Credit Card Fraud Detection Using Machine Learning As Data Mining Technique

Conference Paper · August 2018

CITATIONS
0

READS
6,648

3 authors, including:

Saravanan Sagadevan
Universiti Sains Malaysia
3 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Nurul Malim
Universiti Sains Malaysia
53 PUBLICATIONS **136** CITATIONS
SEE PROFILE

Credit Card Fraud Detection Using Machine Learning As Data Mining Technique

Ong Shu Yee, Saravanan Sagadevan and Nurul Hashimah Ahamed Hassain Malim
School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia.
nurulhashimah@usm.my

***Abstract*—The rapid participation in online based transactional activities raises the fraudulent cases all over the world and causes tremendous losses to the individuals and financial industry. Although there are many criminal activities occurring in financial industry, credit card fraudulent activities are among the most prevalent and worried about by online customers. Thus, countering the fraud activities through data mining and machine learning is one of the prominent approaches introduced by scholars intending to prevent the losses caused by these illegal acts. Primarily, data mining techniques were employed to study the patterns and characteristics of suspicious and non-suspicious transactions based on normalized and anomalies data. On the other hand, machine learning (ML) techniques were employed to predict the suspicious and non-suspicious transactions automatically by using classifiers. Therefore, the combination of machine learning and data mining techniques were able to identify the genuine and non-genuine transactions by learning the patterns of the data. This paper discusses the supervised based classification using Bayesian network classifiers namely K2, Tree Augmented Naïve Bayes (TAN), and Naïve Bayes, logistics and J48 classifiers. After preprocessing the dataset using normalization and Principal Component Analysis, all the classifiers achieved more than 95.0% accuracy compared to results attained before preprocessing the dataset.**

Index Terms—Credit Card; Data Mining; Fraud Detection; Machine Learning.

I. INTRODUCTION

According to Global Payments Report 2015, credit card is the highest used payment method globally in 2014 compared to other methods such as e-wallet and Bank Transfer [1]. The huge transactional services are often eyed by cyber criminals to conduct fraudulent activities using the credit card services. Credit card fraud is defined as the unauthorized usage of card, unusual transaction behavior, or transactions on an inactive card [2]. In general, there are three categories of credit card fraud namely, conventional frauds (e.g. stolen, fake and counterfeit), online frauds (e.g. false/fake merchant sites), and merchant related frauds (e.g. merchant collusion and triangulation) [3].

In the past couple of the years, credit card breaches have been trending alarmingly. According to Nilson Report, the global credit card fraud losses reached $16.31 billion in 2014 and it is estimated that it will exceed $35 billion in 2020 [4]. Therefore, it is necessary to develop credit card fraud detection techniques as the counter measure to combat illegal activities. In general, credit card fraud detection has been known as the process of identifying whether transactions are genuine or fraudulent. As the data mining and machine learning techniques are vastly used to counter cyber-criminal cases, scholars often embraced those approaches to study and detect credit card fraud activities.

Data mining is known as the process of gaining interesting, novel and insightful patterns as well as discovering understandable, descriptive and predictive models from large scale of data collections [5, 6]. The ability of data mining techniques to extract fruitful information from large scale of data using statistical and mathematical techniques would assist credit card fraud detection based on differentiating the characteristics of common and suspicious credit card transactions. While data mining focused on discovering valuable intelligence, machine learning is rooted in learning the intelligence and developing its own model for the purpose of classification, clustering or so on.

