Credit Card Fraud Detection using Machine Learning

Group Member
Md.Radwan(16155028)
Shayan Nandi(16155044)

Course Supervisor
MD. Mynodin
Lecturer, Dept of IT
UITS

Abstract *–Nowadays digitalization gaining popularity because of seamless, easy and convenience use of e-commerce. It became very rampant and easy mode of payment. People choose online payment and e-shopping; because of time convenience, transport convenience, etc.As the result of huge amount of e-commerce use, there is a vast increment in credit card fraud also. Fraudsters try to misuse the card and transparency of online payments. Thus to overcome with the fraudsters activity become very essential. The main aim is to secure credit card transactions; so people can use e-banking safely and easily. To detecting the credit card fraud there are various techniques which are based on Deep learning, Logistic Regression, Naïve Bayesian, Support Vector Machine (SVM), Neural Network, Artificial Immune System, K Nearest Neighbor, Data Mining, Decision Tree, Fuzzy logic based System, Genetic Algorithm etc.*

Keywords: *Machine-learning, Credit card, Fraud detection, Fraud transaction.*

I. INTRODUCTION

In corporate and finance business, financial fraud become very crucial issue. Moreover, financial fraud affect a lot in business, economy instability and it also affects the people's price of living. As shown in figure-1, there are some frauds, which are again classify further, that are the major issues now days. They are credit card fraud, mortgage fraud, money laundering, financial statement fraud, securities and commodities fraud, automobile insurance fraud and healthcare fraud. In this paper, we will focus on Credit card fraud and its detection techniques.

A. Credit Card Fraud Detection:

The dirty use of data for e-commerce referred as credit card fraud. Credit card fraud become rampant, as there is increment in credit card transaction. Nowadays, card transaction is not only for the online purchases; it is beyond that in regular purchases also.

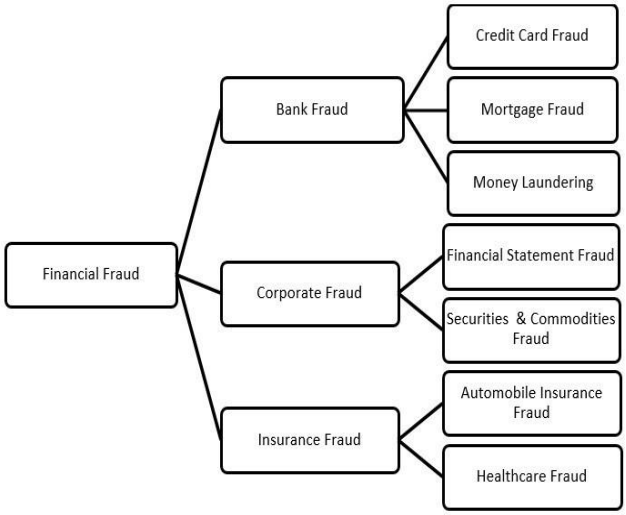

Fig. 1 Types of Common Fraud

Because of financial fraud by credit card transaction, both merchants and shoppers are suffering from economical loss. It is a very important issue; to solve that issue banks and card manufacture organizations pays significant cost. Online purchases and payment services makes e-payment comfortable, seamless, convenient, easy and simple to use; but we cannot ignore the financial losses, which also increases with e-commerce. It is inviting to new type of fraud for criminals. To cope up with these issues, organizations and banks using good security solutions; but fraudsters change their subtle techniques with time. Therefore, to enhance the detection and prevention techniques is very important.

There are two ways of credit card transaction: physically and virtually i.e. CNP (Card not Present). In physical, card is require physically to make a swipe. Whereas in the virtual card, some details are there to swipe a card like CVV number, card holder name, password, security question etc. for net banking.

