ISSN: 2249-7196
IJMRR/ july 2023/ Volume 13/Issue 3/Article No-1/01-10
Deepa Patnaik/ International Journal of Management Research & Review

EVALUATION OF THE EFFECTIVENESS OF MACHINE LEARNING TECHNIQUES FOR DETECTING CREDIT CARD FRAUD

Deepa Patnaik[1], Pravalika J[2],Sruthi J[3],Sathvika G[4]
[1]Assistant Professor ,Department of Electronics and Communication Engineering
[2,3,4] B-Tech Students ,Department of Electronics and Communication Engineering
[1,2,3,4] Department of Electronics and Communication Engineering
swec.deepa.ece@gmail.com
jpravs002@gmail.com
jangampetasruthi217@gmail.com
sathvika360@gmail.com

ABSTRACT: Every year, fraudulent credit card transactions result in losses of billions of dollars. In order to minimise these losses, it is crucial to develop effective fraud detection algorithms. The unpredictable distribution of the data, the extremely imbalanced class distributions, and the scarcity of transactions with fraud investigators' labels make the construction of fraud detection algorithms particularly difficult.In today's financial industry, credit card fraud is a steadily rising issue. Fraudulent actions have increased quickly in frequency during the past few years, costing many businesses, organisations, and government bodies a sizable amount of money.As a result of the predicted growth in numbers, several academics in this area have concentrated on applying cutting-edge machine learning approaches to identify fraudulent behaviours early on.It's simple and easy to fall victim to credit card theft. Researchers began utilising various machine learning techniques to detect and analyse scams in online transactions as fraud rates increased. .The groups are then used to train various classifiers individually in the future. The classifier with the highest rating score can then be selected as one of the most effective ways to detect fraud. Consequently, a feedback system is used to address the issue of notion drift. In this study, we used data on European credit card theft.

Key Words : Decision Tree, Random Forest, Extreme gradient boosting Algorithm.

I. INTRODUCTION

The use of a customer's credit card information without authorization to make transactions or remove money from the cardholder's account is known as credit card fraud. When someone improperly obtains the card's printed number or the vital documents needed to activate the card, the fraudulent extortion begins. The cardholder, the agent who issued the card, and even the card's guarantor may not be aware of the fraud until the record is used to make purchases. The need for a physical card to make transactions has been replaced by the use of internet-based applications for shopping and bill payment.The fraud solutions can be divided into two categories: detection and prevention. Detection refers to the steps conducted after the event has occurred. Due to the volume of transaction data, it is impossible for humans to manually verify each transaction to determine if it is fraudulent or not, necessitating the need for automated fraud detection systems. This thesis is built on a system that uses machine learning to automatically detect fraud.

Fraud is a universal problem that has existed for as long as humanity. Additionally, the advent of new technology offers fraudsters new avenues to exploit. For example, in e-commerce, card information alone is enough to perform a fraud.In today's society, credit cards are widely used, and credit card theft has risen steadily in recent years. Fraud-related financial losses harm not only businesses and banks (such as reimbursements), but also specific customers. When a bank loses money, consumers eventually pay for it by paying increased interest rates, membership dues, etc.

Fraud may also damage a business's brand and image, resulting in non-financial damages that, while hard to measure in the short term, may become apparent over time. A cardholder might select a rival if, for instance, he becomes a victim of fraud with one particular company. A set of credit card transactions is used as the basis for fraud detection, which is the process of determining whether a new authorised transaction falls under the category of fraud or not.

A Fraud Detection System (FDS) should not only effectively identify fraud situations, but also be cost-effective in that the expense incurred for transaction screening shouldn't exceed the loss brought on by frauds. Bhatla demonstrates how fraud losses, which account for 1% of overall transaction value, may be reduced by screening just 2% of transactions.The fraud losses might be dramatically reduced to 0.06%, but the expenses would skyrocket if 30% of the transactions were reviewed. Use expert criteria and statistically based models (like machine learning) to establish a first distinction between actual and probable fraud and instruct the investigators to focus only on the cases with a high likelihood of success in order to reduce detection expenses. It's not as simple as it seems to build a fraud detection system. The practitioner must decide which learning strategy to employ (for example, supervised learning or unsupervised learning), which algorithms to employ (for example, logistic regression, decision trees, etc.), which features to employ, and most importantly, how to handle the class imbalance problem (fraudulent cases are incredibly rare in comparison to legitimate cases).

