Fraud: The Affinity of Classification Techniques to Insurance Fraud Detection

Saliu Adam Muhammad

Abstract- Quite a large number of data mining techniques employed in financial fraud detection (FFD) are seen to be classification techniques. In this paper, we developed an algorithm to find the features of classification techniques (or method) that so much place it (classification techniques) in the heart of researchers in their various efforts in the study of insurance frauds detection. We also got to know the characteristics of insurance frauds data that made data mining classification techniques so much attracted to it (insurance data).

Keywords - affinity, classification techniques, insurance frauds common features.

I. INTRODUCTION

Financial fraud has become a global problem. It ranges from insurance fraud, credit card fraud, telecommunications fraud, and check forgery [1]. Various bodies: security outfits, regulatory agencies, anti-fraud bodies, governments and so on, have reported cases of high profiles of financial fraud. Theses frauds are perpetrated by means of falsifying techniques that appear to be true in the face of the victims. Fraud is a global phenomenon that has been in existence long before now, and has continued to increase geometrically. Fraud is defined as nothing but a wrongful or criminal ploy for financial or personal gains [7]. In US Law, fraud is defined as "a false representation of a matter of fact - whether by words or by conduct, by false or misleading allegations, or by concealment of what should have been disclosed - that deceives and is intended to deceive another so that the individual will act upon it to her or his legal injury" [10].

A striking case of financial fraud is in Ponzi scheme, perpetuated by the former NASDAQ chairman, Bernard Madoff, which led to the loss of about US$50 billion worldwide [8]. Another outstanding case is that of Joseph Hirko, former co-chief executive officer of Enron Broadband Services (EBS), who after pleading guilty to wire fraud, avowed to forfeit approximately US $8.7 million in restitution to Enron victims through the U.S. Securities and Exchange Commission's Enron Fair Fund [8]. In 2007 BBC news report, fraudulent insurance claims cost UK insurers a total of 1.6 billion pounds a year [9].

Happily, data mining and statistical methods have been proved to successfully detect fraudulent activities such as money laundering, e-commerce credit card fraud, telecommunications fraud, insurance fraud, and computer intrusion [6]. Researchers have employed a number of these techniques to detect these fraudulent activities where prevention measures failed. They employed such data mining techniques as Decision Trees, Bayesian Network, Rule Base Network, Support Vector Machines, Neural Network, and so on, in their various attempts to develop models that would help in detecting these frauds in credit card system, telecommunication systems, banking systems, insurance systems and so on. Apparao *et al*, in their analysis of financial fraud statement, discovered that out of the twenty-six (26) data mining based techniques for financial fraud detection, all the 26 (100%) techniques are found to be classification techniques [6]. Ngai et al, reviewed financial fraud detection on the basis of classification framework. They discovered that classification models are mostly used on insurance fraud detection [8]. Clearly, classification techniques of data mining algorithms are seen to be most engaged by researchers in their fraud detection efforts. This development has prompted our attention to find out the features of the classification techniques that make them so amenable to insurance fraud detections and those of insurance fraud data that so much make them so inclined to classification methods.

Manuscript published on 30 April 2014.
*****Correspondence Author(s)
Saliu Adam Muhammad, Department of Computer Technology Application, School of Information Science and Engineering, Hunan University, Changsha, Hunan Province, 410082, P.R. China.

II. REVIEW OF RELATED LITERATURE

Insurance fraud is common in automobile, travel and telecommunication industries, and money laundery. Insurance can be stated as a contract (policy) in which an individual or entity receives financial protection or reimbursement against losses from an insurance company [7]. Classification techniques can be applied easily to categorize crime data and as such have proved to be very useful in fraud detection environment. The proven efficacy of classification methods in fraud detection has justified is applicability in categorizing crime data. A realistic cost model was used to evaluate C4.5, CART, and naïve Bayesian classification models, in a distributed data mining model which was applied to credit card transactions [2]. A classification technique with fraud/legal attribute, and a clustering followed by a classification technique with no fraud/legal attribute were recommended, to credit fraud model development [3]. Maes et al, used STAGE algorithm and backpropagation for Bayesian Belief Network (BBN) and Artificial Neural Network (ANN) respectively, in the study of automated credit card fraud detection [4]. Rekha, employed classification tasks of decision tree and Bayesian Network techniques of data mining to detect frauds in an auto insurance company [1]. Classification, association rules and cluster detection techniques were employed by SAS Enterprise Miner Software to detect fraudulent claims in insurance industry [5].