Fraud prevention and fraud detection both are the way to handle the fraud. In fraud prevention, the main aim is to prevent the fraudulent activity; it spot the transaction and prevent the authorize transactions. While in fraud detection, the aim is to distinguish the fraudulent transaction and legitimate transactions. By historic data, user's

pattern and behavior used to check and verify that the transaction is fraud or not. Sometimes system fails in prevention of fraudulent activity, at that time detection of fraud is take initiative.

As shown in below figure-2, there are four types of Credit Card Fraud: Card not present, Skimming, Phishing, Lost/Stolen Card.

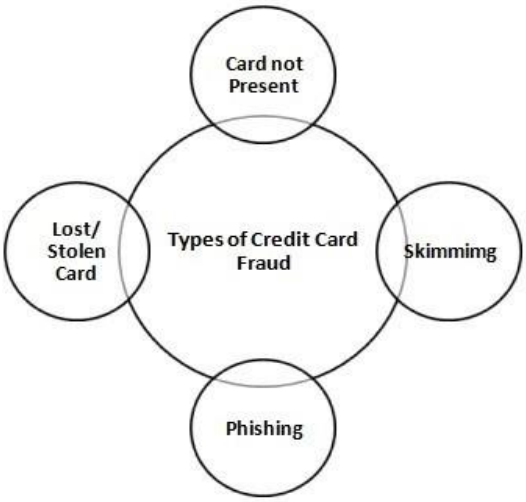

Fig. 2 Types of Credit Card Fraud

In Card not Present fraud, fraudster attempt to mislead the system by dissembling to be some other person. Mail and the web are major routes for fraud against merchants who sell and ship merchandise, and affects legitimate mail order and web merchants. In Skimming, they are obtaining personal data regarding someone else's credit card utilized in an otherwise normal transaction. There is a tiny device (skimmer) which is used to swipe and store huge amount of victim's information. In phishing, Scammers might use a range of schemes to lure users into giving them their card info through tricks corresponding to websites simulation to be of a bank or payment system. When card is steal or lost, there are chances for a thief that he make unauthorized transaction before cardholder block the card.

The remainder of the paper structured as shown below: Theoretical background explained in section-2. In which various data mining techniques are briefly explain. After that, existing techniques explained in section-3. It includes artificial immune system, Bayesian belief network, logistic regression, decision tree, self-organizing map, hybrid methods, etc. In section-4, we presented analysis of classification techniques with their methodology and challenges. In last section-5 and 6 motivation and conclusion of the paper presented.

II. THEORETICAL BACKGROUND

There are mainly six data mining approaches: classification, clustering, prediction, outlier detection, regression and visualization. These approaches of data mining are how useful in credit card fraud detection is given below in this section.

A. Classification:

Classification is very common learning model comes under the data mining techniques. To differentiate various category of object. A model is use called as classification. Classification predicts the labels of object; labels are predefine, unordered and distinct. As per the study in paper authors mentioned that, classification and prediction both are the method of distinguishing the objects which having the similar features. The application like, detection of credit card fraud, healthcare fraud, automobile insurance, corporate fraud, etc. where classification become very useful to detect fraud.

B. Clustering:

Clustering is variant of unsupervised classification. In clustering, objects are divide into conceptually meaningful groups called cluster. In same cluster, the objects are very similar to each other in terms of feature. While dissimilar objects are not become the part of that cluster; it transfer to another cluster group of objects.

C. Prediction:

Prediction is use to predict the continuous value. Based on the historic data, it make patterns, estimate the numeric, and ordered value for future. As per the study author stated that, predicted values are continuous value rather than discrete value.

D. Outlier Detection:

Outlier means the data objects that is completely different from whole remaining dataset. Outlier detection means measure that "distance" between the data points and that outlier object. As per the study in paper, author stated that, "Data points that having different characteristics with compare to reminder of whole dataset are known as outlier". Hence, the authors mentioned that outlier detection is very crucial issue in the field of data mining.

E. Regression:

Regression shows the relationship between more than one dependent and independent variable. Regression is one of the best statistical method. Hence, for several empirical studies regression is like a benchmark.

