Adv. Math. Fin. App., 2023, **8**(4), P. 1409-1428

Advances in Mathematical Finance & Applications

www.amfa.iau-arak.ac.ir
Print ISSN: 2538-5569
Online ISSN: 2645-4610
Doi: 10.22034/AMFA.2022.1950520.1689

Applied-Research Paper

Presenting the Smart Pattern of Credit Risk of the Real Bank Customers Using Machine Learning Algorithm

Hojjat Tajik[a], Ghodratollah Talebnia[b, *], Hamid Reza Vakili Fard[b], Faegh Ahmadi[b]

[a]*Department of Accounting. Qeshm branch, Islamic Azad University. Qeshm, Iran.*
[b]*Department of Accounting. Science and research branch, Islamic Azad University. Tehran, Iran.*

ARTICLE INFO

Article history:
Received 2022-01-27
Accepted 2022-04-10

Keywords:
Smart Pattern,
Bank Customers' Risk,
Credit Risk,
Machine Learning,
Random Forest Algorithm

ABSTRACT

In the past, loan approval decisions for bank customers in Iran were traditionally made based on personal judgments regarding the risk of repayment. However, the increased demand for banking services from economic enterprises and families, coupled with heightened and extended competition among banks and financial institutions in the country to reduce facility repayment risk, has necessitated the adoption of novel methods, including statistical approaches. Today, bankers employ customer credit ranking to predict the risk of default in banking facility repayment and classify candidates. This new approach offers several advantages, including time efficiency, cost-effectiveness, elimination of personal judgments, and enhanced precision when assessing applicants seeking various forms of funding. Numerous statistical methods, including bias analysis, logistic regression, non-parametric parallelism, as well as other techniques such as neural networks, have been applied to credit ranking. In this study, a smart model for real bank customer credit risk, based on the random forest metaheuristic algorithm, is presented, with a focus on the case study of Bank Tejarat. Based on the skewness value, the data can be considered to exhibit a normal distribution. The results reveal that the variable related to the type of facility had the lowest mean, while the maximum value was associated with the facility amount.

1. Introduction

Banking system is one of the main cornerstones of any economic system. Banks as well as financial and credit institutes play a significant role in implementing monetary policies. Therefore, their correct and principal performance can significantly contribute to economic growth and boom of the community [1]. Over their life, banks have faced various risks, for instance, regarding their liquidity, credit status, commercial and financial situations, exchange rates, interest rates, and inflation. Among these risks, credit risk has an especial place. Bankers' profit margin reduction mainly owing to credit risk management inefficiency, have made certain bankers to be under pressure for cost reduction [2]. Credit risk means that fund repayment by bank customers is made either with delay or left unpaid at all. Both cause problems with respect to cash flows of the lending banks and adversely affect their liquidity through

* *Corresponding author. Tel.: +989123360370*
E-mail address: gh_talebni@yahoo.com

marginal investment yield of the banks [3]. Today, with the progressive development of the Internet, the integration of technology and finance proceeds more profoundly, which in turn, leads to dramatic changes in the financing industry. By stimulation of consumer financing, the demand for various credit businesses is growing, so a reasonable and reliable risk assessment model must be established [4]. With increased confrontation of bankers with credit risk, their willingness to experience financial risks is increasing. Creating credit is a major income-generating activity for bankers. However, this process is associated with much risk for both lenders and borrowers. A risk resulted from the commercial partner failure at times of contractual commitments, can significantly compromise the bank commercially. On the other hand, a bank with high credit risk has a high bankruptcy risk, which compromises its investors. Totally, credit risk is defined as the largest risk affecting the bank performance [5]. On the other hand, in the past, commercial banks often assessed credit risks for their real customers, often relying on their risk control personnel. Obviously, this method is prone to inefficient subjective judgments when providing and deciding over customer service, and its evaluation is highly dependent on the mental judgment ability of risk control personnel [4]. There is a possibility of the risk control personnel's internal fraud and their being unable to keep abreast of rapid development of their market economy, to meet their customers' needs, and their risk management needs of online service [6].

The high level of non-acting funds in bank balance sheet, reduces bank profitability and influences its performance. Over the aforementioned risks, banks are subject to credit risk. Therefore, the effective management of credit risk in financial institutions has become crucial for their survival and growth [7]. Credit risk is an indigenous determinant in bank performance. Therefore, risk management on credit side, affects profitability of the banks. Through effective management of credit risk, banks not only support their stability and profitability performance, but they also contribute to economic stability and effective capital allocation in economy. Bankers must ensure that borrowers are able to repay their installments before allocating funds to them [8]. According to BASAL2, every bank requires the organization and development of scoring system for their internal credit, so as to be able to assess borrower's risk. It leads to increased demand for scoring systems which are able to precisely model the risk in high resolution. Some institutions are well accounted for developing such models for banks at their request. Then, such credit scoring techniques can be used as decision supporting tools or as automatic decision algorithms for a wide range of customers.

Given the competitive environment governing the banking system in different countries, banks need to make efforts to maintain and develop themselves in their competitive situations and hence, they need to pay especial attention to customer credit risk and guarantee their capital return resources. Due to economic structure of Iran plus some reasons such as underdevelopment of capital markets and other non-bank networks, the banking system is responsible for financing actual economic sector. Inability of customers to repay their facilities results in blocked bank capital, known as deferred facility, and significant amount of unpaid or deferred facilities suggests non-application of appropriate credit risk measurement models and inappropriate assessment of their borrowers' facility repayment ability, so much so that we witness higher statistics and figures pertaining to different Iranian banks and these statistics have been on the rise over time. Therefore, the banking system should have a basic criterion for appraising customers' ability for the purpose of granting banking facilities.

Credit risk management is to control the credit risk threatening its bank capital. Stabilizing bank assets and ensuring appropriate capital return is another goal of credit risk management. By managing the risk, banks will be able to pay for its liabilities and to create value for its shareholders [9]. To control credit risk, bankers have used both qualitative and quantitative methods to minimize unrealized credit repayment. For this purpose, many of the credit scoring methods have been taken to assess and analyze credit risk. With respect to the economic structure of Iran and for some reasons such as underdeveloped capital markets and other non-banking networks, the responsibility of financing real economic sectors rests on its banking network. Inability of customers to repay facility installments results in blocked bank capital because a considerable amount of granted facilities are left unpaid or deferred, implying lack of applying appropriate models of credit risk. By transmission of progressively larger amounts of data on the Internet as new information, the application of a traditional credit scoring model has been severely curtailed and the core business logic framework has disappeared. Also, by increase of the number of bank customers, provision of online services entails adoption of various methods to reduce manual participation in monitoring and testing process, and automated methods should be used to improve rendering such services [4]. Data analysis method and a part of artificial intelligence comprise machine learning that is able to learn from previous data, identify patterns or distribute data sets, and make decisions. This system builds an analytical model with minimal human intervention and autonomy.

A concern for financial institutions around the world is the widespread growth of non-performing assets. Therefore, identifying, forecasting, and preventing credit risks to achieve sustainable competitive advantages have become the main priorities of bankers. This data is really extensive, very unstructured, and unbalanced. [10]. On the other hand, given the large amount of customer data available to financial companies, adopting traditional statistical approaches, such as regression, to predict real customers' credit risk may not provide the best forecast performance [11].

