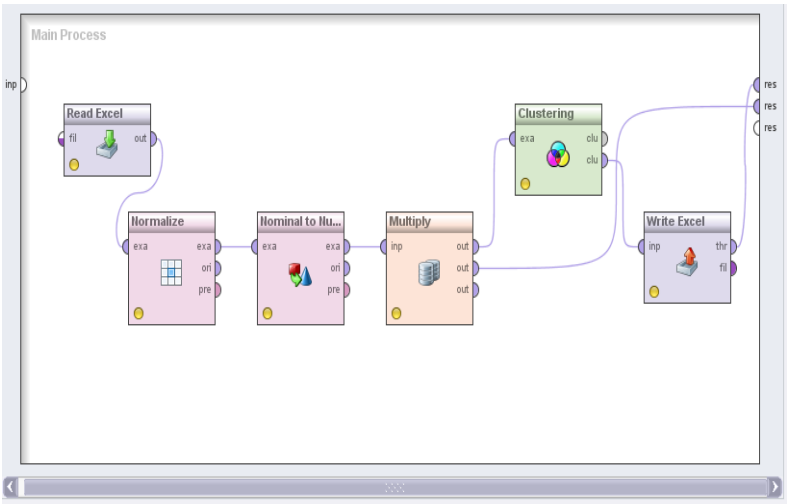

Fig. 3. Average squared error associated with selecting the dataset attributes of the edges

After clustering in fig. 3, the clustered sample outputs are used as the input of the kernel of the boosting algorithms of the neural network, the C4.5 decision tree and KNN.

4.2 Classification by using the neural network algorithm

After implementing the proposed method, the final accuracy of the neural network is 89.55%, its error rate is 10.45%, and the Root Mean Squared Error (RMSE) is 0.301 that is shown in fig. 4.

PerformanceVector

```
PerformanceVector:
accuracy: 89.55% +/- 5.05% (mikro: 89.70%)
ConfusionMatrix:
True:   0.0     1.0
0.0:    279     20
1.0:    18      52
classification_error: 10.45% +/- 5.05% (mikro: 10.30%)
ConfusionMatrix:
True:   0.0     1.0
0.0:    279     20
1.0:    18      52
root_mean_squared_error: 0.301 +/- 0.070 (mikro: 0.307 +/- 0.000)
```

Fig. 4. Accuracy, Error, and Average Squares of the neural network algorithm

4.3 Classification by using the C4.5 decision tree algorithm

One of the most important reasons for choosing the C4.5 algorithm is having a lower classification error rate than other decision tree algorithms such as Chaid, ID3, CART, and so on.

In fig. 5 and 6, the resulting C4.5 decision tree is shown. Finally, after implementing the proposed method, the final accuracy of the C4.5 decision tree is 90.03%, the error rate is 9.97%, and the RMSE is 0.298.

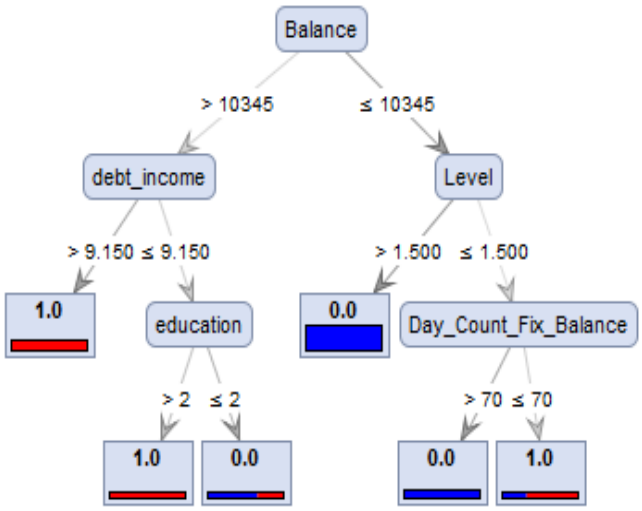

Fig. 5. Generated C4.5 Decision tree

PerformanceVector

```
PerformanceVector:
accuracy: 90.03% +/- 2.90% (mikro: 90.06%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     275     16
1.0:     20      51
classification_error: 9.97% +/- 2.90% (mikro: 9.94%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     275     16
1.0:     20      51
root_mean_squared_error: 0.298 +/- 0.048 (mikro: 0.302 +/- 0.000)
```

Fig. 6. Accuracy, Error, and Average Squares of the C4.5 decision tree

4.4 Classification by using the KNN algorithm

Finally, after implementing the proposed method, the final accuracy of KNN is 93.69%, the error rate is 6.31%, and the RMSE is 0.248 that is shown in fig. 7.

