Int. J. Nonlinear Anal. Appl.
Volume 11, Special Issue, Winter and Spring 2020, 299-309
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2020.4604

Evaluating the Effectiveness of Data Mining Techniques in Credit Scoring of Bank Customers Using Mathematical Models: A Case Study of Individual Borrowers of Refah Kargaran Bank in Zanjan Province, Iran

Abdollah Nazari[a*], Mohammadreza Mehregan[b], Reza Tehrani[b]

[a] *PhD Student of Industrial Management, Alborz Faculty, University of Tehran, Tehran, Iran.*
[b] *Professor, Faculty of Management, University of Tehran, Tehran, Iran.*

Abstract

Loan deferment is a negative consequence of the activities of financial institutions. Increase in the amount of deferred loans can diminish productivity in the banking sector. The purpose of the present research is to cluster bank customers in order to prevent loan deferment and identify and classify customers with varying levels of loan repayment risk. In the proposed method, k-means, two-step, and Kohonen techniques are used for clustering and determining the behavior of each cluster. The results indicate that the k-means model with five clusters has the highest clustering accuracy. Clustering is also used to determine underlying feauture. loan term, loan value, and collateral value are respectively identified as the most influential feauture . Customers are clustered after removing non-significant feautures. Eight different machine learning techniques are used for clustering. These techniques are ranked in terms of efficiency based on certain evaluation criteria and using data envelopment analysis. The results indicated that support vector machine (SVM) and artificial neural networks (ANN) are the most efficient of the examined techniques.

Keywords: Credit Scoring, Clustering, Data Mining, Data Envelopment Analysis (DEA).

*Corresponding Author: Abdollah Nazari
Email address: abd.nazari1353@ut.ac.ir (Abdollah Nazari[a*], Mohammadreza Mehregan[b], Reza Tehrani[b])

Received: February 2020 *Revised:* August 2020

1. Introduction

Debt collection makes up a large part of banks' required funds. Failure to collect debts results in a massive loss of assets and equity. Therefore, banks try a variety of methods to better assess credit applicants in order to reduce credit risk and non-repayment risk. Today, expert estimates and forecasts of applicants' creditworthiness and the future of their economic activities form the basis for lending decisions. One of the main problems with this approach is the prolonged lending process, which selects borrowers with the ability to repay and assesses the necessary collaterals. This increases the cost of lending both for the recipient of the loan and for the bank itself. In addition, assessor's or bank's preferences will have a significant effect on this process, which, along with the multiplicity of approaches instead of a single scientific approach, can cause dissatisfaction in customers and create a breeding ground for corruption. Using an approach that provides a better understanding of the feautures underlying non-repayment of loans can help in making a more accurate risk assessment of new credit applications, which can reduce the duration of the loan approval process and lead to lower collateral pledge by creditworthy borrowers, thus opening up lending opportunities. The present research tries to use appropriate criteria and assess different models of customer credit scoring to propose a model that can properly rank credit applicants and reduce the credit risk of bank customers.

Debt collection by banks is one of the feautures that affect their productivity due to its role in reducing costs and increasing profits. Through debt collection, banks can increase their returns and productivity and ensure their long-term profitability. In addition, debt collection is the least costly method of financing and can increase banks' lending ability. Optimal allocation of banking resources and more lending by banks can have a significant effect on the national economy.

Internal feautures affecting the efficiency and productivity of the customer credit scoring system are more effective than external feautures since they can be modified and controlled by the bank; therefore, these feautures must be the central focus of banks in their planning. Achieving continuous improvement in banking productivity and processes requires the development of new methods and solutions for credit scoring that are tailored to the existing economic conditions and offer the best results with the highest accuracy and at the lowest cost, which will have a significant role in debt collection and improve the efficiency of banks.

The main innovations in the present research are the selection of key features in customer classification using clustering techniques, the ranking of machine learning models using DEA technique and certain criteria such as confusion matrix, precision, accuracy, ROC curve, sensitivity, specificity, recall, and the Gini index.

