Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 15 Issue 6 Version 1.0 Year 2015
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

The Study of Fraud Detection infinancial and Credit institutions with Realdata

By Sodasoltaniziba & Mohammad Alibalafr

Islamic Azad University, Iran Islamic Republic Of

Abstract- This paper presents a review of data mining techniques for the fraud detection. Development of information systems such as data due to it has become a source of important organizations. Method and techniques are required for efficient access to data, sharing the data, extracting information from data and using this information. In recent years, data mining technology is an important method that it has changed to extract concepts from the data set. Scientific data mining and business intelligence technology is as a valuable and some what hidden to provide large volumes of data. This research studies using service analyzes software annual transactions related to 20000 account number of financial institutions in the country.The main data mining techniques used for financial fraud detection (FFD) are logistic models, neural networks and decision trees, all of which provide primarysolutions to the problems inherent in the detection and classification of fraudulent data. The proposed method is clustering clients based on client type. An appropriate rule for each cluster is determined by the behavior of group members in case of deviation from specified behavior will be known among suspected cases. The rules of the C5 have been applied in decision tree algorithm. Model is able to extract about a lot of the rules related to client behavior.

Keywords: *datamining,fraud detection, financialfraud, clustering, classification.*

GJCST-C Classification : *D.4.6 H.2.7*

THESTUDYOFFRAUDDETECTIONINFINANCIALANDCREDITINSTITUTIONSWITHREALDATA

Strictly as per the compliance and regulations of:

The Study of Fraud Detection infinancial and Credit institutions with Realdata

Sodasoltaniziba [α] & Mohammad Alibalafr [σ]

Abstract- This paper presents a review of data mining techniques for the fraud detection. Development of information systems such as data due to it has become a source of important organizations. Method and techniques are required for efficient access to data, sharing the data, extracting information from data and using this information. In recent years, data mining technology is an important method that it has changed to extract concepts from the data set. Scientific data mining and business intelligence technology is as a valuable and some what hidden to provide large volumes of data. This research studies using service analyzes software annual transactions related to 20000 account number of financial institutions in the country.The main data mining techniques used for financial fraud detection (FFD) are logistic models, neural networks and decision trees, all of which provide primarysolutions to the problems inherent in the detection and classification of fraudulent data. The proposed method is clustering clients based on client type. An appropriate rule for each cluster is determined by the behavior of group members in case of deviation from specified behavior will be known among suspected cases. The rules of the C5 have been applied in decision tree algorithm. Model is able to extract about a lot of the rules related to client behavior. Each node in the graph model is built by selecting the corresponding table;chance percent of suspected cases have been identified.

Keywords : datamining,fraud detection, financialfraud, clustering, classification.

I. Introduction

Data has become one of important organizations with the development of information systems. The methods and techniques are required for efficient data access, data sharing, data extraction and use of this information. There are many alternative approaches to fraud detection and deterrence (Brockettet al., 2002). Bolton and Hand (2002) discuss techniques used in several subgroups within fraud detection such as credit card and telecommunications, and related domains such as money laundering and intrusion detection. Kou et al (2004) outline techniques from credit card, telecommunications, and intrusion detection. Weather ford (2002) recommends back propagation neural networks, recurrent neural networks and artificial immune systems for fraud detection. Maes et al., (1993) is an example of straight forward application of existing data mining algorithms to an "ideal" data set: it uses neural networks for credit card fraud detection data. The neural network is a technique that imitatesthe functionality of the human brain using a set of interconnected vertices (Yeh and Lien, 2008; Ghosh and Reilly, 1994). It is widely applied in classification and clustering, and its advantages are as follows. First, it is adaptive; second, it can generate robust models; and third, the classification process can be modified if new training weights are set. Neural networks are chiefly applied to credit card, automobile insurance and corporatefraud.

Unfortunately, details about the used features are not given.Currently, identification of fraudulent claims is achieved using a scoring method to implement a claim auditing strategy. In recent years, data mining technology has become as one of the most important concepts extracted from the data set. Because, its technology has provided as scientific intelligence and valuable commercial and it obscured for a large amount of data. Various fields have been identified for data mining applications and developing.Various data mining techniques have been applied in FFD, such as neural networks (Fanning and Cogger, 1998; Dorronsoro et al,1997; Cerullo and Cerullov, 1999;Kirkos et al, 2007), and decision trees [Kirkos et al,2007; Kotsiant is et al, 2006], among others. Data mining techniques covered by survey papers and bibliographies include outlier detection (Hodge and Austin, 2004), skewed/imbalanced/rare classes (Weiss, 2004), sampling (Domingos et al, 2002), cost sensitive learning, stream mining, graph mining (Washio and Motoda, 2003), and scalability (Provost and Kolluri, 1999).