III. RESEARCH APPROACH

In this research work we proposed and applied the algorithm in Fig. 1 to determine the affinity (a similarity or connection) between data mining classification techniques and insurance fraud data. We carried out a study into some selected classification techniques and insurance fraud data to come out with the relationships between their futures.

To distinguish between objects of different classes, classification builds up and utilizes a model to predict the categorical labels of unknown objects of the different classes [8]. According to Zhang and Zhou, classification and prediction is the procedure followed to identify a set of common features and models that describe and distinguish data classes or concepts [11]. These categorical labels are predefined, discrete and unordered. At different stage of application, eligibility, rating, billing and claims, insurance frauds can be committed by any of consumers, agents/brokers, insurance company employees, healthcare providers, and others [8], [15], and [20].

For our study, insurance fraud includes crop, healthcare, and automobile insurance fraud. Healthcare fraud has been adjourned to be carried out by many segments of the healthcare system through “Billing for Services not Rendered, Upcoding of Services, Upcoding of Items, Duplicate Claims, Unbundling, Excessive Services, Unnecessary Services and Kickbacks” [14]. Purchasers of crop insurance who fake or overstate either the loss of their crops due to natural disasters or the loss of revenue due to declines in the price of agricultural commodities, commit crop insurance fraud [8]. Automobile insurance fraud encompasses a set of fraudulent activities that include thespian accidents, excessive repairs, and fictitious personal injuries [8]. Similarly, we have chosen; neural networks, Naïve Baye’s technique, decision trees and support vector machines as the common classification techniques whose features are determined and the common ones (features) among them determined.

```
1. Select insurance category
    Determine the features of each category
    Relate the features of all categories
    Extract the common features
2. Select classification category
Determine the features of each category
Relate the features of all categories
    Extract the common features
3. Relate (1) and (2)
Determine the relationship between their data
```

Fig. 1: An Algorithm for Determining the Relationship between Classification Models and Insurance Fraud Data

A. The Classification Techniques

Classification is a data mining task of predicting the value of a categorical variable (target or class) by building a model based on one or more numerical and/or categorical variables (predictors or attributes) [16], and [18]. It is: (1) a major data mining operation, (2) can predict the value of a new attribute by means of some of the other available attributes, and (3) applied to categorical outputs.

(i) Neural Networks

Neural Network (NN) is an adaptive general purpose mechanism for training a machine by examples. Neural networks modelled after the human brain, which are perceived as highly connected network of neurons, are classified as artificial intelligence because of their ability to learn and their basis in biological activities. It has three parts (layers) [12]: (1) an input layer consisting of nodes which represent the predictor variables, (2) a hidden layer consisting of nodes that do the computation (3) the output layer consisting of nodes which represents the target variable. It has the learning ability of classification, association, generalization, feature extraction, optimization and noise immunity - fault tolerance.

(ii) Naïve Bayes’ Technique

The Naive Bayesian classifier is a simple probabilistic classifier based on Bayes’ theorem with independence assumptions between predictors. In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature [19]. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms a number of more sophisticated classification methods [13]. Naive Bayes classifier has an apparent merit of a small amount of training data requirement to estimate the parameters (means and variances) necessary for classification [24, 26]. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting: robust to isolated noise points, handles missing values, and vigorous to irrelevant attributes.

(iii) Decision Trees

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed [21]. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor is called root node. Decision trees can handle both categorical and numerical data. Decision trees have many appealing properties [22]: Similar to human decision process, easy to understand, deal with both discrete and continuous features, and highly flexible.

(iv) Support Vector Machines

A Support Vector Machine (SVM) performs classification by finding the hyperplane that maximizes the margin between the two classes. The vectors (cases) that define the hyperplane are the support vectors [30]. In the parlance of SVM literature, a predictor variable is called an attribute, and a transformed attribute that is used to define the hyperplane is called a feature. The task of choosing the most suitable representation is known as feature selection.

A set of features that describes one case is called a vector. So, the goal of SVM modelling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the support vectors [25]. It has the properties of: high accuracy, high flexibility, handling of large dimensional data, and sparse representation of the solutions - fast for making future prediction/classification. In general, every classification model or method uses a set of relevant features or parameters to characterize an object. Here, we consider the approach whereby a set of known objects called the training set is used by the classification program to learn how to classify objects. This approach is called supervised learning. To construct any classifying program (a classifier), two phases are involved [17]: (a) the training phase - the training set is used to decide how the parameters ought to be weighted and combined in order to separate the various classes of objects, (b) the application phase - the weights determined in the training set are applied to a set of objects that do not have known classes in order to determine what their classes are likely to be.

