ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

Data Mining and Knowledge Extraction in the Risk Based Insurance Audit: A Case Study (City of Tehran)

Hossein Amirinia[1], MA Afshar Kazemi[2], Zahra Alipoor Darvish[3]

Islamic Azad Electronic Tehran Branch, Tehran, Iran [1]

Associate Professor, Islamic Azad University, Cantrell Tehran Branch, Tehran, Iran [2]

Department of Management, North Tehran branch, Tehran, Iran [3]

Abstract: Nowadays, data mining techniques are widely used in various industries. The Social Security Insurance Audit, with more than two million audit cases nationwide and limited manpower, cannot respond to all cases. The present research, using data mining techniques, aims to propose an efficient method for the detection of high-risk companies and fraud in the insurance industry. Materials and Methods: After randomly collecting 2000 records from the Social Security Audit database in 2014 and extracting various features for each company, a pre-processing of the data was conducted. The data was then divided into the two categories of Training Set and Test Set and given to the following four algorithms: Neural Network, Decision Tree, Bayesian and Support Vector Machine. Ultimately the accuracy of each algorithm was measured by the confusion matrix. Findings: By comparing the four algorithms mentioned above, it can be seen that all of these algorithms can provide a proper level of accuracy, with the neural network algorithm showing 90.6% accuracy, was found to be the best algorithm for the identification of high-risk companies. The findings of this study enable the Social Security Audit to develop software which can identify companies with high insurance risk in order to achieve higher levels of efficiency in the face of current limitations in manpower.

Keywords: Data Mining; Insurance Audit; Fraud; Neural Networks.

I. INTRODUCTION

Each year, organizations and commercial centers utilize different approaches to avoid paying the premiums of their employees which leads to improper distribution of insurance services to citizens, reducing the quality of customer services.

On the other hand, the Insurance Audit Organization cannot cover the ever-increasing volumes of data which are estimated to be more than two million insurance records. Data mining and predictive methods are great assets to managers and decision makers. Also, the discovery of hidden rules with high certainty can reveal secrets hidden in the data and facilitate efficient auditing.

II. THE DEFINITION OF DATA MINING

According to Han and Kamber [1], data mining, which is known as the most important stage in the process of knowledge discovery, means extracting or exploring knowledge from huge amounts of data. Data mining is discovering new, reliable, and traceable knowledge using artificial intelligence and statistical tools in a high volume of data [2]. The processes involved in discovering knowledge from databases are as follows [3]:

Data Selection: data related to analysis and decision-making are separated from other data.

Data pre-processing: Processing, cleaning and integrating the data are performed.

Data conversion: The selected data are converted in a manner suitable for data mining.

Data mining: At this stage, intelligent methods and decision making approaches are used to extract potentially useful patterns.

Interpretation and evaluation: At this stage, interesting patterns which represent knowledge are identified based on the measures taken and the recently discovered knowledge is made available to the user. It is necessary to utilize visualization in order to assist the user at this point.

III. BACKGROUD OF THE STUDY

To date many investigations have been conducted on fraud detection and identification of financial abuse using data mining algorithms, among which the following can be named; Neural Network, Decision Tree, Bayesian algorithm, Support Vector Machine, etc. The following studies have been conducted on different types of fraud, abuse, and crime detection including areas such as financial issues, credit card abuse detection, computer intrusion detection, and detection of insurance fraud. To this end, Karlys and Migukutsid, using clustering methods, developed a model for the identification and classification of crimes [4]. Similarly, Hang et al., proposed an innovative fraud detection mechanism based on Ziff laws aiming to identify potential fraud in enormous volumes of data [5]. Also, Lio et al., conducted an investigation into Taiwan's National Health Insurance where they used nine variables, such as the average cost of medication, the

DOI 10.17148/IJARCCE.2015.4892

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

average cost of diagnosis, average days of drug distribution, consulting fees, cost of treatment, etc. Using three algorithms, namely, Supporter Regression, Neural Network, and Decision Tree, they found that the decision tree offered the best accuracy [6]. Zhua and Campur, using Logistic Regression, Decision Tree, Neural Network, and Bayesian Network, developed a model capable of conducting more effective fraud detection in financial statements [7].