The application of machine learning techniques spreads widely throughout computer sciences domains such as spam filtering, web searching, ad placement, recommender systems, credit scoring, drug design, fraud detection, stock trading, and many other applications. Machine Learning classifiers operate by building a model from example inputs and using that to make predictions or decisions, rather than following strictly static program instructions. There are many different types of machine learning approaches available with the intentions to solve heterogeneous problems. Due to the nature of this study which was focused on classification, the discussion that follows is based on this topic. Machine learning classification refers to the process of learning to assign instances to predefined classes. Formally, there are several types of learning such as supervised, semi-supervised, unsupervised, reinforcement, transduction and learning to learn [7]. As the interest of this study was to conduct supervised based machine learning classification, the discussions about the rest of the methods are discarded from further elaboration. In most classification studies, supervised-based learning is favoured more than other methods due to the ability to control the classes of the instances with the interventions of human. In supervised learning, the classes of the instances would be labeled prior to feeding into classifiers. Then, by using certain evaluation metrics, the performances of the classifiers could be measured.

In the case of credit card fraud detection, the binary classification technique was employed due to the instances labeled as fraud and non-fraud. The inputs were transformed as Boolean $x = (x1,\ldots, xj)$, where $xj = 1$, if the jth characteristics appeared in the instances, but otherwise, $xj = 0$. A classifier input a training set into (xi, yi), where $xi = (xi, \ldots, xq)$ was an observed input and yi was the corresponding output of the classifier. The rest of the paper is organized into background studies, research methodology, results, discussions and conclusions.

II. BACKGROUND STUDIES

Data mining and machine learning are popular methods to study and combat the credit card fraud cases. There is a large number of studies that exploited the strength of data mining and machine learning to prevent the credit card fraudulent activities. Based on Self-Organizing Map and Neural Network, the study of [8] obtained Receiver Operating Curve (ROC) over 95.00% of fraud cases without false alarms rate. The Hidden Markov Model (HMM) also has been applied in credit card fraud detection with low percentage of false alarm rates [9]. However, transition process of different states and calculating the probability in HMM are very costly and intensive. Furthermore, rather than using single classifiers, some of the credit card fraud detection studies used meta-learning learners based on supervised learning. Stolfo et al. investigated credit card fraud detection system using four types of algorithms namely Iterative Dichotomiser 3 (ID3), Classification and Regression Tree (CART), Ripper and Bayes as base learners and tested with heterogeneous data distributions [10]. Based on 50% / 50% distribution of instances (fraud and non-fraud), the study found that meta-learning using Bayes as a base learner obtained a higher true positive rate compared to other meta learners. However, even though the distribution of 50% / 50% yields good results, it does not reflect real world circumstances where genuine credit card transactions are quite higher than non-legitimate transactions. Researchers have also tested other types of meta learning classifiers such as Adaboost, Logitboost, Bagging and Dagging and yielded interesting outcomes [11].

Through our literature studies, Bayesian Network is one of the classifier types that have been widely applied to detect fraud in the credit card industry. Maes et al examined the true positive and false positive produced by Bayesian Belief Network and Artificial Neural Network on classifying credit card fraud instances. The study found that Bayesian network performed approximately 8% higher than Artificial Neural Network and claimed that the former's classifier processing time is shorter than the latter [12]. Rather than analyzing using traditional classification methods, the investigation by [13] initiated to perform cost sensitive credit card fraud detection based on Bayes Minimum Risk technique. The study measured the performances of Logistic Regression (LR), C4.5 and Random Forest (RF). The study showed that adjusting the probabilities of Bayes Minimum Risk classifier on RF classification yielded consistently better results than LR and C4.5.

Throughout our observation and analysis of previous studies, Bayesian Network classifiers have become one of the popular classifier types that are widely used to classify credit card fraud data. Therefore, this study attempted to investigate the classification by several Bayesian classifiers such as K2, Tree Augmented Naïve Bayes (TAN), and Naïve Bayes. Moreover, this study also measured the performances of Logistics Regression and J48 based on the proposed methodology. A brief discussion about Bayesian Network Classifier and proposed classifiers are stated below.

