ISSN: 2581-3404 (Online) **IF: 5.445 (SJIF)**

IJIRTM, Volume-7, Issue-4, July-2023.

Credit Card Fraud Detection: Survey and Discussion

Hena Naaz[1], Prof. Tanweer Farooki[2]

[1]M. Tech. Scholar, Department of CSE, ASCT, Bhopal, M.P. (India)

[2]Assistant Professor, Department of CSE, ASCT, Bhopal, M.P. (India)

Abstract: *E-commerce has flourished in the recent decades. As an 16 increasing number of people are accustomed to online transactions, this has contributed to the prevalence of card payments. Unfortunately, the prevailing emergence of spending behavior has become an ideal condition for the increase in fraudulent activities. Fraud detection is to identify, monitor, and prevent potentially fraudulent activities from complex data. The recent development and success in AI, especially machine learning, provides a new data-driven way to deal with fraud. Machine Learning is considered as one of the most successful technique used for creating a fraud detection algorithm for fraud identification.*

Keywords: Machine learning methods, Credit Card Fraud Detection, Classification method, Supervised learning, Deep learning.

Introduction

Now these days digital, statistics are very easily available throughout the world because of digital online availability. All the information that also has a large volume, wide range, frequency, as well as importance is stored from small to large organizations over the cloud. The whole information is available from massive amounts of sources such as followers on social media, customer order behaviors, likes, and shares. White-collar crime is the ever-increasing problem with-reaching consequences for the finance sector, business institutions as well as governments. In recent years there has been an increase in financial fraud due to the growth of technologies and paradigms such as the e-commerce and the financial technology (FinTech) sectors [1]. The evolution of these technologies has sparked an increase in the number of credit card transactions. As a result, there has been a rapid spike in the number financial fraud cases that involved credit cards. Credit card Fraud occurs when an unauthorized or undesirable use of a credit card is made by a criminal. This happens when the credit card authentication details are stolen using different types of fraudulent techniques such as intercepting an e-commerce transaction or cloning an existing card [2]. Moreover, the impact of credit card fraud affects institutions such as card issuers, merchants, and small businesses. Fraud can indeed be described as illegal deceit to gain financial benefit Enhanced card transactions had already appreciated a heavy emphasis on communication technology. When credit card transactions are by far the most prevalent form of transaction for offline and online payments, raising the rate of card fraud accelerates as well [4]. Financial fraud is a serious problem that is only getting worse and has far-reaching effects on the financial sector, businesses, and the government. Fraud is defined as criminal deception done with the intention of making money. Credit card transactions have surged thanks to a high reliance on internet technology. The rate of credit card fraud is rising as credit card transactions take over as the preferred method of payment for both online and offline transactions. There are two types of credit card fraud: internal and external [4]. While external card fraud entails using a stolen credit card to obtain money through illegal ways, inner card fraud happens as a result of an agreement between cardholders and the bank and involves using a fake identity to commit fraud. Most credit card frauds are external card fraud, which has been the subject of much investigation. Another classification has been made into three categories: classic card-related frauds (application, stolen, account takeover, fake, and counterfeit), frauds involving retailers (merchant collusion and triangulation), and frauds involving the internet (site cloning, credit card generators, and false merchant sites) [3].

Fraud detection originally relied on rule-based expert system embedding specific domain knowledge, and more recently machine learning has been applied most extensively to mine fraudulent patterns [2]. Due to their time-consuming nature and ineffectiveness, manual methods of fraud detection have become increasingly impracticable with the introduction of big data. The challenge of credit card fraud, however, has drawn the attention of financial institutions to current computational approaches. One significant way for detecting credit fraud is the use of data mining techniques. The technique of separating fraudulent transactions into two categories: legitimate and fraudulent transactions are known as credit card fraud detection. There are several different types of credit card fraud. One of these is stealing a physical card while the other is the stealing of confidential credit card information like account number, CVV key, card type, and many others. The fraudster can try to address the significant amount of money and make a massive amount of payment before the cardholder figures out by manipulating credit card information. Now because of this, businesses are using different techniques for machine learning that identify increasing transactions constitute illegitimate and which are not. Whenever a credit card becomes the more common transaction model (respectively online as well as regular transactions), its frequency of fraud tends may accelerate. Machine learning based research on fraud detection focused on classification techniques such as Logistic Regression, SVM and Bayesian Belief Networks, as well as neural networks, Random forest algorithm, While they are widely used, these classical methods do not account for why they can make sense in imbalanced data and need to collect fraud activities frequently and relearn periodically. Actually, several studies have shown that Logistic Regression, SVM and Random Forests all perform significantly better at detecting legitimate transactions correctly than fraudulent ones [5]. However, fraud detection is a problem with a large difference in misclassification costs: it is typically far more expensive to misdiagnose a fraudulent transaction as legitimate as the reverse.

