Designing a Data Mining Model Based on Predicting Factors Affecting the Improvement of Banking Operations (Case Study: Maskan Bank)

[1]Davoud Ghorbanvatan, [2]Ahmadreza Kasraee*, [3]Mohammad Ali Afshar Kazemei

[1.]department of Industrial management،central Tehran Branch،Islamic azad university،Tehran،Iran
[2.]department of Industrial management،central Tehran Branch،Islamic azad university،Tehran،Iran
[3.]department of Industrial management،central Tehran Branch،Islamic azad university،Tehran،Iran

Abstract
Banks always evaluate various factors to improve their performance. Various research and studies have been carried out to improve the performance of banks. In this study, using data mining technique and combining two algorithms of random forest and Beding method, we have tried to evaluate the criteria for prediction of banks performance. Eight factors of employee productivity, deposit amounts and the number of deposit, the amount of loans and the number of loans facilities, satisfaction of electronic banking services, value added of housing price and percentage of loans for construction of housing have been used to predict the performance of Bank of Housing. The proposed method consists of randomly selected forest blending with Bagging method in 1, 5, 10 and 50 decision trees in each of three bags, suitable performance and appropriate and inappropriate performance, respectively. The results showed that the best result for 10 decision trees with 85.41% accuracy was obtained. The results were also evaluated by the results of other standard data mining algorithms. It was shown that the proposed method had better performance than other algorithms.

Keywords: Data mining; Decision Tree; Bagging Method; Improved Banking Operations.

1. Introduction

The Great Data Age is associated with great opportunities and challenges, nearly all scientific topics experience unpredictably huge volumes and speed of information spillway [1]. In this way, revealing hidden information in large data through data mining techniques has become an emerging trend and the ultimate goal for a wide range of studies [1]. The ability to produce, record and store data has increased greatly in recent years, and the information contained in these data can be very important. Widespread access to large amounts of data and the need to convert these data into knowledge encourages the information technology industry to use data mining [2]. In general, data mining is becoming an important field for many big companies, including the banking industry, and data mining helps bankers make sound decisions [3]. As a data-driven subject, banking has been a popular practice in recent decades of the information science revolution for researchers with data mining skills. Banks have acknowledged that knowledge is the largest new asset rather than financial resources. In addition, the development and prevalence of electronic banking and mobile banking add to the progressive development of

banking information. These continued improvements and rapid increase of access to large data make it one of the most critical tasks of the banking sector to master the tools of large-scale analysis of relevant data.[4] In addition, in the conditions of globalization and tough market competition, banks are trying to gain competitive advantage over each other. Apart from the execution of business processes, creating knowledge-based processes and using it to benefit the bank are becoming a strategic tool for competition [2]. As the main component of the financial system, banks are always faced with different crises, and one of them is the economic crisis. Research results show that a significant amount of donated facilities or banks do not have adequate liquidity due to economic crises in the countries and this indicates the lack of proper crisis management models in the bank network [5]. Therefore, banks are one of those organizations that are in need of the process of managing economic crises due to the nature of their work and on the other hand because of the large amount of data in the bank and the inability to store them and the need for dynamic processing and extraction of information and knowledge contained in the data, the data flow analysis in the bank has been considered very much in recent years [6].

Given that Iran and its banking system has been subject to numerous fluctuations and crises in different fields due to the increase of international sanctions, and the banking system as the most important sector in Iran has also been faced with these crises and is not very ready to handle the economic crisis, it is more than ever necessary to use approaches that are able to increase the dynamic and foresight capability of researchers and also the bank of housing as a government bank which has played a significant role in the economic crisis and the granting of housing and other activities. Reducing bank liquidity and finally reducing investment and granting facilities by the Bank of Housing and in a way affecting banking operations. Therefore, the purpose of this study is to answer this main question: How is the design of data mining model based on factor identification to improve banking operations?

2. Theoretical Foundations

Theoretical literature plays an important role in clarifying the research direction and its applicability. Based on this, the existing literature in data mining field specifies how important data mining is in each field. Data mining approach is based on prediction of factors that affect banking operations. In this paper, data mining and banking literature theories and literatures have been investigated.

2-1. Data Mining

Data mining is an interdisciplinary field that utilizes knowledge of various fields such as database, statistics, machine learning and other fields related to the discovery and extraction of embedded knowledge in large volumes of data. With the rapid growth of computers and its use over the last two decades, almost all organizations have stored large amounts of data in their database. These organizations need to understand these data or to discover useful knowledge of them. In other words, data mining aims to automatically extract patterns in the data [7]. Data mining is an extremely powerful field of carrying out advanced data investigations, as well as incorporating techniques and mechanisms from statistics and machine learning. Business intelligence and advanced analysis applications use data mining data, including data analysis confirmed [8]. Data mining is the process of large data analysis stored in databases for the automated extraction of hidden, unknown, credible,

interesting and practical knowledge such as patterns, abnormalities, associations and changes. Data mining is commonly used in a wide range of areas including marketing, health care, military, environment and education. Also, data mining is becoming increasingly important and necessary for the banking sector, as the volume of data collected by banks has increased significantly, and the need to discover hidden and useful patterns of bank data has increased significantly [9]. Financial analysis of data based on data mining is very important in order to understand the stability and profitability of capital in a business [8]. Data mining is actually a technique that is intended to be solved by that problem. This method works on a large amount of data and performs the required analysis, eventually a series of iterative patterns are extracted which can be a winning leaf, then it's the turn to find relationships between different patterns and finally a series of important solutions to the challenge. Data mining techniques are used to extract hidden patterns and predict future trends and behaviors in financial markets, and advanced statistical, mathematical, and artificial intelligence techniques are usually required to extract such data [8]. Data mining is a powerful science that can penetrate all parts, go through the information, and respond to all questions that arise. This science is so important and so well known in large companies and banks that much of the planning is based on it. Some consider data mining to be equivalent to the common term knowledge discovery of data, and some see it as an essential step in the process of knowledge discovery.

