

Bank Transaction Analyze for Recognize of Money Laundering Using Decision Tree Algorithm (DTA)

Somayeh DadashzadehFahim (✉ sfahim81@aol.com)

IAU: Islamic Azad University

Ramin Karimi

Islamic Azad University

Zahra Faghihrasoul

University of Zahrah

Research Article

Keywords: Bank fraud, Money laundering, Data mining, Decision Tree Algorithm (DTA)

Posted Date: July 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-685083/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Bank Transaction Analyze for Recognize of Money Laundering Using DTA One of the important problems of the banking systems is illegal transactions based on fraud and money laundering that can destroy the economy and financial foundation of a country. Fraud and money laundering are used to escape from tax payment or inject dirty money to the economy cycle. The offenders use lots of transactions to show their illegal funds rightful. One important problem of fraud and money laundering recognize in banking system is high complexity of it. We need to use knowledge discover methods like learning machine and data mining. We can mostly recognize the hidden pattern of illegal transactions by using various learning machines or data mining like Bayesian Network, DT, Support Vector Machine or Artificial Neural Network. To recognize the pattern and classify the banking transactions, ML methods need to classify in two categories Normal and Abnormal to find suitable features for increase their accuracy. Actually, the feature selection is too important in bank fraud recognition and it is a kind of optimization too because the suitable feature selection causes recognize error go down by techniques like DT. In this research we use DTA to select important features related with bank fraud to decide and knowledge discovery which has a good speed in discover of bank fraud. When the features of bank fraud select in right way and with good accuracy then the accuracy of DT will be increased and we must select those features that are more important to get better accuracy in recognition of fraud and money laundering.

Dta (Decision Tree Algorithm)

One of the most important problems in the banking system is the existence of a variety of illegal transactions based on money laundering and fraud that can cause significant damage to the financial infrastructure and economy of a country. Fraud and money laundering are used to tax evasion or dirty money into the economic cycle, and offenders make numerous transactions attempt to legitimately state their illegal capital. One of the most important problems in detecting fraud and money laundering in the banking system is its high complexity that it is needed to use knowledge discovery methods such as machine learning and data research for this purpose. By machine learning or data-mining methods such as Bayesian network, decision tree, support vector machine, and artificial neural network, the hidden pattern of illegal transactions can be largely identified. Machine learning methods for identifying the pattern and proper classification of bank transactions need two normal and abnormal categories that consider appropriate features for learning to increase their accuracy. Feature selection is an important and key issue in the detection of bank fraud and is also considered as an optimization problem because choosing the appropriate features in the detection of bank fraud make it possible to detect them by techniques such as the minimal decision tree. In this papers, to select the important and key features related to bank fraud, the decision tree algorithm, which has a tree structure, is used for decision making and knowledge discovery and has a high speed for discovering bank fraud. However, its accuracy depends on the selection of important features related to bank transactions. In other words, the accuracy of the decision tree in the detection of bank fraud increases when its characteristics are selected correctly

and accurately and features are used for learning that is more important to identify the pattern of fraud and money laundering with high accuracy.

Introduction

Nowadays, most financial and banking services are provided through the Internet, and in other words, e-commerce desperately needs web banking to advance its goals. In e-banking, money transfer services are done through the Internet with great ease [1] and this issue has led to a high level of acceptance of banking services on the Internet and the volume of banking transactions are increasing day by day. Despite the variety of e-banking services [2], this service has various challenges such as fraud and money laundering that limit its use. These challenges can be addressed by using different machine learning methods. One method of machine learning is the decision tree, which uses a tree structure to make decisions and discover knowledge [3] and has a high speed to detect bank fraud. However, its accuracy depends on the choice of important features associated with banking transactions. In other words, the accuracy of the decision tree in detecting bank fraud increases [4] when its features are selected correctly and with high accuracy and features are used for learning that are more important. The purpose of this article is to increase the accuracy of knowledge discovery methods such as decision tree using group intelligence methods and increase trust in electronic banking and combat tax evasion [5].

In this paper, it is shown that with the help of a technique such as decision tree, the average error of distinguishing money laundering transactions from normal can be reduced.

Decision Tree

The decision tree is a supervised learning method used to classify and identify patterns, and in this method data mining of tree structures is used to information analyze.

In this simple form, the decision tree for classifying information defines a set of rules on its nodes, according to which the tree is scrolled and each data or instance is attributed to a sheet, which is a class number, based on the attribute used, is mapped. A specimen scrolls left or right under a tree according to a rule to meet the leaves and place them in its class.

Recommended Flowchart

The recommended flowchart steps for selecting important features in detecting bank fraud and classifying transactions based on the decision tree mechanism based on grasshopper optimization algorithm are shown in Figure 1:

In the recommended flowchart create a population of attribute vectors based on data related to banking transactions which is considered as a grasshopper and evaluated by the objective function. There are two layers in recommended flowchart that in the layer of feature selection algorithm is done by the grasshopper optimization algorithm then, by selection of features and delivering them to the second layer,

which is the decision tree, a classification of transactions is performed and the amount of error in detecting legal and illegal transaction is reported and again, the grasshopper optimization algorithm selects other features to recalculate the model error rate, and the algorithm tries to reduce its error rate by selection important features as much as possible.

SIMULATION REVIEW RESULTS

To analyze the recommended method, its accuracy needs to be compared with several data mining methods in detecting bank fraud. The Bayesian network method, Decision tree, Support Vector Machine and Artificial Neural Network are practical methods for classifying and detecting bank fraud that can be compared with the recommended method. Weka application software can apply classification methods on bank fraud data, several outputs of these techniques can be seen in the figures (11-4), (12-4), (13-4), (14-4) to detect bank fraud in Weka.

According to the above outputs, it can be concluded that among the Bayesian

Network Techniques, Decision Tree, Support Vector Machine and Artificial Neural Network the highest accuracy is related to Bayesian network and the lowest accuracy is related to the backup Vector technique:

The analysis of the graphs shows that the lowest sensitivity is related to the Bayesian Network and the highest sensitivity is related to the Support Vector Machine and the Artificial Neural Network.

