Detection of Credit Card Fraud using Machine Learning and Deep Learning: A Review

Sejal Kharat [1], Sakshi Taur [1], Rucha Khalate [1], Kimaya Kate [1]

[1]*Student, Department of Computer Engineering, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Savitribai Phule Pune University, Pune, India.*

Corresponding Author: sakshisgtaur7840@gmail.com

Abstract: - Nowadays use of credit card is increased due to virtual world, along with its usage there is rapid growth in its misuse and fraud. There are different types of credit card frauds which needs to be identified. Such frauds lead to many financial losses for the card owner as well as for the company. The main aim is to identify whether a particular transaction is a fraud or not. For detecting fraud there is need to access public data, high class imbalance data, change in fraud nature and high false alarm rate. There are many machines learning based approaches like Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and XG Boost. To get more accurate results state of art deep learning algorithms are applied. Analysis of both machine learning and deep leaning algorithms was done to attain meticulous result. The detailed evaluation of European card benchmark is carried out to identify the fraud transactions. Initially machine learning algorithms were applied to the dataset which has increased the accuracy of detecting fraud and later deep learning algorithms were applied to get precise results. Convolution Neural Network (CNN), a deep learning algorithm was applied to enhance the performance. A comprehensive empirical analysis has been carried out by applying variations in the number of hidden layers, epochs and applying the latest models. By performing experiments and balancing the data the false negative rate is also minimized. The main motive is fraud identification.

Key Words- ***Fraud detection, Machine Learning, Deep Learning, Decision Tree, Logistic Regression, Convolution Neural Network.***

I. INTRODUCTION

People can use credit cards for online transactions as it provides an efficient and easy to use facility. With increase in usage of credit cards, the capacity of credit card misuse has also enhanced. Credit card frauds cause significant financial losses for both credit card holders and financial companies. Credit card fraud is a type of identity theft in which someone other than the owner makes an unlawful transaction using a credit card or account details. A credit card that has been stolen, lost or counterfeited might result in fraud.

Manuscript revised May 11, 2023; accepted May 12, 2023. Date of publication May 14, 2023.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

Card-not-present fraud or the use of your credit card number in e-commerce transactions has also become increasingly common as a result of the increase in online shopping. Skimming, Hacking, Phishing, etc. are some of the techniques used in credit card frauds. Generally, credit card frauds are divided into two types card present fraud and card not present fraud. Stolen card or lost card can be reported immediately but it's difficult to identify whether a card is hacked or not.

Identifying frauds is most important to take preventive measures on it. There are several techniques to identify fraud transactions. In this paper machine learning and deep learning techniques are used to identify fraud and legitimate transactions. Using some of the algorithms of machine learning and one algorithm of deep learning we have identified fraud transactions. Cread card frauds leads to huge financial losses, so it's necessary to take proper measures on it. If fraud is done with single customer of bank, then financial loss is less as compared to whole data of bank is hacked. With respect to this bank also requires safety measures to avoid frauds.

II. LITERATURE SURVEY

2.1 Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms

Credit card fraud happens when card is lost, stolen or hacked. There are many machines learning-based approaches like Decision Tree, Random Forest, Support Vector Machine and Logistic Regression to detect fraud transactions. To get more accurate results state of art deep learning algorithms are applied. Initially machine learning algorithms were applied to the dataset which has increased the accuracy of detecting fraud and later deep learning algorithms were applied to get precise results. Convolution Neural Network (CNN), a deep learning algorithm was applied to enhance the performance. To analyse performance of CNN model different layers of CNN are applied, this has increased the accuracy of model. To increase accuracy more layers of CNN are added. According to the results we can increase or decrease the number of hidden layers. Confusion matrix evaluates the performance measures by considering values of accuracy, precision and recall. By performing experiments and balancing the data the false negative rate is also minimized. Features are selected from dataset in order to get important feature. The main motive is fraud identification. Robust classifier can be used to tackle changing nature of fraud.