Therefore, optimal balance and efficient feature rating are required to use machine learning techniques for predicting customers' credit risk. Random forest is the most popular machine learning algorithm, which collects a large number of decision trees from the training data set, and uses a tool called bagging to perform classification and regression tasks. Each decision tree represents a class prediction, this method collects comments from the decision trees, and the class with the most comments in its favor is selected as the final class [12]. Improvement of facility control and repayment process, and also prediction of loss of facility—i.e. without repayment—create a comparative advantage for banks and credit institutions. There was not enough research on assessing credit risk of real bank customers via machine learning algorithm. In this study, we have evaluated real bank customers' credit risk based on random forest algorithm. To validate and apply the proposed model, the data of real customers of Tejarat Bank, which was listed on the Tehran Stock Exchange in the interval from 2015 to 2018, was used.

Based on random forest algorithm, during training, each basic model is created independently by learning from different random samples of data. Samples are pooled by a process called bagging or boot strapping, implying some samples may be used multiple times in a decision tree. By using different sub-samples to teach each decision tree, the whole forest will have low variance but high bias, despite the fact that each tree has a high variance in terms of a specific set of educational data sets. The predictions of each decision tree are meand to extract the general predictions during the test, and this process is known as bagging. This method performs better than other decision tree algorithms because many uncorrelated decision trees can protect each other against individual errors to derive similar predictions, thus overfitting problem can be reduced and the prediction results are unexcelled in terms of accuracy among the current algorithms [4]. This model can also be implemented effectively on a large number of datasets, and can estimate the variables important in the classification as well. Random forest can establish a non-linear relationship between object and attribute, so it works better than linear models [13].

2. Literature Review

As most of the financial resources of banks are used as facilities and the main income of banks is also provided by this section, granting credit is considered as the most important consumption of financial resources of banks. As these resources are constrained, banks must attempt to allocate such resources for the development of the production and services sections with the aim of appropriately acquiring higher profit. In banking system, creation of a customer credit risk system capable of examining the repayment ability of candidates prior to granting the facilities is very importance. There are now many models and various practices to classify the banking candidates, each being based on a special pattern and their objective is to classify candidates into two categories, namely, good candidates with high probability of repaying the facilities and bad candidates with low probability of facility repayment. Banks are the main part of financial systems that have always been facing various risks. Credit risk had been the most important risk among them, causing financial problems for the banks.

Then, its examination requires the advanced modeling techniques. A considerable mass of the unpaid banking facilities indicates a lack of proper models of credit risk examination and risk management system in banking systems. In recent years, delaying claims are a problem that have caused a lack of financial resources of banks and have incurred losses to banks. By rapid growth of banking industry, credit-assessment models are widely used to examine granting or non-granting of credit to candidates. Some of the benefits of credit-assaying models are as follows: 1) reduction of credit analysis cost. 2) Immediate decisions about customer credit-assaying. 3) Credit guarantee and the liberation of possible risks. By analyzing bankers' customer information using data mining process, we are able to examine the fund candidates and to classify them into the good and bad candidates based on smart systems without personal judgements.

2.1 Customer

Customer is an individual or occupation that purchases the goods or services of another firm. Because of driving the income, customers are important. Most public occupations compete with other firms to attract the customer with aggressive advertisement of their own products or lowering the costs to extend their customer bases [14]. Customer refers to audiences who are able and prone to purchase a commodity or a service. In this definition, ability means the possibility of paying the money and being prone to means the perception and recognition of advantages of the commodity and service, causing meeting a part of the audience's requirements. Therefore, the conversion of an audience to customer is met when the two elements of ability and being prone to develop in her/him continually so that the act of "purchasing" is carried out. Having the ability without being prone to purchase, and lacking the ability to repay while being prone to, makes the formation of the purchasing process impossible [14]. The more economic growth in a country, the more contribution of financial and banking services. Basically, with their internal growth, the financial and banking sector of the country play a very important role in the satisfaction of future goals and prospects.

It is the banking system that supports entrepreneurship, provides the possibility of binding thoughts with the capital and financial capabilities, whose product is innovation and advancement in the country. In the situation today, banks should not decide as in the past. Earlier, when the credit decision was made, the capital status, collateral and technical competence of the facility applicant were considered and collateral would take the first place. Today, however, based on the conducted survey, the bankers have come to the conclusion that collateral can no longer be a major factor in repaying facility installments. It is possible that one might not have any collateral but could well repay his own debt and be committed to his obligations. There are many good customers that repay the cost of facilities as their bankers are waiting for their bad customers to repay. It is while some facilities should have been assigned to good customers. From the standpoint of most experts, as many of the bank resources are spent as credit and the main interests of banks are realized from this activity, granting credit is considered as the major bank resource consumption. Therefore, prevention must always be made in prior to treatment in granting banking facilities. Prevention costs are always lower than treatment ones. If banks and financial institutes correctly carry out credit-assaying before granting facilities to the applicants upon a correct credit rating and the facilities are duly repaid, then the costs of granting facilities will be definitely lower as compared to the case of accumulation of delayed claims due for collecting our receivables. In case credit-assay of customers is not carried out appropriately and there is no supervision and control over their ability, the resources will be lost [15].

2.2 Credit Risk

In its general definition, risk is a probability that a certain action or activity leads to a loss or undesirable consequences or outcomes. Almost all human attempts incorporate some degrees of risk. However, some of them are associated with higher risks. In financial literature, risk can be defined as unexpected events that are usually in the form of variation in assets or liabilities value. In various economic and financial contexts, risk means the generation of some conditions in the economy of a country that cause the loss of foreign firms or investors in the host country or causing them not to acquire the expected yield. Economic risk is a type of risk that threatens the economy of any country.

Among the types of risk, the credit risk, financial risk, economic risk, inflational risk, monetary fluidity risk, and systemic or market risk can be addressed. All risk types are important in terms of influencing the attraction of foreign capital [16]. The phenomenon of risk is one of the key indices for the formation of decision in investment area, affairs related to financial markets, and various types of economic activities. In most economic books, labor, land, and capital are the three factors referred to as the main production inputs. However, with some contemplation these three factors are found as the necessary condition for production. In the process of production, the sufficient condition is nothing but the risk factor. In other words, if the three above-said factors are available, yet the manufacturer does not undertake the responsibility for possible losses of this process, the production will never take place. Therefore, in some studies risk factor was called the fourth factor in the process of production [17].

2.3 Similar studies

There are many studies and applications over various contexts of classification to identify banking candidates. In 1930, fisher and Durand created the initial models of credit-assaying. Grace and Williams [18] proposed the new fuzzy GDM model based on a smart factor as an effective MCDA tool to assay the credit risk. They assessed a simple numerical example with three real credit datasets of England, Japan and Germany. The results clearly showed that the proposed fuzzy GDM model outperformed other scalable models. Hsie and Huang [19] proposed a classification system for bank candidates using the similarity classification methods, neural networks, and support vector machine.