B. Insurance as a Policy

Insurance is a way of managing risks. When you buy insurance, you transfer the cost of a potential loss to the insurance company in exchange for a fee, known as the premium [23]. Insurance is an agreement where, for a stipulated payment called the premium, one party (the insurer) agrees to pay to the other (the policyholder or his designated beneficiary) a defined amount (the claim payment or benefit) upon the occurrence of a specific loss [31]. The purpose of insurance is to protect against the event of a financial loss [32]. For example:

(a) Auto insurance could pay the cost of repairs to your vehicle if you have an accident.

(b) Crop insurance could provide finance against loss of crops.

(c) Health insurance could pay for the cost of health problems

But, with the good intent of this system, some fraudsters (at various levels and categories) are seizing the chance to defraud the insurers. This is why researchers/scientist and software practioners are up and doing in developing methods and software products to help combat these fraudsters.

(i) Crop Insurance Fraud

Crop insurance is purchased by agricultural producers, including farmers, ranchers, and others to protect themselves against either the loss of their crops due to natural disasters, such as hail, drought, and floods, or the loss of revenue due to declines in the prices of agricultural commodities [28]. The two general categories of crop insurance are called crop-yield insurance and crop-revenue insurance. Crop insurance frauds were found to be committed through faking and overstating of either the loss of crops due to natural disasters or the loss of revenue due to declines in the price of agricultural commodities [8].

(ii) Healthcare Insurance Fraud

Health insurance fraud is described as an intentional act of deceiving, concealing, or misrepresenting information that results in health care benefits being paid to an individual or group. Fraud can be committed by both a member and a provider. Member fraud consists of ineligible members and/or dependents, alterations on enrolment forms, concealing pre-existing conditions, failure to report other coverage, prescription drug fraud, and failure to disclose claims that were a result of a work related injury [29]. Provider fraud consists of claims submitted by bogus physicians, billing for services not rendered, billing for higher level of services, diagnosis or treatments that are outside the scope of practice, alterations on claims submissions, and providing services while under suspension or when license have been revoked. It is also committed through duplicate claims, unbundling, excessive services, unnecessary services and kickbacks [14].

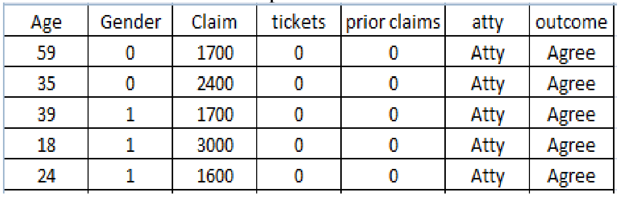
(iii) Automobile Insurance Fraud

Auto insurance fraud is a method by which an individual or insurance company claims more money than they are entitled to [27]. Automobile insurance fraud can involve a staged car accident or a real accident with bills that hide the false claims or state that the costs are higher than they actually are. Insurance companies can also participate in insurance fraud by denying the claims and benefits of victims who deserve them. One in three auto insurance claims are fraudulent. Automobile insurance fraud encompasses a set of fraudulent activities that include thespian accidents, excessive repairs, and fictitious personal injuries [8]. Generally, the datasets for any of these categories (crop, health, and auto) of insurance frauds are mostly derived from claim forms meant for the beneficiaries of the premiums. Such datasets have common features or attributes as: name, amount, rating, attorney class (legal/fraud), etc. The datasets may also contain varied features depending on the insurance fraud category (yield, for crop insurance fraud). Tables 1 and 2 are sample datasets for respectively, real and normalized forms of auto insurance fraud. This shows the convertibility of insurance datasets, which makes them readily open to to classification methods [33], [34].