Rvysnkar et al., identified fraudulent financial statements in companies using algorithms such as neural network and machine support [8]. Proles compared conventional statistical methods and machine learning algorithms, such as neural network and support vector machine in financial fraud detection [9].

Mon et al., using regression, identified the rate of computer use and that of membership in online networks as the two major crime rate predictive variable [10].

Pai et al., designed a linear model to estimate management frauds of a decision function using the SVM algorithm [11]. Shin et al, conducted an investigation in Korea, using the decision tree, into insurance abuses of outpatients in various clinics [12]. Sharma and Pangarahi, assessed data mining methods, such as neural network, Bayesian network, and decision trees in financial fraud detection [13] ,Hung, considering non-financial factors in order to produce a more robust detection, identified financial frauds using the SVM algorithm [14].

Akina et al., used a combination of clustering and Bayesian methods in order to identify potential fraud [15].

Chakarabourti, investigated a number of algorithms regarding fraud detection and found three concepts that can increase detection efficiency, namely the system itself, flexibility, and data quality [16]. Also, bank card abuse as a different type of financial fraud was studied by Bhatachari et al., who used support vector machines and random forest in order to detect such abuses [17]. Regarding online banking fraud detection system, Way et al., conducted an experiment into online banking data which reporting a higher accuracy than previous methods [18].

Computer Intrusion Detection is another application of data mining. Poua et al. investigated e-commerce and e-business and automatic intrusion detection in this area [19]. A study conducted by Masa and Walword focused on e-commerce websites [20]. Also, Dhakar and Tivari developed an innovative method for computer intrusion detection with K2 and TAN classification approaches [21].

Previous studies can be considered from two other perspectives; supervised and unsupervised methods. The study carried out by Ekina et al., which is in the field of health and fraud used the clustering method [15].Tornman and Kepelwin's study can also be mentioned in which they used unsupervised data mining and succeeded in detecting 12 suspicious cases out of 17, with the accuracy of 71% [22]. On misuse and frauds in the field of health, using the supervised method which is properly utilized for prediction, a given data set is divided into two sets, namely, fraudulent and non-fraudulent. For example, we can point to the use of decision tree algorithm by Shin et al. [12] and neural network algorithm by Lio et al. [6], and also support vector machine by Kirlidag and Esok [23].

In this article we aim to investigate supervised data mining methods. Judaki et al also reviewed previous studies [24]. By looking at the literature, we found a gap in detecting insurance frauds and insurance companies using data mining methods, which will be examined in the following sections.

IV. ALGORITHMS

Considering Figure 1: Conceptual Model of Insurance Audit Using Data Mining, the data is given to four algorithms of Neural Network, Decision Tree and Bayesian and Support Vector Machine, results are obtained, and eventually the accuracy of each algorithm is measured. Algorithms are defined in the following section.

A. Neural Network

Neural network is a method that imitates human brain using a set of interconnected nodes. This method is based on computer models of biological neurons. A multi-layer neural network consists of a large number of inter-connected units (neurons) in a pattern of communication. This method is extensively used in classification and clustering and is one of the most useful data mining methods in detecting financial frauds [25]. First, the network is trained for drawing the input and outputs by a set of paired data; then the weight of communication is established between the neurons and the network is utilized for determining the classifications of a new set of data. The advantages of this method are as follows: first, it adaptable. Second, this method creates enduring models and third if training weights change, the classification process can be modified. Neural networks are mostly used for credit cards, car insurance and company frauds [26]. Here in this article we use the multilayer perceptron algorithm.