Chuang and Huang in 2011 [20] proposed a two-stage method: In the first stage, a neural network was used to classify the candidate as accepted or rejected. In the second step, case argument was used to identify the rejected candidates that should have been accepted. Shen et al, [21] proposed the three-stage model of the compound adaptive fuzzy neural deduction system to score the credit that is based on statistical and neural fuzzy techniques. He compared it with conventional models and the commonly used ones. The results displayed that the proposed model performs consistently better than linear distinguishing analysis, LR analysis, and artificial neural network approaches in terms of correct mean classification rate and estimated cost of misclassification. Wang, et al. [4] comparatively evaluated a credit risk model based on machine learning. They used the performance of five classifiers in machine learning: Naive Bayesian Model, Logistic Regression Analysis, Random Forest, Decision Tree, and K-Nearest Neighbor Classifier for credit scoring. They found random forest perform better than others in terms of recall, AUC (area under curve), and accuracy. Lappas and Yannacopoulos [22] proposed a combined strategy by integrating soft computing methods with specialized knowledge. Expert opinions were considered to solve a credit scoring problem in the form of a constrained optimization problem through soft computational methods based on supervised machine learning and evolutionary optimization algorithms. Pandey's, et al. [10] feature rating algorithms were applied to identify the most critical features of credit risk creation. The algorithms were performed on the credit risk data set collected from a bank showed that random forest has the best performance in the optimal balancing ratio of 1: 1335 with a sensitivity of 81.6%, specificity value of 85.3%, accuracy of 83.4%, MCC of 0.669 and AUC of 0.914. Pandimurugan et al. considered [23] lending to individuals in previous years through their extraction and the extracted patterns were utilized to teach the proposed model of random forest. A machine learning model was used to predict whether the loan could be approved by the applicant or not, to effectively reduce risk of lending to financial institutions or banks.

Machado and Karray [24] applied machine learning algorithm to assess commercial customers' credit risk and showed that hybrid models outperform their individual counterparts in predicting business customers' credit scores. García-Céspedes and Moreno [25] investigated the ability of machine learning (ML) techniques to calibrate models that replicate the outputs of the Vasicek's (1987) credit risk model and indicated by using only two variables (confidence level and Gaussian copolyte-based loss distribution estimation), the tree-based models provide real-time and accurate estimates of the real loss distribution. Nazar Aqaei [26] presented a paper titled: Classification of the credit risk of real customers using collective learning (case study: Bank Sepah). It has been conducted using the information of the real customers of Bank Sepah in 2016 and 2017 and modeling of the study has been made using neural network and fuzzy decision tree. The innovation of that study could be collective learning methods used to increase the accuracy of fuzzy decision tree outcomes in this research. The results obtained from the research indicated that income and financial transactions of customers have the highest importance for specification of customers' credit risk. In addition, the results showed that, using bagging method, fuzzy decision tree has higher accuracy than neural network and conventional fuzzy decision tree. In their research, Del Afrooz, et al. [27] presented a appropriate model to manage the credit risk in banks by use of a compound DEMATEL approach and Analysis of Network Process (ANP) for Bank Saderat branches of Gilan province.

The statistical population of the present research is the managers and experts of the bank. The means of data collection was a questionnaire where 8 questionnaires were distributed among the research experts. Analyzing data in DEMATEL process, it was found that among 17 criteria identified to be effective on credit risk in banks, the operational risk has the highest influence on credit risk. Also, among the operational risk criteria, the quality of the investigation process on facilities has the highest influence. In other words, the criterion is the most important one in terms of influencing and being influenced in the entire system. In addition, the concentration risk also has a high importance in the entire system. The results of this research are useable and important for credit risk management of banks.

Khojasteh [28] presented a paper entitled: Credit rating of the real bank customers with compound logistic-symbolic regression approach. The findings showed that among 17 independent variables, the variables of the mean monthly income, number of returned cheques, bank debt record, lifetime of accounts, and the type of collateral have the highest significant impact on dependent variable. In addition, it was found that the accuracy of the compound logistic-symbolic regression model in the classification of good customers had been 0.88 and in the classification of bad customers 0.83. Mohaqeq Nia, et al. [29] investigated the impact of internal and external factors of banking industry on credit risk of the banks in Iran. To investigate the impact of the said factors on the credit risk, a tabular data model (fixed impacts) had been used. The statistical population of the research includes 31 banks. The results implied that among the intra-bank variables, size and capital had a positive impact and development of credit supplying has a negative impact on credit risk. And among the outside-the–bank variables, concentration, liquidity growth rate, and foreign currency rate had a positive impact and banking sector development and economic growth rate had a negative impact on credit risk.

In some research, Heydari, et al. [30] investigated the credit risk modeling of bank credit facilities basket using actuary modeling in Bank Refah. Firstly, using the data related to the number of default cases in various years and initial statistical methods such as mean and standard deviation of default cases, the default risk of credit portfolio of the banks has been estimated in a classic way. Next, using more advanced methods based on actuary modeling, mean and standard deviation as well as distribution type have been estimated more accurately. Therefore, it has been made possible to achieve a clear position over evaluating the credit risk of the portfolio of the facilities granted by Bank Refah.

In their research, Salehi and Kurd Kanouli [31] investigated the selection of optimal characteristics to specify the credit risk of bank customers. This paper presented a compound method composed of colonial competition optimization algorithm and neural network to increase the classification accuracy in the evaluation and assessment of bank customers' credit risk. By identifying a subset of optimal characteristics and the removal of unnecessary characteristics from the entire characteristics existing in data, this method reduced the problem aspects and increased the classification accuracy. The proposed approach was applied to the real dataset of UCI database, as well as the real data of a private Iranian bank for credit-assaying. The obtained experimental results showed that the error amount of neural network for the test set was reduced by selecting effective characteristics and removing low-impact characteristics by zero-an-one optimization algorithm of colonial competition. In addition, for other employed classification methods, the error value of the test data remains within an acceptable limit.

In some research, Ghuli Pour, et al. [32] evaluated the credit risk of some real bank customers using logistic regression method and neural network in tourism bank branches of Tehran. For this purpose, 161 people including 96 good customers and 65 bad customers were selected from among the legal customers having received facilities during 2011-2015. Thirteen indices were identified as independent variables affecting the possibility of default through which a final model was fitted after investigation of the entire regression significance using LR statistics at zero significance level. Then, the variables with significance were considered as input variables in neural network model. The results of the studies were based on the establishment of the hypothesis that there is a significant relation between individual indices of customers and the risk of granting credit to them. In addition, a number of factors causing an increase in the chance of risk of granting credit to real customers were identified and announced.

In their research, Mir Ghafouri and Amin Ashouri [33] evaluated bank customers' credit risk. In this research, a parametric method (logistic regression) and one non-parametric method (division tree and regression) were used to create credit scoring model. To construct the credit scoring model, the data related to 282 small and medium sized borrower firms of one of Bank Tejarat branches of Tehran were used. Thirteen financial rates were employed as the indices determining the financial situation of the selected firms. Using these two methods, the effective rates as well as the accuracy of the named methods in classifying customers were specified. Observing the results obtained from the examination of these methods illustrated that non-parametric methods have competitive accuracy with parametric methods. In some research, Hussein and Zibaei [34] investigated credit risk management in Bank Keshavarzi of Mamasani township using neural network. The required investigations had been conducted over the financial and non-financial information related to a 205-tier sample selected by random multi-stage clustering sampling method among the farmers receiving loan in Mamasani township during 2007-2012. In this research, 17 explanatory variables such as the financial and non-financial variables were investigated and analyzed. The selected variables were used in the model as input vector of multilayered perceptron neural network with three hidden layers. The results implied that neural network model could estimate the observations in reality with a prediction correctness percent of 95.5%, showing high capability of the neural network in predicting customers' credit risk.

In their research, Ghudsi Pour, et al. [35] evaluated the credit risk of borrower firms from bank using fuzzy hierarchal analysis and high degree compound neural network. Using fuzzy hierarchal analysis method, the criteria affecting credit risk for loan-applying firms was weighted and the association between the criteria influencing credit risk and credit amount for the loan-applying firms was extracted as open box model using neural network. Neural network model was applied to historical data of 174 borrower firms that received loans from Bank Mellat of the Islamic Republic of Iran within 2004-2008 period and their repayment period were expired.