Rest of the paper is organized as follows: section II explains background of machine learning techniques in brief, in section III literature review of previous research in the field of credit card fraud detection using machine learning is analyzed, in last we conclude our work in section IV followed by references used in this work.

II. Machine Learning

Machine learning is the innovation of this century that eliminates conventional strategies and also can function on huge datasets where humans can't immediately access. Strategies of machine learning break within two important categories; supervised learning versus unsupervised learning; Tracking of fraud can also be achieved any form and may only be determined how to use as per the datasets. Supervised training includes anomalies to always be identified as before. Many supervised methods are being used over the last few decades to identify credit card fraud. The major obstacle in implementing ML for detecting fraud seems to be the presence of extremely imbalanced databases. Most payments are legitimate in several available evidence sets, with such an extremely small number of fraudulent ones. The significant challenges to investigators are designing the accurate as well as efficient fraud prevention framework that will be low on false positives but efficiently identifies fraud activity.

- Random Forest: This is one of the ensemble methodologies used only to improve its prosperity as well as precision in machine learning algorithms of artificial intelligence. One of this kind classifier is the Random Forest (RF), suggested by Breiman, a researcher. A random forest method may also help identify the genuinely appropriate independent variables such that the system may pick functionality. Also, many findings already demonstrate its significance in selecting several possibilities for each shrub, but in empirical research, this is discovered to also be optimal regarding forecast accuracy [12].
- Naïve Baiyes Classifier: This is indeed a statistical process based on Predictive theory that selects its greatest probability focused ruling. Unidentified outcomes from recognized value systems have been estimated by Bayesian likelihood. This also enables the implementation of previous knowledge as well as logic in unpredictable assertions. That first methodology seems to have a legally binding independence presumption around characteristics throughout the data.
- Logistic Regression: It is another method decided to borrow from either the profession of statistical data by machine learning. It is also the go-to process of issues concerning binary categorization (difficulties with more than just two class moral values). Logistic regression to the mean is used to modeling a class's outcome like actual pass / completely fail, positive and constructive/negative or neutral again and in credit card fraud threat detection cases then we use probability distribution class as fraudulent and not fraud [13].
- Support Vector Classifier: A support vector machine-based machine learning method is also known as SVM; mainly a supervised model. This more or less uses classification learning algorithms also for classification major problems even in two groups and individuals. They really can classify a new document since giving the number of the labeled dataset to each classification on an SVM system.
- K-Nearest Neighbors (kNN): A K-nearest Neighbors (KNN) classifier seems to be a straightforward, simple-to-implement supervised machine learning algorithm that could be used to address respectively classification as well as regression difficulties.
- Classification Trees: The Classification tree marks, records, as well as allocates separate class factors. The Classification tree may provide a charisma measure that perhaps the category becomes accurate. The Classification Tree has been constructed via a process called binary recursive partitioning.
- Artificial Neural Networks: This is a kind of machine learning method patterned on the brain and nervous system. Utilizing historical information, its ANN designs may discover the trends, and also can classify the incoming data.
- Gradient Boosting (GBM): Gradient Boosting is also known as the GB method, is a prominent algorithm of machine learning, always had to conduct classification as well as regression activities. The above model consists of such an amount in fundamental ensemble designs such as feeble decision trees. Such decision trees combine to create a powerful specular reflection-boosting model [15].

III. Related Work

Machine learning approaches play a crucial role throughout numerous efficient areas for data processing; one of them is the identification of card fraud. Through previous research, several methods were suggested to include strategies for detecting fraud through supervised methods, unsupervised methods including a hybrid strategy; that makes it necessary and know some technology involved in identifying credit card fraud and have a better understanding of the types of card fraud. Many strategies were suggested and checked.

[1] This paper proposes an efficient approach to detect credit card fraud using a neural network ensemble classifier and a hybrid data re-sampling method. The ensemble classifier is obtained using a long short term memory (LSTM) neural network as the base learner in the adaptive boosting (AdaBoost) technique. Meanwhile, the hybrid re-sampling is achieved using the synthetic minority oversampling technique and edited nearest neighbor (SMOTE-ENN) method.

[2] In this paper, author proposes deep boosting decision trees (DBDT), a novel approach for fraud detection based on gradient boosting and neural networks. In order to combine the advantages of both conventional methods and deep learning, they first construct soft decision tree (SDT), a decision tree structured model with neural networks as its nodes, and then ensemble SDTs using the idea of gradient boosting. In this way we embed neural networks into gradient boosting to improve its representation learning capability and meanwhile maintain the interpretability. Furthermore, aiming at the rarity of detected fraud cases, in the model training phase we propose a compositional AUC maximization approach to deal with data imbalances at algorithm level. Extensive experiments on several real-life fraud detection datasets show that DBDT can significantly improve the performance and meanwhile maintain good interpretability.