2-2. Data mining and banking

The banking sector is of great importance or value in all countries and everyone uses the banking sector physically and online [10], and due to the increased competitiveness in banking, the relationship between bank and client has become a fundamental strategy for increasing customer satisfaction, and banking systems convert significant amounts of data they collect daily into competitive advantage by data mining [11]. Data mining influences the evolution of banking and the risks associated with financial transactions and costs, replacing traditional banking system with digital banking [12]. In addition, data mining facilitates the process of decision making in banks and reduces the risk of making investment decisions [13]. Bank data mining is important because traditional classification algorithms do not take into account class distribution, often leading to undesirable performance in unbalanced banking data. Therefore, to solve this problem data mining approach can improve banking processes.[9] In addition, banking systems collect huge amounts of data daily, and these data include; Customers' data, details of transactions, details of risks, credit card details, collateral details, compliance and anti-money laundering information, trade financial data, SWIFT and messages, and thousands of decisions made every day by a bank are data that can be used [2]. The main use of data mining in banking has been considered by many researchers and this technique is generally used to predict bank bankruptcy [9, 14, 15], bank customer churn estimation (16), fraud detection [17], and loan approvals [18]. Fraud detection in the banking sector is based on data mining techniques and mass data analysis of past experiences and the probability that scammers can steal from customers and banks [10]. On the other hand, leading banks use data mining tools to classify and profit the customer, rate and credit verification, default payment prediction, marketing, fraudulent transaction identification, etc. [2]. Also classification as one of the most popular data mining techniques in the banking sector for various purposes such as; Forecasting bank customers' collapse, validation, fraud detection, bank failure estimation and telephone marketing prediction are used [9]. Finally, as customers increasingly turn to online banking services, simultaneously the need for banks to data mining also increases [19] and the banking industry around the world is experiencing tremendous changes in the way business is done, and

the banking industry has realized the need for techniques like data mining that can help them to compete in the market [2].

2-3- Evaluation of Banks Performance

Throughout the world, banks and the banking industry in general are one of the most important bases of each country's economy and because of the multiple financial and credit services (microfinance and macro) play a decisive role in the development and economic growth of countries. Banks and financial institutions collect resources by attracting depositors' deposits and provide them to investors in the form of facilities [20]. Banks can provide investment conditions by providing financial capital for the economic sector and increase employment and national production. Banks can also create new job opportunities and better divide incomes at the community level by providing credit to experts. In addition, banks hold cash and facilitate the transfer of that money to protect their personal property, government, and domestic and foreign transactions, and also play an important role in the economic establishment as a factor of monetary policy votes [21].

On the other hand, commercial banking is based on the network of branches that act as a key intermediary between customers and the bank center. Facilities given by bank branches are, in fact, the resources that are granted by attracting various investments through deposits. Also from a marketing point of view, bank branches are a service provider that reflects the extent of the bank's participation and cooperation in understanding the needs of customers and providing the best services. Therefore, banks are under pressure to continue to exist and to compete with other banks in order to increase their share of economic activity in the country, to develop their branches' performance and increase efficiency.

The performance evaluation system of branches represents the amount of customer deposit absorption (resource absorption), cost management (usage) and quality of services. These three scales form a system of performance evaluation of branches which according to the facilities of branches, assets, human resources and... They are considered. However, designing a performance evaluation system in banks is complicated due to the service of these units and their diversity of performance.

Due to these difficult tasks and the more obvious role of banks and credit institutions, in supporting the real sector of the economy and ultimately the development and prosperity of the country, the performance evaluation of banks has become more and more of a focus for authorities and policy makers and a lot of research has been done in this circle [20].

2-3-1- Bank performance evaluation goals

Banks, as the most important financial services institutions, have important and specific reasons for measuring their financial performance. Among the most important of these are measuring the satisfaction of customers and taking feedback from them in the field of services offered, obtaining an appropriate position in the industry by comparing themselves with the higher standard, determining whether the organization has been successful for both their stakeholders and their stakeholders, ensuring that decisions are not based on emotions and assumptions but on the real data, determining the problematic sectors in the organization and providing plans for their solution, specifying the sectors that have the ability to develop and create benefits in the organization [22]. The reasons that drive banks to measure non-financial performance alongside financial performance are the existence of large changes in methods of calculating costs, increased competition in the banking industry, the proper image at national and international levels, constantly changing roles, changing foreign demand, and

increasing unavoidable changes in information technology [23]. Other major reasons include the need for non-financial performance criteria. Such as Competitive Pressure, Technological and Economic Developments, and Legal Regulations among Senior Executives, Political, Social and Economic Status, Corporate Culture and Senior Executives [24].

2-3-2- Bank performance evaluation criteria

Performance measurement can be defined as a system through which banks control their daily operations and evaluate how much they achieve their goals. In order to apply performance measurement task, a set of indicators that reflect the performance of a bank [25]. Banks track their activities for development, especially financial progress, and at the same time value non-financial performance criteria, especially customer satisfaction.

2-4. Effect of Employee Productivity on Bank Performance

Employee productivity depends on the amount of time a person is physically present in a job and the amount of mental presence or efficient performance during the time they are in a job. Companies have to deal with both of these issues in order to maintain high productivity of workers and employees, and this may occur through a variety of strategies that focus on employee satisfaction, health and morale [26].

After the financial and banking crisis of 2008, depositors, lenders to banks and bank asset owners all lost their confidence and sought to save their funds simultaneously with their withdrawal. Often, banking problems have internal causes, such as poor banking oversight, insufficient capital, and political intervention. Profitability ratios show that most commercial banks report a decline in financial performance [27]. Some of these problems, such as poor supervision and incompetence in staff expertise, skills and knowledge experienced by banks, can be resolved through in-service training to improve the performance of banks. Companies with the strongest financial performance often had employees who reported high levels of employee satisfaction. Companies with poor financial performance also had high levels of employee satisfaction [28]. With the current global economic development and the rapid evolution of technology and innovation, organizations are faced with a constant need to learn and develop staff [29]. The resource-based approach claims that the organization can only create sustainable competitive advantage if its activities are created in a unique way of value, which competitors cannot easily copy. Human capital is not easy to copy with the expertise and knowledge required in your workplace.