In the next section, by selecting the appropriate features, an attempt is made to increase the sensitivity and accuracy of the recommended method in detecting bank fraud and to increase the efficiency of the decision tree for detecting fraud.

To analyze the recommended algorithm in distinguishing fake transactions from legal ones based on feature selection and error reduction, it is necessary that the variables and parameters related to the grasshopper optimization algorithm are adjusted and quantified according to Table (1), which are based on the values used in the recommended basic article has been.

Comparison of error, accuracy and sensitivity index between the recommended method with the decision tree and the grasshopper feature selection and the decision tree method are shown in figures 10,11 and 12 respectively:

Conclusion

According to the experiments, it can be seen that error average of the decision tree is about 0.542 and this amount of error for the recommended method is equal to 0.426 and the recommended method was able to reduce error by 18.70%.

The analysis of the accuracy and sensitivity index also shown that the decision tree has an accuracy of 70,63 while the accuracy index in the recommended method is 20,73 and its sensitivity is 20,77 and in other words, the recommended method improves the accuracy of the decision tree, index in detecting fraud by about 21,19% and its sensitivity improved index by 21%.

Our analysis and evaluation shows that the accuracy of the recommended method has the highest possible value then it is in the second place of Bayesian network technique and the worst value of the accuracy index is related to the Support Vector Machine.

On the other hand, the sensitivity index analysis shows that the maximum index is related to the recommended method and the lowest is related to the Bayesian network.

References

1. Chen J, Tao Y, Wang H, Chen T. Big data based fraud risk management at Alibaba. *The Journal of Finance Data Science*. 2015;1(1):1–10.
2. Tsang S, Koh YS, Dobbie G, Alam S. Detecting online auction shilling frauds using supervised learning. *Expert Syst Appl*. 2014;41(6):3027–40.
3. CheHashim R, Mahdzan NS. Fraud in letter of credit transactions: The experience of Malaysian bankers. *International Journal of Law Crime Justice*. 2014;42(3):224–36.
4. Dal Pozzolo A, Caelen O, Borgne L, Waterschoot YA, S., & Bontempi G. Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Syst Appl*. 2014;41(10):4915–28.
5. Sahin Y, Bulkan S, Duman E. A cost-sensitive decision tree approach for fraud detection. *Expert Syst Appl*. 2013;40(15):5916–23.

Tables

Parameters	Amount of
Population size	5 Times
Repeat size	30 Times
Cmax	1 Time
Cmin	0.004 Times
f	0.5 Times
l	1.5 Times
Number of basic features	23 Times

Table 1 – Parameters of the recommended algorithm

Figures

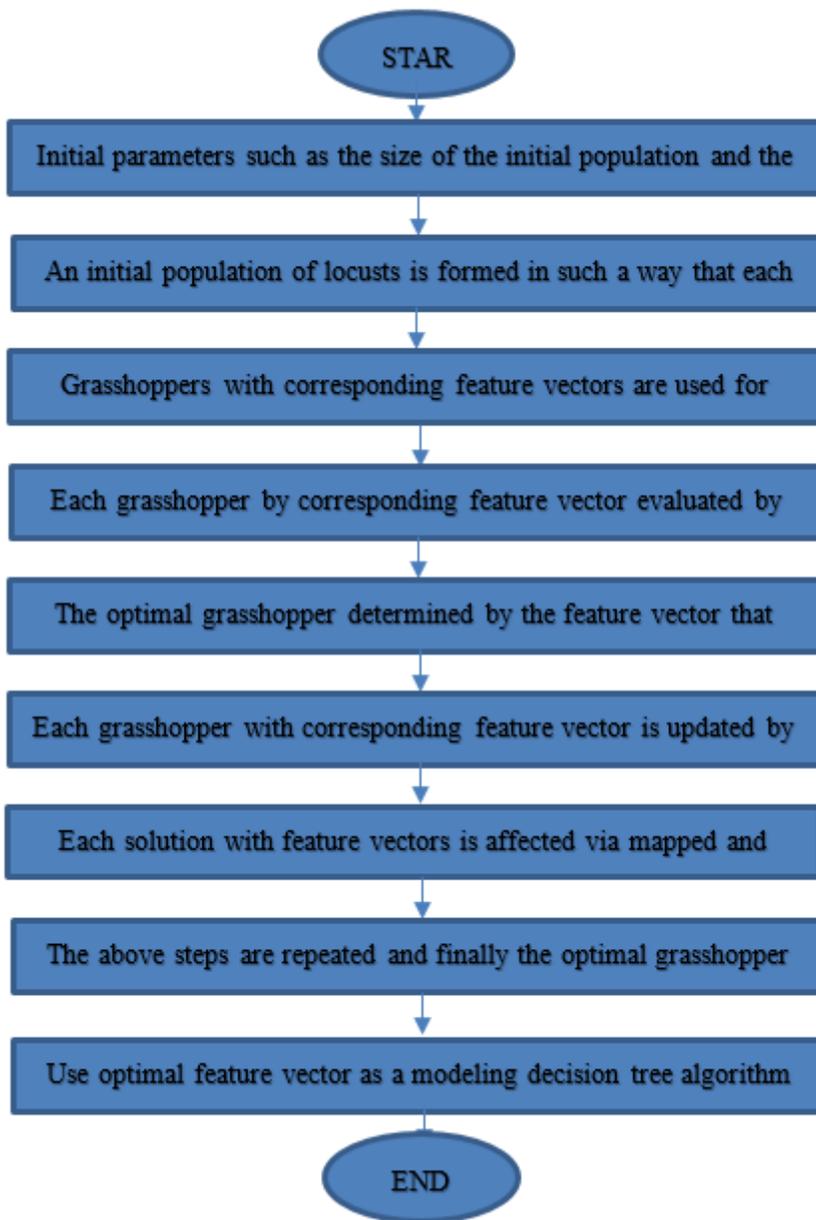


Figure 1

recommended flowchart for detecting fraud and money laundering in the banking system