2.2 An efficient real time model for credit card fraud detection based on deep learning

In the recent years the number of bank transactions via credit cards raised drastically and with it the number of frauds and card theft. Therefore, many solutions and algorithms using machine learning have been proposed in literature to deal with this issue. In the last decades Machine Learning achieved notable results in various areas of data processing and classification, which made the creation of real-time interactive and intelligent systems possible. The accuracy and precision of those systems depends not only on the correctness of the data, logically and chronologically, but also on the time the feed-backs are produced. In banking and financial sectors, machine learning is used actively today for different applications, notably in portfolio management, trading, risk analysis, prevention and fraud detection. Moreover, one of the primary uses of machine learning in the banking industry is the protection against fraud. With the help of ML algorithms, detecting suspicious activities became and easier task. Based on the transaction's history, machine learning showed promising new methods to analyse the behaviour of users and detect if there is a fraud or not. Over the past few years, many solutions have been proposed to cope with the problem of credit card fraud. In this paper author for a real-life data set of Credit Card transactions, using Deep Learning. Deep Learning is presented as a very promising solution to deal with fraud in financial transactions, making to best use of banks big-data. Deep learning is a generic term that refers to machine learning using deep multilayer artificial neural network (ANN). Deep neural networks attracted much attention in the field of machine learning. It's currently providing the best precision and accuracy to many problems; providing promising results in many fields, notably in binary classification.

2.3 Adversarial attacks for tabular data: Application to fraud detection and imbalanced data

Fraud detection plays a crucial role in financial transactional systems such as banks, insurances or online purchases. The ability to detect early whether a transaction is fraudulent has a very high value and big investments have been made to make these systems more effective. It is however important to note that fraudsters are constantly developing new ways of fooling these systems, a phenomenon known as concept drift. A fraud detection system therefore typically has high maintenance requirements. Machine Learning (ML) is a classical approach for fraud detection systems. The ability to retrain the models with new data helps in this need for adaptation to new fraud patterns. However, given the possibility of errors in the models decisions, which could lead to overlooking frauds or blocking licit transactions and sales opportunities, fraud detection systems often do not rely solely on the models but also contain one or more layers involving some form of human intervention. Risky transactions can be manually inspected and a decision is made whether those transactions should go through or should be blocked. Fraudsters may use a wide range of techniques to bypass fraud detection systems. Among these techniques, adversarial attacks are novel and innovative approaches that might be used as a next level of smart financial frauds. The goal of adversarial attacks is to generate adversarial examples, i.e., inputs that are almost indistinguishable from natural data and yet classified incorrectly by the machine learning model. Algorithms to build adversarial examples have recently been shown to be very effective in fooling Machine Learning models, in particular Deep Neural Networks (DNNs) in Image Recognition. In this paper author illustrated the process followed to adapt state-of-the-art adversarial algorithms, that are commonly used in the image classification domain, to imbalanced tabular data.

2.4 Credit card fraud detection from imbalanced dataset using machine learning algorithm

In virtual era, credit card is most common mode of online transaction also it is used many times for offline transactions. Rapid growth of credit card transactions has given rise to fraudulent activities. It provides cashless shopping and less hustle for the users. Along with its use, risk of credit card frauds has also increased in recent decades. Crimes related to credit cards are gaining high momentum leading to the loss of huge amount in finance industry. Many transactions may resemble fraudulent transactions when actually they are genuine transaction. Fraud leads to huge financial losses. The goal is to define that are appropriate and can be adapted by credit card companies for identifying fraud transactions accurately in less time and cost. Therefore, using efficient algorithm to detect frauds is need of all banks to minimize financial losses. The aim of model is to improve fraud detection rather than misclassifying the genuine transaction as a fraud.