2-5. Impact of Capital on Banks' Performance

Financial crises raise fundamental issues with the role of bank stock in particular from the point of view of bank survival. It is not surprising that the public protest for more bank capital increases after the financial crisis, and post-crisis reform proposals tend to focus on how capital regulations should be adopted to avoid future crises. A fundamental assumption of these proposals is that there are external factors due to the safety net provided to banks, and thus social efficiency can be improved by requiring banks to operate with more capital, especially during financial crises. However, bankers often argue that holding more capital will compromise their performance and lead to lower lending.

Survival and market share are two key performance issues that concern bank managers. The bank's survival plays a role not only in the strategic decisions made by banks, but also in the decisions regulating banking stability. Market share is an important goal for most companies [30], and banks often evaluate their performance relative to each other on this basis. Knowing how bank capital affects bank performance, both during financial crises and in normal times, is of great importance to regulators who consider microbanking regulations and macro caution.

This effect varies depending on the size of the bank, and the nature of the crisis is important details for regulators that gage the level and other characteristics of the capital needed to achieve the desired level of bank stability. Although the battle for market share is a zero-sum game, it is important for regulators because it affects the behavior of banks. These issues are also important to how the theory of banking evolves, as they help better understand the logic of assumptions about the channels through which the bank's capital affects the various aspects of the bank's performance.

2-6. The effect of loans on banks performance

The loan portfolio is the biggest asset and the biggest source of revenue for banks. Therefore, most banks offer a large portion of the financial resources to customers as loans [31]. Globally, banks lend to customers as a way to increase financial performance [32]. A significant portion of the loans belong to corporate institutions and the rest belong to private individuals. Banks have a better position to participate or dominate the future of microfinance, as banks are able to offer a wide range of products and financial services, including microfinance loans at a relatively different interest rate [33]. This suggests that banks, in addition to customers in the formal banking sector, offer micro-loans to private customers and provide loans to a larger segment of the population. Banks take different forms of evaluation methods to access the credit value of customers, as well as regulatory strategies for the repayment of loans by customers [34]. Methods for bank evaluation and monitoring include initial screening, evaluating loan offers by the credit committee, and monitoring loan repayments. Assessing customers' credit competency as well as monitoring loan repayment is crucial if banks wish to recover loans paid to customers in due time, as a result, many banks have a credit section that performs these basic tasks.

2-7. The Effect of Electronic Banking Services on Banks Performance

Banking is one of the sectors that closely monitored technology progress and is widely used. Developing new products and services is difficult for the banking market, as is the case in many financial markets. Hence, new shapes of existing banking products that are offered to customers are very important. Technology-based applications such as Internet Banking, Mobile Bank, Telephone Bank, ATM and POS Network offer significant benefits to customers in delivering existing products. The most important benefit is for any bank product with low-cost banking operations for 24 hours to 7 days, which reduces dependence on branches. All of these features encourage banks to consider electronic-based services for their customers, which are increasing day by day and are also expected to continue in the future [35].

3- History of research

Rahmani and Mkhen in 2019, using the partial least squares method, showed that the quality of electronic services as a strong determinant in order to confirm and satisfy the customers and eventually lead to the continued use of electronic banking [36]. Karimi and Ghaemi realized in 2018 that financial fraud can be detected by using meta-heuristic algorithms and data mining techniques [37]. Also, Javadi used a tree to classify a customer's decision based on the use of banking services [38]. Mansoori and Farhudi Nejad were able to discover fraud in financial data in 2018 by eliminating data abnormalities [39]. By using a decision tree, Moradzadeh and Khudaidari were able to identify customers who had turned away from electronic banking services in 2018 [40]. Shahabi Moghaddam and Hosseinzadeh Seadi gave a model by regression lines in 2018 to

identify customers loyal to e-banking [41]. In 2018, Golzar and Rafsanjani used data mining techniques to offer appropriate services to customers such as Internet Bank E-Banking [42].

In 2020, Manthoulis et al. in a study considered data mining to extract meaningful information from banking data [43]. In 2020, Birant conducted data mining research in banking as a tool for predicting bank customers collapse, authentication, fraud detection, bank failure estimation and telephone marketing prediction [9]. In 2021 Mateev and Bachvarov investigated the bank's regulation, ownership and performance in the MENA area [44]. Simoens and Vennet

In 2021, they examined the performance of the bank in Europe and the United States during 2007-2017, the results of which showed that low rates for a long period of time may be harmful to the value of bank privileges [45]. Huong et al. in 2021 at the risk of liquidity and performance of the bank in Southeast Asian countries. The results showed that banks performance was affected by the variable of bank performance, the quality of cash assets, bank size, bank capital, loan losses, GDP growth, money supply, and inflation [46]. The SUYANTO in 2021 also investigated the relationship between credit and liquidity on bank performance with the mediating role of capital adequacy. The results of this study showed that the adverse impact of credit and liquidity on bank performance is not significant. Also the high level of credit is associated with low level of bank performance, and the bank's profit is reduced with low interest [47]. HERSUGONDO et al. in 2021 examined the role of non-current assets, capital, adequacy and bankruptcy risk on bank performance. The results showed that non-current assets have a significant negative impact on bank performance, capital adequacy has a significant negative impact on bank performance, bankruptcy risk means that a bank cannot repay its depositors because its debts are higher than its assets, so it has a significant impact on bank performance [48]. Chong et al. in 2022 in a study showed that banks by single data mining will improve decision making processes in banking [19]

According to the literature and the variables extracted from the articles, the following factors can be identified and used to evaluate the performance of the Bank of Maskan. The highest criteria of evaluation in literature and published articles were labor productivity, the amount of money and the number of deposits, the amount of facilities granted, and the number of facilities and electronic banking services. But according to the literature, other factors such as oil and gas price changes affect the performance of banks. Considering the case study of this research that is Bank of Maskan, this shock can enter the performance of Bank of Maskan through housing price and the percentage of loans for housing construction. Therefore, these eight factors are evaluated in this study.