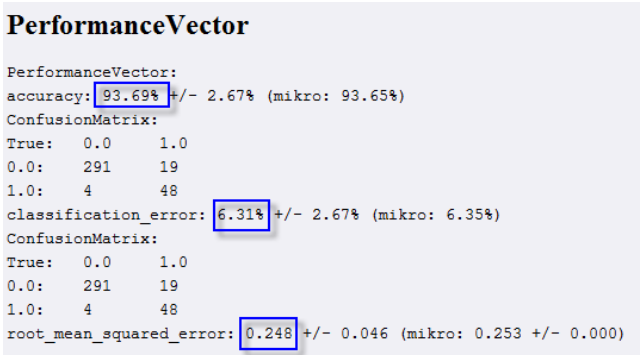

Fig. 7. Accuracy, Error, and average squared error of the KNN algorithm

4.5 Results of proposed algorithms (boosting)

After simulating each algorithm, each of them had errors at some point. In the diagram below, the proposed method is implemented on 22 samples of the data, and the results are examined using the boosting algorithm with the maximum strategy. Therefore, the total error is equal to the number of false classifications relative to the total sample.

Table 1. . Applying the boost algorithm with Max strategy

Row	Main Class	Output Decision Tree C4.5	Neural Network Output	KNN Output	Boosting (Max)
1	0	0	0	0	0
2	0	0	0	0	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
11	1	0	0	1	0 (Error)
12	1	1	1	1	1
13	1	1	1	1	1
14	0	0	0	0	0
15	0	0	0	0	0
16	1	1	1	1	1
17	1	1	0	1	1

18	1	1	1	1	1
19	1	1	1	1	1
20	1	1	1	1	1
21	0	0	0	0	0
22	1	1	1	1	1

As can be seen, in row 11, the type of main class field is 1. In this situation, the C4.5 decision tree algorithm and the neural network are error-free and the KNN algorithm classifies the sample correctly, and since it uses the MAX strategy, the result is selected that identifies its more algorithms and because the two algorithms of C4.5 decision tree and the neural network are identified by 0, so the final class is considered 0.

Finally, according to the data used, the accuracy of the boosting method is 95% and the error rate is 5%. In the next section, the results are fully evaluated.

4.6 Comparison of the proposed method with other methods

In this research, we compare the accuracy criterion, which is one of the most important criteria in determining the credibility of customers, with other methods. In the table below, the accuracy of the proposed method is shown with other popular methods.

Table 2. Comparison of the Accuracy Classification of the Proposed Method with Other Methods (Accuracy Criteria)

Proposed Method	KNN	Neural Network	Decision Tree	Naive Bayes
95%	93.69%	89.55%	90.03%	88.69%

As shown in the above table, the accuracy of the proposed method has improved on average by 1.04% compared to other methods. The average accuracy of classification of all methods, except for the proposed method, is 90.49%. The accuracy of the proposed boosting technique is 95%, which will be achieved by a simple fit of $\frac{95}{90.49} = 1.04$.

The error and RMSE criterions, which is one of the most important Factors in determining customer credibility, is compared with other methods. The table below shows the error rates and RMSE of the proposed method with other popular methods.

Table 3. Comparison of error rate and RMSE prediction of the proposed method with other methods (Error and RMSE Criterions)

	Proposed Method	KNN	Naive Bayes	Random Forest	Linear Regression
Error	5	6.31	10.45	9.97	11.31
RMSE	0.198	0.248	0.301	0.298	0.315

As shown in the above table, the total error of all methods expressed is 38.04%. Therefore, the average errors of the mentioned methods are 9.51%, which ultimately improves the error rate of the proposed method by 1.9 times compared to other methods on average. Finally, the proposed RMSE method improved by an average of 0.72% compared with other methods.

5 Conclusion

By providing the proposed method on Ansar Bank data, each time a new case was run and predicted, the best response to the output was sent, and finally, the optimal response to high accuracy and the least error was witnessed. Therefore, after testing and producing a decision tree, the following rules can be used to determine which group of clients can be offered loans and which customers cannot provide loans.