In a real-world scenario, a fraud detection model predicts whether a class is legitimate or fraudulent and sends the investigators an alert for the transaction that seems the most suspect. Then, investigators conduct additional inquiries while offering comments to the fraud detection system to enhance its functionality.The task of determining the class label of a given data item is known as a classification issue in machine learning. One categorization issue is the detection of fraud, for instance. The objective here is to determine whether a specific transaction is fraudulent or real. The study of credit card fraud detection has been extensively researched. Previously, fraud transactions were manually detected. However, the issue of imbalanced data—a very highly skewed set of data—afflicts the entire problem of credit card fraud detection. Before we train any machine learning model, such as Random Forest or another, for this issue, we must thoroughly process the data.

Credit card fraud is the unauthorised and unwelcome use of a credit card account by a person other than the account owner. Every year, there are more fraudulent credit card data transactions. The major goal is to create innovative methods to identify and stop such credit card data transaction fraud.

II. LITERATURE SURVEY

For the development of this project, various research have been conducted by many researchers.

[1]The amount of money spent on e-commerce globally has been steadily rising over the years, reflecting a clear change in customer interest away from brick-and-mortar stores and towards online retailers. Online markets have recently emerged as one of the major forces driving this expansion. A variety of control and prevention measures are covered along with a detailed analysis of fraudulent e-commerce buyers and their transactions. Merchant fraud refers to another type of fraud that occurs in marketplaces on the seller side. A straightforward example of this kind of fraud is the sale of cheap goods or services that are never delivered.

Online sales have significantly increased in recent years.[2] Online credit card transactions account for a significant portion of this. Applications for credit card fraud detection are therefore highly necessary in banks and other financial institutions. To receive items without paying for them or to withdraw money from an account without authorization are two examples of credit card fraud. The thirst for money led to an increase in credit card fraud incidents. For the cardholder, this causes a significant financial loss.

Machine learning should be utilised to produce fraud protection in e-commerce. This study analyses the best machine learning method, which should be Decision Tree, Naive Bayes, Random Forest, and Neural Network. The accuracy of the neural network that was evaluated using the confusion matrix was 96 percent, followed by Naive Bayes (95 percent), Random Forest (95 percent), and Decision Tree (91 percent). The average F1-Score can rise from 67.9 to 94.5 percent, and the average G-Mean from 73.5 to 84.6 percent, using the Synthetic Minority Over-sampling Technique (SMOTE).[3]

The most common issue in the modern world right now is detecting credit card theft. This is brought on by an increase in e-commerce platforms and online transactions.[4] In general, credit card fraud occurs when the card is used for any unauthorised activity or even when the fraudster utilises the card's details for personal gain.

[5] The popularity of digitalization is growing due to how simple, convenient, and smooth e-commerce is to use. It developed become a widespread and simple method of payment. People prefer to pay and shop online due to the ease it offers in terms of time and transparency. Due to the extensive usage of e-commerce, credit card fraud has also significantly increased.

[6] There are no predictable patterns in frauds. They constantly alter their behaviour, necessitating the use of unsupervised learning. The ability to commit fraud through internet transactions is now known to fraudsters. Fraud tendencies shift quickly, and fraudsters presumably act in accordance with consumer norms.

The use of credit cards for online purchases has significantly expanded as a result of the E-Commerce industry's explosive growth.[7] Due to how challenging it has gotten to identify fraud in the credit card system, credit card fraud has recently grown to be a significant problem for banks. To get around this problem, machine learning is crucial in spotting credit card fraud in transactions.

Credit cards have become the most popular method of payment for both online and off-line purchases as a result of advancements in communication and e-commerce. Thus, it is largely anticipated that the system's security will stop transactions that are fraudulent. Every year, more fraudulent credit card data transactions occur. Researchers are also experimenting with cutting-edge methods in this regard to identify and stop such scams. But there will always be a need for certain methods to accurately and successfully uncover these frauds.[8]

It is essential that credit card companies be able to spot fraudulent credit card transactions in order to prevent customers from being charged for goods they did not order.[9] Data science can be used to solve these issues, and the value of both of these techniques—along with machine learning—cannot be overstated. With the help of machine learning and credit card fraud detection, this research aims to demonstrate data set modelling.

It's simple and easy to fall victim to credit card theft.[10] Online payment options have risen due to e-commerce and numerous other websites, which raises the possibility of online fraud. Researchers began utilising various machine learning techniques to detect and assess scams in online transactions as fraud rates increased.The primary objective of the study is to create and implement a unique fraud detection algorithm for streaming transaction data with the goal of analysing historical customer transaction information and extracting behavioural patterns. wherein cardholders are grouped according to the value of their transactions.