The most common areas can be noted include medical issues, education, production and quality control, retail, and banking and insurance industry as well as marketing and supply chain issues. But one of the applicable the field of data mining is related to client relationship management. Today, there is the high volume of client data in the database and organizations, it is providing the potential for data mining process and hide knowledge extraction. The importance of issues is such as client retention and increase the value and profitability of their companies has rein forced the need to use data mining techniques. The present study is an attempt to analysis of a financial institution and credit

Author α : Department of Computer Engineering, Germi Branch, Islamic Azad University, Germi, Iran. e-mail: akbar.yagmur@gmail.com
Author σ : Department of Communications Engineering, Faculty of Electronic and Computer Enginerring, University of Tabriz, Tabriz, Iran. e-mail: arazazar555@gmail.com

clients, client hide behavioral patterns detection to improve the process of fraud detection in these institutions have taken advantage of it.Financial institutions area mong the organizations tha t interact directly with clients. Therefore,the analysis of client behavior is important to increase their loyalty. In recent years,with increased access to client data and improved data analysis capabilities by intelligent methods, various activities have been carried outto analyze client behavior. One of these activities is the use of intelligent systems for detecting fraud infinancial institutions. Currently fraud is wide range in financial institutions that have been material and immaterial losses many financial institutions and bank clients.

a) Fraud

Fraud involves one or more persons who intentionally act secretly to deprive another of something of value, for their own benefit. Fraud is as old as humanity itself and can take an unlimited variety of different forms. However, in recent years, the development of new technologies has also provided further ways in which criminals may commit fraud (Bolton and Hand, 2002). This approach (fraud detection) makes useof an “operations cycle” and a “development cycle” to detect fraud in health care claims. First, a Peer Group Analysis variant is used to find health care providers which “stand out from the main stream”, which are then presented to a security unit. In the development cycle, rules should be induced based on the expert analysis of the outliers. As in (Kim et al., 2003), these rules are proposed to be cloned and mutated. Details about the development cycle process are not given.In this study,it has tried first;full explanation of the types of fraud was common infinancial institutions, in general and money laundering in particular, and intelligent data mining system to be expressed in the following manner. Fraud from different views is visible, including views of the social, legal, and economic.Fraud in Czech, business, loan and money laundering is instances of fraudin the financial context that our focus is on money laundering (a special type of fraudin the bank with the aim of hiding the true source of money).

There are different definitions for internal fraud, including:

- acts to deceive, exploit or circumvent the law and regulations, with the exception of special events (The Basel Committee on banking supervision, 2006)
- Use of jobs for personal enrichment, intentional abuse or misuse of the resources and assets of the organization (Association of Certified Fraud Examiners, 2008)

According to the Association of FraudExaminers expert, to 959 cases of occupational Fraudas follows: (Association of Certified Fraud Examiners, 2008) organizations in the United States lose to fraudabout 7% of its annual revenue. Analysis of an average of $ 175,000, which lost a quarter of these cases for amounts less than $ 1 million. Instance of fraudsince the beginning of fraudulent behavior to it diagnosis took 2 years.The most common is fraud scheme corruption, fraudulent billing Fraud 27% and 24%, respectively. It seems that the implementation of anti-fraud controls can be largely effective, and etc…

For reasons that are mentioned in the client' employeedriver scould potentially becommitting fraud. (Luell, 2005)

Over the past two decades, the competitive land scape has changedsignificantlyin the banking industry. This is due to factors such asnew regulations, globalization, technology development and product service in to the bank and a significant increase indemand of our clients. Changes in banking activities and the increasing complexity of existing rules in banks are created new topics in the field of bank fraud. Major and Riedinger (2002) describe a workflow and system to setup fraud detection departments with results of its use in the real world. Similar work was done by Ortega et al. (2006), who introduced a data mining based system that decreased the time it takes to detect fraud by 76% from an average of 8.6 months to 2 months. Because Major and Riedinger and Ortega et al., describe real systems that are used to find fraud they cannot go into details of the exact working of the systems. Doing this would give fraud perpetrators an advantage on penetrating the fraud defense.

Fraud detection techniques, in addition to fraud and scams in which an organization has identified and provides analysis ,to some how try to predict the future behavior of their users or clients to will decrease the risk of fraud.

Due to high costs caused by direct or indirect fraud or fraud in financial institutions, banks, financial institutions and money to crooks and fraudsters are aggressively seeking to expedite the recognition activities. The importance of this is due to its direct impact on client service organizations to reduce operating costs and remain as a credible and reliable provider of financial services.

a) Research hypotheses

The main hypothesis of this research is development of data mining combined with pattern matching techniques to construct a scenario with practical and valuable solutions and complete fraud detection system available. Data mining anddata mining algorithms is appropriate to predict large data database.