As it is seen from the rules generated by the proposed model, customers whose income level (account balance) is more than 103.450.000 Rials is more than customers whose income level is less than 103.450.000 Rials. If the customers' income level is more than 103.450.000 Rials, their debt is checked. If the customer's debt is less than 9.150.000 Rials, this customer is credible and can be assigned a loan or facility, and Otherwise, it is decided whether or not to be given a loan by checking the number of times a facility is being acquired. If the number of times a facility exceeds 2 times, that is, more than 2 times the loan has been settled, he will be assigned a loan and otherwise he will not be given a loan or facility. Now, if the customer has an income of less than 100 million Rials, he will only be lent in case his lifetime is longer than 2 months; otherwise, no such loans will be offered to such customers. Therefore, according to the proposed model, the features that have the greatest impact on the provision of loans and facilities to the customer are features such as the amount of income, the duration of the customer's current stay and the level of education.

According to the data used, the accuracy of the Boosting method is 95% and the error rate is 5%. The accuracy of the proposed method improved to 1.04% on average compared to other methods. The total errors of all methods expressed are 38.04%. Therefore, the average of the errors of the mentioned methods is 9.51%, which, finally, the error rate of the proposed method has improved on average by 1.9% compared to other methods and the proposed RMSE method improved by an average of 0.72% compared with other methods.

Some research suggestions for developing and presenting other ideas for current research include:

- Using other algorithms such as SVM, Random Forest and other algorithms to classify and compare with previous methods.
- Using Bagging Techniques instead of Boosting Techniques and comparing the results with the findings in this research.
- Using hierarchical clustering and improved k-means algorithm known as X-Means instead of the clustering algorithm used in this study and comparing the results with the findings in this research.
- Use of feature selection techniques such as PSO, Genetic Algorithm and Bee to select optimal features and use its results in customer ratings with credit and vice versa.
- Use of the proposed method in other applications such as financial institutions, public and private agencies and other organizations and related organizations.

6 References

[1] Bhapkar, M. Y. V., & More, A. D. (2014). Credit Risk Analysis of Bank Customers Using Data Mining Techniques. International Journal of Multifaceted and Multilingual Studies, 1(1).

[2] Kim, J.-B., Song, B. Y., & Zhang, Y. (2015). Earnings performance of major customers and bank loan contracting with suppliers. Journal of Banking & Finance, 59, 384-398. https://doi.org/10.1016/j.jbankfin.2015.06.020

[3] López-Yáñez, I., Sheremetov, L., & Camacho-Nieto, O. (2013). Multivariate Prediction Based on the Gamma Classifier: A Data Mining Application to Petroleum Engineering. Paper presented at the International Conference on Database and Expert Systems Applications. https://doi.org/10.1007/978-3-642-40173-2_3

[4] Horri, M. S., & Mahdavi, K. (2015). Designing a Model for Credit Rating Estimation of Banks Customers Using Fuzzy and Hybrid Multi-criteria Hybrid Fuzzy-Colony Networks (A Case Study of Post Bank Branches in Tehran Province) [Persian]. Management Studies in Iran, 19(1), 91-116.

[5] Azar, A., Ahmadi, P., & SEBT, M. V. (2010). Designing a Human Resource Selection Model with a Data Mining Approach (Case: Recruitment of Entrants' Tests for a Commercial Bank in Iran) [Persian]. IT management, 4(2), 3-22.

[6] Almana, A. M., Aksoy, M. S., & Alzahrani, R. (2014). A survey on data mining techniques in customer churn analysis for telecom industry. Journal of Engineering Research and Applications, 4(5), 165-171.

[7] Balaji, S., & Srivatsa, S. (2012). Customer segmentation for decision support using clustering and association rule-based approaches. International Journal of Computer Science & Engineering Technology, 3(11), 525-529.