III. EXISTING TECHNIQUES

There are several techniques exist based on statistical and computational. Those techniques are apply to the data mining issues. As shown below, this section contains an outline of the functions of every methodology, which is use in the literature. Hence, Table 1 lists the relative strengths and limitations of every methodology.

A. Artificial Immune System:

Artificial immune systems is comes under data mining techniques. To discover antigens, AIS imitates the behavior of a biological immune system. A range of biological characteristics often simulates using AIS. However, most of the models are create detector cells and their ability to detect the foreign bodies. Generation of detector cells are done in random manner. Moreover, the

simulation is perform to check and evaluate the effectiveness of algorithm, i.e. training performed by different classification strategies. There are two basic types of AIS called clone selection and negative selection. In clonal selection, detector cells that frequently generated solely live a short life. In between their life, if a cell detects an antigen it extends its life to fight with an intruder. Then it should mutate as a results of the conflict. At the end of the stimulation, which, cells are best, suited to detecting the antigens called as survival cells. While in negative selection, the creation of cell is arbitrarily. From those cells, it decide which of them react with alternative and non-invasive cells among the whole system. At last, leading to the rest being proficient at detecting intruders are discarded.

B. Bayesian Belief Network:

Bayesian belief networks is a statistical classification technique and based on Bayes theorem. It is work on a principle of, work out the probability that a given hypothesis is true. As per the study, for a hypothesis H, the probability P given by:

$$P(H|X) = \frac{(P(X|H)*P(H))}{P(X)} \quad (1)$$

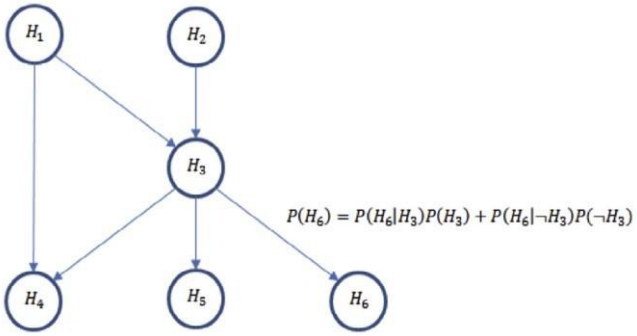

Fig. 3 Graphical Representation of Bayesian Belief Network

A theorem Bayesian belief network uses a classifier to calculate P(Ci|X) for all possible classes Ci. Then it inserts X into the class with it's very best P(Ci|X). During this approach, every sample into the class, which is possible to belong to that class, are classify the network.

As shown in figure-3, a theorem Bayesian belief network is model as a directed acyclic graph. Moreover, nodes are represent as a samples and edges are represents a causal dependency between nodes. Missing edges between nodes represents independency between those two nodes .

C. Logistic Regression:

Logistic regression is use for binary classification, which based on statistical method. It uses linear model. Hence, it is use to perform regression on a group of variables. It is normally used technique for predicting patterns in data with unambiguous or numeric attributes . To calculate probability it uses a series of input vectors and a dependent response variable, using logarithm. Probability is lies among the specific class. For binary classification, the response variable given below:

$$Y_i = \begin{cases} 0 \\ 1 \end{cases} \quad (2)$$

Hence, the formula for calculating that a sample xi belongs in class one given by:

$$P(y_i = 1|X_i) = \frac{\exp(W_0 + W^T X_i)}{1+\exp(W_0 + W^T X_i)} \quad (3)$$

Where, W0 and W are the regression standardization parameters. W0 represents the intercept and W represents coefficient vector .

D. Neural Network:

A neural network is inspired from human brain. It is a computational approximation of the human brain. To represent the neurons and synapses it uses a graph of vertices and edges. As shown in figure-4, modeling of input variables as a layer of vertices performed in network. Then distribution of weight applied to every connection within the graph. Moreover, the other vertices are place into separate levels; and it reflects the distance from the input nodes.

