Identification of Fraud in Banking Data and Financial Institutions Using Classification Algorithms

Ardavan Rajaei
Department of Information
Technology-Computer Network
Engineering,e-campusIslamic Azad
University Tehran, Iran

Abstract
In recent years, due to the expansion of financial institutions, as well as the popularity of the World Wide Web and e-commerce, a significant increase in the volume of financial transactions observed. In addition to the increase in turnover, a huge increase in the number of fraud by user's abnormality is resulting in billions of dollars in losses over the world. Therefore, fraud detection techniques to identify users motivated to do a lot of scientific research. To check the volume of such information data mining techniques used. Data mining, knowledge discovery process that finding patterns in large data sets by combining techniques from statistics, artificial intelligence, database management and etc. In data mining techniques we examine and analyze information to discover uncertain communication and hidden patterns of data. This study aimed to detect fraud in banking data using classification algorithms.

Keywords
Classification Algorithms, Fraud, Detection, Data Mining

Introduction
In recent years, there is a significant increase in the volume of financial transactions due to the increasing presence of banks and financial institutions as well as the World Wide Web and e-commerce. Along with the increase in trading volume, it can be seen a huge increase in the number of cheats by abusive users, resulting in billions of dollars annually causing losses worldwide.so it is necessary to consider the identification and diagnosis of the fraud techniques and this a motivation for doing many researches(1).

Data mining techniques and classification algorithms are used to examine the large amount of information and discovery of data. These methods are typically classified into two groups: Anomaly Detection and Abuse Detection Method. In the first method, customer behavior history is considered as normal behavior, and if a new behavior is seen from the customer that is not compatible with his pattern of behavior, the possibility of fraud is strengthened. In the second method, certain behaviors are considered as fraud and abuse and it is examined which customer behavior is closer to this state(2).

Data mining is a hidden knowledge discovery process in a database in which patterns are derived from a large data set by combining methods of statistics and artificial intelligence in database management. In data mining techniques,it is tryingby data exploration and analysis them,to detect obscure connections and hidden patterns of data. In this paper, three-layer artificial neural network of the perceptron is used as one of the best classification algorithms to detect fraud in financial data.The structure of the article is as follows. In the third part the literature, and some of the work done so far have been presented. In the fourth section, a brief description of the artificial neural network as well as the artificial neural network of the three-layer perceptron is presented. In the next section, and in Section 5, the results of the implementation and the data used in the research are expressed, and finally the research summaries are presented.

The literature
Fraud detection means identifying the fraud in states or situations that previously occurred and is for specific purposes and goals, or it is detecting cases where a fraudulent person intends to misuse the profit that the company makes in return for providing services(3).

In other words, fraud detection involves monitoring user behavior in order to estimate and identify fraud and prevent undesirable behavior(4).

Several methods have been developed to detect fraud. The purpose of this article is to use machine learning and classification to detect fraud in the system. Machine Learning is an extensive field of research that attempts to simulate the behavior and activities of humans. In Machine Learning thepossibility to search, interpret, learn, teach, learn new skills, recognize, etc., for computers have been de-

veloped. As one of the broad and widely used branches of artificial intelligence, machine learning and data classification, regulates and explores the ways and algorithms by which computers and systems can learn and teach. The main purpose of classifying and machine learning is that computers (in the broadest sense) can gradually and with increasing data perform better in the task they are looking for. The range of this task can be from automatic face detection by seeing some examples of faces up to learning how to pick up robots by receiving a reward and punishment signal(5).

In 1999, Christine et al examined the various data mining techniques for detecting fraud(5). The study states that a couple of government agencies are seeking to determine which groups of buyers in their online purchases may be exploited by public funds. These organizations have gathered a lot of information about the features of purchases that have led to misuse of public funds. Such data is found in reports of credit risks associated with purchasing risks. Additional information is available by interviewing experts in this area. Organizations are always seeking to identify what other types of groups are in addition to the groups identified with the available knowledge that there is a risk of fraud occurring there. In fact, the goal of such organizations is to identify all existing behavior patterns for fraudulent groups in order to prevent potential misuse by unauthorized persons from financial resources. These organizations seek to create an effective fraud detection system using existing data. In addition, the study found that the US Federal Aviation Administration had collected transaction data through 40,000 card purchase of its employees. The transaction information includes information on the date of purchase, the amount of the purchase, the seller's name, the seller's address and the standard industrial dealer's sales standard along with other information. Account information includes information on individual accounts such as account holder information, transaction volume from an account, the number of purchases made and the billing cycle per person, and the date of purchase, along with other items. In 2010 Han and colleagues in an article entitled "A Synthetic Immune System for the Detection of Online Fraud," proposed a new hybrid model for detecting online fraud from the demand system on video, with the aim of improving current risk management by adding artificial immune systems (the fraud detection system based on login information) (6).