Given these discussions, the primary question of the present research is:

- What is the most appropriate model for scoring the credit of individual bank customers? This research also tries to answer a number of secondary questions:
- What are the key indicators for credit scoring individual bank customers?
- What are the weights of these indicators?
- How accurate is the proposed model?

2. Theoretical Framework

2.1. Data Envelopment Analysis

Data envelopment analysis (DEA) is a non-parametric approach for evaluating the efficiency of individual units within an organization. This technique is based on linear programming and its goal is to evaluate and compare the efficiency of decision-making units (DMUs) with multiple inputs and outputs. The idea is to find how well a DMU has used its resources for production.

CCR and BCC are the most common models for calculating the efficiency of similar units in DEA. Despite its high discrimination power, CCR has drawbacks due to its assumption of a constant returns to scale. In BCC, the problem of a constant returns to scale is solved, but the number of relatively efficient units increases to an unacceptable level. Therefore, the Andersen-Petersen model is used by adding an ideal and an anti-ideal unit to DMUs (machine learning techniques).

The Andersen-Petersen model or the super-efficiency model, which allows for determining the most efficient DMU, was proposed in 1993 by Per Andersen and Niels Christian Petersen as a procedure for ranking efficient units in DEA. In this approach, efficient units can have a score above one, which allows for having an efficiency rating of efficient units similar to the rating of inefficient units. The procedure involves excluding the unit under investigation from the reference set and implementing the DEA model for all the other DMUs.

$$
\begin{aligned}
&\min \ Z = \theta \\
&S.t: \\
&\sum_{j=1}^{n} \lambda_j x_{ij} \geq x_{i0} \qquad (i = 1, 2, \cdots, m) \\
&\sum_{j=1}^{n} \lambda_j x_{rj} \geq \theta y_{r0} \quad (r = 1, 2, \cdots, s) \\
&\sum_{j=1}^{n} \lambda_j x_{ij} = 1 \qquad (j = 1, 2, \cdots, n) \\
&\lambda_j \geq 0 \quad \theta \text{ is free} \quad (j = 1, 2, \cdots, n)
\end{aligned}
\tag{2.1}
$$

By removing the constraint related to the unit under investigation (an upper limit of one), efficiency in this model can be greater than one and thus efficient units with scores higher than one can be ranked and compared.

2.2. Literature Review

Hu and Ansell [8] compare Naïve Bayes, Logistic Regression, Recursive Partitioning, Artificial Neural Network, and Sequential Minimal Optimization (SMO) to identify the most effective credit scoring technique for measuring retail company performance. Min and Lee [12] used a DEA-based approach for credit scoring in a sample of Korean manufacturing firms. In another study, Xu et al. (2009) used a credit scoring algorithm based on linkage analysis with support vector machine (SVM). Setiono et al. [19] used a genetic algorithm-based neural network algorithm for credit card screening. In addition, Yeh and Lin [23] compared data mining techniques such as logistic regression, nearest neighbor, artificial neural networks, and classification trees.

Zhou et al. [24] used direct search to select the parameters of a support vector machine algorithm. Ping and Yongheng [18] use a neighborhood rough set and SVM-based classifier for credit scoring. In another study, Kao et al. [10] used a Bayesian model with classification and regression tree for behavior and credit scoring. Vukovic et al. [21] proposed a case-based reasoning (CBR) model for credit scoring that employed preference theory functions. Danenas and Garsva [3] used particle swarm optimization to select an optimal linear SVM classifier for credit risk evaluation. Moreover, ensemble credit scoring models have been used in numerous studies.

Tsai and Wu (2008) used a multilayer perceptron (MLP) network for bankruptcy prediction and credit scoring. Nanni and Lumini [16] applied a number of techniques such as rotation forest, random subspace, and class switching to a credit scoring problem. Twala [20] used artificial neural network, decision trees, k-nearest neighbor, and logistic discrimination for credit risk assessment. Hsieh and Hung [7] used an ensemble of classifiers such as SVM, artificial neural network, and Bayesian network to evaluate loan applicant's creditworthiness. Also, Wang and Ma [22] used the ensemble learning approach with SVM as the base learner for enterprise credit risk assessment.