The model output was capable of predicting credit risk with 84% accuracy. Bayazidi [36] presented a paper titled: Credit rating of real customers of Bank Mellat using data mining methods. The results of this study showed that among the variables related to the personality characteristics of candidates, such as gender and occupation, and among the variables related to granted facilities features such as interest rate, punishment rate, number of received facilities, facilities duration and facilities value had the highest importance in the separation of good and bad customers and its significant impact on dependent variable, namely lack of repayment, was confirmed via statistical tests. The criterion for comparing these three models is classification accuracy of good and bad customers. Therefore, the accuracy of each method after computation for logistic regression model via SPSS was 1166%, for tree structure model 1365% and for artificial neural network model 3.66%. Therefore, artificial neural network model fitness was more effective in this field.

Abbasi Estemal and Rahimi [39] Designing an Expert System for Credit Rating of Real Customers of Banks Using Fuzzy Neural Networks. Currently, in Iran's banking system, non-repayment of facilities has become one of the biggest issues, and due to the lack of a proper system for proper allocation of facilities, they face a number of problems, including the problem of allocation of loans, the problem of failure to repay loans of the central bank, or the amount of facilities increased from the amount of reimbursement. The solution of this problem is the credit rating of the customers, which is based on a model based on the theory of fuzzy sets for validation of real customers of the Maskan bank of the East Azer-baijan in Iran in 2016. In this research a structured model was obtained for deter-mination and categorization of input variables for application in the system by factorial analysis then an expert fuzzy system was modelled that consist of six steps. In the first step a fuzzy system is designed that its inputs are financial capacity, support, reliability, repayment record and its outputs is customer credit. In the second step input and outputs are partitioned, in the third step the partitioned inputs and outputs are converted into fuzzy numbers. The fuzzy inference is compiled in step four. In step five the defuzzifier is conducted. Finally, the designed model is tested in step six. These results indicate research model efficiency compared to bank credit measuring experts that they predicate applicants performance according their judgment and intuition. Izadi [40, 41] improved the long-term values of bank shareholders by using the data envelopment analysis model. Given the rapid development of the banking sector, it is reasonable to expect that the performance of banks has become the centre of attention among bank managers, stakeholders, policy makers, and regulators. In order to maximizing the share-holders' satisfactory level, two bank efficiency measurement approaches, i.e. the production approach and the user cost approach, which are financial evaluations, are employed. The evaluations are done by means of data envelopment analysis method. The proposed methodology is run on the 15 privet bank branches in Markazi province. By using this approach, four regions that show the various performances are obtained. In addition, the status of returns to scale for each bank branch is calculated.

Nouraldin Kalantari, et al. [41] presented a fuzzy goal planning model based on roller performance budgeting with an efficiency approach (Case study: Gas refineries in Iran). This research presents a mathematical model for performance-based budgeting and combines it with rolling budget for increased flexibility. The model has been designed by Chebyshev's goal programming technique with fuzzy approach. The parameters or coefficients of the model are derived by measuring the productivity of the organizations considering eight criteria. Data for calculating productivity indicators were collected from gas refineries of Iran in 2011–2015 and analysed by Excel and GAMS software. Then, the model was tested for determining the 2016 budget of those refineries. The model was solved by LINGO software by linking it to Excel. The solution of the model reduced 0.68% of the total refinery's budget compared with the actual budgets for 2016, which is higher than the annual budget of some of the companies in this group. Lack of effective supervision over proper project implementation, use of facilities, and project commissioning prolongation will increase its banking costs.

A large part of potential financial resources of banks includes deferred facilities, which if duly realized, will enhance the financial power of banks, and if not realized, esp. in a competitive atmosphere, the lending banks incur serious damages. Therefore, credit risk assessment can help bank managers with allocating loan and ensuring repayment. According to some previous studies, little research has been conducted on the credit risk of real bank customers upon random forest algorithm. In addition, to validate the proposed model, the data obtained from the real customers of Tejarat Bank have been utilized in this study. On the other hand, one of the advantages of random forest method used in this research was the possibility of managing a large number of input features. The high speed of this algorithm was of its another important feature.

3. Research Method

This study was conducted as applied research for its purpose and was carried out via a descriptive-analytical method. The data was obtained from Tejarat Bank real customers listed on the Tehran Stock Exchange in an interval of time from 2015 to 2018, selected via cluster random sampling method. Data was collected from Tehran Stock Exchange site and Kodal site. Data analysis was performed through MATLAB and SPSS software.

GARCH and ARCH techniques are one of the most important methods used in various econometric branches to estimate uncertainty and instability indices, known as self-regression models under variance heterogeneity. In this method, conditional variance varies based on the previous information and the past prediction error, and is indicative of variable fluctuation. Hence, in this study, firstly, the reliability of the variables was investigated. Then, the behavior of this variable was predicted using ARIMA model. In the next phase, the existence or non-existence of self-correlation and variance heterogeneity was investigated using the related tests and if the model predicting the behavior of the variables has variance heterogeneity, ARCH model was used to estimate the indices related to fluctuations of the variables. Skewness and kurtosis statistics in SPSS were also used to check the data normality.

Table 1: Independent and Dependent Variables

	Variable type	Data type	Reference
Bank interest rate	dependent	Numerical	[4]
Returned check record in banking network	dependent	Interval	[4]
Delayed debt record in banking network	dependent	Interval	[4]
Percentage of collateral value to facilities receivable	dependent	Ordinal	[1]
Facilities value	dependent	Numerical	[1,2]
Duration of facilities repayment	dependent	Interval	[2,26]
Facilities type	dependent	Nominal	[2,4]
Credit risk	Independent	Numerical	[38]

Time series of logarithmic returns

To obtain the return, we use the following formula:

$$r(t) = \ln(p_t) - \mathrm{In}(p_t - 1) \quad (1)$$

Where p_t is the share price at the end of the period T, and $p_t - 1$ is the share price at the end of the period T-1.

3.1 Random Forest

Random forest makes use of a number of the classes of non- pruned decision trees where each of these trees are created using random division in each of the decision tree nodes. Each of the trees are likely less correct relative to the tree created with accurate division. However, their correctness can still be improved by combining several trees of this type in a group. It often causes their correctness to be better than a tree with just the accurate classification. Each of the decision trees are created using the following algorithm. Random forest construct algorithm

1 – In the first iteration, put the value of ring variable While (t) equal to 1.

2- Then, using replacement, sample a μ sample out of S training set and put it in ST set.

3- Using IDT (N), construct MT classifier over ST set.

4- Increase the variable amount of t ring by one unit.

5- Iterate the above until t < T holds.

The algorithm inputs of an Inductor Decision Tree (IDT), are the number of iterations T, and training set S, subsets size μ and the number of the features used in each of the decision tree nodes N. N input parameter indicates the number of input variables that will be used to make decision in a node of decision tree. The parameter must be much smaller than the number of features in S training set. The IDT parameter in the above algorithm, displays any induction algorithm of top-bottom decision tree with the following terms: decision tree must not be pruned and in each of the nodes, instead of selecting the best division among all features, IDT randomly samples N features and selects the best division from among these variables. In addition, classification of label-less samples is done by using majority vote. There are other methods to acquire random forests. For example, instead of using all samples, a subsample of samples can be used to determine the best divider point for any feature. This subsample changes for each of the features. Feature value and divider point are selected as decision in that node based on optimizing a division criterion function. In this study, a version of random forest method has been used that is introduced by Bremen. Another method to randomize decision tree via histograms was introduced by Kamath and Cantu [33]. Using histogram, he disassociated those features to reduce computation time for managing very large data. Usually, a histogram is created for each of the features and the borders of each of these can be used as candidate divider points. In the process, this method of randomization is expressed by selecting random divider points in an interval around the best Bin border.

