PREDICTIVE MODEL ANALYSIS FOR DETECTING FRAUDS IN CREDIT CARD

Dr.J.Suresh Babu
Computer Science and Engineering Narayana Engineering College(JNTUA)
Nellore,India
drjsureshbabu@gmail.com

P.Sudha Revathi
Computer Science and Engineering Narayana Engineering College(JNTUA)
Nellore,India
revathi.pasumarthi4@gmail.com

G.Kowsalya
Computer Science and Engineering Narayana Engineering College(JNTUA)
Nellore,India
kowsalya.garikapati@gmail.com

S.Harini
Computer Science and Engineering Narayana Engineering College(JNTUA)
Nellore,India
sanjamalaharini7@gmail.com

M.Renu Sree
Computer Science and Engineering Narayana Engineering College(JNTUA)
Nellore,India
renuchowdarie1999@gmail.com

ABSTRACT:

Credit card fraud is a significant issue in financial services. Lots of money is getting lost due to credit card fraud every year. There is a scarcity of research studies on analyzing real-world credit card data because of confidentiality issues. In this model, machine learning algorithms are used to detect credit card fraud. Standard models are firstly used. To evaluate the model efficiency, a publicly available credit card data set is used. The experimental results positively indicate that the bulk voting method achieves good accuracy rates in detecting fraud cases in credit cards.The MCC metric has been adopted as a performance measure, because it takes into consideration truth and false or positive and negative predicted outcomes. The majority voting method has yielded the best MCC score to the data set. By using this technique, the high amount of losses due to fraud and the awareness of the relation between loss and also the available limit has got to be reduced.

***Keywords*: Hybrid model, Majority voting, Adaboost**

I.INTRODUCTION:

Machine learning is an application of Artificial Intelligence (AI) that provides system the ability to automatically learn and improve from experience without being explicitly programmed. It focuses on the event of computer programs which will access data and use it learn for themselves. It consists of supervised and unsupervised learning algorithms. Credit card fraud occurs with the illegal use of credit card information for purchases. With the increase of credit card usage, the amount of fraud cases are constantly increased. Although numerous authorization techniques are implemented, credit card fraud cases have not hindered effectively. Hence machine learning algorithms are used in this project which are used to automatically learn and detect frauds. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models.

AdaBoost is a machine learning algorithm in which various weak classifiers can be combined together to form a strong classifier to increase the performance. Majority voting is the process in which every algorithm makes a prediction for each test instance and the final output prediction is the one that receives more than half of the votes. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

II. RELATED WORK

A lot of related work is done to know the disadvantages of the existing system and to have a complete insight of the machine learning algorithms to detect various kinds of frauds occurring in transactions. Several single models such as support vector machine, bayes classification, linear regression, random forest and hybrid models such as majority voting technique and AdaBoost are also studied.

III.EXISTING SYSTEM

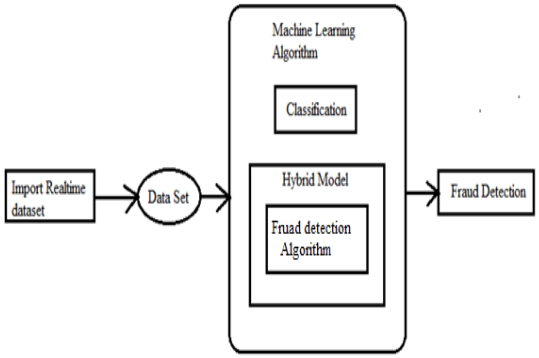
The Dempster–Shafer theory combined various evidential information and created an initial belief, which was accustomed to classify a transaction as normal, suspicious, or abnormal. If a transaction was suspicious, then it was further evaluated using transaction history . Three methods to detect fraud are presented. Firstly, clustering model is employed to classify the legal and fraudulent transaction using data parameter value. Secondly, Gaussian mixture model past behavior and current behavior are often calculated to detect any abnormalities from the past behavior. Lastly, Bayesian networks are used to describe the statistics of a specific user and also the statistics of various fraud scenarios.

Disadvantages:

The above system does not work effectively in detecting frauds. It is time consuming and does not produce efficient results in detecting frauds.

IV.PROPOSED SYSTEM

In the proposed system, machine learning algorithms are used for detecting credit card fraud.