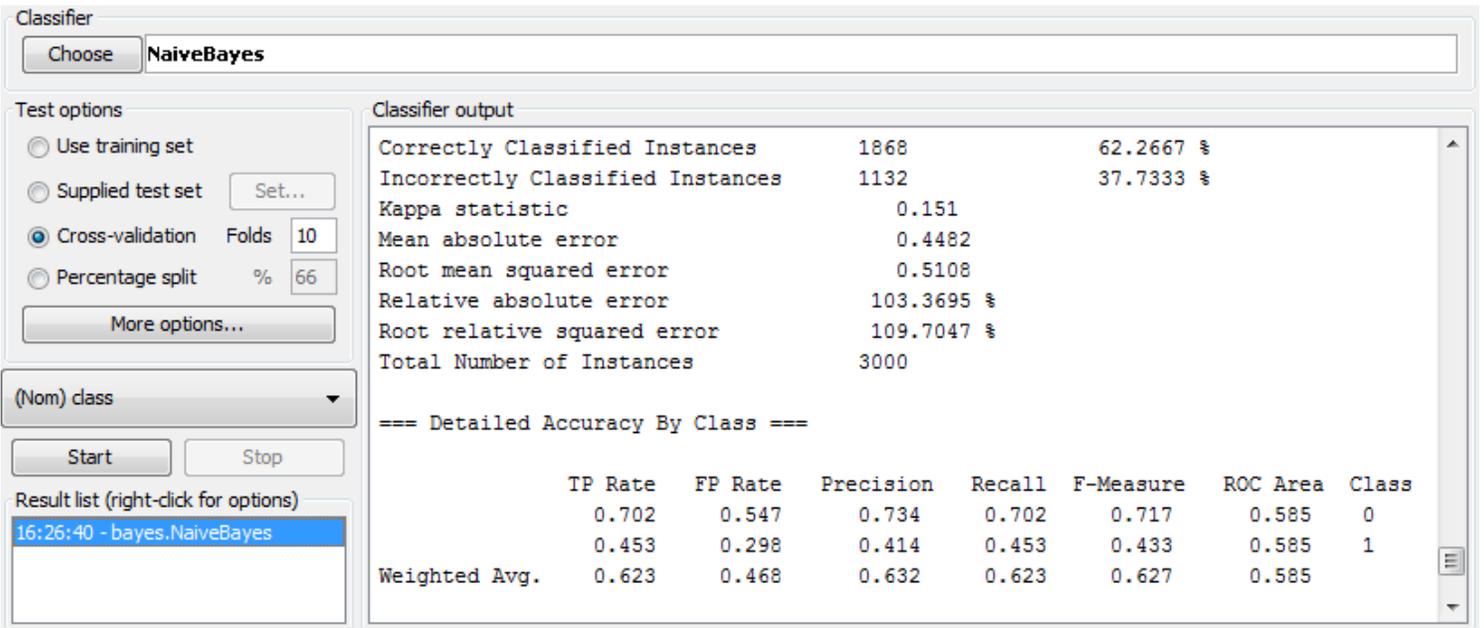


Figure 2

Detection of bank fraud by the Bayesian Network

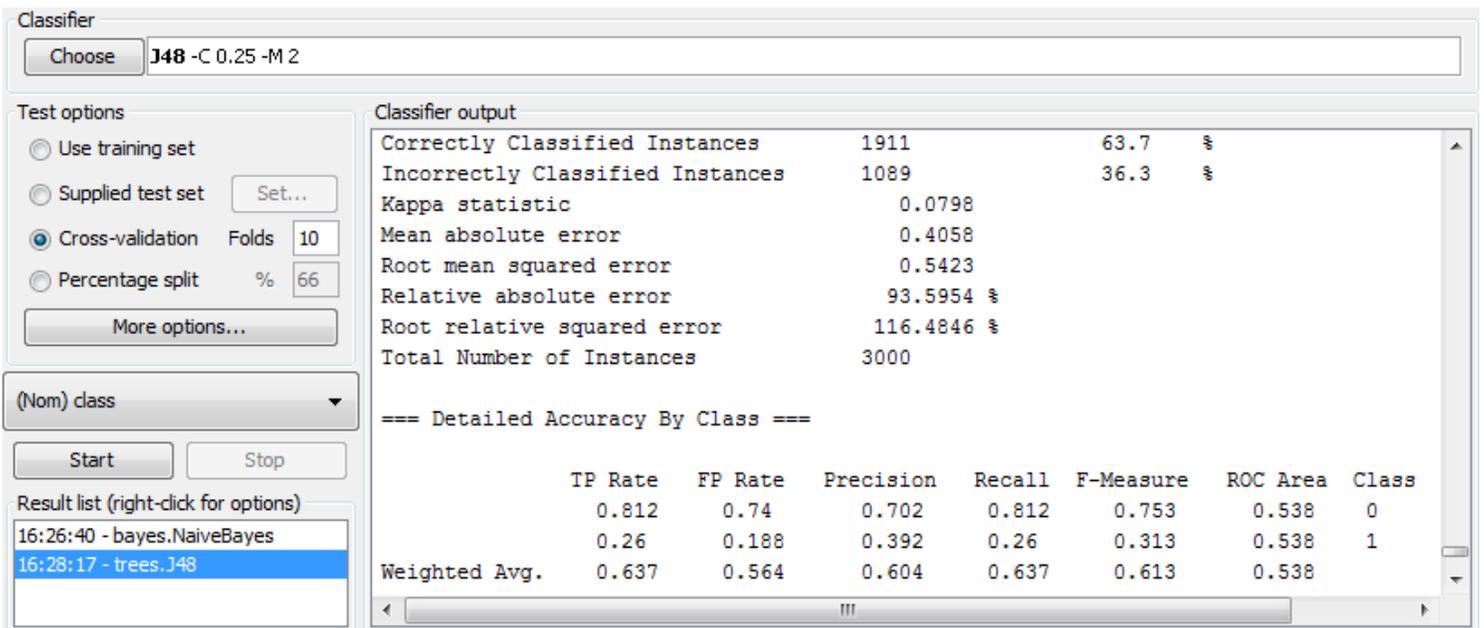


Figure 3

Detection of Bank Fraud by the Decision Tree

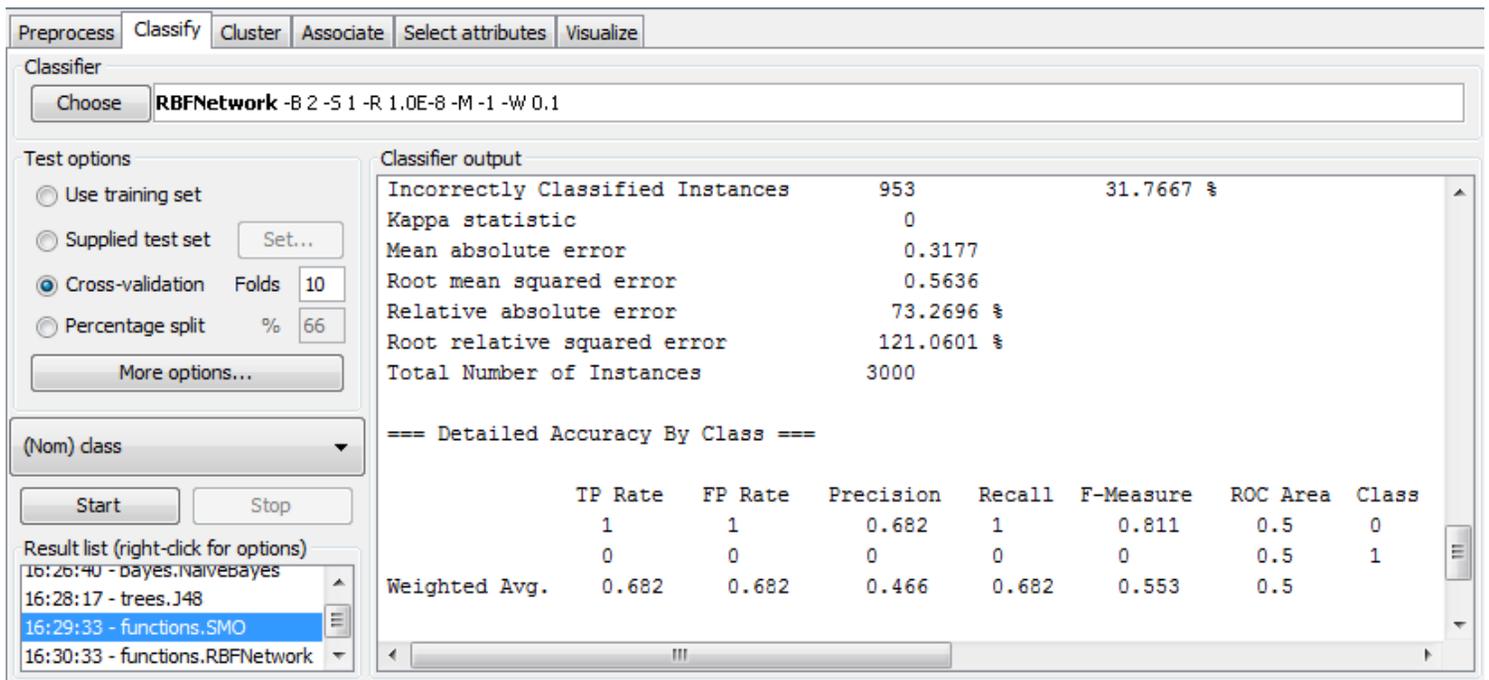