4- Research method

The present study is an applied research considering the objectives and the process of doing the work is a descriptive and field research based on mathematical modeling. In order to collect data, library and field methods will be used. In the library method, the theoretical bases and the literature of the research will be studied. In the field method, the information related to the study of the objectives and questions of the research will be studied. In order to collect data, the database of the Bank of Maskan and the Statistical Center of Iran will be used, as the main data for data mining and determining the model for banking operations in the Bank of Maskan. In the proposed method of this research, an integration of the two algorithms and the proposed method is considered. The beginning of the dataset has different labels that can be used by classification algorithms. Then the number of labels in the data that are 2 labels is created. This is because the number of bags

is chosen because all the labels are possible to be present in the voting and to be considered a separate bag for them. Of course, bags are typically adopted on the number of tags plus one, but since the inside of the bags is supposed to be used for random forest, there is no need to add a bag to determine the vote, because in the decision tree itself and the random forest it happened and eventually a result is returned as output. But according to the above, within each bag, the decision tree or tree is made to test the proposed method, at each bag a decision tree, then five decision trees, at the next step 10 decision trees and finally 15 decision trees are made. In the second to fourth stages, it can be expressed that at each bag a random forest is created with a number of different trees. In this study, the proposed treatment procedure is implemented in Python software and the results are compared together. The comparable parameters in this research are as follows.

- Accuracy:
(1)

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision:
(2)

$$precision = \frac{TP}{TP + FP}$$

- Recall:
(3)

$$recall = \frac{TP}{TP + FN}$$

In the above TP-expression of correct positive data, i.e. data that is correctly identified, TN represents correct negative data, i.e. data that has been properly rejected, FP represents false positive data, which is incorrectly identified, and FN represents negative data, i.e. data that is not incorrectly identified.

5. Results

The data used was initially investigated. In the following figures, eight variables used in this study are evaluated:
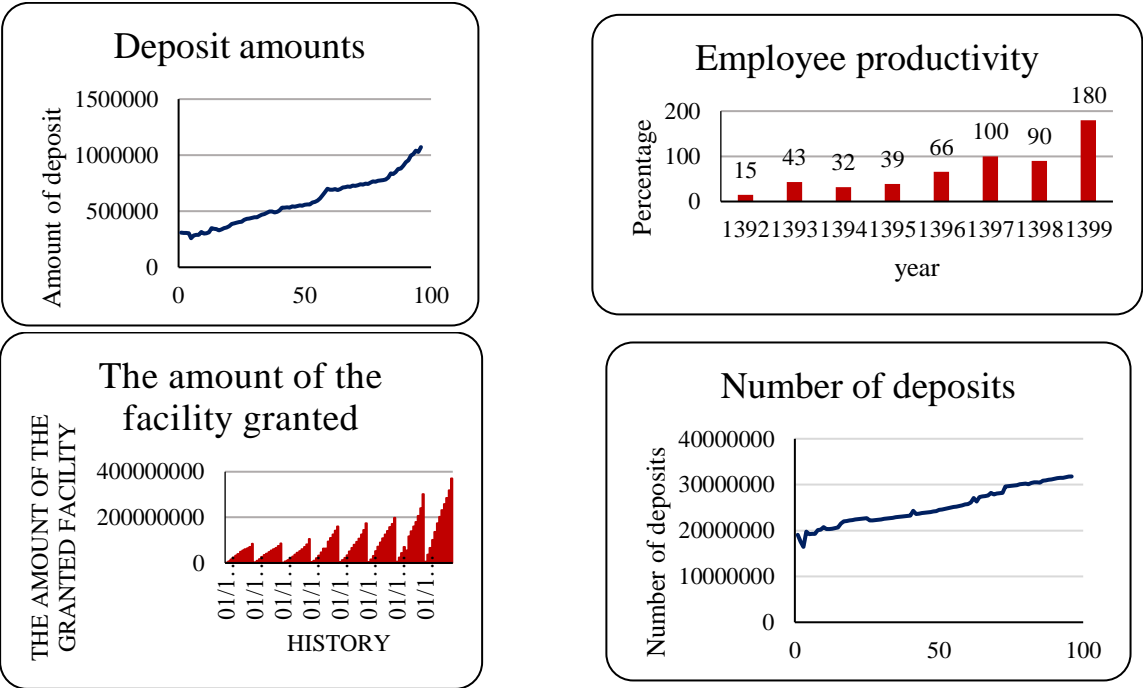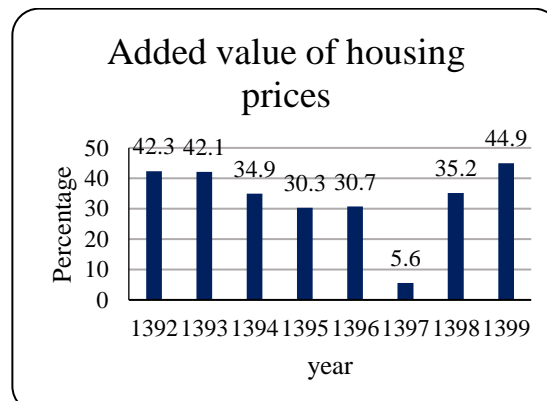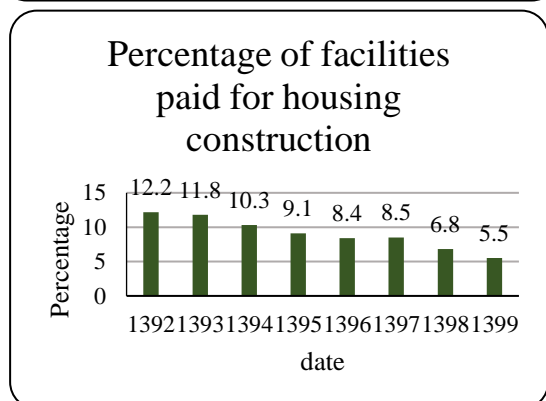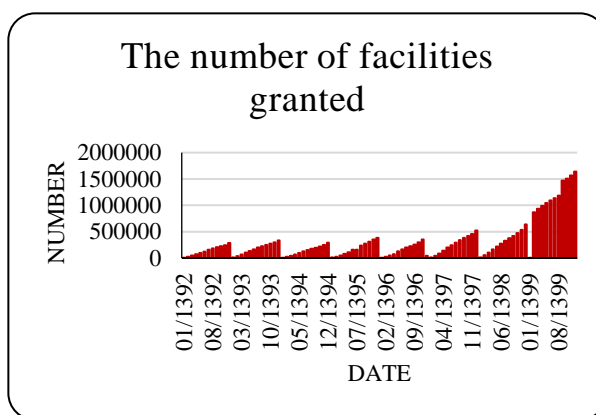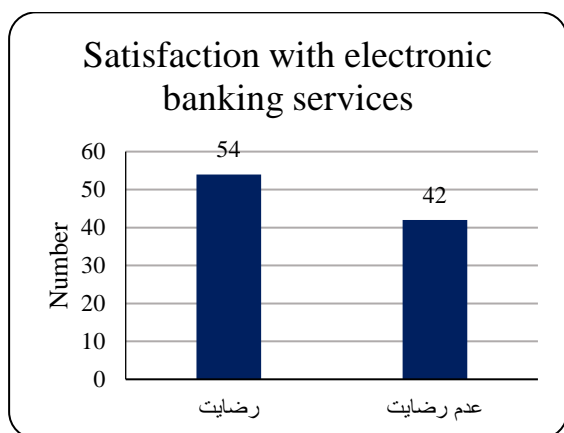
The data used in this study is related to data collected between 2013 and 2020 monthly that were shown in the above diagrams. The results obtained from the data mining technique are then evaluated.

