Industrial Engineering Journal
ISSN: 0970-2555
Volume : 51, Issue 3, March 2022

MODERN APPROACH TO DETECT CREDIT CARD FRAUD THROUGH MACHINE LEARNING

[1]Dr.Y.Narasimha Reddy,[2]Machani Keerthika,[3]Jujaray Preethi,[4]Vadde Divya,[5]Shaik Ayesha Begum,[6]Katika Yogitha Sri

[1]Associate Professor,[2]Students

Department Of CSE

St. Johns College of Engineering & Technology, Errakota, Yemmiganur

Abstract

Credit card fraud is a significant challenge within the realm of financial services. Every year, credit card theft results in the loss of billions of dollars. Insufficient research studies exist for assessing real-world credit card data due to concerns over confidentiality. This article use machine learning techniques to identify instances of credit card fraud. Initially, standard models are used. Subsequently, a combination of Ada Boost and majority voting techniques are used in hybrid approaches. In order to assess the effectiveness of the model, a publicly accessible dataset of credit card information is used. Next, an actual credit card dataset from a financial organization is examined. Furthermore, noise is introduced into the data samples to further evaluate the resilience of the algorithms. The experimental findings unequivocally demonstrate that the majority voting approach attains high accuracy rates in identifying instances of fraud in credit card transactions.

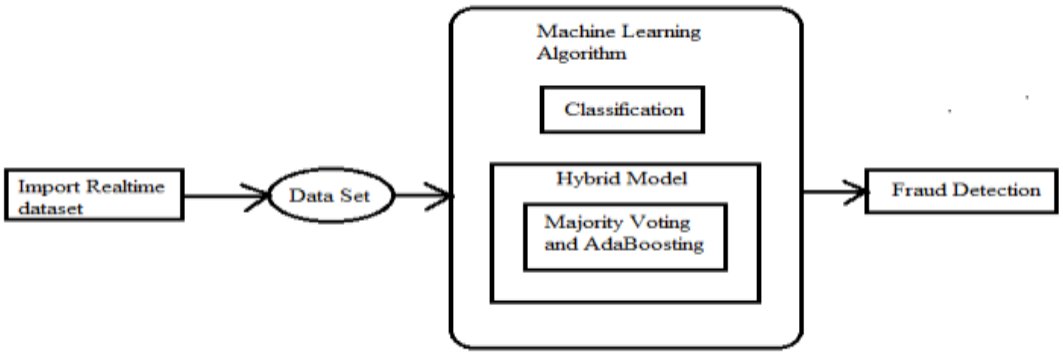
I. INTRODUCTION

Fraud is an act of deliberate deceit or dishonesty with the intention of obtaining financial or personal benefits [1]. To mitigate the risk of financial loss due to fraudulent activities, two strategies may be employed: fraud prevention and fraud detection. Fraud prevention is a preemptive approach that aims to thwart fraudulent activities before they occur. However, fraud detection is necessary when a deceitful transaction is undertaken by a dishonest individual. Credit card fraud refers to the illicit use of credit card data to make unauthorized transactions. Credit card transactions may be conducted using physical or digital means [2]. The credit card is used in the course of physical transactions. During digital transactions, this may occur either by telephone or internet communication. Cardholders often provide their card number, expiration date, and card verification number either by telephone or a website. The use of credit cards has seen a significant surge in recent years due to the rapid growth of e-commerce [3]. In 2011, the total number of credit card transactions in Malaysia was around 320 million. By 2015, this figure had risen to almost 360 million. As credit card use has expanded, the number of fraud instances has also consistently risen. Despite the use of various authentication procedures, credit card fraud incidents have not been adequately prevented. Criminals like to use the internet because it allows them to conceal their