This paper is used to detect various frauds such as

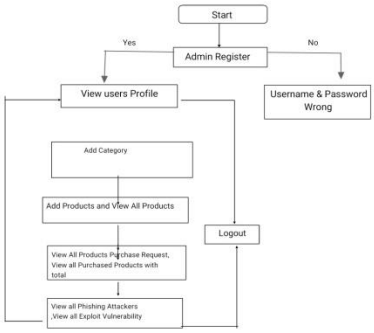
1. Money laundering accounts in social networks
2. Phishing attacks in social networks
3. Finding vulnerable accounts

To overcome above frauds, we use several machine learning algorithms such as support vector machine, linear regression

and neural networks. Additionally, the AdaBoost and majority voting methods are applied for forming hybrid models. Several kinds of transactions are performed using world credit card data sets to provide effective results. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set.

ADVANTAGES

The system is very fast due to AdaBoost Technique. MCC score has been raised when compared to previous results which indicates the decrease in fraud rates.



V. MACHINE LEARNING ALGORITHMS

1. SUPPORT VECTOR MACHINE:

It is a kind of supervised learning algorithm which is used to predict the kind of transaction and to know whether it is a fraudulent transaction or not. It performs both classification and regression. This method is not compatible for large data sets having large amount of noise.

2. LINEAR REGRESSION

It is a supervised machine learning algorithm in which the output can be described based on previous kinds of transactions. It any abnormality can be observed when compared to previous transaction then it can considered as a fraud transaction. It is easy to implement and requires less time to train data. It is less secured and is easily effected .by outliers.

3. NEURAL NETWORKS

This is a method in which it learns about collected data and predict outcome based on it. It consists of three layers namely input layer, hidden layer and output layer. Every layer is connected to one another and can transmit information among themselves to produce efficeient output. It requires high computational power and retraining has to be done for some kinds of frauds.

4. ADABOOST

This algorithm is used to increase the performance of system. It combines weak classifiers in above algorithms and produces effective final outcome. The outputs are combined by using a weighted sum, which represents the combined output of the boosted classifier, i.e.,

$$FT(x) = \sum ft(x)$$

Where x ranges from 1 to t

where every ft is a weak classifier that returns the predicted class with respect to input x.

5. MAJORITY VOTING TECHNIQUE

Majority voting is frequently used in data classification, which involves a combined model with at least two algorithms. Each algorithm makes its own prediction for

every test sample. The final output is for the one that receives the majority of the votes. Given an input x, each classifier provides a prediction with respect to the target class, yielding a total of K prediction, i.e., P1,...,PK. Majority voting aims to produce a combined prediction for input x, P(x) = i,j ∈ ^ from all the K predictions. A binary function can be used to represent the votes, i.e.,

$$Vk(x \in Ci) = \begin{cases} 1, \text{ if } PK(x)=i, i \in N \\ 0, \text{otherwise} \end{cases}$$

Then, sum the votes from all K classifiers for each Ci, and the label that receives highest number of votes is the final predicted class.

VI.CONCLUSION

A study on credit card fraud detection using machine learning algorithms has been presented in this paper. A number machine learning algorithms models which include linear regression, SVM, and neural networks are utilized in the empirical evaluation. The MCC metric has been adopted as a performance measure, because it takes under consideration truth and false positive and negative predicted outcomes. The best MCC score is 0.823, achieved using majority voting. A real credit card data set from a financial organization has also been used for evaluation. The same individual and hybrid models are employed. A perfect MCC score of 1 has been achieved using AdaBoost and majority voting methods. To further evaluate the hybrid models, noise from 10% to 30% has been added into the info samples. The majority voting method has yielded the most effective MCC score of 0.942 for 30% noise added to the info set. This shows that the majority voting method is stable in performance within the presence of noise.

VII.REFERENCES

[1] Gao, J. , Zhou, Z. , Ai, J. , Xia, B. and Coggeshall, S. (2019) Predicting Credit Card Transaction Fraud Using Machine Learning Algorithms. Journal of Intelligent Learning Systems and Applications, 11, 33-63.

[2] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang and C. Jiang, "Random forest for credit card fraud detection," 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, 2018, pp. 1-6.

[3] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, 2017, pp. 1-9.

[4] Maranzato, Rafael & Pereira, Adriano & Neubert, Marden & Lago, Alair. (2010). Fraud detection in reputation systems in e-markets using logistic regression and stepwise optimization.

[5] Pang, S & Kim, Daijin & Bang, S. (2001). Fraud detection using support vector machine ensemble.