Despite the fact that random forest method was defined for decision tree, it can also be used for other classifiers. One of the advantages of random forest method is its ability to manage a very large number of input features. Another important feature of random forest method is its high speed.

In this study, we limited our attention to a class of Bayesian network models whose graph, g, has a tree structure. It means that Bayesian networks exists on random variables set of $x = \{Y, x_1, \ldots, x_d\}$ and nodes have a different situation. Variable Y in graph with tree structure states a node having no parent, namely $\Pi_{[Y]} = \{\emptyset\}$ and is called tree root. The rest of nodes, namely $x_i \epsilon \mathcal{G}$ had parents that were either the root variable of Y or identical to Y or at least one other variable namely x_j. Depending on the complexity of implicit relations between variables, a graph with internal structure can be placed in three classes of naïve, tree-augmented, and k-dependence. N-BN indicates the graph structure with minimum complexity that makes it obligatory that root variable of Y is the parent of x_i variable, namely $\Pi_{[i]} = \{Y\} for\ all\ i = 1, \ldots, \mathrm{d}$. Corresponding probability relation between variables displays that all x_i are conditionally independent. It showed that mixed probability distribution of $Y, x_1, \ldots, x_d$ included the following:

$$\mathrm{P}Y, x_1, \ldots, x_d(\mathrm{y}, x_1, \ldots, x_d) = \mathrm{P}(Y). P(x_1, \ldots, x_d|Y) = \mathrm{P}(Y). \Pi_{i=1}^{d} P(xi|Y) \quad (2)$$

This relation guarantees that in this network structure, the probability of $P(Y|x_1, \ldots, x_d)$ is the main expression determining the converted conditional probability distribution of root variable considering the observed $x_1, \ldots, x_d$ based on the edging probability distribution of $P(xi|Y)$ and hence considering all variables of x_i. Although factorization in N-BN basically simplifies the computations of possibilities, the assumption of conditional dependency among variables is not always realistic. For instance, consider a network structure for modelling risk of fund request: It seems it is not logical to ignore the relations among levels of education, income, and age. In order to receive the possible connections among the variables in BN, we use TAN model that is a tree structure over N-BN graph, providing the grounds for the existence of more graphs among $\{x_1, \ldots, x_d\}$. In a fortified structure, the side extended from x_i to x_j displays that the impact x_i of root variable value depends on x_j examination. The fortified graph structure of the tree can be described by identifying the set of $\Pi_{[i]}$ parents of each group. BN converts the dependencies between $\mathrm{y}, x_1, \ldots, x_d$ by estimating the mixed probability distribution as follows:

$$P_{y,x_1,\ldots,x_d}(y, x_1, \ldots, x_d) = \mathrm{P}(Y). \mathrm{P}(xi|Y). \Pi_{j=1, j\neq i}^{d} P(xj|xi, Y) \quad (3)$$

Three structures of conventional Bayesian network

Eventually, in terms of BN, we discuss K dependency where TAN is extended and any node of x_i there are at least K parent variable and the root variable of Y for any characteristic variable of $\Pi_{[i]} = \{y, x_{i1}, \ldots, x_{ik}\}$ The structure leads to the factoring of the following probability:

$$P_{y,x_1,\dots,x_d}(y,x_1,\dots,x_d) = \mathrm{P}(Y).\mathrm{P}\left(x_1\middle|\Pi_{[1]}\right)\dots.P\left(x_{d-1}\middle|\Pi_{[d-1]}\right).P\left(x_d\middle|\Pi_{[d]}\right) \quad (4)$$

Where $Y \epsilon \Pi_{[i]}\ for\ all\ i = 1,\dots,d$ holds. Figure 1C shows a special case of K-BN with K=1.

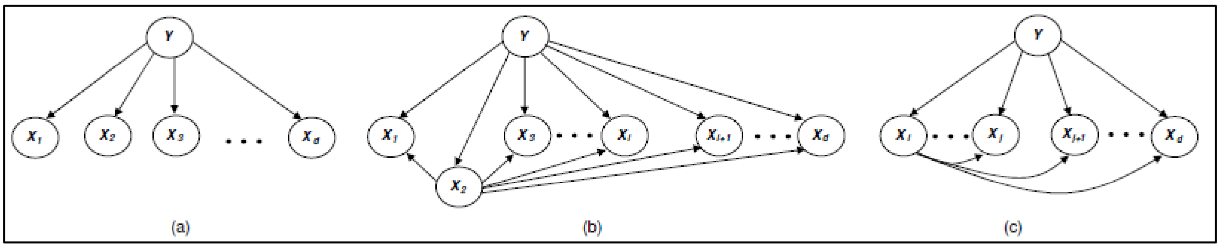

Fig. 1: Displays a Possible Structure for TAN

Note that TAN and K-BN have tree structure graph and TAN models equal K-BN models in which K=1 holds. According to the definition of K-BN, it is observed that the N-BN model is a BN with O dependency. All the G structure types introduced above can be considered as the simplification of multivariate probability distribution. These simplifications are based on conditional dependency relations taken from G and shown with factorization where, in general, require less parameters than mixed probability distribution. Now we investigate rooted graphs and directional tree and show that they provide an effective method to model and evaluate credit risk accumulation.

3.2 Bayesian Rules and Examination of the Previous Credit Risk

Since the quantitative part of BN model is expressed via a set of conditional probabilities, we begin with Bayesian theory that is the essential part of probability computations. This theorem states that two subsets of random variables of x and y, such that $\mathrm{P}(x) \neq 0$ holds, specify the conditional probability distribution of y as follows:

$$\mathrm{P}(Y = y|X = x) = \mathrm{P}(y|x) = \frac{\mathrm{P}(x|y)\mathrm{P}(y)}{\mathrm{P}(x)} \quad (5)$$

A special case in the above relation is obtained when y is a single variable. That is, $Y = \{y\}$ and $x = \{x_1,\dots,x_d\}$ are a subset of variables. In this case, the relation becomes as follows:

$$\mathrm{P}(y|x_1,\dots,x_d) = \frac{P(y)P(x_1,\dots,x_d|y)}{\Sigma_y\, P(y)P(x_1,\dots,x_d|y)} \quad (6)$$

In the credit risk assessment problem, the Bayesian theorem can be stated as follows: financial institute is interested in examining the credit risk caused by the uncertainty of the borrowers' ability to conduct the commitments, that is, to repay the loans. Assume that the borrower's ability to pay the debt, specified with a random binary variable, takes one of the possible values of {solv, lns}. If Y= solv holds, it means that the firm is able to pay the debts. If Y=INS holds, it means that the individual is not able to do so. For simplicity, assume that two indices, namely stock market performance and guaranteeing firm's financial status are related to the credit power of the borrower. Assume that the results of the credit expert investigation illustrate that the borrower's stock market performance is weak and the warrantor lacks a proper health financially. That is, x_1 = Poor, x_2 =Distr holds. Now, considering the evidences, we want to calculate the probability of the firm's inability to pay the liabilities. Using Bayesian theorem, we have:

$$\mathrm{P}(Y=\mathrm{Ins}|X_1=Poor,x_2=\mathrm{Distr})=\frac{1}{k}\ \mathrm{P}(\mathrm{Y}=\ \mathrm{Ins}).\mathrm{P}(X_1=Poor,x_2=\mathrm{Distr}|\mathrm{Y}=\mathrm{Ins}) \tag{7}$$

here we have:

$$\mathrm{k}=\mathrm{P}(X_1=Poor,x_2=\mathrm{Distr}|\mathrm{Y}=\mathrm{Ins}).\mathrm{P}(\mathrm{Y}=\mathrm{Ins})\ \mathrm{P}(X_1=Poor,x_2=\mathrm{Distr}|\mathrm{Y}=\mathrm{Solv}).\mathrm{P}(\mathrm{Y}=\mathrm{Solv}) \tag{8}$$

Note: The probability of P (Y=INS) is the a priori probability of the borrower's inability to pay the debt. That is because it can be obtained prior to being notified of the firm's characteristics. The numerical value of this probability addresses the a priori credit risk that is allocated with the expert's opinion in relation to risk expenditures or is, for example, estimated with the mean of firm's percentage with identical financial level on the entire population whose inability has recently been stated.