Figure 4

Detection of Bank Fraud by backup Vector Machine

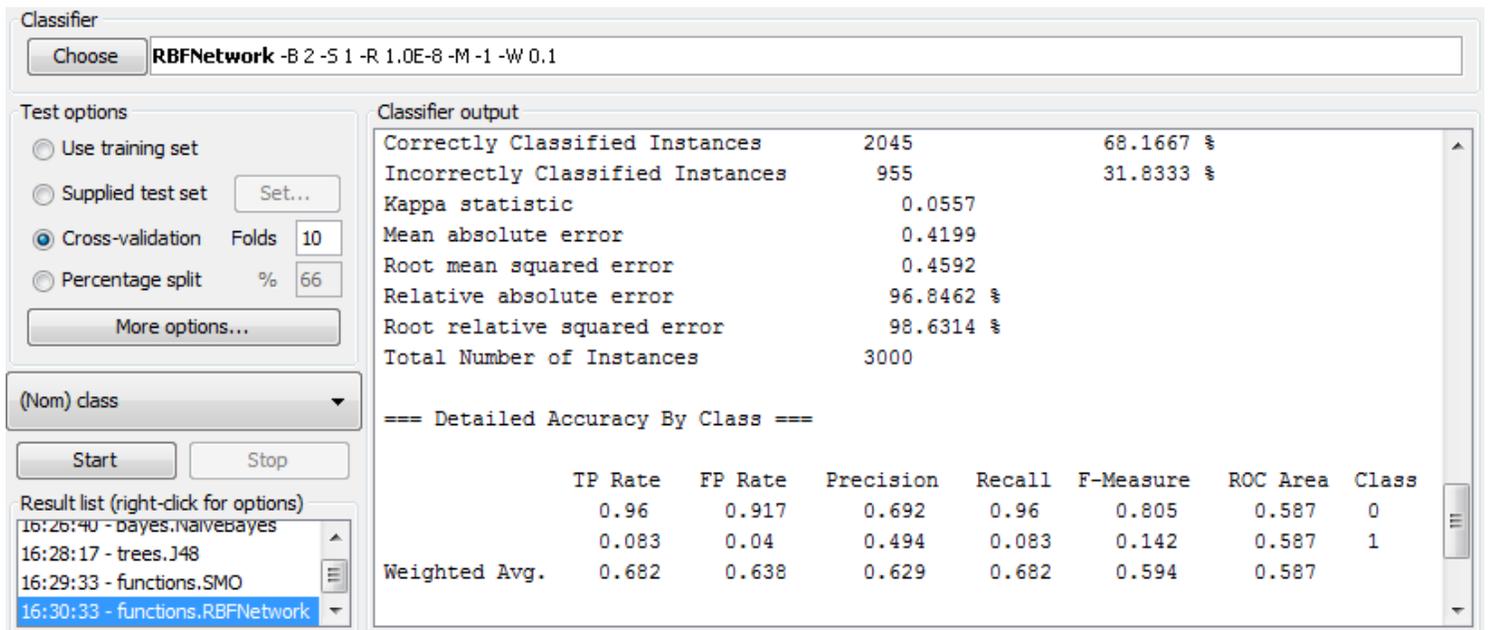


Figure 5

Detection of banking fraud by Artificial Neural Network

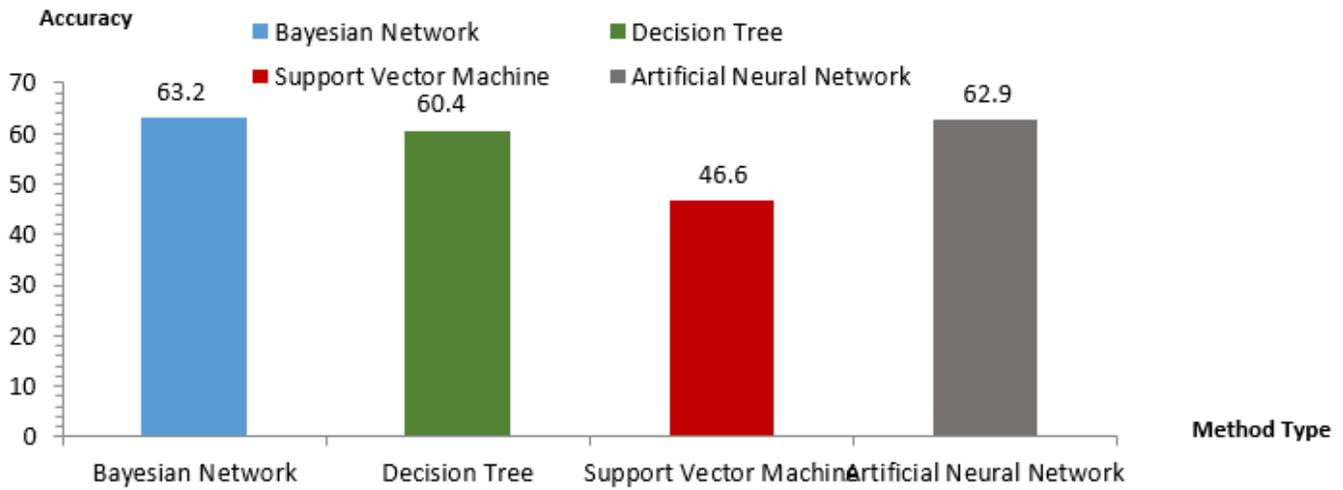