- Classification, is learning function that categories a data item into one of several classes of predefined (for example, a client classified as "cheaters" or "non-cheaters")

- regression, is learning function that a data item is classified into a true predict (e.g forecast of fraud by a client)
- clustering, is a description that seeks to identify a finite set of categories or clusters are used to data describe. (For example, identify target groups of clients)
- Dependency modeling, is focusing on describing the dependencies and relationships between data (for example, find ways unknown to cheat clients)
- a change and deviation, detection in the identify of data significant changes, with a focus on values, principles or the previous measurement. (For example, find ways to unusual usage patterns of clients Institute for fraud detection)
- Decision tree, which is a powerful tool for prediction and classification a similar structure tree.

II. Materials and Methods

The study is based on data collected has been one of the country's financial and credit institutions and the purpose of this proposal predictable patterns of fraud and to prevent fraud by institution profiteer clients. The study is one of the financial institutions includes transactions on accounts of clients, legal and real. Studied data collected from about 20 thousandsclient accounts during the one-year period, of which about 25 million records are for a variety of clients. So thatclients is divided in three groups of clients, legal client and government and non-government agencies related companies as well as actual clients that is the clients majority of financial institutions.

a) Methods

1) The first, study of account behavior should be specified normal behavior procedure for each account.
2) Due to the large number of account numbers, account individual behavior is not true in this case; it should be defined for each account a certain procedure.
3) It can not specify a fixed behavior pattern for each account. Because it is possible for a client's behavior is a normal behavior and abnormal behavior lead to a different account. For example, you may with draw or deposit the amount of 150 million dollars for a client's normal behavior, but the behavior of a group of clients whousually have low amounts transactions ,are considered to be suspicious behavior.
4) So, data clustering should have been used and data clustered to close together in a cluster to take account behavior.

The study data were based on the type of client clustering, so each cluster represent inga certain type of client, the procedure will havea different behavior. In this study, to determine cases of fraud requires a Boolean data typeas “fraud anticipate " would be the initial value will be equal to 0. If exceed behavior of an account from the normal rang, the value of this field will change to 1. This will be implies for abnormal behavior(suspicious behavior) (figuer3.1).

To predict of fraud using adecision tree used of input data including the type of client,the number of deposit transactions, the total deposit transactions, the number of with draw transactions, the total with draw transaction. In this study, 30 percent of data is used as training data and 70 percent of data as the test data.

Decision tree operator used to fraud predict of test data. This function builds a model based on the training data set and the model of is build predicts attribute toespeciallytarget of test data where it is lacking this amount.

Toolsused in thisresearchare:

1. SqlServer 2012
2. AnalysisService2012
3. Excel2012
4. Data Mining tool has been added to Excel software

Various methods have been used to predict Fraud. One of these methods, the use of neural network is more efficient than other algorithms. This algorith m model learns using neural network trained by error back-propagation algorithm. Architecture of this type of artificial neural network is called multi-layer Perceptr on. The foundation of operator is such a multiple layers considered for it.Construct the inner layer neural network can be defined with the help of Hidden Layers parameter list. Each entry in this list defines a new hidden layer. If the user does not specify any hidden layers it will be added to the network by default a hidden layer. Most fraud departments place monetary value on predictions to maximize cost savings/profit and according to their policies. They can either define explicit cost (Chan et al, 1999) or benefit models.Cahill et al (2002) suggests giving a score for an instance (phone call) by determining the similarity of it to known fraud examples (fraud styles) divided by the dissimilarity of it to known legal examples (legitmate telecommunications account).

III. Results

a) The results of the decision tree

Decision Trees is one of the most powerful tools common to classify and predict. In this section the results of the model based on decision tree algorithm reviewed.In Figure 1-4 likelihood of fraudis shown based on entire of statistical population in used model of decision tree.