In 2011, Glenny et al. in an article entitled "Qualitative Model for Fraudulent Financial Reporting", proposed aqualitative Model for Detection of Financial Reporting Fraud. The proposed model tries to disclose hidden or misleading information in US Securities and Exchange Group reports(7).

In 2011 in an article entitled "Data mining for detecting credit card fraud," Battacharya et al. Evaluated advanced data mining, backup vector and random forest methods, along with logistic regression, to determine how they function to detect (control and Prosecution) fraud in credit cards(8).

In 2012, Jia et al. in an article entitled as "Applying a Transaction Accumulation Strategy to Detect Fraud in Credit Cards" the strategy of collecting transactions to detect credit card fraud. In this way, the characteristics of the transactions are collected to identify the consumer's purchasing behavior before the transaction is performed to identify the fraudulent business information collected before the occurrence (9).

In 2013, Sehin et al.in a paper Entitled "An Cost-Deciding Decision Tree for Fraud Detection," presented a new Cost-Effective Tree Decision approach to minimize the total cost of categorization, which addresses the problem of detecting fraud (10).

To deliver better results recently different techniques have been combined together. In 2013, Anil and his colleague presented a hybrid method using a genetic algorithm and a backup vector machine to identify abnormalities in a data set in a paper entitled "A Combined Method Based on Genetic Algorithm and Self-Organizing Algorithm and Support Vector Machine for Better Detection of Abnormalities", in this method, the genetic algorithm is used to select the best subset of attributes that indicate abnormalities. This new data set has been used to support the backup vector machine. The result of the combination of these methods has improved the achieved results(11).

In October 2014, Uluzsky et al proposed a fraud detection method based on user account imaging and threshold type detection in an article entitled "Fraud Detection Using Self-organized Map Based on User Profile Profiling." The illustration method used in this approach is self-organized map(12).

In 2015, Carmine et al. Proposed in an article "Bank-Sealer: A Decision-Making System for Analyzing Online Banking Cheat Sheets", a decision support system for analyzing online banking fraud. During the training phase of this system, easy-to-understand models have been created to understand each customer's spending habits based on past transactions (13).

Method

Learning the machine involves various algorithms for creating intelligent models and classifications for data analysis, among them are a variety of artificial neural networks, linear separation analysis, Hidden Markov model, k-means clustering, fuzzy logic, support vector machine, closest neighbor of algorithms and popular methods in classification. Among these, one of the best algorithms in the classification, which has the ability to learn and teach, is the artificial neural network inspired by how hu-

man neurons work. In this paper, a three-layer perceptron neural network is used to detect fraud. The following is a brief overview of artificial neural networks and three-layer perceptron.

The core idea of this network is inspired by the way the biological nervous system functions, to process data and information in order to learn and create knowledge. The system consists of interconnected processing elements, called neurons. This network is capable of learning. In this case, it can correct its error. Learning in these systems is comparative. That is, using the training examples, the weight of the synaptic changes in such a way that, if given new inputs, the system will produce the correct response (14).

Researchers believe that artificial neural networks include a network of simple processing elements (neurons) that can display the generally determined complex behavior of the relationship between the processing elements and the element parameters. If consider a network to be equal to a graph, the network training process will determine the weight of each edge and the initial bias(14).

In 1958, the Perceptron Network was introduced by Rosenblatt. Usually the perceptron has three layers. The other is the neuronal linear model, created in 1960 by Vidro Huff, the first neural networks used in real issues. The progress made in 1970 to 1980 was critical to attracting attention to neural networks. Some factors also contributed to the escalation of this issue(14). The Multilayer Perceptron Model, or MLP, is one of the most efficient layouts proposed for use in real nerve classification and modeling, consisting of an input layer, one or more hidden layers and an output layer. In this structure, all neurons of one layer are connected to all neurons of the next layer. This arrangement is known as a network with full connectivity. Figure 1 shows the architecture of a perceptron neural network.