3.3 Predictive Performance of Patterns Outside the Estimation Period

Table 2: Impact of Explanatory Variables in the Neural Network

	Class 0 (good customer)	Class 1 (bad customer)	Sum
Cases of recognition as good customer	11	4	12
Cases of recognition as bad customer	8	32	35
Total	18	32	32
Number of correct recognition	13	28	45
Model accuracy percentage	57	92	84
Model error percentage	40	8	25

4. Results

4.1 Respondents' Age

As shown in Table 2, from 384 people selected, 4 people were 20-30 years of age, 250 people were between 31-40, 50 people were between 41-50 and 80 people were over 50 years old. The related results have been presented in Table (3) and Diagram (2).

Table 3: Abundance Distribution and Respondents' Age Percentage (Years of Age)

	age	Abundance	Percentage	Percentage of valid values	Accumulative percentage
Valid data	20-30	4	3.3	3.3	3.3
	31-40	250	53.4	53.4	56.7
	41-50	50	30	30	86.7
	Over 50 Years	80	13.3	13.3	100
	total	384	100	100	---------

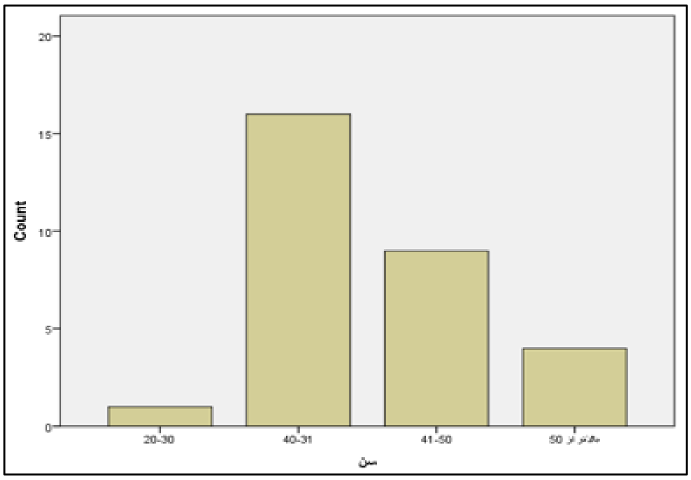

Fig. 2: Respondents' Age Histogram

4.2 Description of research variables

In this section, the statistical indices of mean, standard deviation, and error percentage (independent and dependent variables) were reported. In fact, these indices can contribute to more understanding of research variables. Mean is one of the most frequent criteria of central tendency. In the study of the distribution of a statistical population, the representative value around which other values are distributed is called the central value. Any numerical criterion that is indicative of center of the dataset is a standard of that central tendency. Standard deviation is one of the indicators of variance of the data that shows the distance between each data and mean as compared to an average distance from the mean. If the standard deviation of a set of data close to zero, it indicates that the values are scattered close to their mean and have small dispersion. Whereas large standard deviation shows a large or significant dispersion of data. In fact, standard deviation is also applied in statistical analyses to explain the confidence coefficient. In scientific studies, the data with standard deviation higher than two are considered as outliers and excluded from analysis.

Table 4: Variable Description

	Minimum	Maximum	Mean	Standard deviation	Variance	Kurtosis	skewedness
Bank interest rate	1	4	2.800	1.05176	1.106	-0.939	-0.244
Returned check record in banking network	2	4.75	3.1750	0.66355	0.440	0.323	-0.174
Delayed debt record in banking network	1	4.57	2.7810	1.05695	1.117	-0.167	1.017
Percentage of collateral value to facilities receivable	1	3.60	2.7957	0.49162	0.241	-0.511	0.118
Facilities value	2	4.75	3.2750	0.66465	0.430	0.313	-0.174
Duration of facilities repayment	1	4.75	2.7710	1.05985	1.127	-0.157	1.017
Facilities type	1	3.60	2.7657	0.49862	0.233	-0.522	0.118

Given the above table, it is observed that the lowest mean value is related to facilities type variable and its highest value is related to facilities cost variable. In addition, as the skewedness value is within the interval of (-3, +3), hence it can be stated that the data have normal distribution.

5. Discussion and examination of outcomes

Table 5: Artificial Neural Network Ranking Indices

Validation index		System validation		
Sum square error	SSE	Simple model	Compound model	Results improvement (%)
Root-mean square Error	RMSD or RMSE	231.1799717	212.173712	8.22%
Recognition coefficient	R^2	0.57468	0.55055	4.20%
Adjusted recognition coefficient	Adjusted-R^2	0.896748561	0.90523729	0.95%
Mean square error	MSE	0.893861738	0.90223343	0.94%
Medium mean error	MAE	0.330257102	0.3031053	8.22%
Normal mean square deviation	NRMSD	0.642008631	0.6203509	3.37%
Coefficient of variation (normal root-mean square deviation)	CV(NRMSD)	0.082097143 0.129141573	0.07865 0.1237191	47.20% 4.20%

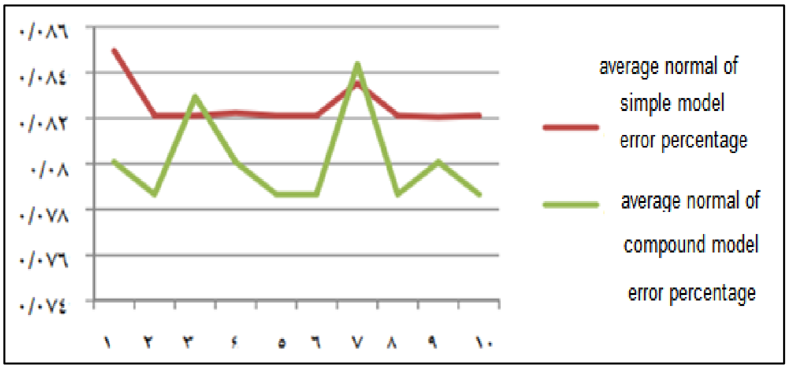

Fig. 3: Comparison of the Diagram with and Without Artificial Neural Network

Table 6: Ranking Criteria

Validation criterion	English equivalent	Values interval	Optimal value
Adjusted recognition coefficient	Adjusted-R^2	From zero to one	Between 0/6 to 0/8 is proper/ Between 0/8 to 1 is excellent
Medium squared error	MSE	From zero to infinity	In comparison of two systems, the lower it is, the better
Medium absolute error	MAE	From zero to infinity	In comparison of two systems, the lower it is, the better
Normal root-mean square deviation	NRMSD	From zero to infinity	the lower it is, the better
Coefficient of variation (normal root-mean square deviation)	CV(NRMSD)	From zero to infinity	the lower it is, the better

6. Conclusion

By optimizing the values for the three root-mean square error indices and the corrected detection coefficient and detection coefficient, the values of other indices were optimized normally as well. However, if there was an insignificant difference, priority was given to these three indicators. The bottom line comprises that except for the use of artificial neural networks; other values set for training the systems were the same. In Table 5, the results of 10 different runs of the model were presented comparatively. The diagrams related to other errors had trend scales and shapes similar to this one and showed a functional distance between the two models in scales.