[8] Mitik, M., Korkmaz, O., Karagoz, P., Toroslu, I. H., & Yucel, F. (2017). Data Mining Approach for Direct Marketing of Banking Products with Profit/Cost Analysis. The Review of Socionetwork Strategies, 1-15. https://doi.org/10.1007/s12626-017-0002-5

[9] Karimi-Majd, A.-M., & Mahootchi, M. (2015). A new data mining methodology for gene lprating new service ideas. Information Systems and e-Business Management, 13(3), 421-443. https://doi.org/10.1007/s10257-014-0267-y

[10] 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(2), 2592-2602. https://doi.org/10.1016/j.eswa.2008.02.021

[11] Chalmeta, R. (2006). Methodology for customer relationship management. Journal of systems and software, 79(7), 1015-1024.

[12] Miguéis, V. L., Camanho, A. S., & e Cunha, J. F. (2012). Customer data mining for life style segmentation. Expert systems with applications, 39(10), 9359-9366. https://doi.org/10.1016/j.eswa.2012.02.133

[13] Golmah, V., & Mirhashemi, G. (2012). Implementing A Data Mining Solution to Customer Segmentation For Decayable Products–A Case Study For A Textile Firm. International Journal of Database Theory and Application, 5(3), 73-90.

[14] Van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2020). Predicting customer demand for remanufactured products: A data-mining approach. European Journal of Operational Research, 281(3), 543-558. https://doi.org/10.1016/j.ejor.2019.08.015

[15] Danaee, H., Aghaee, Z., Haghtalab, H., & Salimi, M. P. (2013). Classifying and Designing Customer's Strategy Pyramid by Customer Life Time Value (CLV) (Case study: Shargh Cement Company). J. Basec Appl. Sci. Res, 3(7), 473-483.

[16] Zephaniah, C. O., Ogba, I.-E., & Izogo, E. E. (2020). Examining the effect of customers' perception of bank marketing communication on customer loyalty. Scientific African, e00383.https://doi.org/10.1016/j.sciaf.2020.e00383

[17] Anshari, M., Al-Mudimigh, A., & Aksoy, M. (2009). CRM initiatives of banking sector in Saudi Arabia. Int J Comput Internet Manag, 19(SP1), 56.51-56.57.
[18] Yong Wang, D. S. W. (2011). Research of the Bank's CRM Based on Data Mining. School of Economics and Business Administration Chongqing University Chongqing China, CISME, 30-35.
[19] Li, D.-C., Dai, W.-L., & Tseng, W.-T. (2011). A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business. Expert systems with applications, 38(6), 7186-7191. https://doi.org/10.1016/j.eswa.2010.12.041
[20] Raju, P. S., Bai, D. V. R., & Chaitanya, G. K. (2014). Data mining: techniques for enhancing customer relationship management in banking and retail industries. International Journal of Innovative Research in Computer and Communication Engineering, 2(1), 2650-2657.
[21] Thelen, S., Mottner, S., & Berman, B. (2004). Data mining: On the trail to marketing gold. Business Horizons, 47(6), 25-32. https://doi.org/10.1016/j.bushor.2004.09.005
[22] Herrera-Restrepo, O., Triantis, K., Seaver, W. L., Paradi, J. C., & Zhu, H. (2016). Bank branch operational performance: A robust multivariate and clustering approach. Expert systems with applications, 50, 107-119. https://doi.org/10.1016/j.eswa.2015.12.025
[23] Wang, Y., Ma, X., Lao, Y., & Wang, Y. (2014). A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization. Expert systems with applications, 41(2), 521-534. https://doi.org/10.1016/j.eswa.2013.07.078
[24] Ng, K.-H., & Khor, K.-C. (2017). StockProF: a stock profiling framework using data mining approaches. Information Systems and e-Business Management, 15(1), 139-158. https://doi.org/10.1007/s10257-016-0313-z
[25] Quan, P., Liu, Y., Zhang, T., Wen, Y., Wu, K., He, H., & Shi, Y. (2018). A Novel Data Mining Approach Towards Human Resource Performance Appraisal. Paper presented at the International Conference on Computational Science.
[26] Sebt, M. V., & Yousefi, H. (2015). Comparing data mining approach and regression method in determining factors affecting the selection of human resources. Cumhuriyet Science Journal, 36(4), 1846-1859.
[27] Fletcher, S., & Islam, M. Z. (2019). Decision tree classification with differential privacy: A survey. ACM Computing Surveys (CSUR), 52(4), 1-33. https://doi.org/10.1145/3337064
[28] Islam, M., & Habib, M. (2015). A data mining approach to predict prospective business sectors for lending in retail banking using decision tree. arXiv preprint arXiv:1504.02018.
[29] Elsalamony, H. A. (2014). Bank direct marketing analysis of data mining techniques. International Journal of Computer Applications, 85(7).
[30] Bahari, T. F., & Elayidom, M. S. (2015). An efficient CRM-data mining framework for the prediction of customer behaviour. Procedia Computer Science, 46, 725-731. https://doi.org/10.1016/j.procs.2015.02.136
[31] Azar, A., Sebt, M. V., Ahmadi, P., & Rajaeian, A. (2013). A model for personnel selection with a data mining approach: A case study in a commercial bank. SA Journal of Human Resource Management, 11(1), 1-10. https://doi.org/10.4102/sajhrm.v11i1.449
[32] Farid, D., Sadeghi, H., Hajigol, E., & Parirooy, N. Z. (2016). Classification of Bank Customers by Data Mining: a Case Study of Mellat Bank branches in Shiraz. International Journal of Management, Accounting & Economics, 3(8), 534-543.
[33] Pan, Y., & Tang, Z. (2014). Ensemble methods in bank direct marketing. Paper presented at the Service Systems and Service Management (ICSSSM), 2014 11th International Conference on. https://doi.org/10.1109/icsssm.2014.6874056
[34] Ling, C. X., & Li, C. (1998). Data mining for direct marketing: Problems and solutions. Paper presented at the KDD.

