JES
Journal of Engineering Sciences

Vol 13, Issue 06, June/2022
ISSN NO:0377-9254

CREDIT CARD FRAUD DETECTION USING ADABOOST AND MAJORITY VOTING

[1]T. Aditya Sai Srinivas [2]P. Pradeep [3]Repalle.Raju [4]Pesala.Preetham

[1,2,3,4]U.G. SCHOLAR

G. Pullaiah College of Engineering and Technology

ABSTRACT

Credit card fraud is a serious problem in financial services. Billions of dollars are lost due tocredit card fraud every year. There is a lack of research studies on analyzing real-world credit card dataowing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit cardfraud. Standard models are firstly used. Then, hybrid methods which use AdaBoost and majority votingmethods are applied. To evaluate the model efficacy, a publicly available credit card data set is used. Then,a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to thedata samples to further assess the robustness of the algorithms. The experimental results positively indicatethat the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

I.INTRODUCTION

Fraud is a wrongful or criminal deception aimed to bring financial or personal gain [1]. In avoiding loss from fraud, two mechanisms can be used: fraud prevention and fraud detection. Fraud prevention is a proactive method, where it stops fraud from happening in the first place. On the other hand, fraud detection is needed when a fraudulent transaction is attempted by a fraudster. Credit card fraud is concerned with the illegal use of credit card information for purchases. Credit card transactions can be accomplished either physically or digitally [2]. In physical transactions, the credit card is involved during the transactions. In digital transactions, this can happen over the telephone or the internet. Cardholders typically provide the card number, expiry date, and card verification number through telephone or website. With the rise of e-commerce in the past decade, the use of credit cards has increased dramatically [3]. The number of credit card transactions in 2011 in Malaysia were at about 320 million, and increased in 2015 to about 360 million. Along with the rise of credit card usage, the number of fraud cases have been constantly increased. While numerous authorization techniques have been in place, credit card fraud cases have not hindered effectively. Fraudsters favour the internet as their identity and location are hidden. The rise in credit card fraud has a big impact on the financial industry. The global credit card fraud in 2015 reached to a staggering USD $21.84 billion [4]. Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges [5]. Since the merchants need to bear the loss, some goods are priced higher, or discounts and incentives are reduced. Therefore, it is imperative to reduce the loss, and an effective fraud detection system to reduce or eliminate fraud cases is important.

There have been various studies on credit card fraud detection. Machine learning and related methods are most commonly used, which include artificial neural networks, rule-induction techniques, decision trees, logistic regression, and support vector machines [1]. These methods are used either standalone or by combining several methods together to form hybrid models. IEEE In this paper, a total of twelve machine learning algorithms are used for detecting credit card fraud. The algorithms range from standard neural networks to deep learning models. They are evaluated using both benchmark and realworld credit card data sets. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set. The key contribution of this paper is the evaluation of a variety of machine learning models with a real-world credit card data set for fraud detection. While other researchers have used various methods on publicly available data sets, the data set used in this paper are extracted from actual credit card transaction information over three months

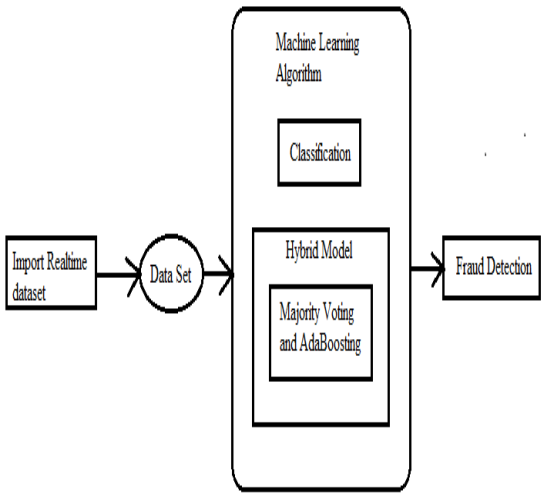
II.Existing System

Three methods to detect fraud are presented. Firstly, clustering model is used to classify the legal and fraudulent transaction using data clusterization of regions of parameter value. Secondly, Gaussian mixture model is used to model the probability density of credit card user's past behavior so that the probability of current behavior can be calculated to detect any abnormalities from the past behavior. Lastly, Bayesian networks are used to describe the statistics of a particular user and the statistics of different fraud scenarios. The main task is to explore different views of the same problem and see what can be learned from the application of each different technique.

III. Proposed System

Total of twelve machine learning algorithmsare used for detecting credit card fraud. The algorithmsrange from standard neural networks to deep learningmodels. They are evaluated using both benchmark and realworldcredit card data sets. In addition, the AdaBoost andmajority voting methods are applied for forming hybridmodels. To further evaluate the robustness and reliability ofthe models, noise is added to the real-world data set. Thekey contribution of this paper is the evaluation of a varietyof machine learning models with a real-world credit carddata set for fraud detection.