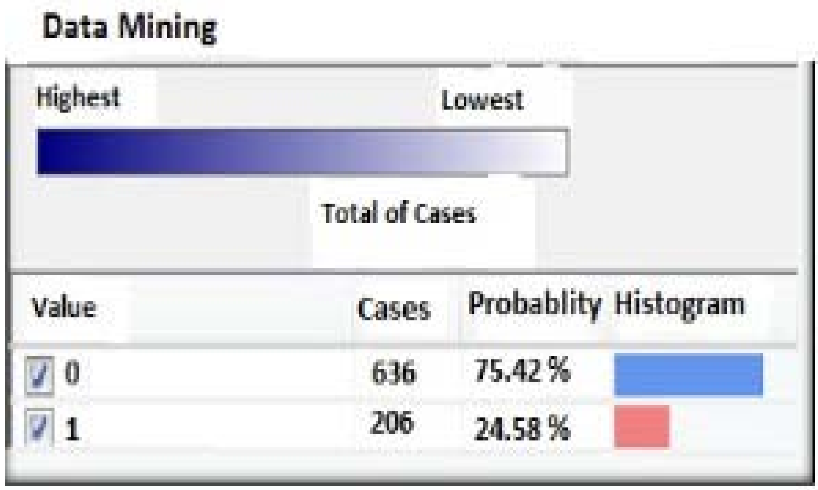

Figure 1 : Possibility of fraud in entire of statistical population

It can be identify fraud probability studied based on type of client in Figure 1. In a survey conducted by the type of client fraud took place is shown in Table 1. According to Table 1in the type of special clients are most likely to fraud and so on based on the legal and real clients will be less likely to fraud. High probability of fraud among clients is due to transactions and amounts of high for this type of clients.

Table 1 : Fraud probability based on client in decision tree

Type ofClients	Fraud Probability (percent)
Special	62.5
Legal	44.85
Real	11.44

Case studies are checked for each node and each branch:

1. there are 62.5% of fraud probability for special clients on the basis of the proposed legislation that it will different according to the amount of the withdraw transaction.

A. fraud probability will be approximately 62.27%, If the number of transactions is equal to 365.

B. If the number of transactions take less or more than 365 transactions, it is likely to reach 73.75%.

2. There are 44.85% of fraud probability for legalclients on the basis of the proposed legislation that it will different according to the amount of the withdraw transaction

A. Fraud probability will be approximately 97%, If the number of transactions is equal to 2920.

B. If the number of transactions take less or more than 2920 transactions, it is likely to reach 44.72%.

3. Thereis 11.44% of fraud probabilityfor realclients on the basis of the proposed legislation that it will different according to the amount of the total ofdeposit transaction.

A. Fraudpossibility is very low, about 0.4 percent, If the total of deposit transaction is less than 414359160 RialIRR.

B. Fraud probability will be 19.84%, If the sum of deposit transaction is less than 569946764 Rial IRR or 414359160 Rials IRR is greater than or equal.

C. Fraud probability will be 69.72%, If the sum of deposit transaction is more than 569946764 Rial IRR.

Figure 2 : Chart of clients'fraud

It can be observed the effective information in model made to study of made tree network model properly. Network diagram created in the tree is presented at figure 3. Greatest impact on the fraud prediction is related to information of withdraw transaction number, client type and sum of deposit transactions in this model.

.

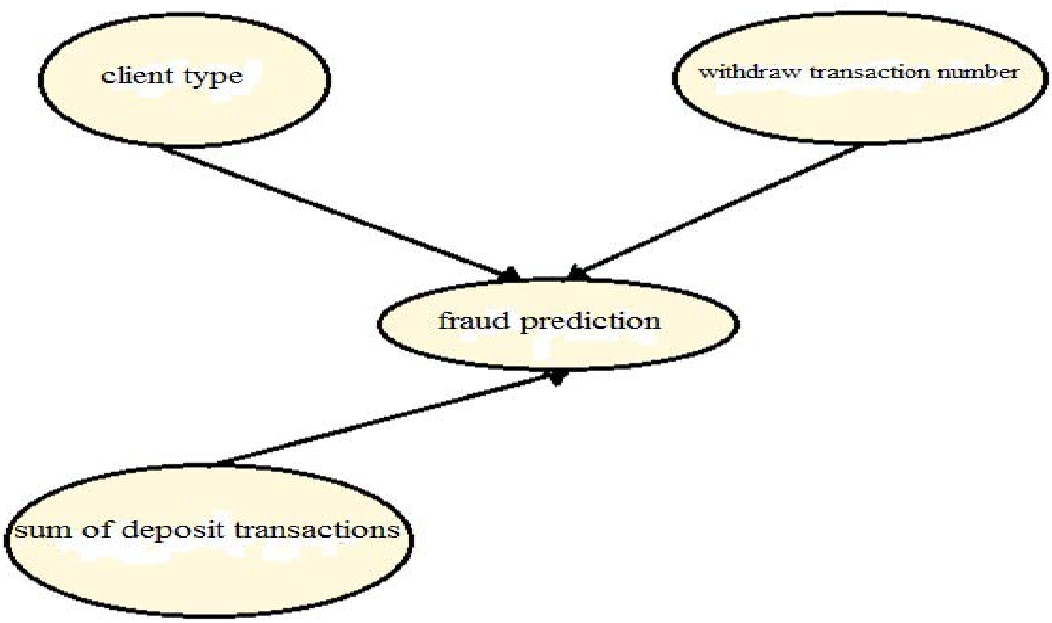

Figure 3 : Diagrams createda network model

The evaluation model is shown in Figure 4. The green line shows in the graph ratio of cheaters population to total of statistical population on the basis of the model (figure 4).As figure 4 shows,possibility of frauds person will be increased with increasing of institution clients'population, as well as, the amount of fraud will be increased in the institutions, accordingly.