Mousavi and Gholipour [15] ranked credit scoring criteria using the Delphi method. The goal was to identify underlying feautures and prioritize credit scoring criteria for credit risk assessment in corporate bank customers. Data were collected from lending experts and officials from selected banks using a questionnaire. The results supported economic theories about credit risk, indicating that feautures underlying credit risk do not have an equal weight and some feautures play a more significant role in credit risk assessment.

Alborzi et al. [1] used genetic algorithm in decision tree optimization for credit scoring bank customers. The proposed classification model, which had the least number of leaves, the smallest size, and the lowest complexity, was more accurate than all the other decision trees compared in this paper.

Armashi [2] used logistic regression to identify the feautures that affect the credit risk of individual customers of Bank Saman in Northern Iran. The relationship between certain financial and demographic feautures and customers' credit risk was examined. Data were extracted from customer records of different branches of Saman Bank. The results showed that gender, income, type of residence, marital status, age, and job had a significant effect on loan non-repayment, while no significant effect was observed for loan size and loan term.

Kamali [9] used logistic regression, probit regression, and neural networks for credit risk assessment in a case study of the customers of Sina Bank in Iran. The qualitative and financial data of a sample of 349 customers who received loans from Sina Bank between 2007 and 2009 were collected and examined. 19 explanatory feautures were identified by reviewing the credit files of the applicants. Using this approach, eight feautures that had a significant effect on credit risk and distinguishing creditworthy from non-creditworthy customers were identified.

3. Methods

The present research is an applied, developmental research with a non-experimental descriptive design. The proposed method includes the following three phases:

- **Phase I-clustering customers to find key criteria:** At this stage, 18 criteria were identified for bank customers using the conceptual model of Doumpos [4]. These criteria were localized based on expert views. Based on initial clustering results, the number of criteria dropped to eight. Several clustering techniques were used to classify customers, including classical k-means, Kohonen network, and hierarchical clustering with different parameters.

- **Phase II-application of machine learning techniques to classify customers:** At this stage, machine learning techniques such as logistic regression, decision trees, neural networks, Bayesian networks, and SVM were used to classify bank customers.

- **Phase III-ranking these techniques using data envelopment analysis:** Finally, DEA is used to evaluate and rank customer classification techniques.

3.1. Data Collection

The population of this research consists of all the customers of Refah Bank in Zanjan Province, Iran. Stratified random sampling was used to select the customers. In this sampling technique, first the population is divided into homogenous groups called strata and then random samples are taken from each stratum in a number proportional to the stratum's size when compared to the population. The following formula is used for sampling:

$$N = \frac{Nt^2pq}{Nd^2 + t^2pq} \tag{3.1}$$

where N is the population size, t is Student's t, d is the degree of freedom, p is the probability of a success, and q is the probability of a failure. A sample size of 1000 customers was obtained at the 95% confidence interval (CI) and the 6% error rate. The database of customers included 18 input fields, including the applicant's age, gender, marital status, education and job, loan term, loan size, collateral type, collateral value, average six-month checking account balance, experience in the current job, credit history, credit rating, nominal capacity matching debt obligations, ownership of the workplace, and reputation/public image (Table 1).

Table 1: Model's input fields [4]

Credit Capacity	Financial	Guarantee	Asset - Debt	Type of collateral the customer can provide
				Nominal capacity matching credit obligations
				Average balance for the past six months
				Direct obligations to the bank
			Net Assets	Total value of movable and immovable assets
		Cash Flow	Cost - Income	Average monthly income
	Personal	Ability	Mental Ability	Age
				Gender
				Reputation and public image
				Activity (main job)
			Motivation	Field of study
				Credit score
				Repayment quality
				How obligations to the bank are fulfilled
				Number of deferred days during the total repayment period
				Number of returned checks that have been restituted (due to insufficient balance)
				Amount of deferred debt
		Business Behavior	Economic Mindset	Ownership of the workplace
				Residential status (homeowner or renting)
				Type of contract (leasing, partnership, unilateral contract)
				Facility amount (in a million rials)
				Interest rate
				Loan repayment term
				Number of installments
			Compliance with Social and Economic Standards	Duration of activity in the bank (i.e., contact with the bank)
				Work experience in the current job
				The relationship between occupation and education
				Recognition based on the duration of activity