Hence, each node considers input as a function of vertices; and that vertices connected to the previous layer. For every neuron j the signal received is as given bellow:

$$u_j = \sum W_{ij} * X_i \quad (4)$$

Where, Wij is that the weight of the link between neurons i and j and Xi is that the input of i. If the result is, greater than a predicate value the present neuron is "fires". Then it becomes an input for consequent layer .

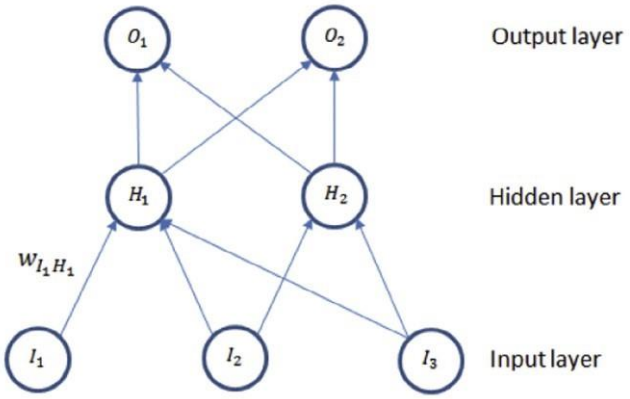

Fig. 4 Simple Neural Network with One Hidden Layer

In back propagation neural network, set of training data given to the network; and then network compares it with targeted results. Weights of the edges given arbitrarily for the very first iteration. After first iteration, every weight adjusted for consecutive sequence. This process is continues until network reduced its error to the certain value; or else when the predetermined limit of iteration has been reached. Once after completion of training, the performance is test on the set of validate data. Overtraining is a very common issue while working with a back propagation neural network. This issue may cause the network to concentrate on tendencies particular to the training set.

E. Support Vector Machine :

Support vector machines transforms the linear issue into a higher dimensional feature space. This allows sophisticated, non-linear issues like credit card fraud

detection to be solved using linear classification, without increasing the computational complexity. Kernel function is use to remodel the dataset. It is consider as a mapping between the input space point and a higher dimensional space point. The kernel function outlined by:

$$k(X_1, X_2) = < Ø(X_1), Ø(X_2) > \quad (5)$$

Where, Ø: X→H is use to map the point between input space X and higher dimensional space H. When the kernel function is, apply to the dataset, a hyperplane employed for class separation; which is in the form of:

$$< W, X > + b = 0 \quad (6)$$

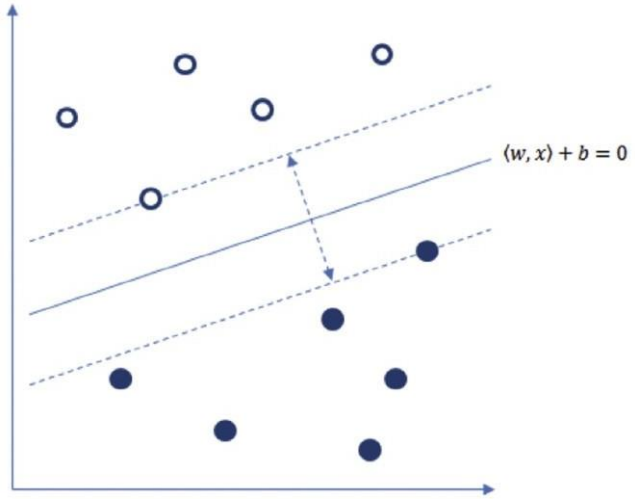

Fig. 5 Separation between Support Vector and Hyperplane

As shown in figure-5, the hyperplane is built in the simplest way on maximize the separation between both classes. After that, it helps in reducing the potential errors, which generated because of overtraining. SVM classification is outline as below:

$$\sum_i \alpha_i Y_i k(X_i, X) + b = 0 \quad (7)$$

There are several kernel function are there; like Gaussian radial basis function, polynomial function. Dataset and necessity of classification defines that, which function we have to apply.

