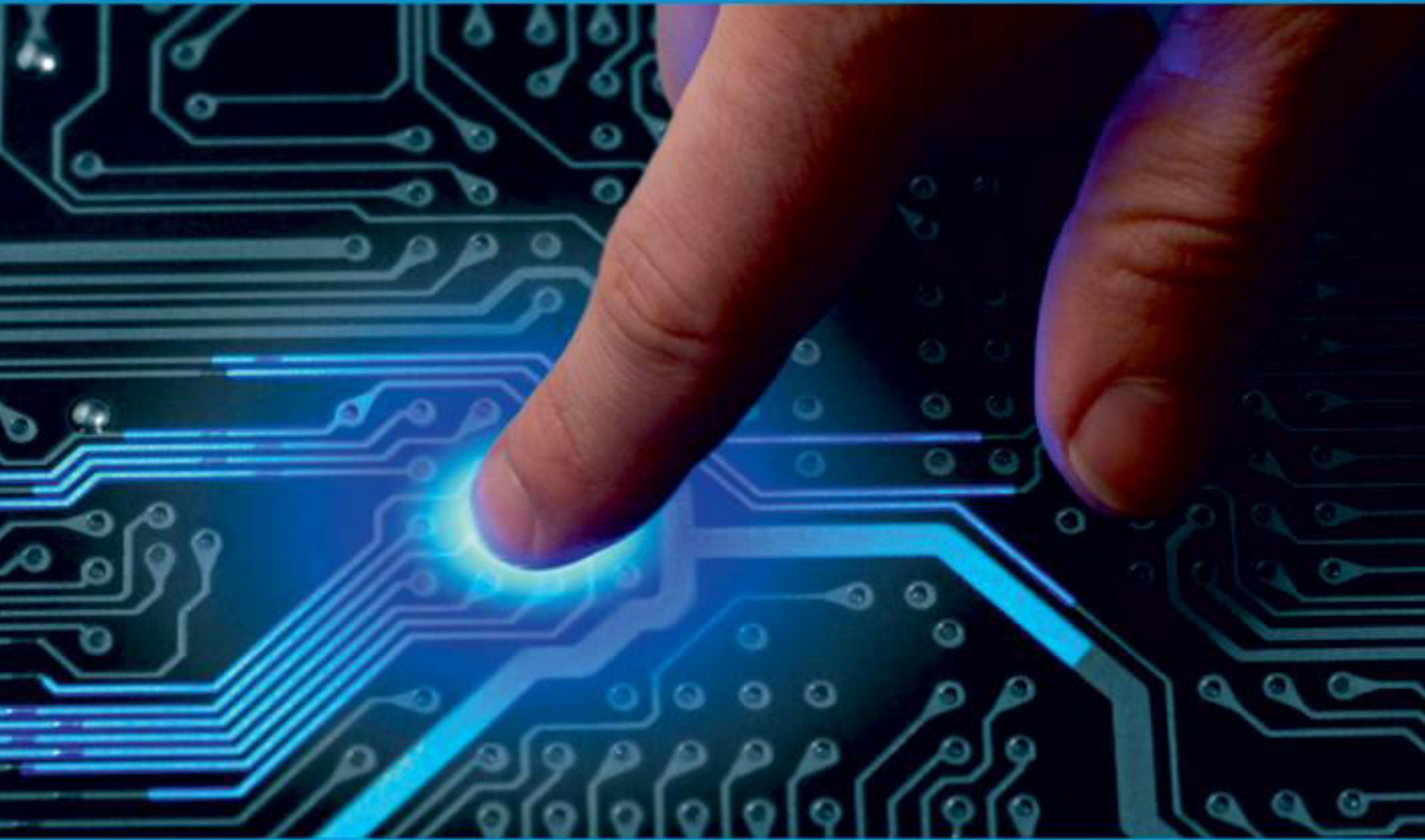
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

Credit Card Fraud detection using Machine Learning: A Survey

Damle Varad S[1], Dudhade Nikita S[2], Bhavar Akshada U[3], Ghenand Akanksha M[4], Prof. Masira Kulkarni[5]

Student, Dept. of Computer Science, JSPM's Imperial College of Engineering & Research, Pune, Maharashtra, India