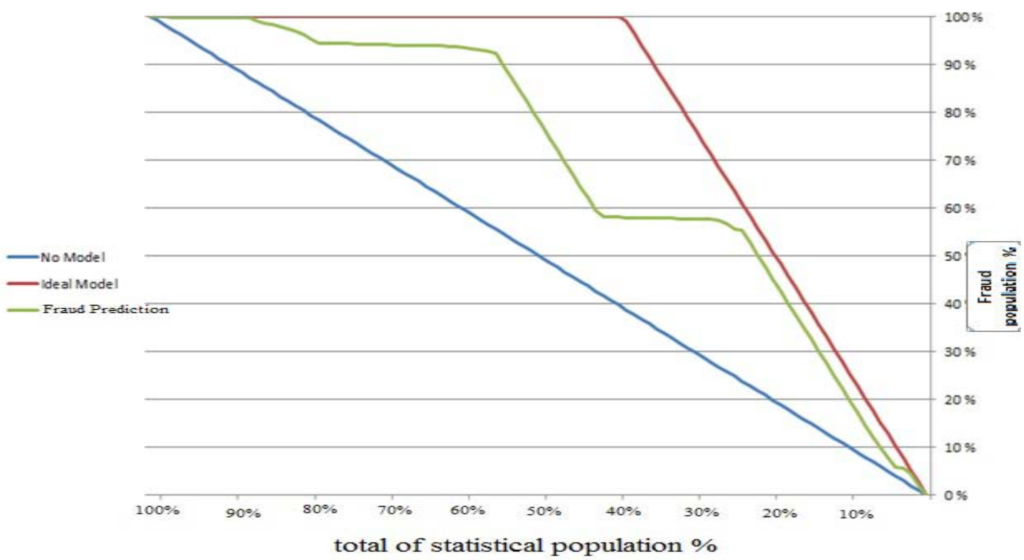

Figure 4 : Ratio of cheaters population to total of statistical population in decision tree model

As figure 5 shows,red and green colorrepresent legal,real and special clients. In this study, the blue and green colors have been allocated the highest and the lowest rate among the total of clients population.

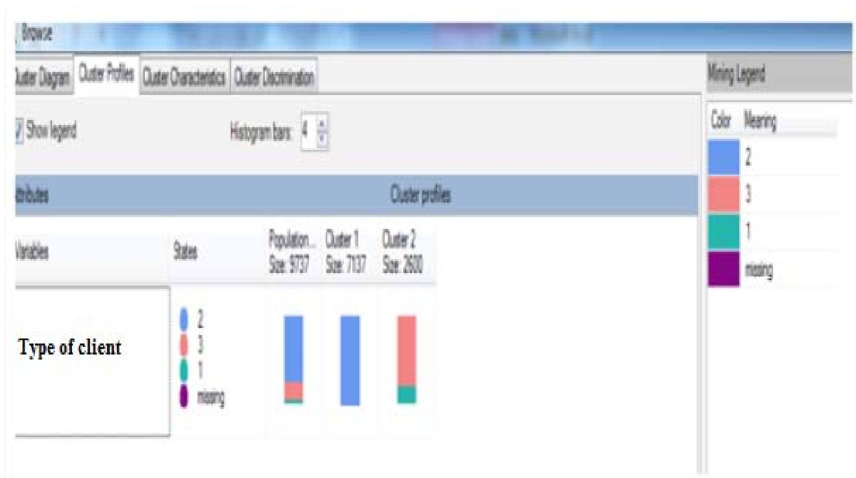

Figure 5 : The total of used population inmodeling

b) The results of the modelbased on neural network

In this section, the applications of neural network have been investigated using Neural Network Algorithm to fraud predict. input data is including the type ofclient,the number of deposit transactions, the total of deposit transactions, the number of with draw transactions, the total of withdraw transaction. Created model is evaluated based on Neural Network operator. As figure 6-4 shows, number of fraudsters will be increase with increasing of total of statistical population. However, the ratio of fraudsters' population to total of statistical population is less comparison with the decision tree model. Breakpoints is much lower In this model (neural network) than decision tree model Which indicates that growth of fraudulent population based on neural network moves a lower slope than the model based on decision tree.