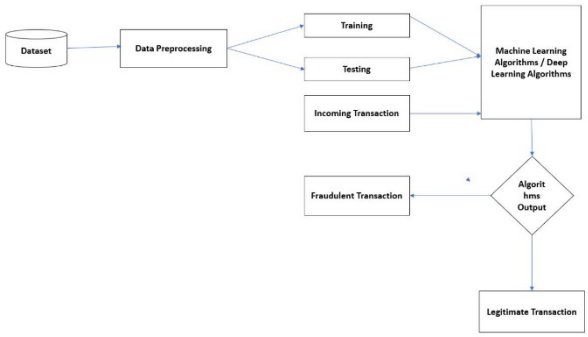
2.5 Credit card fraud detection using machine learning algorithms

Nowadays, credit cards are used for online as well as regular purchase. It allows users to purchase goods and services within credit card limit and withdraw cash in advance. Due to increase in use of credit card fraud related to credit cards are also increased. So, our aim is to detect these frauds. Various supervised and semi-supervised machine learning algorithms such as Random Forest and Logistic Regression are used to detect these frauds. Random Forest is used to solve classification as well as regression problems. These algorithms can work with large datasets. It improves the efficiency of system, and avoids the overfitting problem. Logistic Regression is mostly used for classification. By using dependent variables LR predicts the categorical variables. The implementation of an efficient fraud detection system is essential for all credit card issuing companies and their clients to minimize their losses.

III. SYSTEM ARCHITECTURE

3.1 Data Set

Dataset used is European Card Dataset which has attributes from V1 to V28 which are PCA converted values along with v1 to v28 it also has time, amount and class attributes, the last attribute class will classify the transaction as fraudulent or legitimate transaction. The dataset has 2.8 lakhs rows (i.e. Transactions). The dataset has noise like null values, duplicate values, etc.



Therefore, it is required to preprocess the dataset. Due to the issue of concealment, we cannot offer the structures of the original dataset and the data more background information.

IV. RESULTS AND DISCUSSION

By using various machine learning algorithms such as logistic regression, decision tree and random forest we are trying to get accurate results also, to increase the accuracy obtained through machine learning algorithms one deep learning algorithm i.e., CNN is used. Accuracy of Decision tree algorithm is 99.91%, accuracy of Random Forest is 99.96% and that of CNN is 99.93%. In real world, fraudulent and non-fraudulent classes are not balanced due to nature of problem. For instance, if here are huge transactions performed per day then very few could be fraudulent. Numerous sampling techniques are used to increase the performance of existing algorithms but they significantly decrease on unseen data.

V. CONCLUSION

In the proposed system, we have implemented machine learning and deep learning algorithms to detect credit card frauds. At the end our model can distinguish fraud transaction and non-fraud transactions. We can use this model with various bank dataset as data of any bank can be given as input and our model can identify fraud transactions. This is very useful for banks to save themselves and their customers from financial losses. It's better for banks to opt this model to increase security. It can help to identify legitimate transactions and fraud transactions so that further measures can be taken against fraud transaction. As there are very less fraud transactions as compared to legitimate transactions our model gives good efficiency. Machine learning algorithms give good results but to increase its efficiency a deep learning algorithm is used. We get accurate results through our model.

REFERENCES

[1]. Fawaz Khaled Alarfaj, Iqra Malik, Hikmat Ullah Khan,"Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," April 18, 2022..

[2]. Y. Abakarim, M. Lahby, and A. Attioui, "An efficient real time model for credit card fraud detection based on deep learning," in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 1–7.

[3]. F. Cartella, O. Anunciacao, Y. Funabiki, D. Yamaguchi, T. Akishita, and O. Elshocht, "Adversarial attacks for tabular data: Application to fraud detection and imbalanced data," 2021.

[4]. S. Warghade, S. Desai, and V. Patil, "Credit card fraud detection from imbalanced dataset using machine learning algorithm," Int. J. Comput. Trends Technol., vol. 68, no. 3, pp. 22–28, Mar. 2020.

[5]. V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," Proc. Comput. Sci., vol. 165, pp. 631–641, Jan. 2019.

[6]. H. Abdi and L. J. Williams, ''Principal component analysis,'' Wiley Interdiscipl. Rev., Comput. Statist., vol. 2, no. 4, pp. 433–459, Jul. 2010.

[7]. I. Mekterović, M. Karan, D. Pintar, and L. Brkić, ''Credit card fraud detection in card-not-present transactions: Where to invest?'' Appl. Sci., vol. 11, no. 15, p. 6766, Jul. 2021.