B. Decision Tree

The decision tree is a tree in which the samples are classified in a way that they grow from the roots downward and eventually reach the nodes of the leaves. A tree usually consists of roots, branches and nodes from which the branches spread out. Decision trees are support tools for predicting decisions that create an image of the observations for the possible outcomes [1]. They classify topics based on the amount of attributes. The leaves are symbols of prediction; each node in a tree is the decision of a representative of an attribute in a classification topic and each branch is the representative of the amount that a node can obtain, and it in fact shows sharing features. Decision trees are usually utilized in credit cards, car insurance and company frauds [26]. The classification rules are obtained from the patterns of "if"-"then" of the decision tree; it reduces making a decision about a

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

complicated matter to deciding about a large number of simple matters [27]. J48 algorithm is one of the most widely used algorithms of decision tree, which will also be used in this article

C. Support Vector Machine

The support vector machine seeks a solution to separating two different classes with minimum fault, which was established by Cortes and Vapnik and their team at AT&T Bell [28]. The purpose of SVM was to find the best function for classification such that the members of the two classes are distinguishable from each other. The criterion for the best classification is determined geometrically. For data sets that are disintegrable linearly, the border that is defined as a part of the space or the line between the two classes is defined by hyperplane intuitively. This geometric definition allows us to find out how to maximize the borders, even if we have an infinite number of hyperplanes and only a few deserve a solution for support vector machine. The best dividing line is the one that has the least distance from the nearest point [29].

In linear division of data, the purpose is to reach a function that determines the hyperplane with the largest margin. With the maximization of the margin of this hyperplane, the division among the levels is maximized. In support vector machine, the set of points can be divided by linear and non-linear methods [30]. Considering the assumption that the sets are dividable linearly, hyperplanes with maximum margin are obtained so that the sets can be divided. In cases where the data is not dividable linearly, the data assumes a space with greater dimensions, so that they can be divided linearly in this new space. We use the libSVM algorithm in this article.

D. Bayesian

The Bayes algorithm is also a proper classification algorithm, with proper functioning in facing extensive entry dimensions [1]. Naive Bayes utilizes the Bayes conditional rules [13]. Assuming that there are C classes for the new X sample, the classified section foretells that X belongs to the class that has the highest posterior conditional probability on X. This means that the classified section considers the data X as belonging to Ci class; therefore the new X sample belongs to the class that has the highest probability condition [31] and it also uses the Bayes theory for probability. In this research we use naïve Bayes.

V. DATA DESCRIPTION

In this part we describe the data obtained from the audit of Social Security insurance, which includes 2000 data related to Tehran province. As it is mentioned in Table I: Obtained Features from Each Insurance Paying Company, various features of the data are stored in the data bank, including the number of the insured, company age, company ownership, company type, field of activity, application date, application repetition, inbox circular priority and investigable period.

These features and the description for each is given in the following table.

TABLE I: OBTAINED FEATURES FROM EACH INSURANCE PAYING COMPANY

Feature Name	Feature Description
Number of Insured People	It refers to the number of the people on the insurance list of any company or organization, and naturally, the higher this number the more value it has for supervision.
Background or Company Age	The number of years in which the company or organization has checked an insurance list.
Company Type	The working function of the company that is divided into the three classes of manufacturing, servicing and trade.
Field of Activity	The industry that the company is working in, such as oil industry, or motor industry.
Application Date	When this case has been presented for inspection.
Application Repetition	The more the number of repetitions, the higher its importance.
Inbox Circular Priority	This section deals with inter-company priorities for inspection of cases.
Investigable Period	One or a few years of the records must be investigated, and obviously the larger the number of the years the slower the investigation.
The Type of Ownership	Institution ownerships could be public or private.

VI. METHOD

Considering ***Fig 1. Conceptual Model of Insurance Audit Using Data Mining***, the conceptual insurance audit model obtains raw data from the companies using data mining, which are delivered to insurance audit companies as audit applications. Then the audit application is stored in the database, and data cleaning is done considering our need. In order to clean the data in this stage, invalid data are cleared from the training data. Faulty or incomplete data are instances of redundant data that must be cleaned.

There are two ways to confront redundant data; one is identifying and removing redundant data as a part of pre-processing stage, and the other is presenting a model that is resistant to this data. Incomplete resampled data and redundant data such as application repetition or a company's application in different fiscal years are removed, summed up, and changed so that they be understandable for data mining algorithms.