[11] Fraud with credit cards is a serious criminal offence. Billions of dollars are spent each year on it by people and financial institutions. The Federal Trade Commission (FTC), a consumer protection organisation, claims that over the previous two years, the number of theft reports has risen. Because of this, detecting and stopping fraudulent actions is crucial for financial institutions. With respectable accuracy, machine learning algorithms offer a preventative mechanism for credit card fraud. In order to detect fraudulent transactions, machine learning methods such Logistic Regression, Naive Bayes, Random Forest, K-Nearest Neighbour, Gradient Boosting, Support Vector Machine, and Neural Network algorithms are employed in this article. For the purpose of determining the best option, various algorithms are compared.

III. Evaluation of the effectiveness of machine learning techniques for detecting credit card fraud

With this model, we significantly get around the problems. We can identify fraud and figure out how to reduce it in order to generate an optimised result that will make a better forecast using the random forest classifier, XG boost classifier and Decision tree classifier algorithms. We are able to spot fraud based on client behaviour. The local outlier factor is applied here. We have two classification class which is named as class 0 and class 1. If there is legal transaction then the result will stored in class 0 and if there is a fraudulent transaction then the result will stored in class 1.

Figure 1 : Block diagram of credit card fraud detection

Systems for detecting fraud are harder than they seem. The practitioner must decide in practise which classification method to useand how to address the issue of class imbalance (suspicious cases are disproportionately few in comparison to valid ones).The wealth gap is not the only factor making fraud detection difficult also underprivileged. Many machine learning algorithms struggle with classification because there is a mismatch between the true and fraudulent classes due to a lack of transaction data.

In a real-world fraud detection situation, a model that makes use of artificial intelligence to identify suspicious transactions and notify the proper authorities when one of those transactions is found to be either authentic or fraudulent would be used to detect fraud. Investigators that carry out investigations and feed the system with their results help to improve the fraud detection system. As a result, only a small number of transactions are timely certified by investigators using this method. The prediction model is frequently less accurate when it receives limited feedbacks.

It is very challenging to locate the real financial statistics because financial companies rarely provide consumer data due to privacy concerns. Overcoming this difficulty is a key issue with fraud detection systems.

ALGORITHMS:

1. Random Forest

The supervised learning method includes the well-known machine learning algorithm Random Forest. It can be applied to ML Classification and Regression issues. Its foundation is the idea of ensemble learning, which is the process of mixing various classifiers to solve a challenging problem and enhance the performance of the model. As its name implies, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead than depending on a single decision tree, the random forest considers the prediction from each tree and predicts the outcome based on the majority votes of predictions.Since there are more trees in the forest, the accuracy is higher and the overfitting issue is avoided. Some decision trees may predict the correct output, while others may not, because the random forest combines numerous trees to forecast the class of the dataset. But when all the trees are combined, they forecast the right result.

2. Decision Tree

The actual world is full of analogies for trees, and it turns out that these analogies have inspired a large portion of machine learning, including both classification and regression. A decision tree is a visual and clear representation of decisions and decision-making in decision analysis. It employs a decision-tree structure, as suggested by the name. Although an often employed method in data mining for determining a plan of attack to accomplish a specific objective.

Placed reversed with the root at the top, is a decision tree. In the left image, a condition or internal node, based on which the tree divides into branches or edges, is represented by the bold text in black. The decision or leaf, in this example whether the passenger lived or died, is the end of the branch that doesn't divide any more and is shown as red or green text, depending on the outcome.

3. **XG Boost**

XGBoost stands for “Extreme Gradient Boosting”. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements Machine Learning algorithms under the Gradient Boosting framework. It provides a parallel tree boosting to solve many data science problems in a fast and accurate way.

Boosting is an ensemble learning strategy that creates a strong classifier out of a number of successively weak classifiers. In order to address the bias-variance trade-off, boosting techniques are essential. Boosting algorithms regulate all the components of a model—bias and variance—and are thought to be more effective than bagging algorithms, which solely account for excessive variance.

IV. RESULTS AND ANALYSIS

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

MODEL PERFORMANCE DATASET

Models	Accuracy	Precision	Recall
XG Boost	99.824445	95.4556	75.2336

Table 1: shows the overall classification results in terms of accuracy, precision, recall

Random Forest	99.937580	92.179487	82.55556
Decision Tree	99.925877	71.260870	71.000000

Displaying Output

User sees the output of all the algorithms. A pycharm window is used to display the output of all the fraud process by using the XG boost, Random forest and Decision tree algorithms.