F. ***Genetic Algorithm:***

Genetic algorithms inspired by the idea of population evolution. It improves solution iteration by iteration of the problem. In first iteration, generation of population done in random manner. Reproduction of population generation done using various techniques, and it continues until best survivors found. The process of finding best survivors depends on their fitness. For the reproduction process, pair of parents is require from the current generation. After reproduction, crossover is apply on pair of parents. At the end of the process, we get the single part of the resulting child; and the ability of that child is measure by the fitness function. After that using result, they will determine that, which pair of parents and child chosen for successive generation.

To measure the fitness of child is simple and straightforward as to measure the percentage of samples they classify properly. There is one similarity between Genetic algorithms and neural network is that, both does not require prior knowledge of the problem domain . They are also capable to find underlying relationships between the samples.

G. Decision Tree:

Decision trees are using a binary tree with successive nodes to classify the data. As the sample travers the tree, according to that node being created. Tree divided into the subset until it stored in to mutually exclusive subgroups eventually [16]. It is also refer as classification and regression tree (CART) .

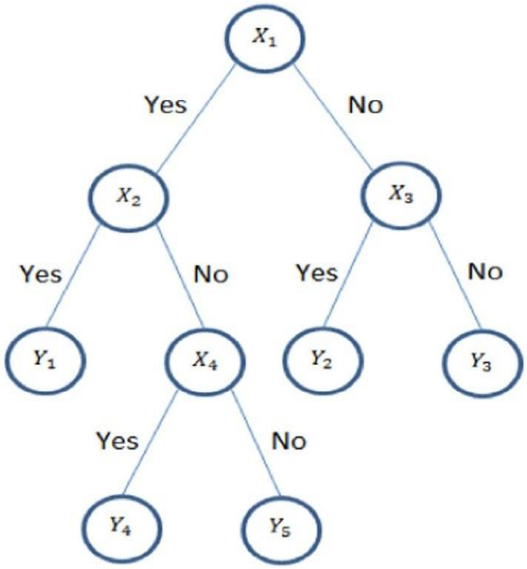

Fig. 6 Binary Decision Tree

Apart from this technique, to overcome with the issue of overfitting authors defines the method called pruning . Pruning removes the tree nodes without reducing the overall accuracy.

H. Self-organising Map:

Self-organising map is very similar to artificial neural network. Both contains the one matrix of neurons. To map the input from higher-dimension to two-dimension array, a non-linear algorithmic program is use. The aim of mapping is, to model the similar input vector as neurons, which are close to target matrix; provides the visualization of input. Then various distance function is employ on the set of nodes. Distance functions such as, Euclidean distance formula, Gaussian formula, etc.. To every neurons, there is a clustering function applied which given by:

$$Y_{i+1} = Y_i + \alpha (X_i - Y_{i-1}) \quad (8)$$

Where, yi represents current weight of a specific node, Xi represents current input vector, and α represents distance function. Clustering is perform on a set of iterations before algorithm terminate.

I. ***Hybrid Methods:***

Hybrid techniques made for a particular problem domain. It is a mixture of two or more traditional technique, which is choose base on their benefits, to make an algorithm

superior. Hybrid methods are construct in a various ways: in highest-level techniques and in preprocessing stage. In highest-level techniques, it simply apply linearly. It means the output of first is becomes input for another one. In hybrid technique, individual steps of traditional algorithm may intertwine to make something essentially original. Moreover, hybrid strategies are often use for a specific problem domain. Completely different aspects of performance is target, classification ability, simple use, and efficiency in term of computation, etc.

Similarly, in pre-processing step the data is modify while it prepare for classification, or we can say at the lower level .

IV. ANALYSIS OF MACHINE LEARNING TECHNIQUES

In previous section, we see the algorithms based on credit card fraud detection. Moreover, goes through its introduction and working. Now in this section we will analyze those algorithms.