Considering the existent indexes and the input data, 60% of data is used by data mining algorithms for learning, and the rest of the data is allocated to testing data mining algorithms; the companies will be divided into two types:

- Companies with a low insurance risk (Positive)
- Companies with a high insurance risk (Negative)

Based on **Fig 2.Information Classification by each of the Algorithms**, if the algorithm puts the company exactly in the position where it should, that is, if a company with a low risk be correctly classified under low risk companies, true positive will occur, and if a company with a high risk be correctly classified, true negative will occur.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

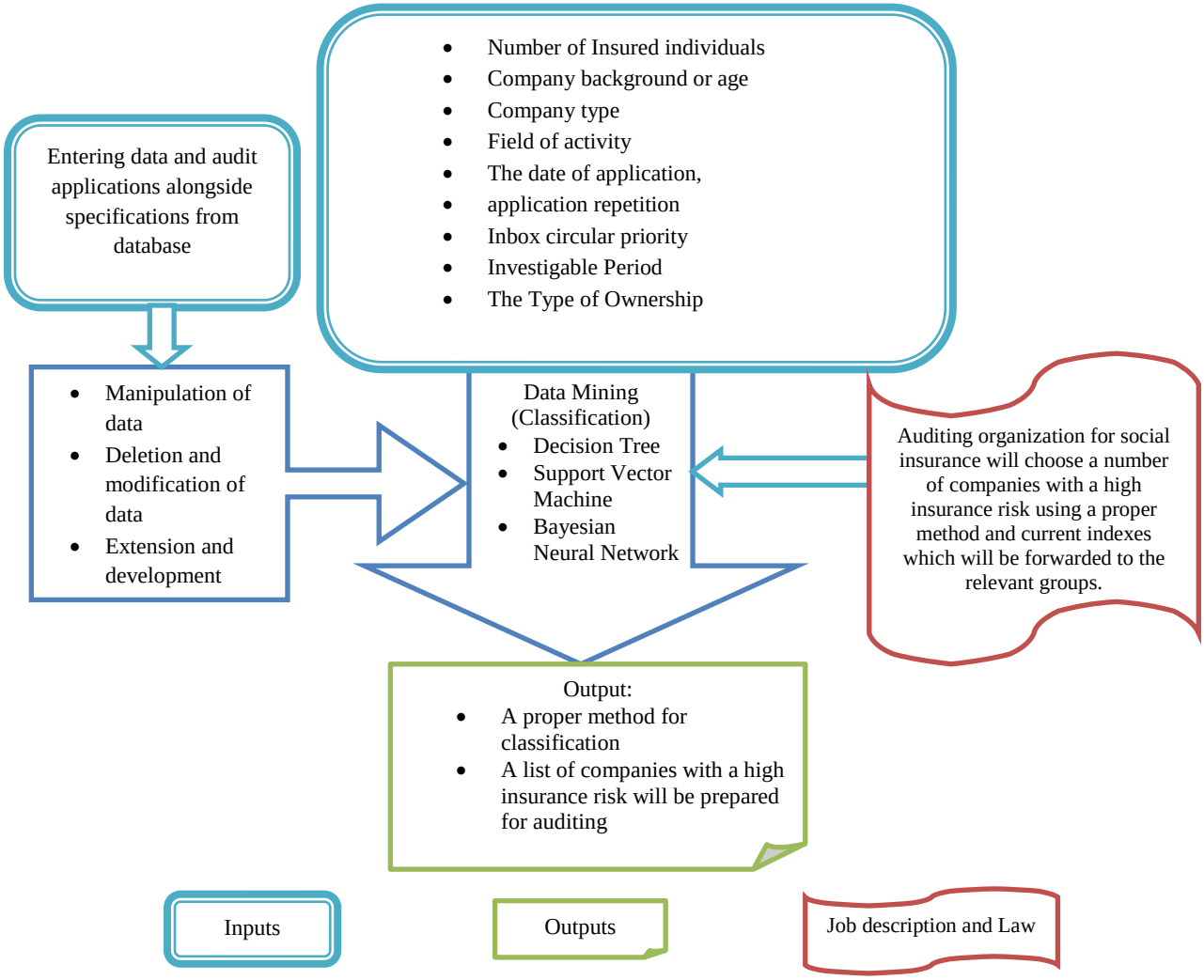

Fig 1. Conceptual Model of Insurance Audit Using Data Mining

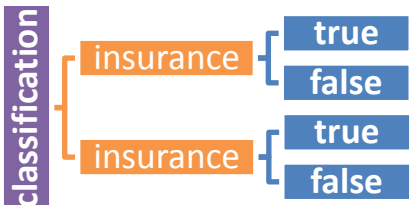

Fig 2.Information Classification by each of the Algorithms

However, if a high risk company is classified as low risk or a low risk one is classified as high risk, false positive and false-negative will occur, respectively. Raj and Pourita used true positive and negative in order to measure the accuracy of the algorithm in identifying credit card frauds [27].The algorithm which has the least fault in our classification, that is, the least amount of false-positive and false-negative, will be our optimal algorithm. When a new application enters the organization it is referred to as a data mining algorithm, and if it is classified as companies with a high insurance risk, it will be sent to the relevant departments to be audited, which saves both time and costs for the organization.

In this article WEKA software is utilized for data mining. Sixty percent of the existent data is used as the training set and the rest of them as the test set; that is initially the model is determined using the training set, and then its accuracy is tested by the test set. The reason for using two sets of data in classification is to prevent over-fitting. The phenomenon occurs when one model is too dependent on a certain data set. In this case, although the model showcases a very detailed and exact performance on that set, it lacks it on other sets of data. In fact, in this situation the model remembers the data set