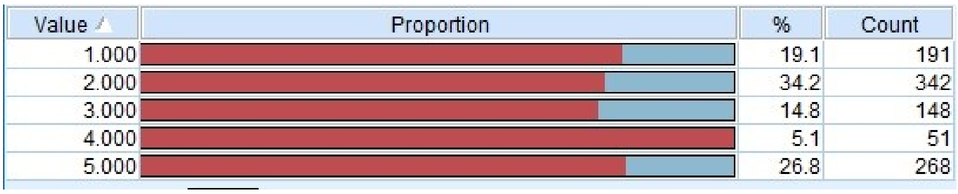
The output field of this database indicates whether or not the loan applicant is creditworthy. Since the inputs of machine learning techniques must be coded in numbers, first textual values are converted to numerical values based on their frequency.

4. Evaluation and Results

Data analysis is conducted in SPSS Modeler 18.0 and MATLAB.

4.1. Data Preprocessing and Identification of Outliers

An overview of the data is obtained using the Data Audit tool in SPSS Modeler and the distribution of creditworthy and non-creditworthy customers in all the fields is calculated using the Distribution operator. For instance, as shown in Figure 1, significant imbalance can be observed for the occupation field, making it a key player in the clustering and machine learning techniques.



Value	Proportion	%	Count
1.000		19.1	191
2.000		34.2	342
3.000		14.8	148
4.000		5.1	51
5.000		26.8	268

Figure 1: Distribution of credit capacity of one of the indicators.

The Anomaly Index (AI) diagram is illustrated in Figure 7 for detecting outliers. In this figure, 10 values are detected as outliers. Data with an AI greater than 1.5 are considered outliers. Examining outliers indicates that all of these data are related to non-creditworthy customers. Given that only non-significant data must to be removed at this stage, detected outliers are not removed so that model precision does not increase at the cost of lower accuracy.

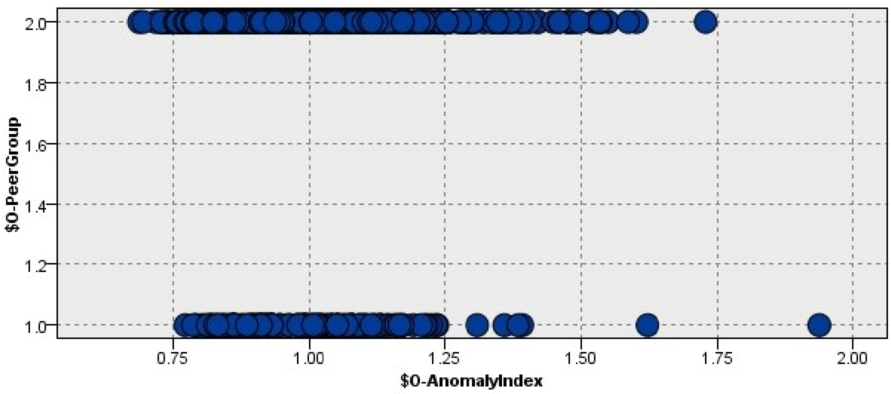


Figure 2: Outlier detection by the software.

4.2. Customer Clustering

Using the Auto Cluster operator in SPSS Modeler, 100 models are clustered with a number of clustering techniques, including k-means, Kohonen, and two-step models. At this stage, 18 feautures are considered for each customer. The objective feauture is the creditworthiness of the customer. These models are then evaluated using the profile index and the best scenario was identified based on the conceptual model and the profile index. The results showed that k-means with 5 clusters was the best model.

Figure 3 shows the profile index for the superior model, which is obtained using 18 feautures before removal of non-significant ones, while Figure 4 provides the weight of each characteristic. As shown in Figure 3, clustering quality is poor and requires the removal of certain non-significant feautures in order to improve model accuracy.