Figure 6

Comparison of the accuracy index of learning methods in detecting fraud bank

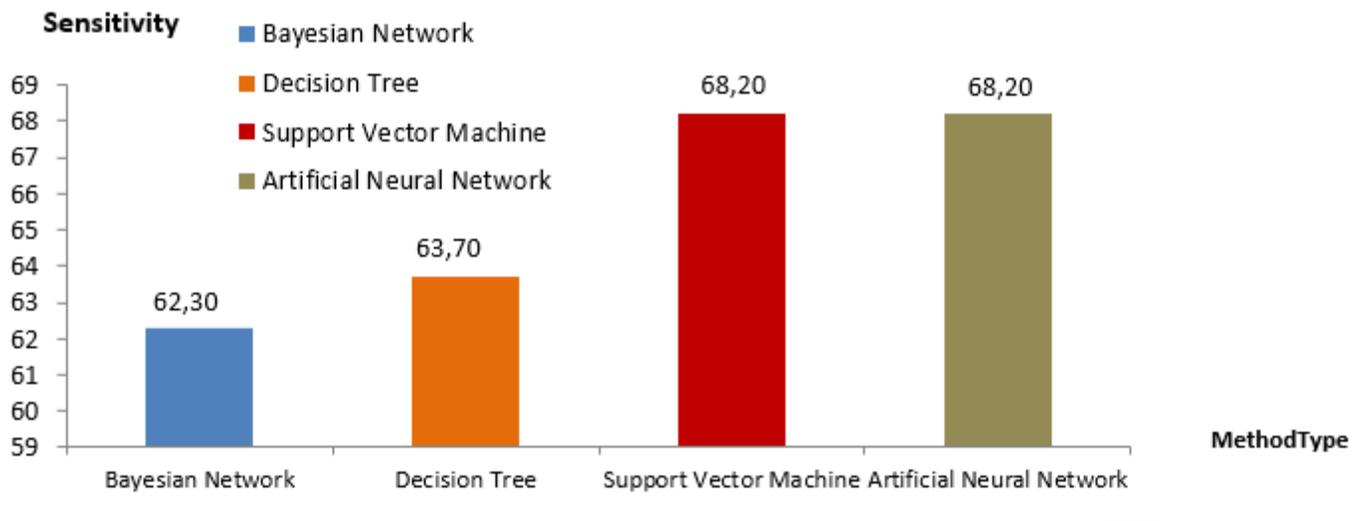


Figure 7

Comparison of the Sensitivity index of learning methods in detecting banking fraud