instead of learning it. Considering the fact that our original purpose in creating such a model was to provide proper predictions for new data, we have to try to reduce over-fitting as much as possible. Therefore, we use the test set after creating the model in order to test if its accuracy is reduced in confronting new data. We also pay attention to FP=False Positive and FN=False Negative in order to compare classification algorithms. False positive occurs when a statistical examination denies a true assumption. For instance, consider someone who does not have diabetes, but the statistical examination shows that he suffers from

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

it. False negative also occurs when the statistical examination accepts a false assumption. The creator of a model must determine to what degree the proportion of these models is acceptable. For example, receiving a spam email in the inbox (false negative) may be not so bothersome, but a non-spam email going to the spam folder can cause a great deal of trouble (false positive). As a result, it must be determined at the very start how much is the proportions of false negative to false positive, which in this study are considered equal. They represent the algorithm with the least amount of false positive and negative and are considered as the best algorithm. As it can be witnessed in **Fig 3. Knowledge Flow**, it is possible to create a knowledge process and run the algorithms in parallel in WEKA software. The data, after being recalled in a 60 to 40 proportion is given to the four algorithms of decision tree, neural network, Bayesian, and support vector machine in order to learn, and then to start predicting. The outputs are presented in the format of a text file in WEKA software, in which the output file of all four algorithms is presented as follows.

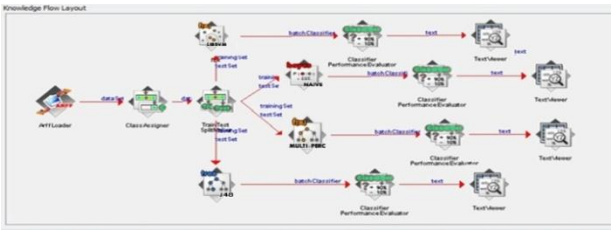

Fig 3. Knowledge Flow

As it is mentioned in TABLE II: Confusion Matrix, the two criteria of correctly and incorrectly classified instances are considered in order to compare the algorithms. The confusion matrix which includes false positive and false negative and true positive and true negative, was used by Dehkar and Tivari in identifying computer intrusion to compare algorithms [21].