Figure 2: showing home page

Figure 3: Uploading of dataset

Figure 4: Preprocessing the data

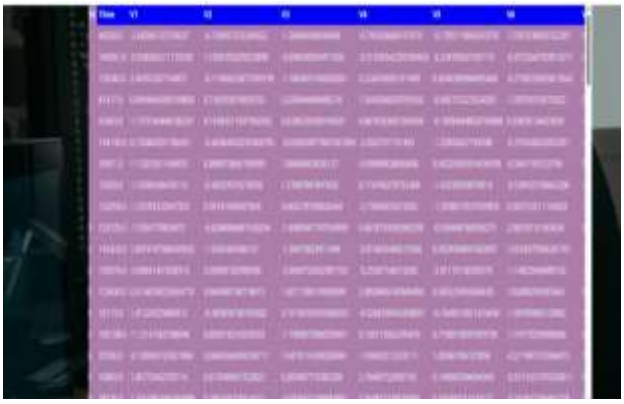

Figure 5: Uploaded dataset values

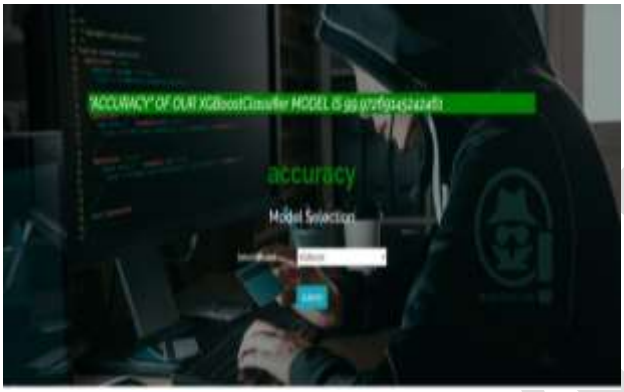

Figure 6: Model selection of XG boost classifier

Figure 7: Model Selection of Random Forest Classifier

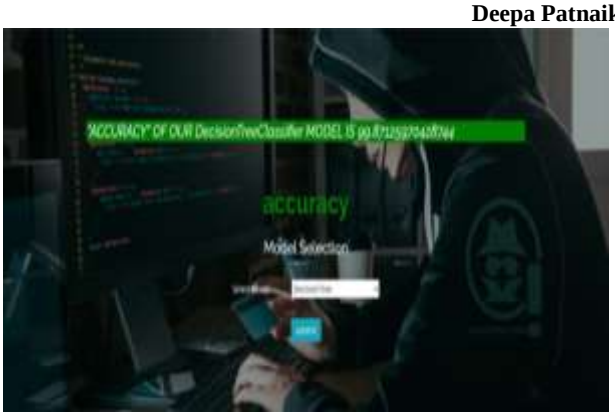

Figure 8: Model selection of Decision tree classifier

Figure 9:Predicting the data as Normal Card

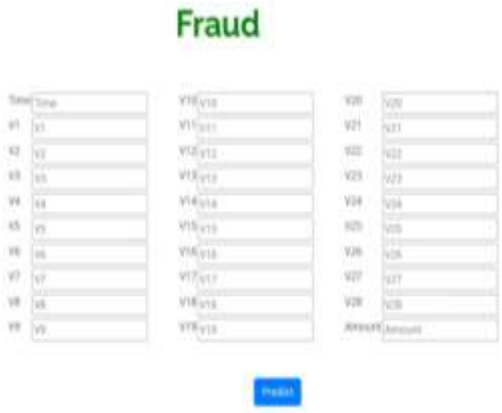

Figure 10: Predicting the data as fraud card

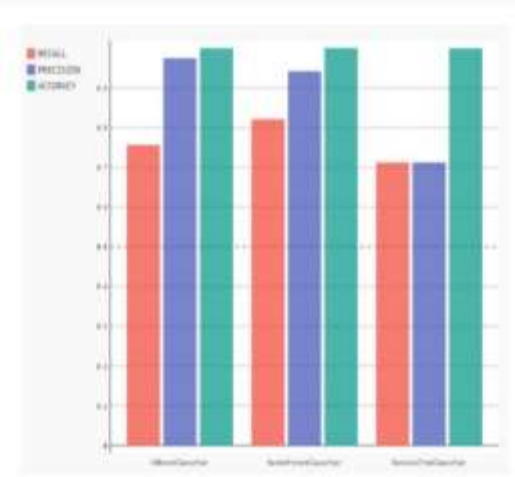

Figure 11: Accuracy, precision,recall values of the algorithms

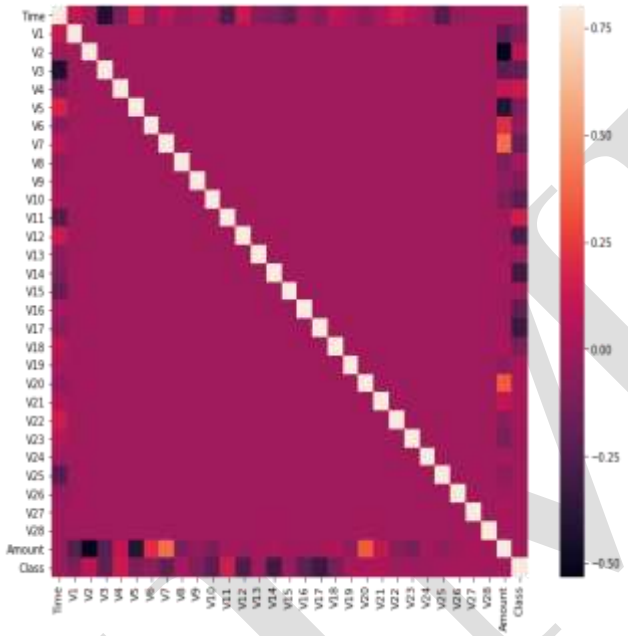

Figure 12: Correlation of matrix of credit card fraud detection

First, we use a credit card data collection that contains information about credit cards in our designing. But in this case, we only allocate a small amount of transaction effort. The cycle of data cleaning comes after, where each copy of the database is examined and any empty values are eliminated by the database in use. The next step is data partitioning, where the credit card database will be split into two portions for training and testing.

V. CONCLUSION

Fraudulent use of credit cards is an example of delinquency. The research has dug down the almost universally known scamming strategies along with their detection. Strategies were examined, as well as recent finds in the meadow. The essay also addressed how artificial intelligence might be used to enhance fraud prevention alongside algorithms, protocols, explanations of its application, and testing results. When the 10th of this dataset is taken into account, the algorithm's accuracy only remains at 29%. However, this algorithm's exactness increases to 34% when the entire dataset is considered.Because of the huge discrepancy between the number of substantial and certifiable dealings, that high level of faithfulness may be counted upon. As can be seen, this entire data set

only contains the results of two days' worth of transactions. If the protude had been used in mercantile measurement, more data would have been available. Depending on artificial intelligence approaches, this code will only increase their rate of productivity because the majority of data is exposed to it.

REFERENCES