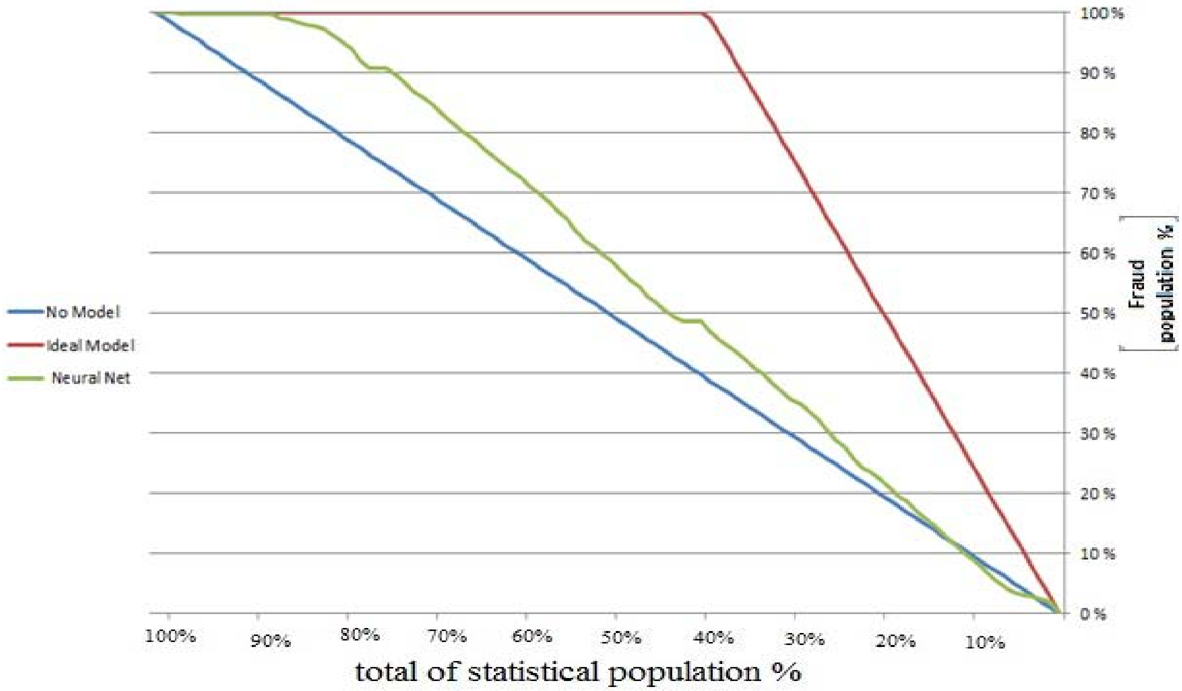

Figure 6 : Fraudulent population ratio to total of population in the neural network model

input data is including the type of client,the number of deposit transactions, the total of deposit transactions, the number of with draw transactions, the total of withdraw transaction. Chart created is shown by label classes as fraudor none-fraud,in Figure7.Figure7 shows redandblue diagram represents thefraud and none-fraud ,basis of input data, respectively.as Figure7 shows,sum and number of deposit transactions have more effective at predicting of fraud.