TABLE II: CONFUSION MATRIX

True Positive = TP Classified correctly as true	False Negative = FN Wrongly considered false
Positive False = FP Wrongly considered true	True Negative =TN Classified correctly as false

TP + TN = accuracy
FP + FN = inaccuracy

Considering the figure of knowledge process, and Table III : Algorithm Comparison, all four algorithms are usable and have acceptable performances in insurance auditing prediction, in a way that the weakest algorithm, i.e., Bayesian algorithm with 82% accuracy was placed fourth; the third place was taken by decision tree algorithm with 89.6% accuracy; support vector machine algorithm with 90% accuracy was ranked as second, and as mentioned earlier, the neural network algorithm with the classification accuracy of 90.6% and inaccuracy of 9.4% gained the highest of indexes. This result displays the proper performance of all algorithms in identifying companies with high insurance risks, of course with unequal priorities.

Algorithm	Accuracy	Inaccuracy	Confusion matrix
Decision Tree	**89.6%**	**10.4%**	**a** **b** **<-- as** **336** **128** **a = YES** **80** **1456** **b = NO**
Bayesian	**82.8%**	**17.2%**	**a** **b** **<-- as** **152** **312** **a = YES** **32** **1504** **b = NO**
SVM	**90%**	**10%**	**a** **b** **<-- as** **300** **160** **a = YES** **40** **1500** **b = NO**
Neural Network	**90.6%**	**9.4%**	**a** **b** **<-- as** **356** **108** **a = YES** **84** **1452** **b = NO**

As examined in Table III: Algorithm Comparison, the neural network algorithm with accuracy of 90.6% and inaccuracy of 9.4%, and also the proportion of true positive and negative of 1808 out of 2000, is the best algorithm in predicting high insurance risk companies.

VII. CONCLUSION AND FUTURE WORKS

In this article, companies with a high insurance risk were identified using a new method proposed based on data mining, which is a first of its kind in insurance auditing. Fortunately, all the algorithms could meet our needs with high indexes and prove the importance of data mining in today's world. In this article, after collecting the data and performing preprocessing and comparing the obtained results on the accuracy of each algorithm, the neural network algorithm with the accuracy of 90.6% showed the best results.

Since the data of this study is from the social security insurance auditing in Iran, it is suggested that a study, utilizing the same methods, be attempted in insurance or other similar organizations and its results be compared with this study and studies using other methods.

In the end, our suggestion is that in future studies, insurance auditing can consider more features of companies and use them. Also a system can be developed based on the findings of this study, in order to reduce the costs of auditing and predict the behavior of new

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

customers in using insurance services. Scheduled periodical updating of the training set can be very useful in order to enhance the performance of this system to even a greater extent.