[3] The credit card has become the most popular payment method for both online and offline transactions. The necessity to create a fraud detection algorithm to precisely identify and stop fraudulent activity arises as a result of both the development of technology and the rise in fraud cases. This paper implements the random forest (RF) algorithm to solve the issue in the hand. A dataset of credit card transactions was used in this study. The main problem when dealing with credit card fraud detection is the imbalanced dataset in which most of the transaction are non-fraud ones. To overcome the problem of the imbalanced dataset, the synthetic minority over-sampling technique (SMOTE) was used. Implementing the hyperparameters technique to enhance the performance of the random forest classifier.

[4] In this paper, author introduce an effective credit card fraud detection mechanism including a feedback system, dependent on machine learning methodology. Its feedback approach contributes to enhancing the classifier's detection rate as well as cost-effectiveness. Afterward examined the performance of different methodologies incorporates random forest, tree classifiers, artificial neural networks, support vector machine, Naïve Baiyes, logistic regression and gradient boosting classifier strategies, on a slightly skewed credit card fraud data sets. These data sets include transaction data through credit card emerges from European account holders with 284,807 trades. Similar approaches apply towards both raw including and pre-processed content. The efficiency of the approaches has always been evaluated depending on just the performance assessment dimensions for different classifiers, which will include precision, recall, F1-score, accuracy, and FPR percentage.

[5] This research was conducted on the IEEE-CIS Fraud Detection Dataset provided by Vesta Corporation. Based on the logic of labeling for converting the entire account to ``FraudD1" once the credit card has fraud, we navigate the research process towards predicting fraudulent credit cards rather than fraudulent transactions. The key idea behind the proposed model is user separation, in which they divide users into old and new people before applying CatBoost and Deep Neural Network to each category, respectively. In

addition, a variety of techniques to improve detection accuracy, namely handling heavily imbalanced datasets, feature transformation, and feature engineering, are also presented in detail in this paper.

[6] This research seeks to detect credit card fraud and make attempts to cut down on it. Financial institutions place a high priority on identifying and stopping fraudulent activity. Fraud prevention and detection are pricey, time-consuming, and labor-intensive processes. Several machine learning algorithms can be utilized for detection. In order to evaluate past customer transaction information and identify behavioral traits, the study's main goal is to develop and apply a special fraud detection algorithm for simulcasting transaction data. Through the research, try to give a genuine solution to Credit card users and make their transactions secure. This research aims to propose a trustworthy and efficient way for identifying credit card fraud. The accuracy of several autonomous classifiers using machine learning that were employed for recognition is compared and examined.

[7] The detection of fraudulent transactions has become a significant factor affecting the greater utilization of electronic payment. Thus, there is a need for efficient and effective approaches for detecting fraud in credit card transactions. This paper proposes an intelligent approach for detecting fraud in credit card transactions using an optimized light gradient boosting machine (OLightGBM). In the proposed approach, a Bayesian-based hyper parameter optimization algorithm is intelligently integrated to tune the parameters of a light gradient boosting machine (LightGBM). To demonstrate the effectiveness of our proposed OLightGBM for detecting fraud in credit card transactions, experiments were performed using two real-world public credit card transaction data sets consisting of fraudulent transactions and legitimate ones.

[8] In this research study, the main aim is to detect such frauds, including the accessibility of public data, high-class imbalance data, the changes in fraud nature, and high rates of false alarm. The relevant literature presents many machines learning based approaches for credit card detection, such as Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and XG Boost. However, due to low accuracy, there is still a need to apply state of the art deep learning algorithms to reduce fraud losses. The main focus has been to apply the recent development of deep learning algorithms for this purpose. Comparative analysis of both machine learning and deep learning algorithms was performed to find efficient outcomes. The detailed empirical analysis is carried out using the European card benchmark dataset for fraud detection. A machine learning algorithm was first applied to the dataset, which improved the accuracy of detection of the frauds to some extent. Later, three architectures based on a convolutional neural network are applied to improve fraud detection performance. Further addition of layers further increased the accuracy of detection. A comprehensive empirical analysis has been carried out by applying variations in the number of hidden layers, epochs and applying the latest models.

[10] The fraud detection system in banking organization relies on data-driven approach to identify the fraudulent transactions. In real time, detection of each and every fraudulent transaction becomes a challenging task as financial institutions need aggressive jobs running on the log data to perform a data mining task. This paper introduces a novel model for credit card fraud detection which combines ensemble learning techniques such as boosting and bagging. Their model incorporates the key characteristics of both the techniques by building a hybrid model of bagging and boosting ensemble classifiers. Experimentation on Brazilian bank data and UCSD-FICO data with our model shows sturdiness over the state-of-the-art ones in detecting the unseen fraudulent transactions because the problem of data imbalance was handled by a hybrid strategy.