A. Bayesian Network Classifier

Bayesian Network is a threshold-based model that computes the sum of the output accumulated from child nodes. The reasons behind the creation of such model is the ability of child nodes to operate independently without interrupting other child nodes and particularly influence the probability of root node. Basically, a Bayesian Network A= <N, B, Θ>, is a directed acyclic graph that consists a set of random variables, where, DAC= <N, B>, and each node n ∈ N represents the variable of the data. Each arc a ∈ A in between nodes represents probability dependency. Bayesian network is able to compute the conditional probability of a node based on given values assigned to other nodes. There are several advantages of Bayesian Network such as the ability to handle incomplete inputs, the learning of causal relationship and so on [17]. As illustrated in Figure 1, there are minor differences between Naïve Bayes, TAN and general framework of Bayesian Network. Naïve Bayes is a very popular classifier as it is simple, efficient and yields better performance in solving real world problems. Naïve Bayes is a probabilistic classifier based on Bayes rules with strong independent assumptions. In simple term, a descriptive "independent feature model" based on probability will allow NB to make assumptions that the presence or absence of a peculiar feature of a class is not related to the presence of absence of other features. K2 as one of Bayesian type classifiers used scoring functions to compute the joint probability of any instantiation of all the variables in a belief network as the product of probabilities [18]. In WEKA, K2 classifiers used hill climbing methods in order to develop the Bayesian beliefs. On the other hand, TAN classifier used Bayesian scoring function to develop the Bayesian Belief. As illustrated in Figure 1, TAN classifier allows arcs between the children of the classification node xc. Therefore, the TAN classifier is able to compute the probability from each child and eventually identify the appropriate classes of the children based on computed probability. Although the information channeled by TAN looks better than Naïve Bayes, none of the studies found to be investigating the performances of TAN on credit card fraud detection domain. Then, compared to Naïve Bayes as generative model, Logistics as discriminative classifier predicts the probability using direct Bayes Functional Form. Logistics uses conditional probability and iterative based estimation in order to estimate the classes of the instances. J48 is an open source Java implementation in WEKA based on C4.5 algorithm. C4.5 algorithm was developed by Ross Quinlan to generate the decision tree based on a set of labeled input data [19]. J48 is a predictive machine-learning classifier that determines the target values of new samples based on various attribute values in the data. The internal decision tree nodes represent the different attributes or features while the branches between nodes denote the possible or viable attributes that could be included in the observed samples or classes. The terminal nodes depict the final classification attributes of the target value.

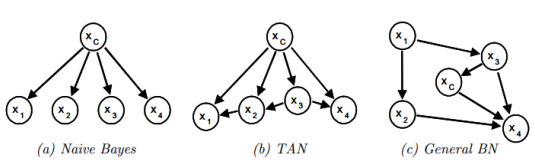

Figure 1: Illustration of Naïve Bayes, TAN and General BN structures

III. RESEARCH HYPOTHESIS

Based on the review from past studies, two main conclusions are made on the evaluation of credit card fraud detection investigations. The first conclusion is that credit card data plays essential roles in identifying fraudulent and

non-fraudulent characteristics. However, the process of getting the real credit card fraud related data is very hard due to record privacy and sensitivity. Therefore, as to mimic the real data, the authors of this study used a dummy data created based on manipulating certain features that were expected to have significant impact for fraud detection. For instance, if the customer entered a wrong pin number from an actual or shipping address that was different than billing address or transaction date and time that were too close with large sum of transactions from previous actions, it could be suspected as suspicious affairs. Furthermore, some countries such as Yugoslavia, Lithuania, and Pakistan have a very high number of fraud incidents with unverifiable addresses. Based on such indicators, the data was developed using several attributes such as credit card number, reference number, terminal id, actual pin, entered pin, transaction amount, transaction date and time, location, billing address and shipping address. Those attributes were the common variables that were used to study the credit card fraud activities. The data was developed manually using spreadsheet and GNU auto data generation script derived from generatedata.com. The instances were labeled as fraud based on the presence of correlations among the attributes as stated in Table 1. The rest of the correlation was defined as non-fraud.