You can start by getting a strong idea about how the startup works. We can create a new instance of the dataset by randomly selecting rows from the dataset and adding them to the new list. We can repeat this task for the fixed number of rows or until the new data set size matches the size ratio of the original data set. By not deleting the row that has been selected, sampling with a replacement is allowed to be available for future choices. To do this, the data is divided into two parts of training and testing data. 20% of the data are used as test data and 80% of the data is used as training data. This choice occurs in total randomness. We will then apply the Random Forest algorithm to the dataset used. First the data set is loaded, the values of the string are converted to numerical, and the output column of the strings is converted to the correct values of 0 and 1, which indicate the proper or non-horse performance. We will use k-fold mutual credit to estimate the performance of the model taught on the data used. This means that we build and evaluate k models and estimate performance as the average error of the model. Classification accuracy will be used to evaluate each model. We will also use adaptive classification and regression trees algorithm (CART) to create bags, including auxiliary functions to divide a dataset into groups, to assess the split point, to find the optimal split. It also has a predictive task function with each decision tree and combining forecasts in a single recursive value. This is achieved by choosing the most common prediction of the forecasts made by decision trees in the bags. Finally, a task function is to create examples of the training data set, train the decision tree about each, then predict the test data set using a list of trees to make a decision in the bags. k=3 value used to validate. Deep trees with a maximum depth of 6 and minimum training rows per node were made of 2. Samples of training data set with 80% of original data set size were created. This was to create variation within the data subsets used to train each tree. The default choice of

the bag is that the size of the sample data set matches the size of the original training data set. A series of 3 different numbers of trees were evaluated to show the behavior of the algorithm. The resolution is equal to each and the average precision for each configuration is printed. As the number of trees increases, it is possible to observe a slightly increased yield rate. According to the above facts, the final result of the formation of the bags and the random forest is shown in Table 1:

Table 1- The results obtained from the data-mining algorithm

Tree1			
Score	78.125	75	84.375
Mean Accuracy		79.167	
Tree5			
Score	81.25	78.125	71.875
Mean Accuracy		77.083	
Tree10			
Score	87.5	81.25	87.5
Mean Accuracy		85.417	
Tree50			
Score	78.125	78.125	93.75
Mean Accuracy		83.333	

As it is clear from Table 1, the proposed method for extracting information from data has the best performance with ten decision trees and has an accuracy of 85.42%. Now according to the obtained results, the obtained values will be compared with other algorithms. The compared algorithms are KNN, LDA and neural network. The evaluated criteria that were discussed in the third chapter are compared in the following.
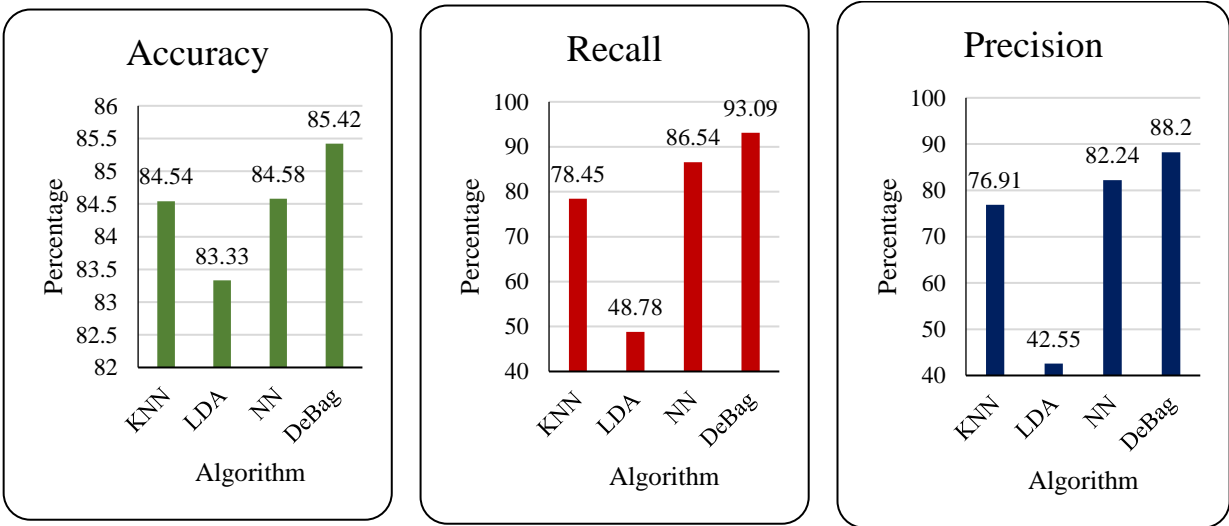

As it is clear from the above figure, the proposed algorithm (DeBag) has good performance compared to other algorithms and no algorithm has performed better than the proposed method. Also, among the different algorithms and the proposed method, the worst performance in the Precision evaluation criterion is related to

Davoud Ghorbanvatan et al.