[1] L. Bhavya, V. Sasidhar Reddy, U. Anjali Mohan, S. Karishma, 2020, Credit Card Fraud Detection using Classification, Unsupervised, Neural Networks Models, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020)

[2] Renjith, Shini. Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine ApproachIJETT, V57(1),48-53 March 2018. ISSN:2231-5381.

[3] Saputra, Adi &Suharjito, Suharjito. Fraud Detection using Machine Learning in e-Commerce(2019).

[4] RuttalaSailusha, V. Gnaneswar, R. Ramesh, and G. Ramakoteswara Rao (2020). Credit Card Fraud Detection using Machine Learning.Published in: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS).

[5] Rimpal R. Popat and Jayesh Chaudhary (2020). A Survey on credit card fraud detection using machine learning.Published in: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI).

[6] A. K. Rai and R. K. Dwivedi, "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC).Published in: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC).

[7] S P Maniraj; Aditya Saini; Swarna Deep Sarkar Shadab Ahmed contributed "Credit card fraud detection using machine learning and data science" published by IJERT in September 2019.Volume & Issue **:** Volume 08, Issue 09 (September 2019)

[8] Masoumeh Zareapoor; Pourya Shamsolmoali wrote "Application of credit card fraud detection: based on Bagging Ensemble Classifier" Published by Elsevier in 2015

[9] Mohammed Azhan; Shazli Meraj contributed there work for "Credit card Fraud detection using Machine learning and deep learning techniques" IEEE Explore 2020. Published in: 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS).

[10] Pratyush Sharma; Souradeep Banarjee; Devyanshi Tiwari; and Jagadish Chandra Patni developed "Machine Learning Model for credit card fraud Detection-A Comparative Analysis" February 2021.Received June 10, 2020; accepted February 17, 2021

[11] Apoorva; Ashwini S H; Bindushree S; Sinchana N; Prof. K N Prashanth Kumar published "Review of credit card fraud detection using machine learning" June 2020.International Journal For Technological Research In Engineering Volume 7, Issue 10, June-2020.

[12] Sanobar Khan; Sanovar; Suneel Kumar; Mr. Hitesh Kumar worked on "Credit card fraud detection using machine learning" published on June 2021.International Journal of Scientific and Research Publications, Volume 11, Issue 6, June 2021 60 ISSN 2250-3153.