Table 1
Rules to Labeling Fraud Instances (T=TRUE, F=FALSE)

Case	Similar Pin	Blacklisted country	Exceed Threshold	Time Interval	Similar Address	Fraud
1	T	T	T	T	T	T
2	T	T	T	T	F	T
3	T	T	T	F	F	T
4	T	T	F	T	F	T
5	T	T	F	F	F	T
6	T	F	T	T	T	T
7	T	F	T	F	F	T
8	T	F	T	T	F	T
9	T/F	T/F	T/F	T/F	T/F	T

In order to evaluate the validity of the dummy data, the first experiment was conducted to verify the authentication of the corpus to be used in the credit card fraud detection. The second conclusion is most of the previous studies attempted to use heterogeneous types of the classifiers to measure the performances on detecting genuine and non-genuine transactions. On the intention to contribute further to body of the knowledge, the second experiment was conducted to evaluate the performances of the proposed classifiers in the classification of credit card fraud activities. Therefore, the first and second hypotheses that reflect the former and latter experiments are stated as follows:

Hypothesis (1) : The dummy dataset that was created based on suspicious behaviors can be used for classification in data mining.

Hypothesis (2) : The performances on the dataset which undergo data preprocessing are better than the raw dataset.

IV. RESEARCH METHODOLOGY

The overview of the research methodology illustrated in Figure 2.

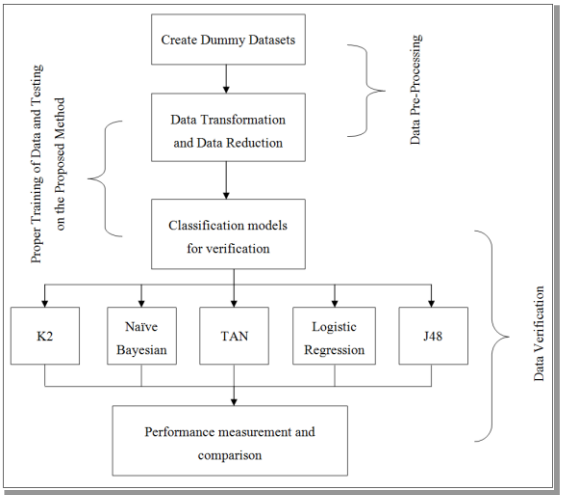

Figure 2: A simple illustration on the flow of methodology in this work

In the classification process, a prominent data mining and machine learning tool namely WEKA was used in order to measure the performances of the classifiers. WEKA is one the open source prominent tools that is used widely to study many real world problems such as sentiment analysis, personality detection, spam filtering, and fraud detection. The classification was run using 10-fold cross validation techniques. The 10-fold cross validation technique is widely applied in data mining and machine learning studies due to the training and testing process that occurred on the entire dataset. Through 10-fold cross validation, the dataset was splitted into ten parts, each part was held out in turn, and eventually the average results were computed. In other words, each data point in the dataset was used once for testing and 9 times for training. Then, in order to measure the performances of the classifiers, several evaluation metrics were employed in this study. Primarily, the output of the metrics depended on the results obtained by True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP refers to the number of fraud transactions predicted as fraud while FP is the number of legal transactions predicted as fraud. TN refers to the number of fraud transactions predicted as legal transactions while FN is the number of legal transactions predicted as fraud. This study evaluated the performances of the classifiers using True Positive Rate (TPR), False Positive Rate (FPR), Precision, Recall, F-Measure, and accuracy.. The description and formula for each evaluation metrics are defined in Table 2.