Designing a Data Mining Model Based on Predicting Factors Affecting the Improvement of Banking Operations (Case Study: Maskan Bank)

the LDA algorithm. In the Recall evaluation criterion, the proposed algorithm has a good performance and the reason for this can be due to the use of bagging method, which returns the final answer according to the amount of voting. But among the existing methods, no algorithm has a better performance than the proposed method. In terms of accuracy, the proposed method has a suitable and acceptable performance compared to other methods. Now, in the table below, the criteria of TP, TN, FP and FN have been examined.

Table 2- Evaluation criteria of the proposed method

Percentage	Criteria
70.16	TP
15.26	TN
9.38	FP
5.2	FN

As it is clear from Table 2, the false positive data criterion has a value of 9.38%, which is a relatively large amount and is considered a big error for evaluating the bank's performance. But the remarkable thing is that as the number of decision trees increases and the random forest is more crowded, the algorithm execution time increases. Also, the performance of the proposed method is different in different bugs, which will be evaluated in the following. First, a comparison will be made between the accuracy of detection in different bugs, which consist of 3 tags applied to the data.
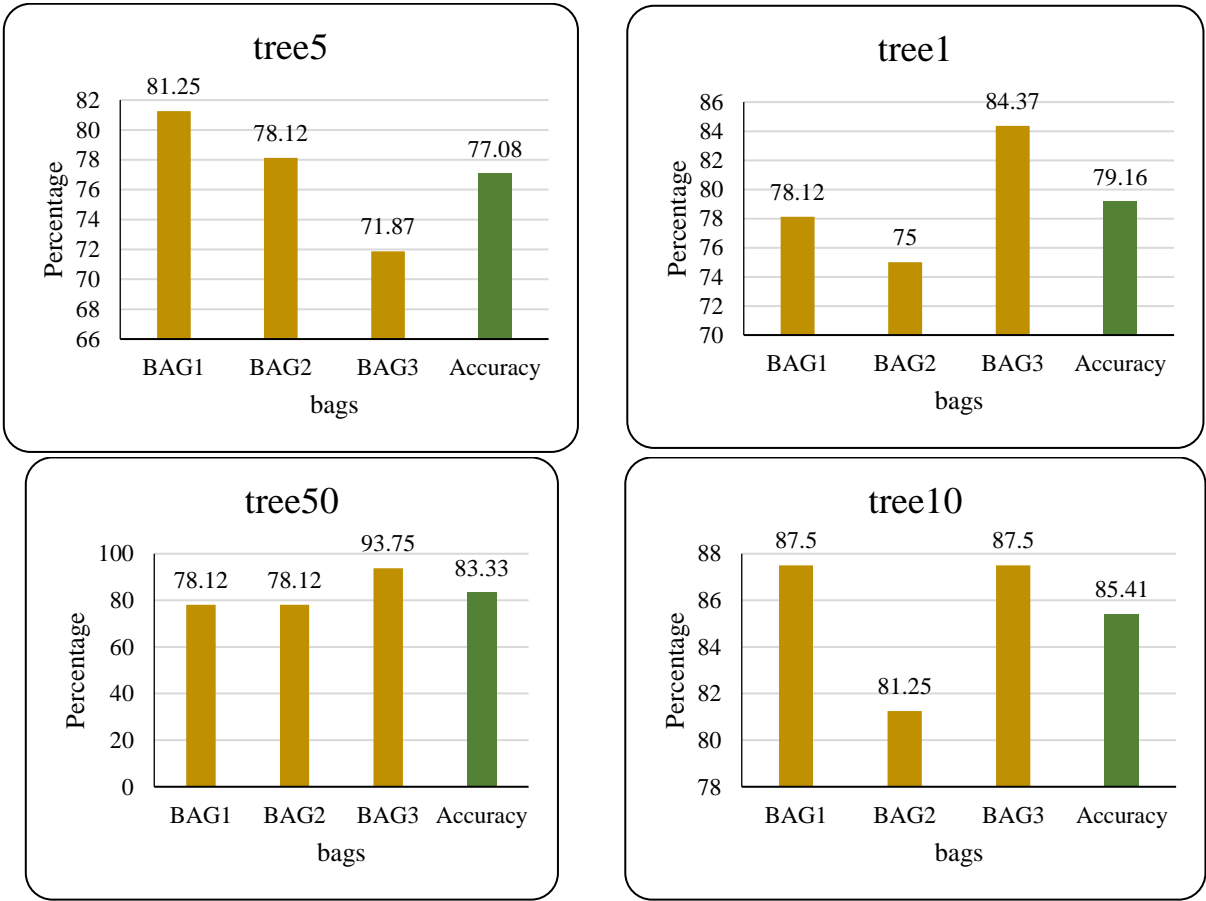
According to the above figure, for a decision tree Bag3 has the highest accuracy in detecting and selecting as the final output, and the worst measure of accuracy is Bag2. For the five decision trees, Bag1 has the best performance and the most choice made in the final vote. The worst performance is also Bag3. For ten decision trees, Bag3 and Bag1 have the highest accuracy in diagnosis and selection as final output, and the worst accuracy is Bag2. Ultimately, for the 50 decision trees, the Bag3 has the best performance and the most choice made in the final vote. The worst performance is also related to Bag2 and Bag1.