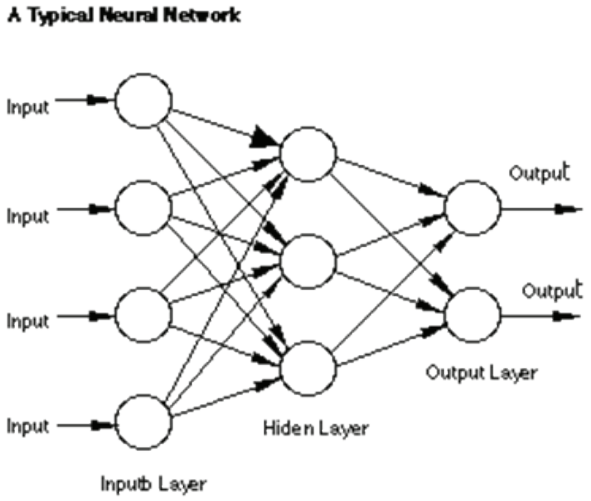

Figure 1: Architecture of a three-layer perceptron network

Findings and outcomes

In this research, the system is implemented and tested from a German credit card dataset that is available in this area of research. This dataset has been discussed by many researchers. The German Credit Card Dataset is one of UCI's data mining and acquisition data set.

The selected dataset contains 1000 samples, each containing a number of attributes. Existing examples are ultimately categorized into two classes of the right class and the wrong class (fraud). Among the available data in this dataset, 700 samples are for normal customers and 300 for non-traditional customers. As stated in this dataset, several features have been suggested that will be used in our research to detect fraud. These features include the age of the individual, the number of accounts held by the bank, the number of months since the opening of the account, the amount of credit in the bank, the number of individuals affiliated with the person and the period during which the person is resident in the city. In this paper, three-layer perceptron artificial neural networks were used to detect fraud. Table 1 shows the number of test data and test data.

Table 1: Number of samples in the database

Dataset type	Total number of samples	Number of normal samples	Number of cheat saples
Education	500	300	200
Test	200	100	100

Also in Table 2, the accuracy of the MLP algorithm detection is presented

Table 2: Accuracy of correct algorithm detection

	Education	Test
Total number of samples	500	100
Number of correct cases detected	454	86
Accuracy	90.8 %	86 %

In Figure 2, the ROC graph and the collision matrix associated with the MLP algorithm are shown in the test data.