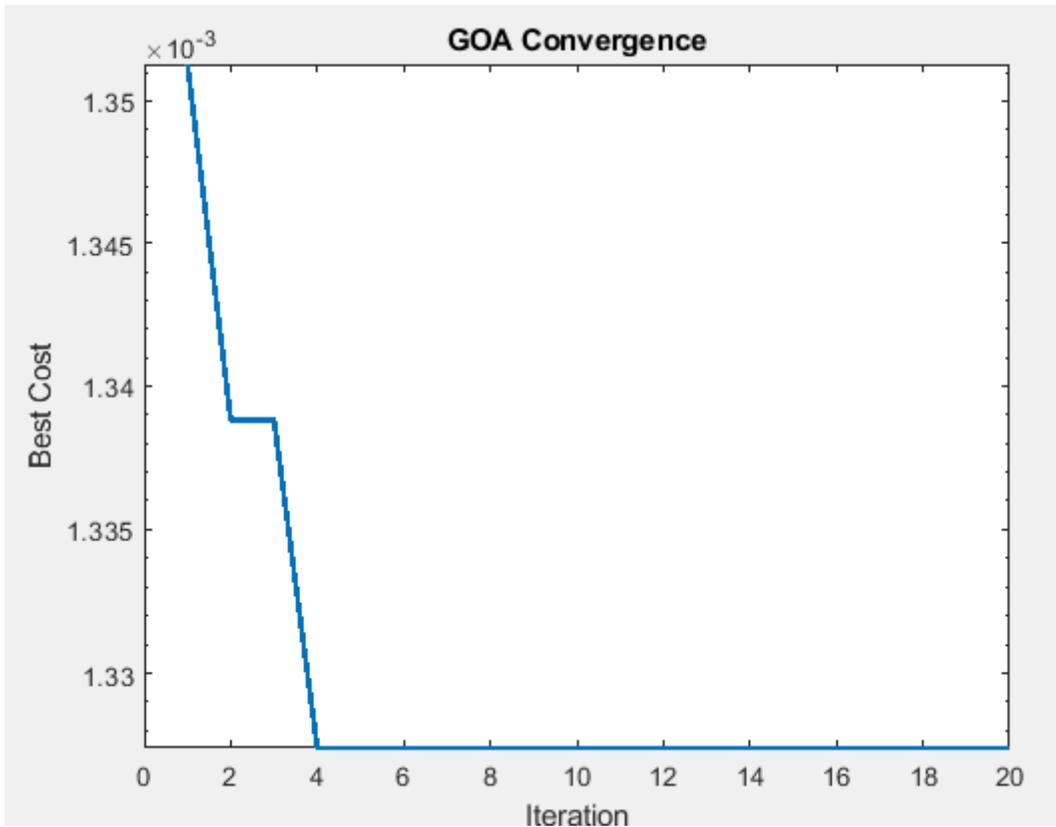


Figure 8

Detection of bank fraud with the recommended method and population equal to 5

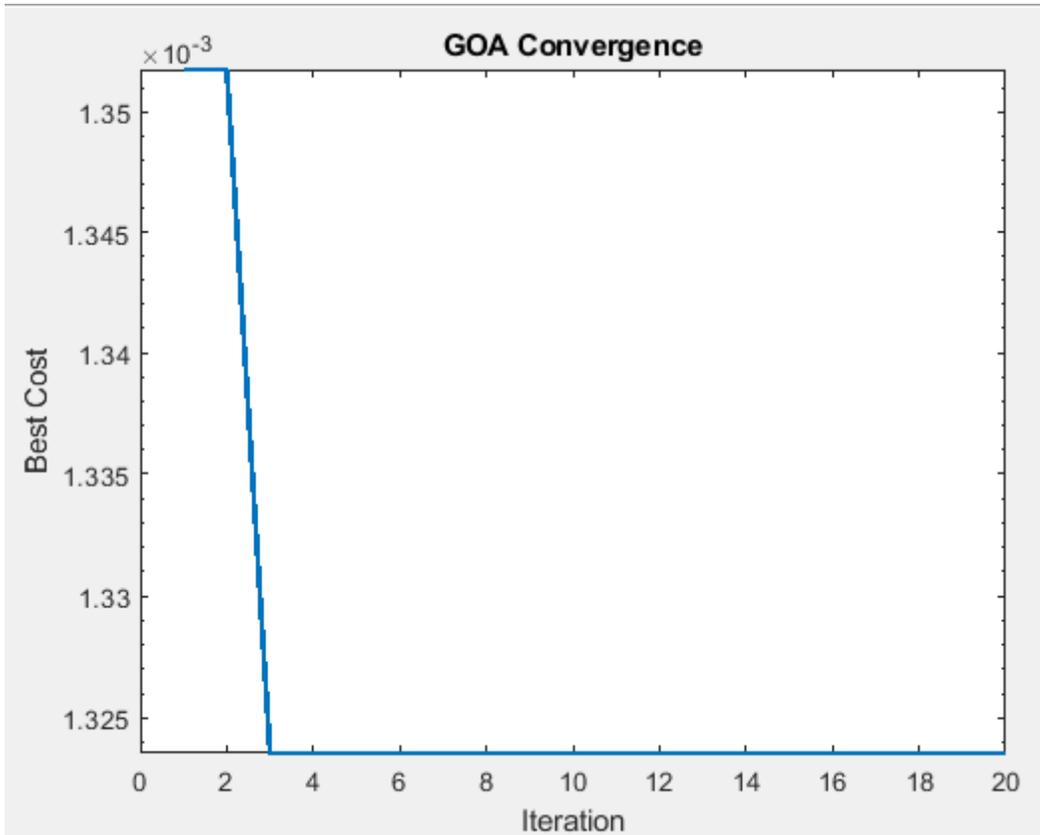


Figure 9

Detection of bank fraud with the recommended method & a population equal to 10

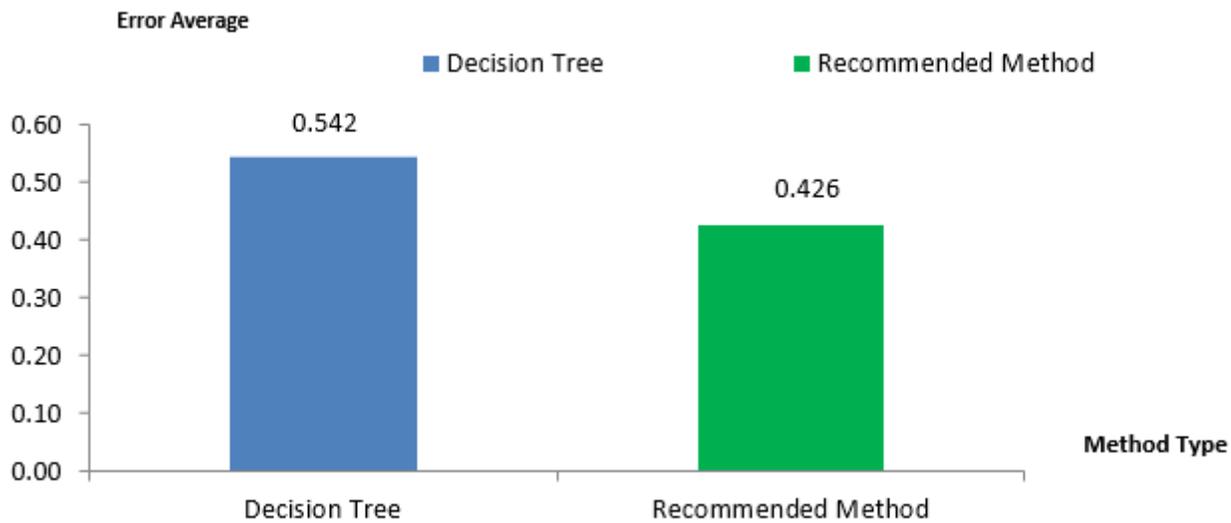


Figure 10