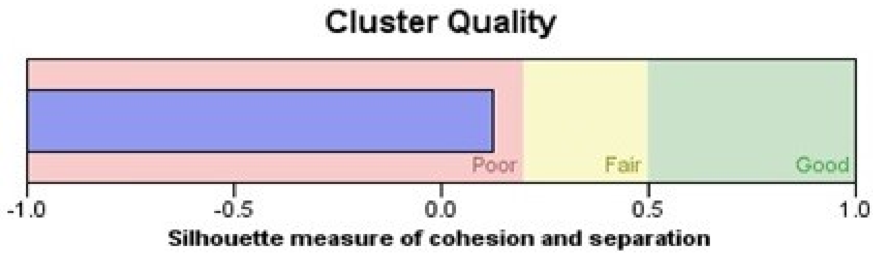


Figure 3: Evaluation of clustering quality using k-means.

The diagram below indicates the weight of the indicators. As shown, the first six feautures have the highest weight (Figure 4).

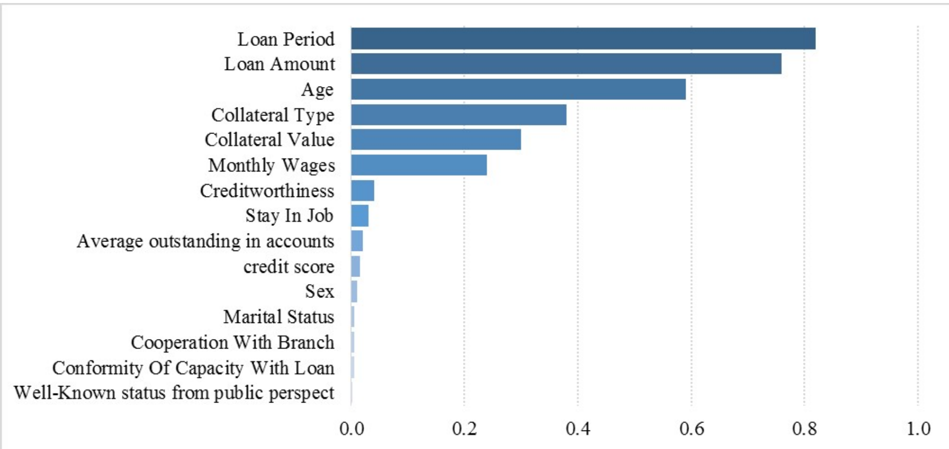


Figure 4: Weight of indicators in clustering with k-means.

To increase clustering accuracy, we try to work with a limited number of feautures that exhibit the highest variation or effect. The profile diagram for the superior model is illustrated in Figure 5 after removing non-significant indicators, reducing them to a total of eight feautures. In addition, the weight of each feautures is shown in Figure 6.

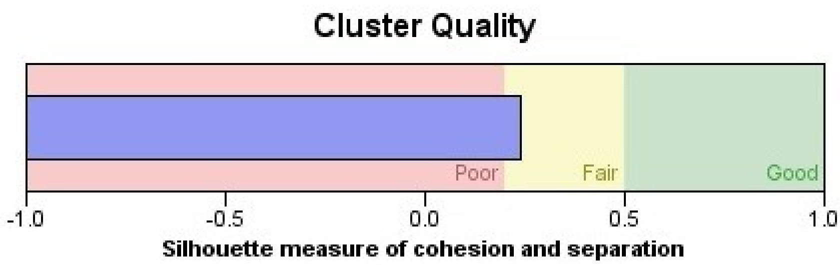


Figure 5: Clustering quality evaluation after removal of non-significant indicators.

As shown in Figure 6, loan size, collateral value, loan term, monthly income, age, collateral type, work experience, and initial credit score have the highest weight.