6- Discussion and conclusion

Today, banking systems collect data at a higher rate by increasing the number of channels: Internet banking, telephone banking, retail banking, mobile banking, automation, and rapidly automated banking, and bank data is now generated from a variety of sources, including bank account transactions, credit card details, loan requests, and text messages. Hence, data mining can use this collected bank data to derive meaningful information so that banking institutions can make better decision-making processes.[43] In this regard, the purpose of this study was to use data mining technique to assess and predict the performance of the bank. For this purpose, combining the two techniques of random forest and Bagging method was used. For Beging's Three Bag 1: Fit, Bag 2:Inappropriate and Bag 3: Suitable and Inappropriate. This method is in voting, i.e., at the end of the process, an answer is displayed as the output that is reserved for the most votes. The reason for using a third bag is also for this purpose, so that the answer to be voted on as output can be the maximum number of votes. For this purpose, at each stage a number of decision trees were put in each bag. In the first step, one decision tree per bag, in the second phase each decision five trees at each bag, in the third phase each decision ten trees and in the last step fifty each bag made the decision. The highest accuracy of Bag Three was obtained with 84.37% accuracy and the total accuracy of the method was 79.16% in one decision tree. Combining Beding method with five decision trees, the highest precision was related to Bag I with 81.25% accuracy, and the total accuracy of the method with five decision trees was 77.03%. The overall accuracy of the method with ten decision trees was 87.5%. Also, the total accuracy of the method with ten decision trees was 85.41%. The highest accuracy of Bag Three was 93.75%, and the total accuracy of the method with 50 decision trees was 83.33%. Finally, it can be argued that the combination of Begg's method with ten decision trees had the best performance among all the proposed methods.

References

[1] H. Hassani and E. S. Silva, "Big Data: a big opportunity for the petroleum and petrochemical industry," *OPEC Energy Review,* vol. 42, no. 1, pp. 74-89, 2018.

[2] S. Pulakkazhy and R. S. Balan, "Data mining in banking and its applications-a review," 2013.

[3] R. Farooqi and N. Iqbal, "Effectiveness of Data mining in Banking Industry: An empirical study," *International Journal of Advanced Research in Computer Science,* vol. 8, no. 5, pp. 827-830, 2017.

[4] H. Hassani, X. Huang, and E. Silva, "Digitalisation and big data mining in banking," *Big Data and Cognitive Computing,* vol. 2, no. 3, p. 18, 2018.

[5] E. K. Ozyirmidokuz, K. Uyar, and M. H. Ozyirmidokuz, "A Data Mining Based Approach to a Firm's Marketing Channel," *Procedia Economics and Finance,* vol. 27, pp. 77-84, 2015.

[6] H. Haleh, A. Ghaffari, and A. K. Meshki, "A combined model of MCDM and data mining for determining question weights in scientific exams," *Applied Mathematical Sciences,* vol. 6, no. 4, pp. 173-196, 2012.

Davoud Ghorbanvatan et al.

Designing a Data Mining Model Based on Predicting Factors Affecting the Improvement of Banking Operations (Case Study: Maskan Bank)

[7]   J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[8]   N. Bilotta and S. Romano, *Tech Giants in Banking: The Implications of a New Market Power*. JSTOR, 2019.

[9]   D. Birant, "Data Mining in Banking Sector Using Weighted Decision Jungle Method," in *Data Mining-Methods, Applications and Systems*. IntechOpen, 2020.

[10]   R. Rambola, P. Varshney, and P. Vishwakarma, "Data mining techniques for fraud detection in banking sector," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 2018, pp. 1-5: IEEE.

[11]   A. I. Pascu, "DATA MINING. CONCEPTS AND APPLICATIONS IN BANKING SECTOR," *Annals of Constantin Brancusi'University of Targu-Jiu. Economy Series,* no. 1, 2018.

[12]   M. Laboure and N. Deffrennes, "Democratizing Finance," in *Democratizing Finance*. Harvard University Press, 2022.

[13]   I. G. N. N. Mandala, C. B. Nawangpalupi, and F. R. Praktikto, "Assessing credit risk: An application of data mining in a rural bank," *Procedia Economics and Finance,* vol. 4, pp. 406-412, 2012.

[14]   P. Carmona, F. Climent, and A. Momparler, "Predicting failure in the US banking sector: An extreme gradient boosting approach," *International Review of Economics & Finance,* vol. 61, pp. 304-323, 2019.

[15]   Z. Jing and Y. Fang, "Predicting US bank failures: A comparison of logit and data mining models," *Journal of Forecasting,* vol. 37, no. 2, pp. 235-256, 2018.

[16]   A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," *Financial Innovation,* vol. 2, no. 1, pp. 1-13, 2016.

[17]   F. Lv, J. Huang, W. Wang, Y. Wei, Y. Sun, and B. Wang, "A two-route CNN model for bank account classification with heterogeneous data," *Plos one,* vol. 14, no. 8, p. e0220631, 2019.

[18]   W. Jie, Y. Zeng-lei, Y. Dong-hui, L. Jiao, L. Zhi, and L. Jinfu, "Predicting non performing loan of business bank with data mining techniques," *International Journal of Database Theory and Application,* vol. 9, no. 12, pp. 23-34, 2016.

[19]   S. Chong, J. Chin, and W. Loh, "Data mining in retail banking: A case study of branch classification," in *AIP Conference Proceedings*, 2022, vol. 2465, no. 1, p. 030006: AIP Publishing LLC.

[20]   A. Abdali, Sh. Dosti, and A. Nikbakht, "Designing a model for evaluating the efficiency and grading of commercial bank branches using the data coverage analysis technique," Bank and Economy vol. 99, 1388.

[21]   Dadger, Yadaleh, and N. Nemat, "The use of DEA model in the evaluation of the efficiency of economic units; a case study of the supervision of Tejarat Bank," two-quarterly Journal of Economic Studies, vol. 4, no. 7, pp. 11-54, 2007.

[22]   C. Parker, "Performance measurement," *Work study,* 2000.