Comparison of RMSE index of the recommended method and decision tree in detecting bank fraud

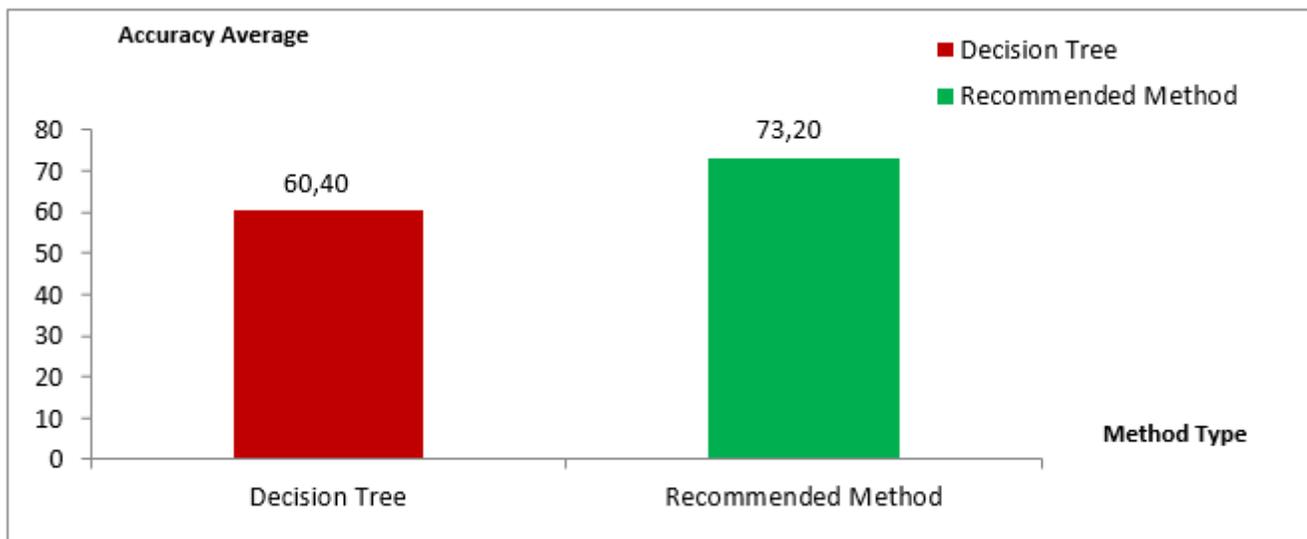


Figure 11

Comparison of decision tree accuracy index in detecting bank fraud

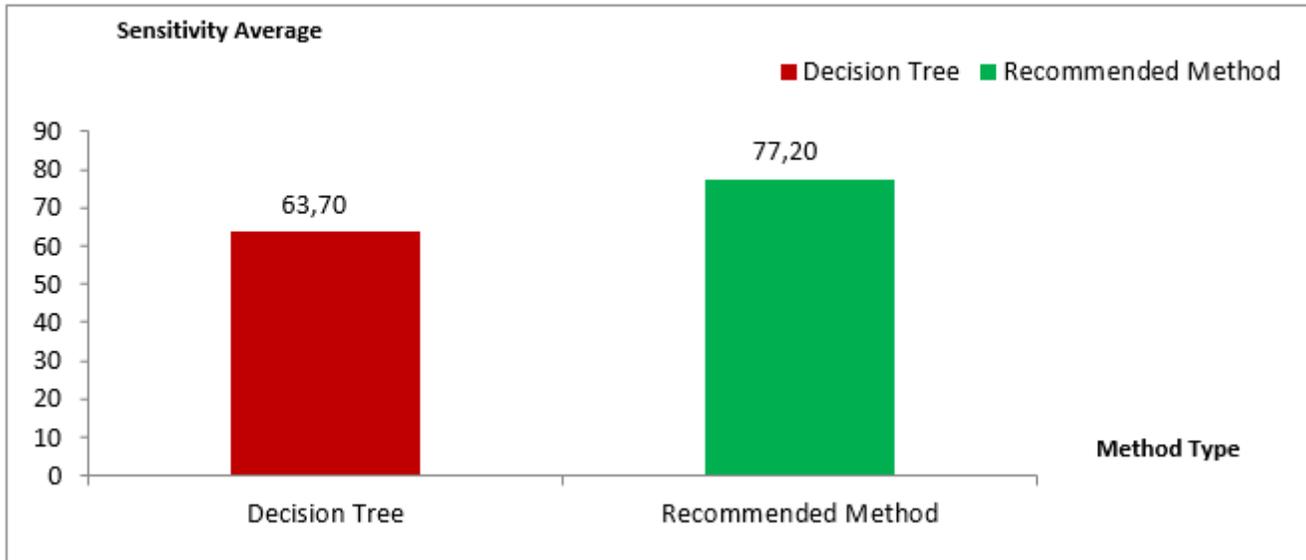


Figure 12

Comparison of the sensitivity index of the recommended method and the decision tree in detecting bank fraud