[35] Chan, C. C. H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. Expert systems with applications, 34(4), 2754-2762. https://doi.org/10.1016/j.eswa.2007.05.043
[36] Isakki, S. (2011). The Expert System Designed to Improve Customer Satisfication. Adv. Comput. An Int. J. (ACIJ), 2(6), 69-84. https://doi.org/10.5121/acij.2011.2607
[37] Yang, G. X. (2013). The research of improved Apriori mining algorithm in bank customer segmentation. Paper presented at the Advanced Materials Research. https://doi.org/10.4028/www.scientific.net/amr.760-762.2244
[38] Chitra, K., & Subashini, B. (2013). Data mining techniques and its applications in banking sector. International Journal of Emerging Technology and Advanced Engineering, 3(8), 219-226.
[39] Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. Expert systems with applications, 38(10),13274-13283. https://doi.org/10.1016/j.eswa.2011.04.147
[40] Hooshdar Mahjoob, A., & Minaei Bidgoli, R. (2013). Customer credit clustering to provide convenient facilities [Persian]. Management Studies in Iran, 17(4), 1-24.
[41] Hasanzade, A., Ghanbari, M. H., & Elahi, S. (2012). Mobile Bank Users Categorization Using Data Mining Approach: Comparison between Artificial Neural Networks Technique and Simple Bayes Technique [Persian]. Management Studies in Iran, 16(2), 57-71.
[42] Lin, S.-W., Shiue, Y.-R., Chen, S.-C., & Cheng, H.-M. (2009). Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks. Expert systems with applications, 36(9), 11543-11551. https://doi.org/10.1016/j.eswa.2009.03.029
[43] Pulakkazhy, S., & Balan, R. (2013). Data mining in banking and its applications-a review.
[44] Kahreh, M. S., Tive, M., Babania, A., & Hesan, M. (2014). Analyzing the applications of customer lifetime value (CLV) based on benefit segmentation for the banking sector. Procedia-Social and Behavioral Sciences, 109, 590-594. https://doi.org/10.1016/j.sbspro.2013.12.511
[45] Čermák, P. (2015). Customer profitability analysis and customer life time value models: Portfolio analysis. Procedia Economics and Finance, 25, 14-25. https://doi.org/10.1016/s2212-5671(15)00708-x
[46] Saeed, S., Shaikh, A., Memon, M. A., & Naqvi, S. M. R. (2018). Impact of Data Mining Techniques to Analyze Health Care Data. Journal of Medical Imaging and Health Informatics, 8(4), 682-690. https://doi.org/10.1166/jmihi.2018.2385

7 Authors

Mohammad Vahid Sebt is working in Kharazmi University, Tehran, Iran, Email: sebt@khu.acir.

Elahe Komijani is working in Islamic Azad University (South Tehran Branch), Tehran, Iran.

Article submitted 2020-06-07. Resubmitted 2020-07-03. Final acceptance 2020-07-04. Final version published as submitted by the authors.