REFERENCES

[1] J. Han and M. Kamber, Data Mining: Concepts and Technique. 2006.
[2] B. Fang and S. Ma, "Data Mining Technology and Its Application in CRM of Commercial Banks," 2009, pp. 243–246.
[3] S. . Deshpande and V. . Thakare, "Data Mining System and Applications: A Review," Int. J. Distrib. Parallel Syst., vol. 1, no. 1, pp. 32–44, Sep. 2010.
[4] D. Karlis and L. Meligkotsidou, "Finite mixtures of multivariate Poisson distributions with application," J. Stat. Plan. Inference, vol. 137, no. 6, pp. 1942–1960, Jun. 2007.
[5] S.-M. Huang, D. C. Yen, L.-W. Yang, and J.-S. Hua, "An investigation of Zipf's Law for fraud detection (DSS#06-10-1826R(2))," Decis. Support Syst., vol. 46, no. 1, pp. 70–83, Dec. 2008.
[6] F.-M. Liou, Y. Tang, and J.-Y. chen, "Detecting hospital fraud and claim abuse through diabetic outpatient services.," Health Care Manag. Sci., vol. 11, no. 4, p. pp 353–358, 2008.
[7] W. Zhou and G. Kapoor, "Detecting evolutionary financial statement fraud," Decis. Support Syst., vol. 50, no. 3, pp. 570–575, Feb. 2011.
[8] P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," Decis. Support Syst., vol. 50, no. 2, pp. 491–500, Jan. 2011.
[9] J. Perols, "Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms, Auditing," J. Pract. Theory, vol. 30, pp. pp. 19–50, 2011.
[10] B. Moon, J. D. McCluskey, and C. P. McCluskey, "A general theory of crime and computer crime: An empirical test," J. Crim. Justice, vol. 38, no. 4, pp. 767–772, Jul. 2010.
[11] P.-F. Pai, M.-F. Hsu, and M.-C. Wang, "A support vector machine-based model for detecting top management fraud," Knowl.-Based Syst., vol. 24, no. 2, pp. 314–321, Mar. 2011.
[12] H. Shin, H. Park, J. Lee, and W. C. Jhee, "A scoring model to detect abusive billing patterns in health insurance claims," Expert Syst. Appl., vol. 39, no. 8, pp. 7441–7450, Jun. 2012.
[13] A. Sharma and P. K. Panigrahi, "A review of financial accounting fraud detection based on data mining techniques," ArXiv Prepr. ArXiv13093944, 2013.
[14] S. Yan Huang, "Fraud Detection Model by Using Support Vector Machine Techniques," Int. J. Digit. Content Technol. Its Appl., vol. 7, no. 2, pp. 32–42, Jan. 2013.
[15] T. Ekina, F. Leva, F. Ruggeri, and R. Soyer, "Application of Bayesian Methods in Detection of Healthcare Fraud," Chem. Eng. Trans., vol. 33, 2013.
[16] S. Bhattacharya Chakraborty and M. Z. Shaikh, "A comprehensive and relative study of detecting deformed identity crime with different classifier algorithms and multilayer mining algorithm," Int. J. Adv. Res. Co Mputer Co Mmu Nication Engine Ering, vol. 3, no. 3, 2014.
[17] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decis. Support Syst., vol. 50, no. 3, pp. 602–613, Feb. 2011.
[18] W. Wei, J. Li, L. Cao, Y. Ou, and J. Chen, "Effective detection of sophisticated online banking fraud on extremely imbalanced data," World Wide Web, vol. 16, no. 4, pp. 449–475, Jul. 2013.
[19] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," ArXiv Prepr. ArXiv10096119, 2010.
[20] D. Massa and R. Valverde, "A Fraud Detection System Based on Anomaly Intrusion Detection Systems for E-Commerce Applications," Comput. Inf. Sci., vol. 7, no. 2, Apr. 2014.
[21] M. Dhakar and A. Tiwari, "A Novel Data Mining based Hybrid Intrusion Detection Framework," J. Inf. Comput. Sci., vol. 9, no. 1, pp. 037–048, 2014.
[22] D. Thornton and Guido van Capelleveen, "Outlier-based Health Insurance Fraud Detection for U.S. Medicaid Data:," 2014, pp. 684–694.
[23] M. Kirlidog and C. Asuk, "A Fraud Detection Approach with Data Mining in Health Insurance," Procedia - Soc. Behav. Sci., vol. 62, pp. 989–994, Oct. 2012.
[24] H. Joudaki, A. Rashidian, B. Minaei-Bidgoli, M. Mahmoodi, B. Geraili, M. Nasiri, and M. Arab, "Using Data Mining to Detect Health Care Fraud and Abuse: A Review of Literature," Glob. J. Health Sci., vol. 7, no. 1, Aug. 2014.
[25] D. Yue, X. Wu, Y. Wang, Y. Li, and C.-H. Chu, "A review of data mining-based financial fraud detection research," in Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on, 2007, pp. 5519–5522.
[26] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," Decis. Support Syst., vol. 50, no. 3, pp. 559–569, Feb. 2011.
[27] S. B. E. Raj and A. A. Portia, "Analysis on credit card fraud detection methods," in Computer, Communication and Electrical Technology (ICCCET), 2011 International Conference on, 2011, pp. 152–156.
[28] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
[29] K.-S. Shin, T. S. Lee, and H. Kim, "An application of support vector machines in bankruptcy prediction model," Expert Syst. Appl., vol. 28, no. 1, pp. 127–135, Jan. 2005.
[30] E. Avci, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm–support vector machines: HGASVM." Expert Systems with Applications, 2009.
[31] S. Theodoridis and K. Koutroumbas, "Pattern recognition and neural networks," in Pattern Recognition, Institute for Space Applications & Remote Sensing, National Observatory of Athens, Greece, 2009.