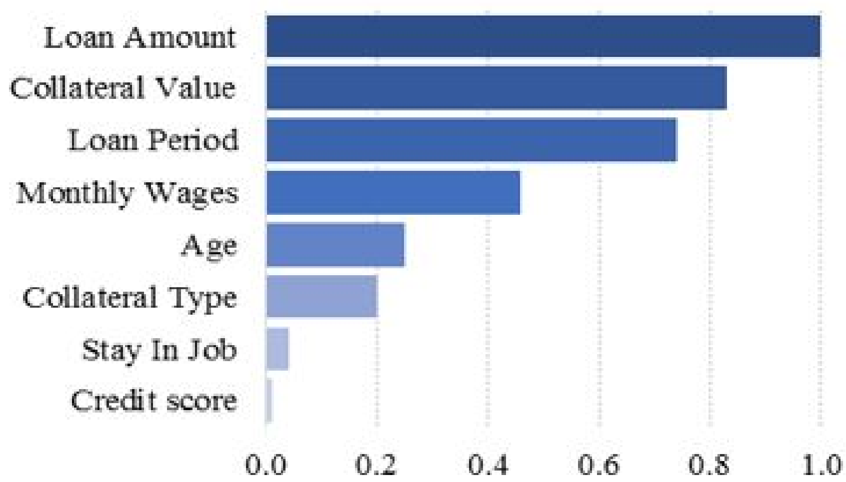


Figure 6: Weight of indicators after removal of non-significant indicators.

4.3. Customer Classification

Eight machine learning techniques are used to classify bank customers. These include artificial neural networks (ANN), support vector machine (SVM), logistic regression, Bayesian networks, k-nearest neighbors, decision trees, C5.0, and classification and regression tree (CART). Three types of neural networks are used. The first one, ANN1, is a single-layer perceptron (SLP), the second neural network, ANN2, is a three-layer perceptron, and the third neural network, ANN3, is a five-layer perceptron.

A number of criteria are used to evaluated the applied machine learning techniques. These include precision, accuracy, sensitivity, recall, Gini index, area under the ROC curve, and the F-measure. These criteria are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{4.1}$$

$$Recall = TPR = \frac{TP}{TP + FN} \tag{4.2}$$

$$F\_Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4.3}$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4.4}$$

TN (true negative) denotes the number of real negative cases in the data, which are correctly identified by the classifer as negative. TP (true positive) denotes the number of real positive cases in the data, which are correctly identified by the classifer as positive. FP denotes the number of real negative cases that are incorrectly identified as positive by the classifer. FN denotes the number of real positive cases that are incorrectly identified as negative by the classifer. Table 2 provides the results of evaluating machine learning techniques using the above-mentioned criteria.

Table 2: Comparison of machine learning techniques based on performance measures

	+	+	+	+	+	-	-
	Recall=TPR	Precision	Specificity	F-Measure	AUC	Gini	Accuracy Different
C5	0.907	0.903	0.641	0.905	0.679	0.136	0.036
ANN 1	0.867	0.912	0.743	0.889	0.717	0.082	0.255
B N	0.950	0.894	0.684	0.921	0.777	0.123	0.056
SVM	0.934	0.914	0.722	0.924	0.808	0.071	0.059
ANN 3	0.941	0.961	0.855	0.951	0.769	0.088	0.031
CART	0.935	0.874	0.655	0.904	0.829	0.114	0.062
ANN 2	0.948	0.987	0.956	0.967	0.889	0.083	0.079
R L	0.941	0.905	0.718	0.923	0.617	0.143	0.056
KNN	0.868	0.944	0.815	0.905	0.853	0.129	0.154

The ROC curve is a method of illustrating the organization and selection of classifiers based on their efficiency. These curves are two-dimensional, where TPR is detection rate of true positive cases plotted on the Y axis and FPR is the detection rate of true negative cases plotted on the X axis. The ROC curve illustrates the relative trade-off between benefits and costs. In the present research, machine learning techniques are used for performance evaluation and the ROC curve is used to show the effectiveness of these techniques for training and testing data. The resulting plot is shown in Figure 7.