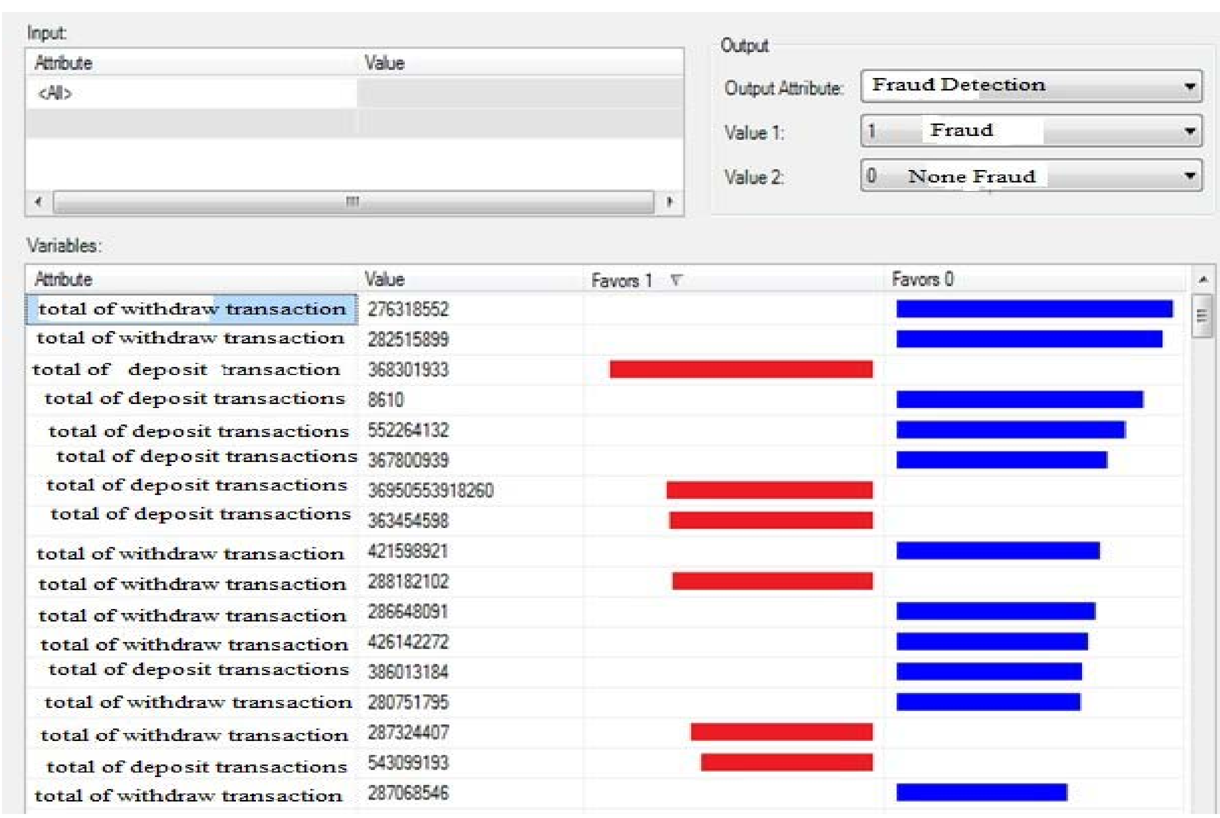

Figure 7 : Chartcreated by classeslabelfor neural network model

IV. Conclusion

Details of transactions is an important source related to the 20,000 financial institution clients account number that it is an important source for data mining. Because, very high volume of information are related to transactions of client's account that represents their behavior and source of high information is dynamic. One of the main problems of financial and credit institutions by extending the jobber clients, increased behavioral diversity and the loss of financial resources and it will be reduce institute consumer confidence. Therefore, terrorist actions of some jobber clients should have been identified and possibly to submit proposals to prevent a fall in the confidence of clients and increased the security of the financial institutions. Inthis study suggest that the use of decision trees because the goal is to predict consumer fraudand financial institutions, after mining the type of anticipation. Modelbased on decision tree algorithm better, it has higher accuracy and speed than neural network. When the decision tree due to prune the tree after the tree branches that the risk of fraud is minimal, they will be removed.This method makes it easier to evaluate the model. Missing data very little in model created in detail of transactions on client accounts, but the accumulation of information and the use of CIF s data preparation phase is very important. If the predictive of fraud is respect to client account transactions behavior it should be based on the type of client, transactions are aggregated. After this stage, the aggregated data should be normalized. Because there is no specific measure to analyze the obtained data are very hard, but if the data has a high or low level that they are better understood. Therefore,data should be normalized. One of the benefits of the details of transactions on client accounts too ther clients of the financial institution and credit characteristics, including characteristics such as age, place of residence, education, address, etc. The data mining techniques of outlier detection and visualization have seen only limited use. The lack of research on the application of outlier detection techniques to FFD may be due to the difficulty of detecting outliers. Indeed, Agyemang et al. (2006)point out that outlier detection is a very complex task a kin to finding a needle in a hay stack. Distinct from other data mining techniques, out lierdetection techniques are dedicated tofinding rare patterns associated with very few data objects. In thefield of FFD, outlier detection is highly suitable for distinguishing fraudulent data from authentic data, and thus deserves more investigation. Fanning and Cogger (1998) highlight the challenge of obtaining fraudulentfinancial statements, and note that this creates enormous obstacles in FFD research.it is concluded that its attention towardfinding more practical principles and solutions for practitioners to help them to design, develop, and implement data mining and business intelligence systems that can be applied to FFD.The data is quite dynamic, and the data revealed, if the behavior of client changes the above-mentioned. Data normalization is better after aggregation of information on the different patterns of clients' behavior. In this study,it concluded that the possibility of fraud was high for many special clients and it may have transport with different way so such as dirty money through the financial institution, and legal clients through the creation of fictitious institution want to make money launde ring that registration of the company or institution should be considered very care fully to prevent such acts. Conclusions are as follows:

1) The presence of special clients in each institutionwill benefittoinstitutionsand it is very useful to riseof institutere sources. On the other hand,based on the results, the possibility of fraudis more among clients.There fore, more focusis needed to uncover cases of fraud this group of clients.
2) Measures should be considered in addition trust of legal clients, transactions of these clients more control, as well as the delivery of documents when opening an account for legal clients more controlin order to reduce the amount of fraud of group of clients.
3) Measures should beconsidered to determine actual client behavior groups. In this case the behavior of a client can be more easily analyz e dandidentified suspicious.