Table 2
The Formula of Metrics Used in the Study

Metric	Formula and Description
True Positive Rates (TPR)	$TPR = TP / (TP + FN)$
False Positive Rates (FPR)	$FPR = FP / (FP + TN)$
Precision	$Precision = TP / (TP + FP)$
Recall	$Recall = TP / (TP + FN)$
F-Measure	$F\text{-}Measure = 2TP / (2TP + FP + FN)$
Accuracy	$Accuracy = (TP + TN) / (TP + TN + FP + FN)$

The following paragraphs will elaborate on data transformation and data reduction. Generally, data transformation and data reduction are referred to as data pre-processing phase, where the raw data is cleaned and

transformed into appropriate forms (or standardization) to be evaluated and fed into machine learners. Data transformation process involves activities such as normalization, smoothing, aggregation, attributes construction and generalization of the data. While data reduction is to reduce the number of attributes such as data cube aggregation, removing irrelevant attributes and principle component analysis. For instance, during data transformation, the format of transaction date and time were standardized into a uniform state so that it was identical to machine learners to interpret it as date and time attributes. Then, Principal Component Analysis technique was employed to detect the anomaly transactions. Principal Component Analysis is a method to transform the correlated variables into a smaller number of uncorrelated attributes called Principal Components. The objective of applying the method was to identify and reduce the dimensionality of the dataset and discover new meaningful underlying attributes. The advantage of Principal Component Analysis is during reducing the dimensions of the data using eigenvector, the losses to the information of the data are insignificant. Furthermore, the losses could be trace back by decompressing the eigenvalue.

V. Results & Evaluation

This study used two datasets to run through the experiments. The raw dataset and the new dataset were created by data transformation and data reduction.

A. Results and Analysis for Experiment 1

For Experiment 1, the raw dummy dataset was used to evaluate the integrity of the data for credit card fraud detection. The result (see Table 3) showed that the TPR (75.0%), precision (73.0%), recall (75.00%), F-Measure (68.5%) and accuracy (84.0%) of TAN are the highest among the classifiers on the evaluations. The minimal FPR rate of TAN showed the ability of TAN to process the raw data better than other classifiers even though the classifier's speed was higher than K2, Naïve Bayesian, and Logistics. This could be due to the heavy processes such as finding the probability and creating the tree model which caused the processing of the data to take too long. The J48 that was also based on tree model as TAN achieved TPR (73.0%), precision (69.4%), recall (67.5%), and F-Measure (67.4%) which were slightly lower than TAN. Moreover, the processing speed for J48 was also slower than TAN although the processes involved in latter classifier were more heavy/costly than former classifier. From the point of views of the authors, the underperformances of Logistics, Naïve Bayesian and K2 showed that the raw data with high number of noises affects the modeling and evaluation of the raw data. As the worst performer, K2 even obtained very poor results in terms of TPR (31.0%), precision (21.0%), recall (32.0%), F-Measure (32.2%) and accuracy (41.8%). Even though some of the classifiers obtained poor results, the ability of the learners to classify the data showed the reliability of the dummy data being used to test the credit card fraud detection. To further improve the classification results, in experiment 2, the raw dummy dataset was fed into data transformation and data reduction techniques as mentioned above.

Table 3
Results of Classification Using Raw Dummy Dataset

Metric	K2	Naïve Bayesian	TAN	Logistic	J48
True Positive Rate (%)	31.0	50.3	75.0	60.3	73.0
False Positive Rate (%)	69.0	49.7	25.0	39.7	27.0
Precision (%)	21.0	45.7	73.0	44.7	69.4
Recall (%)	32.0	60.3	75.0	47.8	67.5
F-Measure (%)	32.2	34.3	68.5	44.9	67.4
Processing Speed (seconds)	10.0	10.0	56.0	25.0	84.0
Accuracy	41.8	53.7	84.0	67.3	80.0