[11] The advance in technologies such as e-commerce and financial technology (FinTech) applications have sparked an increase in the number of online card transactions that occur on a daily basis. As a result, there has been a spike in credit card fraud that affects card issuing companies, merchants, and banks. It is therefore essential to develop mechanisms that ensure the security and integrity of credit card transactions. In this research, we implement a machine learning (ML) based framework for credit card fraud detection using a real world imbalanced datasets that were generated from European credit cardholders. To solve the issue of class imbalance, they re-sampled the dataset using the Synthetic Minority over-sampling technique (SMOTE). This framework was evaluated using the following ML methods: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Decision Tree (DT), and Extra Tree (ET). These ML algorithms were coupled with the Adaptive Boosting (AdaBoost) technique to increase their quality of classification. The models were evaluated using the accuracy, the recall, the precision, the Matthews Correlation Coefficient (MCC), and the Area Under the Curve (AUC).

IV. Conclusion

New advances in electronic commerce systems and communication technologies have made the credit card the potentially most popular method of payment for both regular and online purchases; thus, there is significantly increased fraud associated with such transactions. Increase in online transactions using payment methods like credit card has also increased the fraudulent activities. Every year, a large amount of financial losses are caused by these illegal credit card transactions. No system is 100% secure and there is always a loophole in them. Therefore there is need to solve the issues of detecting fraud in transactions done by credit cards. In this paper we study different data mining and machine learning techniques for credit card fraud detection, Future work associated may explore the use of more state of art machine learning methods to improve the performance of the existing method study, and enhance the performance of an existing system.

References

[1] Ebenezer Esenogho, Ibomoiye Domor Mienye, “A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection”, IEEE Access, 2022, pp. 16400-16408.

[2] Biao Xua, Yao Wang, “Efficient Fraud Detection Using Deep Boosting Decision Trees”, 2023, pp. 1-34.

[3] AlsharifHasan Mohamad Aburbeian, “Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data”, 2022, pp. 1-11.

[4] Naresh Kumar Trivedi, Sarita Simaiya, “An Efficient Credit Card Fraud Detection Model Based on Machine Learning Methods”, International Journal of Advanced Science and Technology, 2020, pp. 3414 - 3424.

[5] Nghia Nguyen, Truc Duong, “A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network”, IEEE Access, 2023, pp. 96852-96861.

[6] Anik Malaker, “An Approach to Detect Credit Card Fraud Utilizing Machine Learning”, Int. J. Advanced Networking and Applications,2023, 5619-5625.

[7] Altyeb Altaher Taha, Sharaf Jameel Malebary, “An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine”, IEEE Access, 2020, pp. 25579-25588.

[8] Fawaz Khaled Alarfaj, Iqra Malik, “Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms”, IEEE Access, 2022, pp. 39700-39715.

[9] Rashmi S. More, Chetan J. Awati, “Credit Card Fraud Detection Using Supervised Learning Approach”, International Journal Of Scientific & Technology Research, 2020, pp. 216-220.

[10] V. S. S. Karthik, Abinash Mishra, “Credit Card Fraud Detection by Modelling Behaviour Pattern using Hybrid Ensemble Model”, Arabian Journal for Science and Engineering, Springer 2021, pp. 1-12.

[11] Emmanuel Ileberi, Yanxia Sun, “Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost”, IEEE Access, 2021, pp. 165286-165295.

[12] Dileep M R, Navaneeth A V, “A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms”, IEEE, 2021, pp. 1025-1028.

[13] Mosa M. M. Megdad, Bassem S. Abu-Nasser, “ Fraudulent Financial Transactions Detection Using Machine Learning”, International Journal of Academic Information Systems Research, 2022, pp. 30-39.

[14] Pumsirirat, Apapan, and Liu Yan. "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine." International Journal of advanced computer science and applications 9.1 (2018): 18-25.

[15] G. Sudha Sadasivam, Mutyala Subrahmanyam and Dasaraju Himachalam, Bhanu Prasad Pinnamaneni, “Corporate governance fraud detection from annual reports using big data analytics”, Int. J. Big Data Intelligence, Vol. 3, No. 1, 2016

[16] Ophir Gottlieb, Curt Salisbury, Howard Shek, Vishal Vaidyanathan, “Detecting Corporate Fraud: An Application of Machine Learning”, December 15, 2006

[17] Renjith, S. Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach. International Journal of Engineering Trends and Technology (2018).

[18] Roy, Abhimanyu, et al. "Deep learning detecting fraud in credit card transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2018.