Table-1 Analysis of Credit Card Fraud Detection Techniques

Sr. No.	Title, Publication & year	Learning Paradigm	Method	Challenges
1	**Title:** A Novel Model for Credit Card Fraud Detection using Artificial Immune System [2] **Publication:** ELSEVIER – Journal 2014	Supervised	- Artificial Immune System.	-Weighting dataset fields -Memory generation phase & calculation of affinity is time consuming. -Misuse of cloud computing using AIRS. - Distance Function depends on dataset.
2	**Title:** Detecting Credit Card Fraud by Modified Fisher Discriminant Analysis [1] **Publication:** ELSEVIER - Journal 2015	Supervised	- Profit based modified discriminant linear classifier.	- Cannot effectively handle false negative.
3	**Title:** APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection using Network based Extensions [4] **Publication:** ELSEVIER - Journal 2015	Supervised	-Automatic detecting model for online fraud transaction. -Intrinsic Feature Extraction. -Network based Feature Extraction.	- Data imbalance is too high.
4	**Title:** Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier [21] **Publication:** ELSEVIER – Conference 2015	Supervised	- bagging ensemble classifier design to improve the stability and accuracy of machine learning algorithms.	- Classification slow. - Unbalanced Dataset. - Large Size of the Dataset. - Determining the appropriate evaluation parameters.
5	**Title:** Feature Engineering Strategies for Credit Card Fraud Detection [5] **Publication:** ELSEVIER - Journal 2016	Supervised	-Logistic regression. -Von Mises Distribution to analyze periodic behavior.	- Response and Calculation time of the different features.
6	**Title**: Horse Race Analysis in Credit Card Fraud – Deep Learning, Logistic Regression, and Gradient Boosted Tree [23] **Publication:** IEEE - Conference 2017	Supervised	- Logistic regression. - decision Tree - Neural network	- Less predictive power - Large size of dataset. - Feature selection.
7	**Title:** Adversarial Learning in Credit Card Fraud Detection [24] **Publication:** IEEE - Conference 2017	Supervised	- Logistic regression.	- Velocity variables in effort to discover more characteristic of the algorithm. - Cost of retraining the classifier.

V. MOTIVATION

The use of machine learning in fraud detection has been an interesting topic now days. A credit card fraud detection algorithm consists in identifying those transactions with a high probability of being fraud, based on historical fraud patterns. Machine learning having three types, from that also the supervised and hybrid approach is more suitable for fraud detection. Here, recent algorithms that suitable for credit card

fraud detection is explain and compare in section of

VI. CONCLUSION

Credit card detection is a fascinating domain. From this survey, we analyzed machine learning is best in compare to prediction, clustering, outlier detection etc., that earlier used. Machine-learning techniques are mostly preferred in fraud detection, because of its high accuracy and detection rate. Still researchers are struggling to get more accuracy and detection rate. Moreover, organizations are interested in finding methods that can reduce cost and increase the profit; they can find and select the method from above studies.

REFERENCES

1N. Mahmoudi, E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis", Elsevier Expert System with Application, 2015, pp. 2510-2516.

2N. Halvaiee, M. Akbari, "A novel model for credit card fraud detection using Artificial Immune System", Elsevier Applied Soft Computing, 2014, pp. 40-49.

3M. Zareapoor, K. Seeja, M. Alam, "Analysis of credit cardfraud detection techniques: based on certain design criteria",. International Journal Compututer Application, 2012, pp. 35–42.

4 V. Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, B. Baesens, "APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection using Network based Extensions", ELSEVIER Decision Support Systems, 2015, pp. 38-48.

5A. Bahnsen, D. Aouada, A. Stojanovic, B. Ottersten, "Feature Engineering Strategies for Credit Card Fraud Detection", ELSEVIER Expert System with Applications, 2016, pp. 134-142.

6J. Quah, M. Shriganesh, "Real-time credit card fraud detection using Computational Intelligence", Expert System Application, 2008, pp. 1721-1732.

