Detection of Credit Card Fraud Using AdaBoost and Majority Voting

Vinutha D[1], Dr. Mohammed Rafi[2]
[1]*Department of Studies in Computer Science & Engineering University BDT College of Engineering (A Constituent College of VTU, Belagavi), Davanagere, Karnataka*
[2]*Professor, Department of Computer Science & Engineering, University BDT College of Engineering (A Constituent College of VTU, Belagavi), Davanagere, Karnataka*

***Abstract*- Credit card is one of the popular modes of payment for electronic transactions. With the developments in the information technology, fraud is spreading all over the world, resulting in huge financial losses. Credit card fraud is a serious and growing problem. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are used first. Then, hybrid methods which use AdaBoost and majority voting methods are applied. We present detailed experimental results to show the effectiveness of our approach and compare it with other techniques available in the literature. This paper provides a picture of recent trend in credit card fraud detection.**

Index terms- credit card, fraud detection, electronic transaction, AdaBoost, majority voting, classification.

I. INTRODUCTION

Credit-card-based purchases can be categorized. In to two types:1)physicalcardand2)virtual card. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has
to steal the credit card [1]. Invention of credit cards has made online transactions seamless, comfortable and convenient. However, it has also provided new fraud opportunities for criminals, and in turn, increased fraud rate. The global impact of credit card fraud is alarming, millions of US dollars have been lost by many companies and individuals. This paper presents a review of improved credit card fraud detection techniques. Precisely, this paper focused on recent Machine Learning based credit card fraud detection techniques[2]. Data mining is widely used in financial and internet fields. Contrasting with the whole credit card trade, credit card fraud transactions are the few anomalies[3].

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. Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges [4]. The traditional majority voting in RF was replaced with the potential nearest neighbor method. A total of 12 different data sets were used in the experimental study. The PCA-based model produced a higher classification accuracy and a lower variance, as compared with those from RF and DT methods[5].

In the existing technology 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.

The organization of this paper is as follows. Literature survey in section II. Methodology in III, Implementation in section IV, results and analysis in

section V. Conclusion for work are given in Section VI.

II LITERATURE SURVEY

The paper [1] states that different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. We have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM. We have suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions.

The paper [2] states that the 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.

The paper [3] states the feasibility of credit card fraud detection based on outlier mining, applies outlier detection mining based on distance sum into credit card fraud detection and proposes this detection procedures and its empirical process. And finally this method proves accurate in predicting fraudulent transactions through outlier mining emulation experiment of credit card transaction data set of one certain commercial bank. The experiment shows that outlier mining can detect credit card fraud better than anomaly detection based on clustering when anomalies are far less than normal data. If this algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions by the banks. And a series of anti-fraud strategies can be adopted to prevent banks from great losses before and reduce risks.

III METHODOLOGY

To overcome the limitations of existing technology 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.

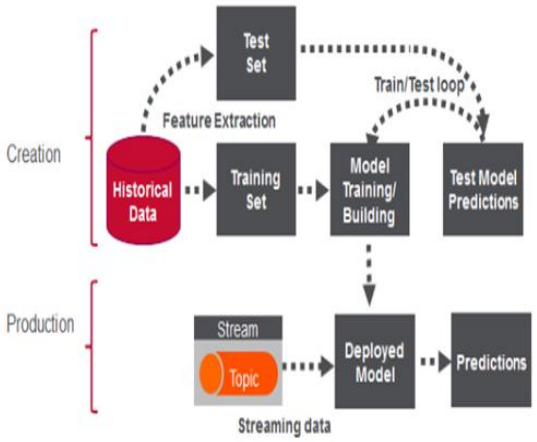

Fig. 1: Architecture of proposed system

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.

IV IMPLEMENTATION

Fraud detection is done using Adaboost and majority voting methods. 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. 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.

V RESULT AND ANALYSIS

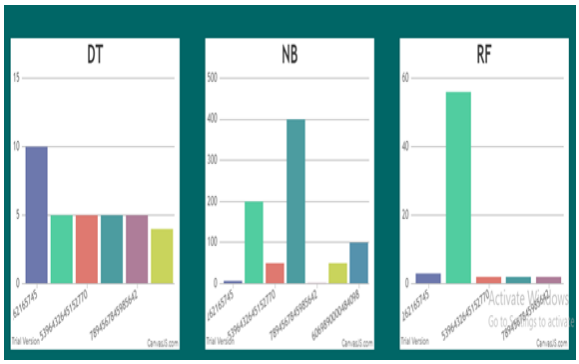

Figure1. Fraud detection using adaboost and majority voting

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 data samples. The majority voting method has yielded the best MCC score of 0.942 for 30% noise added to the data set. It can be represented by the pie chart, column chart, bar chart, spline chart, line chart, area chart using DT, NB, RF. This shows that the majority voting method offers robust performance in the presence of noise.

VI CONCLUSION

A study on credit card fraud detection using machine learning algorithms has been presented in this paper. A number of standard models which include NB, SVM, and DL have been used in the empirical evaluation. A publicly available credit card data set has been used for evaluation using individual (standard) models and hybrid models using AdaBoost and majority voting combination methods. The MCC metric has been adopted as a performance measure, as it takes into account the true and false positive and negative predicted outcomes. A real credit card data set from a financial institution has also been used for evaluation.

For future work, the methods studied in this paper will be extended to online learning models. In addition, other online learning models will be investigated. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector.

REFERENCES

[1] Credit Card Fraud Detection Using Hidden Markov Model, Abhinav Srivastava, Amlan Kundu, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar, Senior Member, IEEE.

[2] A. O. Adewumi and A. A. Akinyelu,‘‘A survey of machine-learning and nature-inspired based credit card fraud detection techniques,’’ Int. J. Syst. Assurance Eng. Manage., vol. 8, no. 2, pp. 937–953, 2017.

[3] Research on Credit Card Fraud Detection Model Based on Distance Sum. Wen-Fang YU Na Wang Computer Science and Information Engineering College Zhejiang Gongshang University Hangzhou, China Ywf_1@163.com

[4] J. T. Quah and M. Sriganesh, ‘‘Real-time credit card fraud detection using computational intelligence,’’ Expert Syst. Appl., vol. 35, no. 4, pp. 1721–1732, 2008.

[5] S. Subudhi and S. Panigrahi, ‘‘Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection,’’ J. King Saud Univ.-Comput. Inf. Sci., to be published, doi: 10.1016/j.jksuci.2017.09.010.