IV.Architecture



V.System Requirements

H/W System Configuration:-

Processor : Intel (R) Pentium (R)

Speed : 1.1 Ghz

RAM : 2GB

Hard Disk : 57 GB

Key Board : Standard Windows Keyboard

Mouse : Two or Three Button Mouse

Monitor : SVGA

S/W System Configuration

- Operating System : Windows 8/7/95/98/2000/XP
- Application Server : Tomcat5.0/6.X/8.X
- Front End : HTML, Java, Jsp
- Scripts : JavaScript.
- Server side Script : Java Server Pages.
- Database Connectivity : Mysql.
- Java Version : jdk 1.8

VI.Module Implementation

1. Standard Neural Networks To Deep Learning

The Feed-Forward Neural Network (NN) uses thebackpropagation algorithm for training as well. Theconnections between the units do not form a directed cycle,and information only moves forward from the input nodes tothe output nodes, through the hidden nodes. Deep Learning(DL) is based on an MLP network trained using a stochasticgradient descent with backpropagation. It contains a largenumber of hidden layers consisting of neurons with tanh,rectifier, and maxout activation functions. Every nodecaptures a copy of the global model parameters on local data,and contributes periodically toward the global model usingmodel averaging.

2. Forming Hybrid Models

Adaptive Boosting or AdaBoost is used in conjunction withdifferent types of algorithms to improve their performance.The outputs are combined by using a weighted sum, whichrepresents the combined output of the boosted classifier, AdaBoost tweaks weak learners in favor of misclassifieddata samples. It is, however, sensitive to noise and outliers.As long as the classifier performance is not random,AdaBoost is able to improve the individual results fromdifferent algorithms.Majority voting is frequently used in data classification,which involves a combined model with at least twoalgorithms. Each algorithm makes its own prediction forevery test sample. The final output is for the one thatreceives the majority of the votes,

3. Evaluate The Robustness And Reliability

To further evaluate the robustness of the machine learningalgorithms, all real-world data samples are corrupted noise,at 10%, 20% and 30%. Noise is added to all data features.It can be seen that with the addition of noise,the fraud detection rate and MCC rates deteriorate, asexpected. The worst performance, i.e. the largest decrease inaccuracy and MCC, is from

majority voting of DT+NB andNB+GBT. DS+GBT, DT+DS and DT+GBT show gradualperformance degradation, but their accuracy rates are stillabove 90% even with 30% noise in the data set.

VII.Algorithm

1. Machine Learning Algorithm

A total of twelve algorithms are used in this experimentalstudy. They are used in conjunction with the AdaBoost andmajority voting methods.Naïve Bayes (NB) uses the Bayes' theorem with strong ornaïve independence assumptions for classification. Certainfeatures of a class are assumed to be not correlated to others.It requires only a small training data set for estimating themeans and variances is needed for classification.The presentation of data in form of a tree structure is usefulfor ease of interpretation by users. The Decision Tree (DT) isa collection of nodes that creates decision on featuresconnected to certain classes. Every node represents a splittingrule for a feature. New nodes are established until the stoppingcriterion is met. The class label is determined based on themajority of samples that belong to a particular leaf. TheRandom Tree (RT) operates as a DT operator, with theexception that in each split, only a random subset of featuresis available. It learns from both nominal and numerical datasamples. The subset size is defined using a subset ratioparameter.

VIII. Conclusion

A study on credit card fraud detection using machinelearning algorithms has been presented in this paper. Anumber of standard models which include NB, SVM, and DLhave been used in the empirical evaluation. A publiclyavailable credit card data set has been used for evaluationusing individual (standard) models and hybrid models usingAdaBoost and majority voting combination methods. TheMCC metric has been adopted as a performance measure, as it takes into account the true and false positive and negativepredicted outcomes. The best MCC score is 0.823, achievedusing majority voting. A real credit card data set from afinancial institution has also been used for evaluation. Thesame individual and hybrid models have been employed. Aperfect MCC score of 1 has been achieved using AdaBoostand majority voting methods. To further evaluate the hybridmodels, noise from 10% to 30% has been added into the datasamples. The majority voting method has yielded the bestMCC score of 0.942 for 30% noise added to the data set.This shows that the majority voting method is stable inperformance in the presence of noise.

REFERENCES

[1] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013.

[2] A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," International Journal of System Assurance Engineering and Management, vol. 8, pp. 937–953, 2017.

[3] A. Srivastava, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detection using hidden Markov model," IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.

[4] The Nilson Report (October 2016) [Online]. Available: https://www.nilsonreport.com/upload/content_promo/The_Nilson_R eport_10-17-2016.pdf

[5] J. T. Quah, and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, 2008.

[6] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C., "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011. [

7] N. S. Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using Artificial Immune Systems," Applied Soft Computing, vol. 24, pp. 40–49, 2014.

[8] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning," Information Fusion, vol. 10, no. 4, pp. 354–363, 2009. [9] N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified Fisher discriminant analysis," Expert Systems with Applications, vol. 42, no. 5, pp. 2510–2516, 2015.

[10] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," Expert Systems with Applications, vol. 36, no. 2, pp. 3630–3640, 2009.