[23]   A. Neely, "The performance measurement revolution: why now and what next?," *International journal of operations & production management,* 1999.

[24]   M. Hussain and A. Gunasekaran, "An institutional perspective of non-financial management accounting measures: a review of the financial services industry," *Managerial Auditing Journal,* 2002.

[25]   H.-Y. Wu, G.-H. Tzeng, and Y.-H. Chen, "A fuzzy MCDM approach for evaluating banking performance based on Balanced Scorecard," *Expert systems with applications,* vol. 36, no. 6, pp. 10135-10147, 2009.

[26]   G. Ron and O. Ronald, "Health Productivity Management Assists Benefits Business Strategy"," 2002.

[27]   A. M. Kithinji and N. M. Waweru, "Merger restructuring and financial performance of commercial banks in Kenya," *Journal of Economics, management and financial markets,* vol. 2, no. 4, pp. 10-32, 2007.

[28]   P. Bruce and K. Ira, "The Hidden Human Resource: Shareholder Value Finding the Right Blend of Rewards," 2002.

[29]   R. A. Swanson, E. Holton, and E. F. Holton, *Foundations of human resource development*. Berrett-Koehler Publishers, 2001.

[30]   P. Aghion and J. C. Stein, "Growth versus margins: Destabilizing consequences of giving the stock market what it wants," *The Journal of Finance,* vol. 63, no. 3, pp. 1025-1058, 2008.

[31]   J. P. Bonin and Y. Huang, "Dealing with the bad loans of the Chinese banks," *Journal of Asian Economics,* vol. 12, no. 2, pp. 197-214, 2001.

[32]   S. Ongena, V. Roşcovan, W.-L. Song, and B. J. Werker, "Banks and bonds: the impact of bank loan announcements on bond and equity prices," *Journal of Financial Management, Markets and Institutions,* vol. 2, no. 2, pp. 131-156, 2014.

[33]   J. Asantey and S. Tengey, "An empirical study on the effect of Bad Loans on banks' lending potential and financial performance: The case of SMEs lending in Ghana," *International Journal of Research in Business Management,* vol. 2, no. 11, pp. 1-12, 2014.

[34]   G. Pollio and J. Obuobie, "Microfinance Default Rates in Ghana: Evidence from Individual-Liability Credit Contracts: SME financing in frontier markets in Africa," *SME Financing in frontier markets in Africa,* no. 20, pp. 8-14, 2010.

Davoud Ghorbanvatan et al.

Designing a Data Mining Model Based on Predicting Factors Affecting the Improvement of Banking Operations (Case Study: Maskan Bank)

[35] I. Akhisar, K. B. Tunay, and N. Tunay, "The effects of innovations on bank performance: The case of electronic banking services," *Procedia-Social and Behavioral Sciences,* vol. 195, pp. 369-375, 2015.

[36] M. Rahmani and R. Mobarhan, "Modelling the continuous use of electronic banking (case study: Maskan Bank)," presented at the 5th International Conference on Web Research, Tehran, University of Science and Culture, 2018.

[37] c. Karimi and R. Qaemi, "Investigation and analysis of classification techniques in data mining to detect financial fraud in bank transactions," presented at the 15th conference of modern researches in science and technology, in electronic form, Asman Science and Technology Company, 2017.

[38] Q. Javadi, "Electronic banking and data mining," presented at the International Conference on Technology and Innovation in Science, Engineering and Technology, Tehran, Arvin Alborz Conference Company, 2017.

[39] N. Mansouri and A. Farhoudi, "Fraud detection method in electronic payment systems using data mining algorithms," presented at the third national conference in computer engineering, information technology and data processing, Tehran, Payam Noor University, 2017.

[40] Z. Moradzadeh and B. Khodayari, "Presentation of a model for predicting customer reversion to electronic banking services using the decision tree method and data mining," presented at the 4th International Conference on Management, Entrepreneurship and Economic Development, Takestan, Takestan Institute of Higher Education, 2017.

[41] M. Sh. Moghadam and E. A. H. Zadeh, "Analysis of rules extracted from data mining in loyal customers to electronic banking with the help of regression," presented at the 3rd National Conference on New Ideas in Computer Electrical Engineering, Jiroft, Gostaran Conference, 2017.

[42] c. Golzar and Sh. Rafshanjani, "Investigation of the provision of appropriate services to bank customers in the field of internet banking and electronic banking using data mining algorithm (case study of the Bank of Agriculture)," presented at the second international conference on management, accounting and economics in sustainable development, Mashhad , Knowledge-based Cooperative Institute, Kemarash, 2017.

[43] G. Manthoulis, M. Doumpos, C. Zopounidis, and E. Galariotis, "An ordinal classification framework for bank failure prediction: Methodology and empirical evidence for US banks," *European Journal of Operational Research,* vol. 282, no. 2, pp. 786-801, 2020.

[44] M. Mateev and P. Bachvarov, "Regulation, ownership and bank performance in the MENA region: Evidence for Islamic and conventional banks," *Emerging Markets Review,* vol. 47, p. 100789, 2021.

[45] M. Simoens and R. Vander Vennet, "Bank performance in Europe and the US: A divergence in market-to-book ratios," *Finance Research Letters,* vol. 40, p. 101672, 2021.

[46] T. T. X. Huong, T. T. T. Nga, and T. T. K. Oanh, "Liquidity risk and bank performance in Southeast Asian countries: a dynamic panel approach," *Quantitative Finance and Economics,* vol. 5, no. 1, pp. 111-133, 2021.

[47] S. SUYANTO, "The effect of bad credit and liquidity on bank performance in Indonesia," *The Journal of Asian Finance, Economics and Business,* vol. 8, no. 3, pp. 451-458, 2021.

[48] H. HERSUGONDO, N. ANJANI, and I. D. Pamungkas, "The Role of Non-Performing Asset, Capital, Adequacy and Insolvency Risk on Bank Performance: A Case Study in Indonesia," *The Journal of Asian Finance, Economics and Business,* vol. 8, no. 3, pp. 319-329, 2021.