JES
Journal of Engineering Sciences
Vol 11, Issue 4 ,April/2020
ISSN NO:0377-9254

CREDIT CARD FRAUD DETECTIONUSING ADABOOST AND MAJORITY VOTING

KOLUSU. SOWMYA SREE[1], G. ANITHA[2], GADAGONI.LIKHITHA[3],
KOTA. PRIYANKA[4], DARAVENI.VYSHNAVI[5], JOIS.SRUTHI[6]
UG SCHOLAR[1,3,4,5&6], ASSOCIATE PROFESSOR[2]
DEPARTMENT OF CSE, AVN INSTITUTE OF ENGINEERING AND TECHNOLOGY,
KOHEDA ROAD, IBRAHIMPATNAM(M), R.R.DIST-501510, HYDERABAD

ABSTRACT

Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are first used. Then, hybrid methods which use AdaBoost and majority voting methods 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 the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

INTRODUCTION:

Fraud is a wrongful or criminal deception aimed to bring financial or personal gain . 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 . 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. 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 favor 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. Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges. 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 [7] [8]. 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. These methods are used either standalone or by combining several methods together to form hybrid models.

EXISTING SYSTEM

Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges. 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. These methods are used either

standalone or by combining several methods together to form hybrid models.

Disadvantages:

- Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges
- Billions of dollars are lost due to credit card fraud every year.
- credit card fraud cases have not hindered effectively

PROPOSED SYSTEM

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 real-world 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.

Advantages:

- In avoiding loss from fraud, two mechanisms can be used: fraud prevention and fraud detection
- he 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.
- The MCC metric has been adopted as a performance measure, as it takes into account the true and false positive and negative predicted outcomes.

MODULE IMPLEMENTATION

1. **Fraud Detection**

- **Decision Tree (DT)**

The presentation of data in form of a tree structure is useful for ease of interpretation by users. The Decision Tree (DT) is a collection of nodes that creates decision on features connected to certain classes. Every node represents a splitting rule for a feature. New nodes are established until the stopping criterion is met. The class label is determined based on the majority of samples that belong to a particular leaf. The Random Tree (RT) operates as a DT operator, with the exception that in each split, only a random subset of features is available. It learns from both nominal and numerical data samples. The subset size is defined using a subset ratio parameter.

The Random Forest (RF) creates an ensemble of random trees. The user sets the number of trees. The resulting model employs voting of all created trees to determine the final classification outcome. The Gradient Boosted Tree (GBT) is an ensemble of classification or regression models. It uses forward-learning ensemble models, which obtain predictive results using gradually improved estimations. Boosting helps improve the tree accuracy.

- **Naïve Bayes (NB)**

Naïve Bayes (NB) uses the Bayes' theorem with strong or naïve independence assumptions for classification. Certain features of a class are assumed to be not correlated to others. It requires only a small training data set for estimating the means and variances is needed for classification.

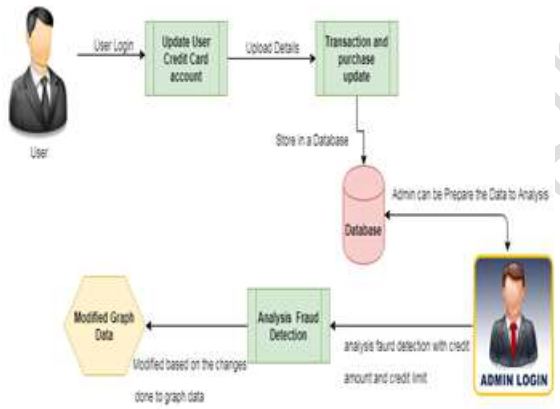
- **The Random Forest (RF)**

The Random Forest (RF) creates an ensemble of random trees. The user sets the number of trees. The resulting model employs voting of all created trees to determine the final classification outcome. The Gradient Boosted Tree (GBT) is an ensemble of classification or regression models. It uses forward-learning ensemble models, which obtain predictive results using gradually improved estimations. Boosting helps improve the tree accuracy. The Decision Stump (DS) generates a decision tree with a single split only. It can be used in classifying uneven data sets.

2. AdaBoost and Majority Voting

Adaptive Boosting or AdaBoost is used in conjunction with different types of algorithms to improve their performance. The outputs are combined by using a weighted sum, which represents the combined output of the boosted classifier. AdaBoost tweaks weak learners in favor of misclassified data 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 from different algorithms.AdaBoost helps improve the fraud detection rates, with a noticeable difference for NB, DT, RT, which produce a perfect accuracy rate. The most significant improvement is achieved by LIR. 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.The majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

Architecture



ALGORITHM

1. Machine Learning Algorithms

Machine learning is the science of designing and applying algorithms that are able to learn things from past cases. It uses complex algorithms that iterate over large data sets and analyze the patterns in data. The algorithm facilitates the machines to respond to different situations for which they have not been explicitly programmed. It is used in spam detection, image recognition, product recommendation, predictive analytics etc. Significant reduction of human effort is the main aim of data scientists in implementing ML. Even with modern analytics tools, it takes a lot of time for humans to read, collect, categorize and analyze the data. ML teaches machines to identify and gauge the importance of patterns in place of humans. Particularly for use cases where data must be analyzed and acted upon in a short amount of time, having the support of machines allows humans to be more efficient and act with confidence.

CONCLUSION Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results. While the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. However, when the entire dataset is fed into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.

FUTURE ENHANCEMENTS While we couldn't reach out goal of 100% accuracy in fraud detection, we did end up creating a system that can, with enough time and data, get very close to that goal. As with any such project, there is some room for improvement here. The very nature of this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result. This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project. More room for improvement can be found in the dataset. As demonstrated before, the precision of the algorithms increases when the size of dataset is increased. Hence, more data will surely make the model more accurate in detecting frauds and reduce the number of false positives. However, this requires official support from the banks themselves.

REFERENCES

[1] "Credit Card Fraud Detection Based on Transaction Behaviour -by John Richard D. Kho, Larry A. Vea" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[2] CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 " A Comprehensive Survey of Data Mining-based Fraud Detection Research" published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800, Australia

[3] "Survey Paper on Credit Card Fraud Detection by Suman" , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

[4] "Research on Credit Card Fraud Detection Model Based on Distance Sum – by Wen-Fang YU and Na Wang" published by 2009 International Joint Conference on Artificial Intelligence

[5] "Credit Card Fraud Detection through Parenclitic Network AnalysisBy Massimiliano Zanin, Miguel Romance, ReginoCriado, and SantiagoMoral" published by Hindawi Complexity Volume 2018, Article ID 5764370, 9 pages

[6] "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018

[7] Mohammad Gandhi Babu, Pravin Kshirsagar, Boyini Mamatha, Pranav Chippalkatti, "A Machine Learning Approach for Credit Card Fraud Detection",Test Engineering and Management, January-February 2020 ISSN: 0193-4120, Vol. 82 Page No. 5237 – 5244,2020

[8]. K Ashok Kumar, C Jagadeesh, Pravin Kshirsagar, Swagat M Marve "Sentiment Analysis of Amazon Product Reviews using Machine Learning", January-February 2020 ISSN: 0193-4120 , Vol. 82 Page No. 5245-5254,2020