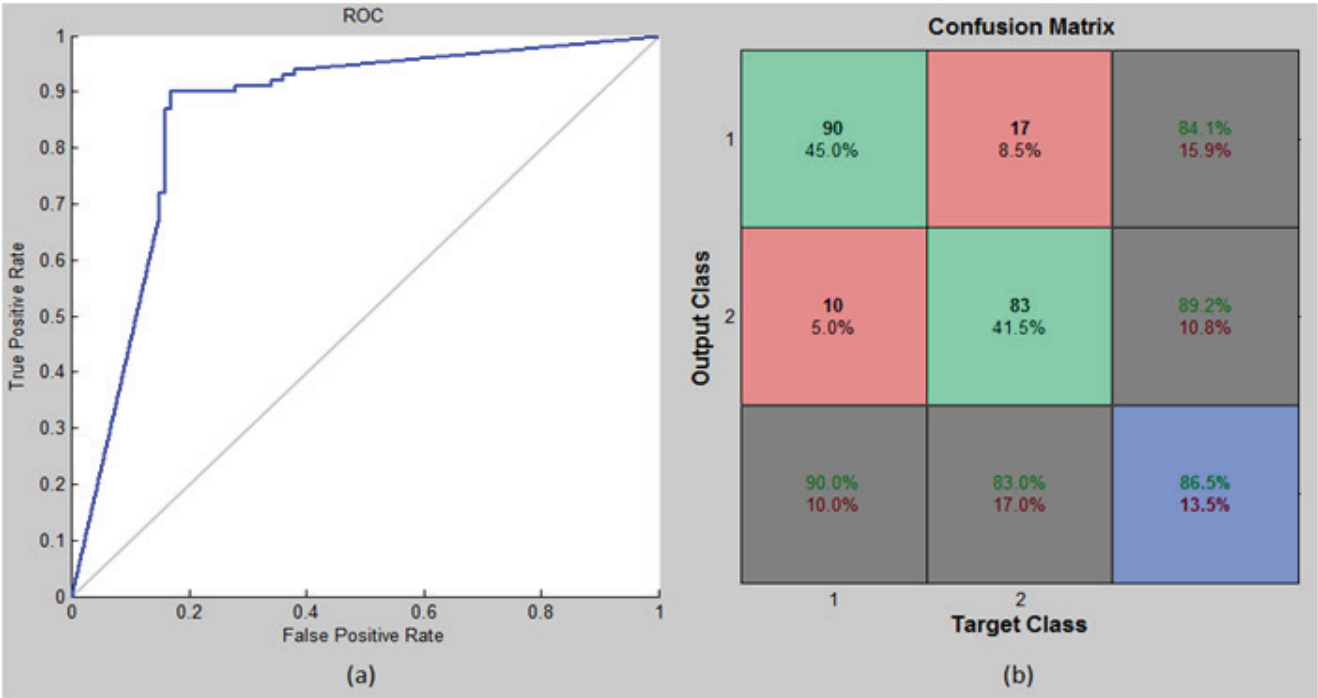

Figure 2: ROC Chart and Collapse Matrix

Conclusion

The main purpose of this research is to provide a method based on data mining and classification techniques for detecting fraud in financial institutions and banks. In this study, an artificial neural network of three perceptron layers (MLP) was used to detect fraud in the UCI data set. The obtained results show that the use of data mining and classification methods makes it possible to establish a safe hospitalization and to detect an offense with minimal error, without using the human element and smart models in this area. Considering the good performance and the desirable accuracy of creating a learningable model using the artificial neural network, it can be suggested to use the functional vision of the system provided by financial institutions and banks and use it to predict and detect fraud. It is also possible to use other intelligent and combined methods in machine learning and data classification to improve the accuracy of diagnosis.

References

1. Caldeira, E., Brand˜ao, G, and Pereira, A, C, M. "Fraud Analysis and Prevention in e-Commerce Transactions", Web Congress (LA-WEB), 9th Latin American, IEEE: 42-49.2014.

2. Mishra, P., Neelamadhab P, and RasmitaPanigrahi. "The survey of data mining applications and feature scope." Asian Journal of Computer Science & Information Technology, 2.4: 68-77,2013.

3. Sharma, A., and Panigrahi K, P. "A review of financial accounting fraud detection based on data mining techniques." International Journal of Computer Applications, Volume 39– No.1: 37-47, 2012.

4. Green, Brian Patrick, Choi, Jae Hwa. "Assessing the risk of management fraud through neural-network technology", Auditing: A Journal of Practice and Theory, 16(1): 14–28,1997.

5. Kristin, R. N., and I. P. Matkovsky. "Using Data Mining Techniques for Fraud Detection." SAS Institute Inc. and Federal Data Corporation: 1-32,1999.

6. Huang, Rentian, HissamTawfik, and Atulya K. Nagar.

"A novel hybrid artificial immune inspired approach for online break-in fraud detection." Procedia Computer Science (ICCS), Elsevier, 1.1: 2733-2742,2010.

7. Glancy, F, H., Yadav, S, B. "A computational model for financial reporting fraud detection, Decision Support Systems", Elsevier, Volume 50, Issue 3: 595-601, 2011.

8. BhattacharyyaS., Jha, S,Tharakunel, K, and Westland J, C. "Data mining for credit card fraud: A comparative study", Decision Support Systems, Elsevier,Volume 50, Issue 3: 602-613, 2011.

9. Jha, Sanjeev, Montserrat Guillen, and J. Christopher Westland. "Employing transaction aggregation strategy to detect credit card fraud." Expert systems with applications, Elsevier, 39.16: 12650-12657,2012.

10. Sahin, Y., Serol, B, and Ekrem D. "A cost-sensitive decision tree approach for fraud detection." Expert Systems with Applications, Elsevier, Volume 40, Issue 15: 5916-5923,2013.

11. S. Anil and R. Remya, "A hybrid method based on genetic algorithm, self-organised feature map, and support vector machine for better network anomaly detection,"- Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1-5. doi: 10.1109/ICCCNT.2013.6726604, 2013.

12. Olszewski, D. "Fraud detection using self-organizing map visualizing the user profiles." Knowledge-Based Systems, Elsevier, 70: 324-334, 2014.

13. Carminati, Michele, et al. "BankSealer: A decision support system for online banking fraud analysis and investigation." Computers & Security, Volume 53: 175-186,2015.

14. Y. Li and W. Ma, "Applications of Artificial Neural Networks in Financial Economics: A Survey,"International Symposium on Computational Intelligence and Design, Hangzhou, pp. 211-214,2010.