identity and whereabouts. Credit card fraud has a significant influence on the financial sector. In 2015, the worldwide credit card theft amounted to an astonishing USD $21.84 billion [4]. Merchants incur all expenses, such as card issuer fees, charges, and administrative fees, as a result of credit card fraud, leading to financial losses [5]. Due to the merchants' responsibility for covering the loss, they may increase the prices of certain items or decrease the availability of discounts and incentives. Hence, it is crucial to minimize the loss, and the implementation of an efficient fraud detection system to mitigate or eradicate instances of fraud is of utmost significance. Multiple research have been conducted on the detection of credit card fraud. The most often used approaches in this field include machine learning and its associated techniques, such as artificial neural networks, rule-induction techniques, decision trees, logistic regression, and support vector machines [1]. These strategies are used either alone or in conjunction with other methods to create hybrid models. IEEE This article employs a total of twelve machine learning algorithms to identify instances of credit card fraud. The algorithms include a variety of techniques, including conventional neural networks and advanced deep learning models. They undergo evaluation using both benchmark and real-world credit card datasets. Furthermore, the Ada Boost and majority voting techniques are used to create hybrid models. In order to assess the strength and dependability of the models, additional noise is introduced into the real-world dataset. The primary significance of this work is in the assessment of several machine learning models using an actual credit card data set for the purpose of detecting fraud. While previous researchers have used diverse methodologies on publically accessible data sets, the data set utilized in this study is derived from authentic credit card transaction data over a period of three months. The structure of this document is as follows. Section II provides an overview of previous research on both individual and combined machine learning methods used in financial applications. The paper outlines the machine learning methods used in Section III. The studies using both benchmark and real-world credit card datasets are outlined in Section IV. Section V provides concluding observations and ideas for further study.

II. SYSTEM ARCHITECTURE

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 51, Issue 3, March 2022



III. 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.

IV. PROPOSED SYSTEM

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 Ada Boost 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.

V. IMPLEMENTATION

1. Standard Neural Networks To Deep Learning

The Feed-Forward Neural Network (NN) uses the back propagation algorithm for training as well. The connections between the units do not form a directed cycle, and information only moves forward from the input nodes to the output nodes, through the hidden nodes. Deep Learning(DL) is based on an MLP network trained using a stochastic gradient descent with backpropagation. It contains a large number of hidden layers consisting of neurons with tan h, rectifier, and max-out activation functions. Every node captures a copy of the global model parameters on local data, and contributes periodically toward the global model using model averaging.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 51, Issue 3, March 2022

2. Forming Hybrid Models

Adaptive Boosting or Ada Boost 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, Ada Boost 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, Ada Boost is able to improve the individual results from different algorithms. 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,

3. Evaluate The Robustness And Reliability

To further evaluate the robustness of the machine learning algorithms, 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, as expected. The worst performance, i.e. the largest decrease inaccuracy and MCC, is from majority voting of DT+NB and NB+GBT. DS+GBT, DT+DS and DT+GBT show gradual performance degradation, but their accuracy rates are still above 90% even with 30% noise in the data set.

Algorithm

1. Machine Learning Algorithm

A total of twelve algorithms are used in this experimental study. They are used in conjunction with the Ada Boost and majority voting methods. 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 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.

VI. CONCLUSION

This article presents a research on the use of machine learning techniques for credit card fraud detection. Several conventional models, such as Naive Bayes (NB), Support Vector Machines (SVM), and Deep Learning (DL), have been used in the empirical assessment. An openly accessible

dataset including credit card information has been used for assessment purposes. Both individual (standard) models and hybrid models, using AdaBoost and majority voting combination approaches, have been employed for analysis. The MCC metric is used as a performance measure since it considers both the true and erroneous positive and negative expected outcomes. The highest MCC (Matthews Correlation Coefficient) score obtained is 0.823, which was accomplished by the use of majority voting. Additionally, a genuine credit card dataset obtained from a financial institution has been used for assessment purposes. Both individual and hybrid models have been used. An optimal MCC score of 1 has been attained by the use of AdaBoost and majority voting techniques. In order to conduct a more comprehensive assessment of the hybrid models, several levels of noise ranging from 10% to 30% have been introduced into the data sets. The majority voting technique achieved the highest Matthews Correlation Coefficient (MCC) score of 0.942 after 30% noise was introduced to the data set. This demonstrates that the majority voting technique exhibits consistent performance even in the face of noise. In further research, the techniques used in this study will be expanded to include online learning models. Furthermore, other online learning models will be examined. Online learning may facilitate the swift identification of fraudulent activities, perhaps in real-time. Consequently, this will aid in identifying and thwarting deceitful transactions prior to their occurrence, thus reducing the daily count of financial sector losses.

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.

[11] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063, 2011.

[12] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining 60techniques," Decision Support Systems, vol. 50, no. 2, pp. 491–500, 2011.