IV. RESEARCH FINDINGS AND CONCLUSION

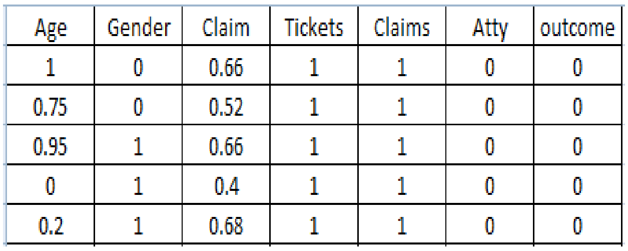
We applied our algorithm (Fig.1) to the selected classification algorithms of subsection A, and the insurance fraud data of subsection B; we obtained summarized results as shown in Tables 3 and 4 respectively. We are able to note that the common features (supervised learning method, two-phase process of training and testing, and applicability to categorical data) identified with these classification algorithms (Table 3) are the features of the classification algorithms that are applicable to features of the insurance datasets (data values being categorical or convertible to categorical forms, dataset divisible into training and testing sets, dataset divisible into predictors set and a class set), for the purpose of separating them (the insurance dataset) into different groups. This, perhaps, explains the “affinity” nature of classification algorithms to the fraud data, particularly the insurance fraud data, as our case in this study. Thus, the inclination of researchers towards the employment of classification models to fraud datasets can be anchored on the basis of this relationship between classification algorithms and the categorical nature of insurance fraud datasets.

Table 1: A real sample dataset for insurance fraud.



Age	Gender	Claim	tickets	prior claims	atty	outcome
59	0	1700	0	0	Atty	Agree
35	0	2400	0	0	Atty	Agree
39	1	1700	0	0	Atty	Agree
18	1	3000	0	0	Atty	Agree
24	1	1600	0	0	Atty	Agree

Table 2: A normalized sample dataset for insurance fraud.



Age	Gender	Claim	Tickets	Claims	Atty	outcome
1	0	0.66	1	1	0	0
0.75	0	0.52	1	1	0	0
0.95	1	0.66	1	1	0	0
0	1	0.4	1	1	0	0
0.2	1	0.68	1	1	0	0

Table 3: Common features of the selected algorithms

Selected classification algorithms	**Common features**
Neural network, Bayesian network, decision tree, and support vector machine	a) Supervised learning method b) Two-phase process of training and testing c) Applicability to categorical data

Table 4: Common features of the selected insurance frauds category.

Selected insurance frauds category	Common features
Auto insurance, crop insurance, and health insurance	a) Data values being categorical or convertible to categorical ones b) Dataset divisible into training and testing sets c) Dataset divisible into predictors set and a class set (target)

Classification models are designed to group data into their various categories and insurance datasets are mostly categorical in nature or convertible into categorical forms. This confirms the reason behind the use of classification methods by researchers in detecting insurance frauds. While we encourage that more efforts are expended in the standardization of classification models in the realm of insurance fraud detection, efforts should also be deployed by researchers to other algorithms to see their light in the realm of insurance fraud detection in particular, and financial fraud detection in general.