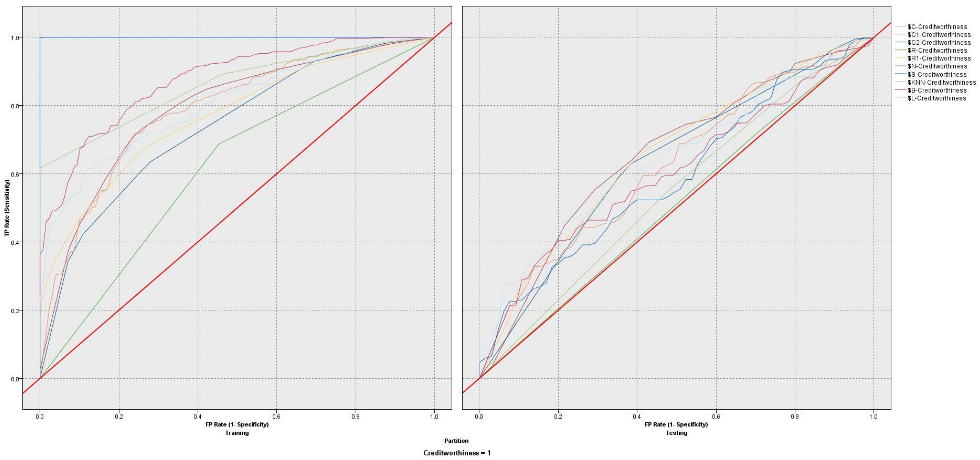


Figure 7: The ROC curve for all the techniques used in the model.

Various points in the ROC space are noteworthy. The bottom left point (0,0) represents a classifier that predicts all cases to be negative. It does not raise false alarms and does not provide a correct detection. The opposite point on the top right corner (1,1) predicts every case to be positive and raises a false alarm on every negative case. Point (0,1) is the perfect classifier, predicting all positive and negative cases correctly. Point (1,0) is the classifier that is incorrect for all classifications. Any point in the ROC space is better than every other point that is in the southeast. If located in the northwest of this space, the ROC curve allows for a visual comparison of a set of techniques.

4.4. Comparison and Ranking of Classification Techniques

The Andersen-Petersen Model, one of the most efficient DEA models, is used to rank the machine learning techniques applied in this study. This ranking is based on various criteria such as precision, accuracy, ROC curve, sensitivity, specificity, recall, and Gini index. A comparison of models is provided in Table 3, indicating that the SVM-ANN3 model is the best one.

Table 3: Ranking of machine learning techniques using Andersen-Petersen DEA Model

Technique	Efficiency	RANK
C5	0.8683	3
ANN1	0.8659	4
BN	0.5772	7
SVM	1	1
ANN3	1	1
CART	0.6226	6
ANN2	0.8554	5
RL	0.5566	8
KNN	0.5504	9
C5	0.8683	3

5. Findings and Recommendations

5.1. Conclusion

The general purpose of the present research was to provide a model for credit scoring individual customers of Refah Bank in Iran. This was done in three phases. In the first phase, 18 feautures were identified for assessing the creditworthiness of bank customers using the model proposed by Doumpos et al. [4] and expert views. Later, the number of key feautures was reduced to eight by applying clustering techniques. Loan size, collateral value, loan term, monthly income, applicant's age, and loan type were respectively the most important feautures. In the second phase, eight machine learning techniques were used to categorize customers into creditworthy and non-creditworthy classes. Finally, in the third phase, one of the most effective DEA models called the Andersen-Petersen Model was used to rank machine learning techniques. The results indicated that SVM has the best performance based on various measures.

5.2. Recommendations

- Developing a credit allocation policy based on different categories of customers.
- The category to which a customer belongs should determine the maximum loan size, number, type and value of collaterals, loan type, interest rate and repayment type.
- When offering credit facilities to bank customers, attention must be paid to the parameters extracted from credit assessment models such as occupation, age, monthly income, collateral type, collateral value, and credit score.
- Credit scoring indicators often does not include willingness to repay credit facility installments (behavioral and cognitive indicators). More weight is often given to repayment capacity (indicators of income and cash flows) in various models.
- The proposed indicators and their respective weights can be adjusted during different time periods depending on economic conditions and technology.