References Références Referencias

1. Agyemang, M.; Barker, K.; Alhajj, R. (2006): A comprehensive survey of numeric and symbolic outlier mining techniques, Intelligent Data Analysis 10 (6) : pp.521–538.
2. Association of Certified Fraud Examiners. (2008): Report to the Nation on occupational Fraud and abuse, p 4.
3. Bolton, R.; Hand, D. (2001): Unsupervised profiling methods for fraud detection. Credit scoring and credit control VII.
4. Bolton, R.; Hand, D. (2002): Statistical fraud detection: A review. Statistical Science, 17(3):pp.235–255.
5. Brockett, P. L.;Derrig, R. A.; Golden, L. L.; Levine, A.;Alpert,M. (2002): Fraud classification using principal component analysis of ridits. The Journal of Risk and Insurance, 69:pp.341–371.
6. Cahill, M.; Chen, F.; Lambert, D.;Pinheiro, J.;Sun, D. (2002): Detecting fraud in the real world. In Handbook of Massive Datasets, pp. 911–930. Kluwer Academic Publishers.
7. Chan, P.; Fan, W.; Prodromidis, A.; Stolfo, S. (1999): Distributed data mining in credit card fraud detection. IEEE Intelligent Systems, 14:pp.67–74.

8. Cerullo, M..J.; Cerullo, V. (1999): Using neural networks to predict financial reporting fraud, Computer Fraud and Security May/June 14–17.
9. Core Principles for Effective Banking Supervision. (2006): Basel committee on banking supervision, Basel, October 2006.
10. Domingos, C.; Gavalda, R.; Watanabe, O. (2002). Adaptive sampling methods forscaling up knowledge discovery algorithms. Data Mining and Knowledge Discovery 6: pp. 131-152.
11. .Dorronsoro, J.R.;Ginel, F.; Sánchez, C. ;Cruz, C.S. (1997) :Neural fraud detection in credit card operations, IEEE Transactions on Neural Networks 8 (4).pp. 827–834.
12. Fanning, K. ;Cogger, K. O. (1998): Neural network detection of management fraud using published financial data. In International Journal of Intelligent Systems in Accounting, Finance and Management, pp. 21–24.
13. Ghosh S. ; Reilly, D.L.(1994) : Credit card fraud detection with a neural-network, 27th Annual Hawaii International, Conference on System Science 3.pp. 621–630.
14. Hodge, V; Austin, J. (2004):A Survey of outlier detection methodologies. Artificial Intelligence Review 22: pp.85-126.
15. Kim, J.; Ong, A.;Overill, R. E. (2003): Design of an artificial immune system as a novel anomaly detector for combating financial fraud in the retail sector. In IEEE Congress on Evolutionary Computation (1), pp. 405–412.IEEE.
16. Kirkos, E.; Spathis, C.;Manolopoulos, Y. (2007) :Data mining techniques for the detection of fraudulent fi nancial statement s, Expert Systems with Applications 32 (4). pp.995–1003.
17. .Kotsiantis, S.;Koumanakos, E. ;Tzelepis, D.;Tampakas, V.(2006) :Forecasting fraudulent financial statements using data mining, International Journal of Computational Intelligence 3 (2). pp.104–110.
18. Kou, Y.; Lu, C.; Sir wongwattana, S.; Huang, Y. (2004): Survey of Fraud Detection Techniques. International Conference on Networking, Sensing, and Control. 2004, pp. 749-754.
19. Luell, J. (2005): Analytical fraud detection. Master's thesis, University of Zurich.
20. Maes, S.; Tuyls, K.; Vanschoenwinkel, B.;Manderick, B. (1993): Credit card fraud detection using bayesian and neural networks. In In: R.J.Maciunas, editor. Interactive image-guided neurosurgery. American Association Neurological Surgeons, pp. 261–270.
21. Major, J. A.; Riedinger, D. R. (2002): A hybrid knowledge/statistical-based system for the detection of fraud. Journal of Risk and Insurance.
22. Ortega, P. A.; Figueroa, C.J.;Ruz, G.A.(2006): A medical claim fraud/abuse detection system based on data mining: a case study in Chile. DMIN 2006; 6, pp.26-29.
23. Provost, F.; Kolluri,V. (1999): A survey of methods for scaling up inductive algorithms. Data Mining and Knowledge Discovery 3 (2): pp. 131-169.
24. Washio, T.; Motoda, H. (2003): State of the art of graph-based data mining. SIGKDD Explorations 5(1): pp. 61-70.
25. Weatherford, M. (2002). Mining for fraud. IEEE Intelligent Systems July/August: pp. 4-6.
26. Weiss, G. (2004). Mining with rarity: A Unifying Framework. SIGKDD Explorations 6(1): pp. 7-19.
27. Yeh, I .; Lien, C.(2008):The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients, Expert Systems with Applications 36 (2). pp. 2473–2480.

Global Journals Inc. (US) Guidelines Handbook 2015

www.GlobalJournals.org