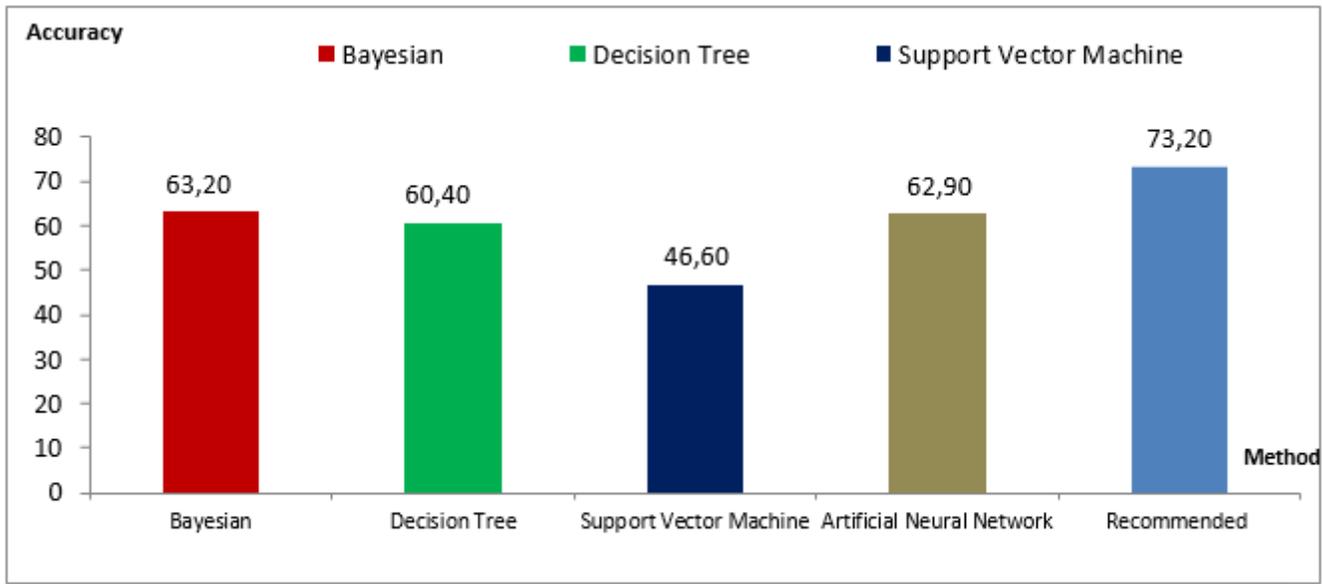


Figure 13

Comparison of the accuracy index of the recommended method and other methods in detecting bank fraud

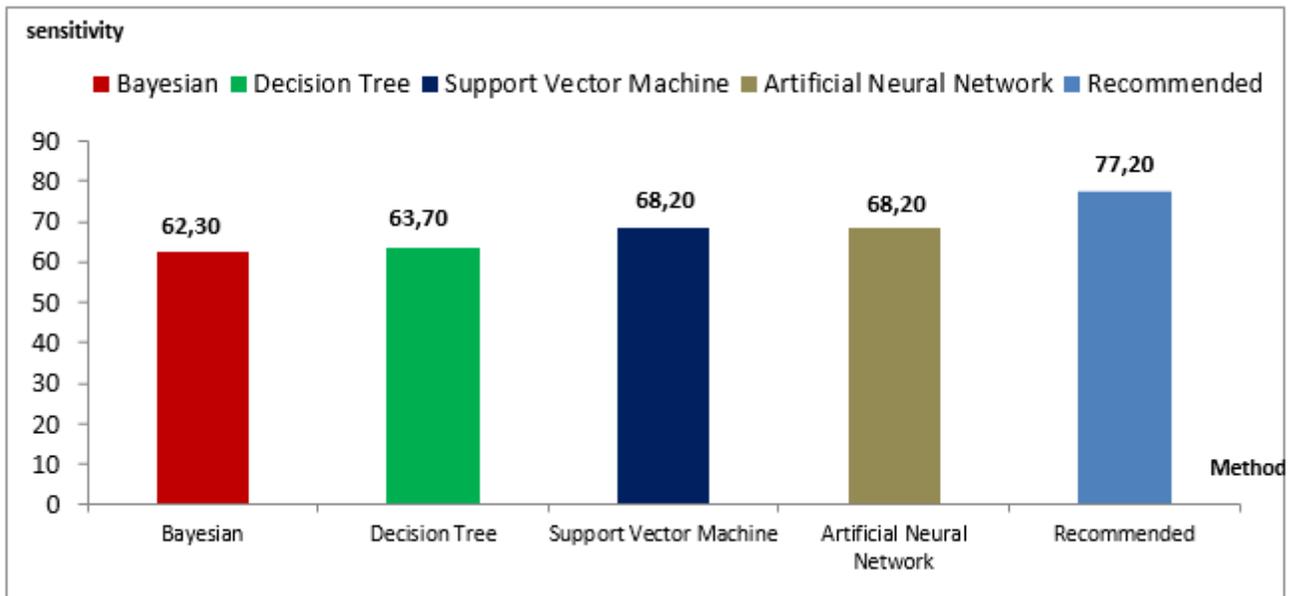


Figure 14

comparison of the sensitivity index of the recommended method and other methods in detecting bank fraud.