References

[1] M. Alborzi, M.E. Pourzarandi, M. Khanbabei, Application of genetic algorithm in decision tree optimization for credit scoring bank customers, Iranian Journal of Information Technology Management, 2.4 (2010).

[2] M. Armashi, Identifying influential factors in the credit risk of individual customers: A case of Saman Bank branches in Northern Iran, Master's Thesis, Payame Noor University, Mazandaran, Iran, (2011).

[3] P. Danenas, G. Garsva, Selection of support vector machines based classifiers for credit risk domain, Expert Systems with Applications, 42.6 (2015): 3194-3204.

[4] M. Doumpos, C. Lemonakis, D. Niklis, C. Zopounidis, Analytical techniques in the assessment of credit risk. New York: Springer, 2019.

[5] H. Ghodsipour, Discussions in Multicriteria Decision-Making: The Analytic Hierarchy Process (AHP). Tehran: Amir Kabir University Press, (2002): 12-20.

[6] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques (Third Edition), San Francisco, CA: Morgan Kaufman, 2012.

[7] N.C. Hsieh, L.P. Hung, A data driven ensemble classifier for credit scoring analysis, Expert systems with Applications 37.1 (2010): 534-545.

[8] Y.C. Hu, J. Ansell, Measuring retail company performance using credit scoring techniques, European Journal of Operational Research, 183.3 (2007): 1595–1606.

[9] A. Kamali, Factors affecting customers' credit rating and a model for ranking customers: A case of Sina Bank. Faculty of Management and Accounting, Islamic Azad University, Central Tehran Branch, 2011.

[10] L.J. Kao, C.C. Chiu, F.Y. Chiu, A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring, Knowledge-Based Systems 36 (2012): 245-252.

[11] D. Liang, Ch.F. Tsai, H.T. Wu, The effect of feature selection on financial distress prediction, Knowledge-Based Systems 73 (2015): 289-297.

[12] J.H. Min, Y.C. Lee, A practical approach to credit scoring, Expert Systems with Applications 35.4 (2008): 1762-1770.

[13] H. Moghasami, B. Alizadeh Savareh, Data Mining with MATLAB and IBM SPSS Modeler. Tehran: Young Computer Explorers, (2015).

[14] M. Momeni, Clustering Data (Cluster Analysis), Tehran, Moallef publication.(In Persian) (2011).

[15] R. Mousavi, E. Gholipour, Ranking measures of bank customer credit scoring using the Delphi method. Proceedings of the First International Conference on Service Marketing in the Banking Sector, Tehran, Iran, (2009).

[16] L. Nanni, A. Lumini, An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring." Expert systems with applications 36.2 (2009): 3028-3033.

[17] G. Paleologo, A. Elisseeff, G. Antonini, Subagging for credit scoring models, European journal of operational research, 201.2 (2010): 490-499.

[18] Y. Ping, L. Yongheng, Neighborhood rough set and SVM based hybrid credit scoring classifier, Expert Systems with Applications, 38.9 (2011): 11300-11304.

[19] R. Setiono, B. Baesens, C. Mues, A note on knowledge discovery using neural networks and its application to credit card screening, European Journal of Operational Research, 192.1 (2009): 326-332.

[20] B. Twala, Multiple classifier application to credit risk assessment, Expert Systems with Applications, 37.4 (2010): 3326-3336.

[21] S. Vukovic, B. Delibasic, A. Uzelac, M. Suknovic, A case-based reasoning model that uses preference theory functions for credit scoring, Expert Systems with Applications, 39.9 (2012): 8389-8395.

[22] G. Wang, J. Ma, A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine, Expert Systems with Applications, 39.5 (2012): 5325-5331.

[23] I.C. Yeh, C.h. Lien, The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients, Expert Systems with Applications, 36.2 (2009): 2473-2480.

[24] L. Zhou, K.K. Lai, L. Yu, Credit scoring using support vector machines with direct search for parameters selection, Soft Computing, 13.2 (2009): 149-155.