As many old and probability-, and random-based techniques cannot appropriately interact with data of ambiguous and fuzzy nature, there is a wide range of such methods in the literature that already studied the subject. Therefore, this method was validated by collecting sufficient amount of data and multi-running of the data via software and training. Compared to similar studies conducted on this issue, the vantage point of this research was its combined application of all three methods by adopting an appropriate approach through which the advantages of one method were utilized in favor of making up for the disadvantages or weaknesses of the other, although each of these methods served various attitudes, for example, towards multi-criteria decision-making, data-mining for knowledge extraction, as well as compound optimization algorithm. The results suggested that this method could be employed to evaluate the current situation, and also to present appropriate estimates for future.

7. Summary and suggestions

The results of the research can be presented as follows:

1 – Based on the given variables, there is a statistically significant relationship for determining the credit risk status of the real bank customers.

2- The significance and the independent variable coefficients signal that the model imply the economic and financial theories in the context of the determinants of credit risk were verified.

3- Based on the qualitative and financial variables, it is possible to classify and score the real customers of Bank Tejarat in terms of credit risk.

4- Among the financial rates, the cash rate and current rate have the highest share in dividing customers into two groups of high credit risk and low credit risk individuals.

5- The factors affecting the credit risk of the real customers of Bank Tejarat had much similarities with the factors affecting credit risk of the real customers of other banks (including Bank Mellat and Saderat).

Conducting practical research on the financial context, credit risk, in particular, can be effective in the order of optimal allocation of bank resources. In turn, conducting such pieces of research is a step towards this direction and it presents the following suggestions to the beneficiaries of this research, Bank Tejarat, in particular:

1 – Given the fact that the results indicated use of compound models to validate the real customers is effective, then it is suggested that an internal validation system is introduced based on quantitative methods, highlighting the variables introduced in this research to rate candidates of the facilities in terms of credit to be able to be used.

2- Given the fact that the results showed the credit risk can be inhibited to a large extent by correctly validating the customers, increasing the professional knowledge level of credit commission in terms of credit risk discussions, and the factors affecting them, on the one hand, and training the validation and analyzing the personality of candidates of facilities, on the other hand, is proposed.

3- Given the fact that the bail is established as one of the important factors to reduce the credit risk, it is proposed that the credit commission of the branches receive the required training over taking compound bails from candidates and be obliged to take compound bails from credit customers. In addition, entering into practical bank agreements with other organizations to validate the candidates' membership of that organization and the creation of commitment to pay the facility installments and transferring risk into those organizations can be an effective step in this regard.

4- Given the fact that the results proved that the candidates' income is an important factor for installments repayment, it is proposed to analyze the occupational situation of the customer and his actual income amount considering the economic, social, and other pertinent conditions and to adapt it to the requested amounts of facilities which shall be in line with the income of the individual.

5- Given the fact that the results showed the bank liability, returned check record are among the most important factors affecting credit risk, it is proposed to seriously investigate the previous customer records in this context. Since the bank liability is eliminated by paying the deferred and delayed installments, along with the returned check capable of being cleared from adverse impact, it is proposed to conduct special investigations in this regard and to find the roots for the reasons of this highly risky credit behavior by the credit commission of the branches.

Since the results obtained from this research cannot be generalized to the entire banking system of the country owing to the limitation of the statistical population, it is proposed to estimate similar models at a wider scope in future research based on the patterns presented to fit a wider range of variables for increase of the confidence level and the prediction correctness.

References

[1] Rajabzadeh Moghani, N., Lotfalipour, M., Seifi, A., Razmkhah, M., *The Study of Factors Affecting on Credit Risk of Bank Customers Using Non-Parametric and Semi-Parametric Survival Analysis Models*, Monetary and Financial Economics, 2017, **24**(14), P.88-123. Doi: 10.22067/pm. v24i13.52294.

[2] Shirin Bakhsh Masooleh, Sh., Yousefi, N., Ghornan Zadeh, J., *Investigating the factors affecting the probability of non-repayment of credit facilities (case study of legal customers in Iran Export Development Bank),* Quarterly Journal of Securities Analysis, 2011, **4**(12), P. 111-137.

[3] Chen, W., Xiang, G., Liu, Y., Wang, K., *Credit risk Evaluation by hybrid data mining technique*, Systems Engineering Procedia, 2012, **3**, P. 194-200. Doi: 10.1016/j.sepro.2011.10.029.

[4] Wang, Y., Zhang, Y., Lu, Y., Yu, X., *A Comparative Assessment of Credit Risk Model Based on Machine Learning-a case study of bank loan data*, Procedia Computer Science, 2020, **174**, P. 141-9. Doi: 10.1016/j.procs.2020.06.069.

[5] Tsai, C.F., Wu, J.W., *Using neural network ensembles for bankruptcy prediction and credit scoring*, Expert systems with applications, 2008, **34**(4), P.2639-49. Doi: 10.1016/j.eswa.2007.05.019.

[6] Mhlanga, D., *Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment*, International Journal of Financial Studies, 2021, **9**(3), P. 39. Doi: 10.3390/ijfs9030039.

[7] Shahari, F., Zakaria, R.H., Rahman, M.S., *Investigation of the expected loss of sharia credit instruments in global Islamic banks,* International journal of managerial finance, 2015. Doi: 10.1108/IJMF-12-2014-0196.

[8] Louzada F, Ferreira-Silva PH, Diniz CA. On the impact of disproportional samples in credit scoring models: An application to a Brazilian bank data. Expert Systems with Applications. 2012, **39**(9), P. 8071-8. Doi: 10.1016/j.eswa.2012.01.134.

[9] Singh, S., Murthi, B. P. S., Steffes, E., *Developing a measure of risk adjusted revenue (RAR) in credit cards market: Implications for customer relationship management*, European Journal of Operational Research, 2013, **224**(2), P. 425-434. Doi: 10.1016/j.ejor.2012.08.007.

[10] Pandey, MK., Mittal, M., Subbiah, K., *Optimal balancing and efficient feature ranking approach to minimize credit risk*, International Journal of Information Management Data Insights, 2021, **1**(2), P: 100037. Doi: 10.1016/j.jjimei.2021.100037.

[11] Doko, F., Kalajdziski, S., Mishkovski, I., *Credit risk model based on central bank credit registry data,* Journal of Risk and Financial Management, 2021, **14**(3), P.138. Doi: 10.3390/jrfm14030138.

[12] Smith, P.F., Ganesh, S., Liu, P., *A comparison of random forest regression and multiple linear regression for prediction in neuroscience,* Journal of neuroscience methods, 2013, **220**(1), P. 85-91. Doi: 10.1016/j.jneumeth.2013.08.024.

[13] Polamuri, S., *How the random forest algorithm works in machine learning?* Retrieved December, 2017, P. 21.