B. Results and Analysis for Experiment 2

The second experiment used the data that was filtered with normalization and Principal Component Analysis. From Experiment 2, all the five classifiers showed better results compared to Experiment 1. All the classifiers achieved accuracy more than 95.0% with better processing speed than Experiment 1. The minimal FPR showed the preprocessing techniques employed by this study which had increased the reliability of the data by removing the unusable attributes. The results of J48 and Logistics showed that both classifiers gained maximum strengths upon preprocessing of the dataset. It is a huge classification improvement showed by K2 compared to the previous experiment. The classifiers achieved almost 195.80% increase of TPR after data transformation and data reduction process. Furthermore, besides the improvement to the TPR, precision, recall, F-Measure and accuracy, the processing speed for all the classifiers also improved significantly compared to the previous experiment. The authors were curious about attributes that were removed during data preprocessing, hence the cleaned dataset was observed. During the observation, the authors noticed that the terminal_id attributes were reduced significantly. Based on the results shown in Table 4, the hypothesis of experiment 2 was proven where the performances of the classifiers on the preprocessed dataset are better than the raw dataset after undergoing data preprocessing tasks.

Table 4
Results of Classification Using Transformed Dataset

Metric	K2	Naïve Bayesian	TAN	Logistic	J48
True Positive Rate (%)	91.7	99.6	99.7	100.0	100.0
False Positive Rate (%)	8.3	0.4	0.3	0.0	0.0
Precision (%)	92.6	95.6	98.4	100.0	100.0
Recall (%)	91.7	99.6	99.6	100.0	100.0
F-Measure (%)	95.7	89.3	99.0	100.0	100.0
Processing Speed (seconds)	2.0	2.0	30.0	5.0	32.0
Accuracy	95.8	96.7	99.7	100.0	100.0

C. Discussion and Future Work

The detection of credit card fraud using data mining and Machine Learning techniques have become one of the reliable approaches to counter this illegal activity. However, the process to gather real time credit card fraud data is very hard. Therefore, to mimic the real data, the development of dummy data may assist the detection process. However, the creation and credibility of dummy data must be ascertained prior to conducting the classification processes. Based on the results from Experiment 1, the credibility of the data could be

ensured by noticing the ability of the WEKA to produce non-zero results. Generally, WEKA would not be able to process the data if the data is highly unstructured and would return N/A (Not Applicable) results, errors, or freeze during modeling process. However, it did not happen to our dummy dataset. Furthermore, the development of the dummy dataset was based on attributes commonly used for credit card fraud detection and created automatically by using GNU data generation scripts. Then, as always emphasized by many data mining researchers, the preprocessing of raw dataset is an essential factor to improve the classification results. This has been proven by observing the differences between results of Experiment 1 and Experiment 2. The improvement on Experiment 2 after data transformation and data reduction significantly improve the classification performances. As mentioned earlier, the strength of Principal Component Analysis that reduced the dimensionality, losing much the information from the attributes was one of the major factor that improved the classification process. Therefore, we believed that Principal Component Analysis technique is the better filtering approach to be considered and to be used in credit card fraud detection processes. Then, our classification process also proved that Bayesian based classifiers such as K2, Naïve Bayesian, Tan, Logistics and J48 were able to classify and predict the credit card fraud activities better if the data was preprocessed using reliable filtering techniques. Moreover, after the dimensionality of the raw data was reduced by using Principal Component Analysis, the authors of this study found that the *terminal_id* attributes were largely reduced.. Therefore, we made the assumptions that *terminal_id* information contribute less to the credit card fraud detection. However, the investigation of credit card hacking based on physical methods (e.g. hardware stressing) has to use *terminal_id* attributes as the reference to identify the illegal activity.

In the future, this study will attempt to explore more credit card fraud detections using real time data. Then, since the Bayesian Networks classifiers showed better results, the comparisons with other types of classifiers such as Hyperplane based may contribute further to the body of the knowledge.

VI. CONCLUSION

This paper tested classification metrics by using five Bayesian classifiers namely Naïve Bayes, K2, TAN, Logistics and J48. The evaluations conducted using two datasets, where, the first dataset was a dummy dataset that represented the characteristics of credit card data and a newly transformed dataset using data normalization and Principal Component Analysis techniques. Overall, all the Bayesian classifiers achieved significantly better results after being fed with filtered data.