7 Y. Sachin, E. Duman, "Detecting Credit Card Fraud by Decision Tree and Support Vector Machine", In Proceedings of the international multi Conference of Engineers and Computer Scientists, Hong Kong, 2011, pp. 1-6.

8J. West, M. Bhattacharya, "Intelligent Financial Fraud Detection: A Comprehensive Review", ELSEVIER Computer & Security, 2016, 47-66.

9J. Han, M. Kamber, "Data Mining: Concepts and Techniques", Second ed, Morgan Kaufmann Publishers, 2006,pp. 285–464.

10 Yue, X. Wu, Y. Wang, Y. Li, C. Chu, "A review of data mining-based financial fraud detection research", international conference on wireless communications Sep, Networking and Mobile Computing, 2007, pp. 5519–5522.

11K. Yamanishi, J. Takeuchi, G. Williams, P. Milne, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms", Data Mining and Knowledge Discovery, 2004, pp. 275–300.

existing classification techniques.

12E. Turban, J.E. Aronson, T.P. Liang, R. Sharda, "Decision Support and Business Intelligence Systems", Eighth ed, Pearson Education, 2007.

13W. Sx, W. Banzhaf, "Combatting financial fraud: aco evolutionary anomaly detection approach", ACM In proceedings of the 10th annual conference on genetic and evolutionary computation, 2008, pp. 1673-1680.

14E. Ngai, Y. Hu, Y. Wong, Y. Chen, X. Sun, "The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature", Elsevier Decision Support Systems, 2011, pp. 559–569.

15 P. Ravisankar, V. Ravi, G. Rao, I. Bose, "Detection of financial statement fraud and feature selection using datamining techniques", Elsevier Decision Support Systems, 2011, pp. 491–500.

16E. Kirkos, C. Spathis, Y. Manolopoulos, "Data mining techniques forthe detection of fraudulent financial statements", ExpertSystems with Applications, 2007, pp. 995–1003.

17 S. Bhattacharyya, S. Jha, K. Tharakunnel, J. Westland,"Data miningfor credit card fraud: a comparative study", Decision SupportSystems, 2011, pp. 602–613.

18 D. Olszewski,"Fraud detection using self-organizing mapvisualizing the user profiles", Knowledge-Based Systems, 2014, pp. 324–334.

19E. Duman, M. Ozcelik,"Detecting credit card fraud by geneticalgorithm and scatter search", Expert Systems withApplications, 2011, pp. 13057–13063.

20 M. Jans, D. van,J. Werf, N. Lybaert, K. Vanhoof,"A businessprocess mining application for internal transaction fraud mitigation", Expert Systems with Applications, 2011, pp. 13351–13359.

21M. Zareapoor, P. Shamsolmoali, "Application of Credit card Fraud Detection: Based on Bagging Ensemble Classifier", Elsevier International Conference on Intelligent Computing, Communication & Convergence, 2015, pp. 679-685.

22A. Pozzolo, O. Caelen, Y. Borgne, S. Waterschoot, G. Bontempi, "Learned Lessons in Credit Card Fraud Detection from Practitioner Perspective", ELSEVIER Expert System with Applications, 2014, pp. 4915-4928.

23 G. Rushin, C. Stancil, M. Sun, S.Adams, P. Beling, "Horse Race Analysis in Credit card Fraud- Deep Learning, Logistic Regresion, and Gradient Boosted Tree", IEEE, 2017, pp. 117-121.

24 M. Zeager, A. Sridhar, N. Fogal, S. Adams, D. Brown, P.Beling, "Adversarial Learing in Credit card Fraud Detection", IEEE, 2017, pp. 112-116.

[25] A. Srivastava, A. Kundu, S. Sural, A. Majumdar, "Credit Card Fraud Detection using Hidden Markov Model", IEEE Transaction on Dependable and Secure Computing, 2008, pp. 37-48.