[14] Dong, G, Lai, KK., Yen, J., *Credit scorecard based on logistic regression with random coefficients,* Procedia Computer Science, 2010, **1**(1), P.2463-8. Doi: 10.1016/j.procs.2010.04.278.

[15] Witzany, JM., Rychnovský, M., Charamza, P., *Survival Analysis in LGD Modeling*, European Financial and Accounting Journal, 2012, **7**(1), P. 6-27, Doi: 10.18267/j.efaj.12.

[16] Addo, P.M., Guegan, D., Hassani, B., *Credit risk analysis using machine and deep learning models Risks*, 2018, **6**(2), P. 38. Doi: 10.3390/risks6020038.

[17] Witkowska, D., *Discrete choice model application to the credit risk evaluation*, International Advances in Economic Research, 2006, **12**(1), P.33-42. Doi: 10.1007/s11294-006-6124-0.

[18] Grace, AM., Williams, SO., *Comparative analysis of neural network and fuzzy logic techniques in credit risk evaluation*, International Journal of Intelligent Information Technologies (IJIIT), 2016, **12**(1), P.47-62. Doi: 10.4018/IJIIT.2016010103.

[19] Huang, C.L., Chen, M.C., Wang, C.J., *Credit scoring with a data mining approach based on support vector machines*, Expert Syst, Appl, 2010, **33**(4), P.847–856. Doi: 10.1016/j.eswa.2006.07.007.

[20] Yurdakul, F., *Macroeconomic modelling of credit risk for banks*, Procedia-Social and behavioral sciences, 2014, **109**, P. 784-93.

[21] Shen, F., Ma, X., Li, Z., Xu, Z., Cai, D., *An extended intuitionistic fuzzy TOPSIS method based on a new distance measure with an application to credit risk evaluation*, Information Sciences, 2018, **428**, P. 105-19. Doi: 10.1016/j.ins.2017.10.045.

[22] Lappas, P.Z., Yannacopoulos, A.N., *A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment*, Applied Soft Computing, 2021, **107**, P.107391. Doi: 10.1016/j.asoc.2021.107391.

[23] Pandimurugan, V., Usha, D., Guptha, MN., Hema, MS., *Random forest tree classification algorithm for predicating loan*, Materials Today: Proceedings, 2021, Doi: 10.1016/j.matpr.2021.12.322.

[24] Machado, M.R, Karray, S., *Assessing credit risk of commercial customers using hybrid machine learning algorithms*, Expert Systems with Applications, 2022, P. 116889. Doi: 10.1016/j.eswa.2022.116889.

[25] García-Céspedes, R., Moreno, M., *The generalized Vasicek credit risk model: A Machine Learning approach*. Finance Research Letters, 2022, P. 102669. Doi: 10.1016/j.frl.2021.102669.

[26] Nazar Aghaei, M., Ghiasi, H., Asgharkhah, M., *Credit risk categories of real customers using collective learning (Case study of Sepah Bank)*, Money-Banking Journal, 2019, **12**(39), P. 129-166.

[27] Del Afrooz, N., Homayoon Far, M., Taghi Poor Tamijani, M., *Credit Risk Management in Banks Using Combined Approach*, Quarterly Journal of Financial Engineering and Securities Management, 2019, **10**(38), P. 94-116. Doi: 20.1001.1.22519165.1398.10.38.5.1.

[28] Khojasteh, G., Daee Karimzade, S., Sharifi Ranani, H., *Credit Rating of Real Customers of the Bank with a Combined Approach of Logistic-Symbolic Regression (Case Study: Ghavamin Bank of Shiraz)*, Resource Management in Police Journal of the Management Dept, 2019, (3), P. 117-48.

[29] Mohaghegh Nia, M., J., Dehghan Dehnavi, M., A., Bayi, M., *The effect of internal and external factors of banking industry on bank credit risk in Iran*, Journal of Financial Economics, 2019, **13**(46), P. 127-144.

[30] Heidari, M., S., Ebrahimi, B., Mohebi, N., *Credit Risk Modeling of Basket Bank Credit Facility Basket using Actuary Modeling*, Journal of Financial science of securities analysis, 2017, **10**(34), P. 55-71.

[31] Salehi, M., Kurd Kanouli, A., *Choosing optimal features to determine the credit risk of bank customers,* Quarterly Journal of Smart Business Management, 2017, **6**(22), P. 129-154.

[32] Gholi Poor, S., Amoozadeh Khalili, H., Haji Aghaei, M., *Credit risk assessment of the real customers of the bank using logistic regression and neural network (case study of Tourism Bank Branch)*, The third World Conference on Management, Economics of Accounting and Humanities at the beginning of the third millennium, Shiraz, in cooperation with the Institute of Higher Education Institute of Allameh Khoyi University, Zarghan University of Research in the Conference, 2016.

[33] Mir Ghafoori, H., Ashoori, Z., Credit risk assessment of bank customers. *Business Management Explorations,* 2015, **7**(13), P. 147-166.

[34] Hoseini, A., Zibaei, M., *Credit Risk Management at the Mamasani Agricultural Bank using the Neural Network Model*, Journal of Agricultural Economics, 2015, **9**(2), P. 103-119 .

[35] Ghodsi Poor, H., Salari, M., Delavari, V., *Credit risk assessment of borrowing companies from the bank using fuzzy hierarchical analysis and high grade neural network*, International Journal of Industrial Engineering and Production Management, 2012, **23**(1), P. 44-54.

[36] Bayazidi, F., Mohamadi, E., Mohamadi, M., *Credit ranking of the real customers of the Mellat bank using data mining methods*, First National Conference on Development of Monetary and Bank Management, Tehran, Permanent Secretariat of Monetary and Bank Management Conference, 2013.

[37] Cantú-Paz, E., Kamath, C., *Combining evolutionary algorithms with oblique decision trees to detect bent-double galaxies*, In Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation III, 2000, **4120**, P. 63-71. Doi: 10.1117/12.403609.

[38] Tang, L., Cai. F., Ouyang, Y., *Applying a nonparametric random forest algorithm to assess the credit risk of the energy industry in China*, Technological Forecasting and Social Change, 2019, **144**, P. 563-72. Doi: org/10.1016/j.techfore.2018.03.007.

[39] Abbasi Astamal M., Rahimi R., *Designing an Expert System for Credit Rating of Real Customers of Banks Using Fuzzy Neural Networks*, Advances in Mathematical Finance and Applications, 2019, **4**(1), P. 89-102. Doi: 10.22034/amfa.2019.577561.1128.

[40] Izadikhah. M., *Improving the Banks Shareholder Long Term Values by Using Data Envelopment Analysis Model*, Advances in Mathematical Finance and Applications, 2018, **3**(2), P.27-41. Doi:10.22034/AMFA.2018.540829.

[41] Izadikhah, M. *Financial Assessment of Banks and Financial Institutes in Stock Exchange by Means of an Enhanced Two stage DEA Model*. *Advances in Mathematical Finance and Applications*, 2021, **6**(2), P. 207-232. Doi: 10.22034/amfa.2020.1910507.1491

[42] Kalantari, N., Rahmatolah Mohammadi Pour, R., Seidi, A., Shiri, A., Azizkhani M., *Fuzzy Goal Programming Model to Rolling Performance Based Budgeting by Productivity Approach (Case Study: Gas Refiner-ies in Iran*, Advances in Mathematical Finance and Applications, 2018, **3**(3), P. 95-107. Doi: 10.22034/AM FA.2018.544952.