ACKNOWLEDGMENT

We are grateful to Universiti Sains Malaysia for providing support to this work.

REFERENCES

[1] WorldPay. (2015, Nov). Global payments report preview: your definitive guide to the world of online payments. Retrieved September 28, 2016, from http://offers.worldpayglobal.com/rs/850-JOA-856/images/GlobalPaymentsReportNov2015.pdf.

[2] Federal Trade Commision. (2008). consumer sentinel network - data book for January - December 2008. Retrieved Oct 20, 2016. From https://www.ftc.gov/.

[3] Bhatla, T.P., Prabhu, V., and Dua, A. (2003). understanding credit card frauds. Crads Business Review# 2003-1, Tata Consultancy Services.

[4] The Nilson Report. (2015). Global fraud losses reach $16.31 Billion. Edition: July 2015, Issue 1068.

[5] Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines", *Proceedings of the International Multi-Conference of Engineers and Computer Scientists 2011 Vol I, IMECS 2011, March 2011.*

[6] Elkan, C. (2001). Magical thinking in data mining: lessons from COIL challenge 2000. Proc. of SIGKDD01, 426-431.

[7] Mohammed, J. Zaki., & Wagner, Meira Jr. (2014). Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press. ISBN 978-0-521-76633-3.

[8] F. N. Ogwueleka. (2011). Data mining application in credit card fraud detection system. *Journal of Engineering Science and Technology, Vol. 6, No. 3 (2011) 311 - 322.*

[9] V. Bhusari & S. Patil. (2011). Application of hidden markov model in credit card fraud detection. *International Journal of Distributed and Parallel Systems (IJDPS) Vol.2, No.6.*

[10] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan. (1998). Credit card fraud detection using meta-learning: issues and initial results, *Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90.*

[11] Sen, Sanjay Kumar., & Dash, Sujatha. (2013). Meta learning algorithms for credit card fraud detection. *International Journal of Engineering Research and Development Volume 6, Issue 6, pp. 16-20.*

[12] Maes, Sam, Tuyls Karl, Vanschoenwinkel Bram & Manderick, Bernard. (2002). Credit card fraud detection using bayesian and neural networks. *Proc. of 1st NAISO Congress on Neuro Fuzzy Technologies. Hawana.*

[13] A.C. Bahnsen, Aleksandar, Stojanovic., D. Aouada & Bjorn, Ottersten. (2013). Cost sensitive credit card fraud detection using bayes minimum risk. *12th International Conference on Machine Learning and Applications*.

[14] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar. (2009). Credit card fraud detection: a fusion approach using dempster–shafer theory and bayesian learning. *Special Issue on Information Fusion in Computer Security, Vol. 10, Issue No. 4, pp.354-363.*

[15] Lam, Bacchus (1994). Learning bayesian belief networks: an approach based on the MDL principle. *Computational Intelligence, Vol. 10, Issue No. 3, pp.269–293.*

[16] M. Mehdi, S. Zair, A. Anou and M. Bensebti (2007). A bayesian networks in intrusion detection systems. *International Journal of Computational Intelligence Research, Issue No. 1, pp.0973-1873 Vol. 3.*

[17] R.Najafi & Afsharchi, Mohsen. (2012). Network intrusion detection using tree augmented naive-bayes. *The Third International Conference on Contemporary Issues in Computer and Information Sciences (CICI) 2012.*

[18] G. Cooper, E. Herskovits (1992). A bayesian method for the induction of probabilistic networks from data. Machine Learning. 9(4):309-347.

[19] Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

[20] Friedman, N. and Goldszmidt, M. (1996). Building classifiers using bayesian networks. Proc. 13th National Conference on Artificial Intelligence.Vol. 2, pp 1277-1284.

[21] Friedman, N., Geiger, D. and Goldszmidt, M. (1997). Bayesian network classifiers. machine learning,Vol. 29, pp 131-163. Kluwer Academic Publishers, Boston.