REFERENCES

1. B. Rekha, "Detecting auto insurance fraud by data mining techniques", Journal of Emerging Trends in Computing and Information Sciences, Volume 2 No.4, Page: 156 – 162, APRIL 2011.
2. R. Chen, M. Chiu, Y. Huang, L. Chen, "Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines". In: IDEAL 2004, Page: 800 - 806, (2004).
3. R. Groth,. Data Mining: A Hands-on Approach for Business Professionals, Prentice Hall, pp. 209-212(1998).
4. S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, "Credit card fraud detection using Bayesian and Neural Networks". Proc. of the 1st International NAISO Congress on Neuro Fuzzy Technologies (2002).
5. SAS, e-Intelligence Data Mining in the Insurance industry: Solving Business problems using SAS Enterprise Miner Software. White Paper (2000).
6. G.Apparao, S. Arun, G.S. Rao, B. LalithaBhavani, K. Eswar, D. Rajani, "Financial statement fraud detection by data mining", Int. Journal of Advanced Networking and Applications, Volume: 01 Issue: 03 Pages: 159-163 (2009).
7. H.S. Lookman and T. Balasubramanian, "Survey of insurance fraud detection using data mining techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-3, Page: 62 – 65, February 2013.
8. E.W.T. Ngai, H. Yong, Y.H. Wong, C. Yijun and S. Xin, "The application of data mining techniques in financial fraud detection", A Classification Framework and an Academic Review of Literature, Decision Support Systems, Elsevier, 50, Page: 559–569, (2011).
9. I. Bose, and R. K. Mahapatra, Business Data Mining — A Machine Learning Perspective, Information Management 39 (3), Page: 211–225, (2001).
10. Fraud – The Free Dictionary, Online LL.M – In US Law. http://legal-dictionary.thefreedictionary.com/fraud
11. D. Zhang, and L Zhou., "Discovering golden nuggets: data mining in financial application", IEEE Transactions on Systems, Man and Cybernetics 34 (4) (2004) Nov.
12. A. P. Sampson, "Comparing classification algorithms in data mining", Central Connecticut State University, 2012.
13. http://www.google.com.hk/#newwindow=1&q=naive+bayesian+-+Classification%2C+Dr.+Saed&safe=strict
14. FBI, "Federal bureau of investigation, financial crimes", Report to the Public Fiscal Year, Department of Justice, United States, 2007, http://www.fbi.gov/publications/financial/fcs_report2007/financial_crime_2007.htm.
15. Coalition against Insurance Fraud, "Learn about fraud," http://www.insurancefraudorg/learn_about_fraud.htm
16. W. M. Andrew, "Decision trees". www.cs.cmu.edu /~cga/ai-Course/dTree, (2001).
17. Methods for Classification - http://www.sundog.stsci.edu/rick/SCMA/node2.html
18. Classification, http://www.saedsayad.com/ classification.htm
19. Naïve Bayes Classify, http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Naive_Bayes_classifier.html
20. J.L. Kaminsk, "Insurance Fraud", OLR Research Report, http://www.cga.ct.gov/2005/rpt/2005-R-0025.htm. 2004.
21. www.chem-eng.utoronto.ca/~datamining/.../Decision_Trees
22. www.classes.engr.oregonstate.edu/eecs/.../decisiontree-5-part2
23. http://www.cooperators.ca/en/Answer-Centre/how-does-insurance-work/why-do-we-need-insurance.aspx
24. P. Bhargavi, S. Jyothi., "Applying naive Bayes data mining technique for classification of Agricultural Land Soilś", International Journal of Computer Science and Network Security, vol.9, No.8, Page: 117-122, (2009).
25. SVM – "Support vector machines", Introduction to Support Vector Machine (SVM) Models, http://www.dtreg.com/svm.htm
26. D. Qingshan, "Detection of fraudulent financial statements based on Naïve Bayes Classifier", The 5th International Conference on Computer Science & Education, Hefei, China. August 24–27, Page: 1032 – 1035, (2010).
27. wiseGEEK, "What is auto insurance fraud?", http://www.wisegeek.com/what-is-auto-insurance-fraud.htm#
28. Crop Insurance, Wikipedia – The Free Encyclopedia http://en.wikipedia.org/wiki/Crop_insurance
29. Insurance Fraud, Wikipedia – The Free Encyclopedia http://en.wikipedia.org/wiki/Insurance_fraud#Health_care_insurance
30. http:// www.chem-eng.utoronto.ca/~datamining/Presentations/SVM.pdf
31. F. A. Judy and L. B. Robert, (FSA), Education and Examination Committee of the Society of Actuaries Risk and Insurance, Copyright 2005 by the Society of Actuaries, P-21-05 Printed in U.S.A. Second Printing.
32. FCAC, Financial Consumer Agency of Canada, Understanding Insurance Basics, February, 2011.

33. L. Jenn-Long and Chien-Liang C., Application of Evolutionary Data Mining Algorithms to Insurance Fraud Prediction, 2012 IACSIT Hong Kong Conferences, IPCSIT vol. 25 (2012) © (2012) IACSIT Press, Singapore.
34. D. Olson, and Y. Shi, Introduction to Business Data Mining, McGraw-Hill Education, 2008, pp. 75 – 77.

AUTHOR PROFILE

Saliu A. M., received B. Tech. Mathematics/Computer Science from Federal University of Technology, Minna, Niger State - Nigeria, MSc. Computer Science from Abubakar Tafawa Balewa, Bauchi, Bauchi State – Nigeria. He is currently a PhD. Student in Computer Science Technology Application at School of Information Science and Engineering, Hunan University, Changsha, Hunan Province – PR. China. He was a lecturer in the Department of Mathematics/Computer Science and currently a lecturer in the Department of Computer Science, School of Information & Communication Technology, Federal University of Technology, Minna, Niger State – Nigeria. He has authored and co-authored *nine papers* in Journals (National & International). He has also participated in *ten Conferences* (all national) – with *four papers* in Book of Proceedings, three presentations and *three* without presentation.