Assistant Professor, Dept. of Computer Science, JSPM's Imperial College of Engineering & Research, Pune, Maharashtra, India

ABSTRACT: when the term fraud is used nowadays; credit card fraud is the first thing that springs to mind. Credit card fraud has increased dramatically in recent years as a result of the large increase in credit card transactions. Fraud detection entails keeping track of users'/customers' spending habits in order to identify, detect, or prevent fraudulent activity. As credit cards become the most popular method of payment for both online and offline purchases, fraud related to them is on the rise. Fraud detection is concerned with detecting not only fraudulent events, but also such activity as fast as feasible. Credit cards are widely used in today's society. Fraud is a multibillion-dollar industry that is growing every year. Fraud costs our economy a lot of money all over the world. For detecting credit card fraud, modern approaches such as data mining, machine learning, sequence alignment, fuzzy logic, genetic programming, and artificial intelligence have been introduced. This study demonstrates how data mining approaches can be successfully integrated to provide high fraud coverage and a low or high false alarm rate.

KEYWORDS: Credit card fraud, spending pattern; credit card, fraud detection techniques, on-line

I. INTRODUCTION

For credit card transactions, 'frauds/scams' refer to unapproved and unwanted use of a record by someone other than the account's owner. Essential foresight estimates might be made to prevent this mistreatment, and the conduct of such fraudulent activities could be concentrated to limit it and ensure that similar incidents do not occur in the future. Credit Cards Scam is defined as a circumstance in which an individual uses another person's credit cards for personal purposes while the owner and card issuer are unaware. Examined using pre-programmed devices to determine which transactions should be approved.

The method that the card is used by someone is unknown to experts. Scam detection is studying the activities of large groups of clients in order to detect, see, or avoid suspicious behaviour such as fraud, interruption, and defaulting. This is an important issue that necessitates the use of networks, such as machine/soft learning and data science, to automate the solution to this problem. This issue is particularly challenging from the standpoint of studying, as it is described in a variety of ways, such as class imbalances. The number of valid transactions considerably outnumbers the number of fraudulent transactions. In the same way, transaction designs frequently modify their factual properties over time. In any case, these are far from the last issues with using a legitimate fraud detection framework. The huge torrent of installment demands is promptly reviewed in actual models by programmed equipment that determines which transaction to approve. A fraud/scam detecting or identification method is continual prepared to prevent crime person for adopting to their fraud planning. These scams are categorized as: Credit Cards scams: Online and Offline Cards Theft Account Bankrupt gadgets Intrusion solicitation Fraud Counterfeit Card Telecommunication scam. one of recently applied techniques to detect or identify these scams are.

- Artificial Neural Network
- Fuzzy Logic Genetic Algorithms
- Logistic Regression algorithm
- Decision tree algorithm

- K-Nearest Neighbor

Types and techniques of credit card frauds

a. Traditional techniques

Financial fraud

This happens when someone tries to get more credit than they are entitled to. Under his or her own name, a person will apply for a credit card. In this case, the person will provide false information about his or her financial situation. A person's income is frequently exaggerated, while his or her outgoings are undervalued. Banks try to prevent this type of fraud by demanding the submission of papers to back up individual financial claims. For example, a card issuer may request three months' worth of up-to-date account statements or mortgage bills from an individual. Banks have also been known to call the individual's workplace to confirm their employment. Fraudsters, on the other hand, have been known to circumvent all security measures. Fraudsters have and will continue to falsify documents and even phone numbers. Credit checks are another security check that card issuers perform to protect themselves. Credit checks disclose a person's financial situation as well as his or her current residence. It is already clear that card issuers are fighting a losing battle against fraudsters.

b. Modern Techniques

Triangulation

Triangulation is a very young phenomena as well. When a merchant offers a product at an extremely low price through a website, this is known as triangulation. When a consumer requests to purchase a product, the merchant instructs the customer to pay via e-mail after the item has been delivered. The merchant purchases a product from a website using a fraud card number and provides it to the buyer, who then sends the merchant his or her credit card information by e-mail. The merchant continues to operate using credit card details given by customers to purchasethings, giving the impression of being a legitimate business for a brief period before closing the site and starting a new one.

Credit card generators

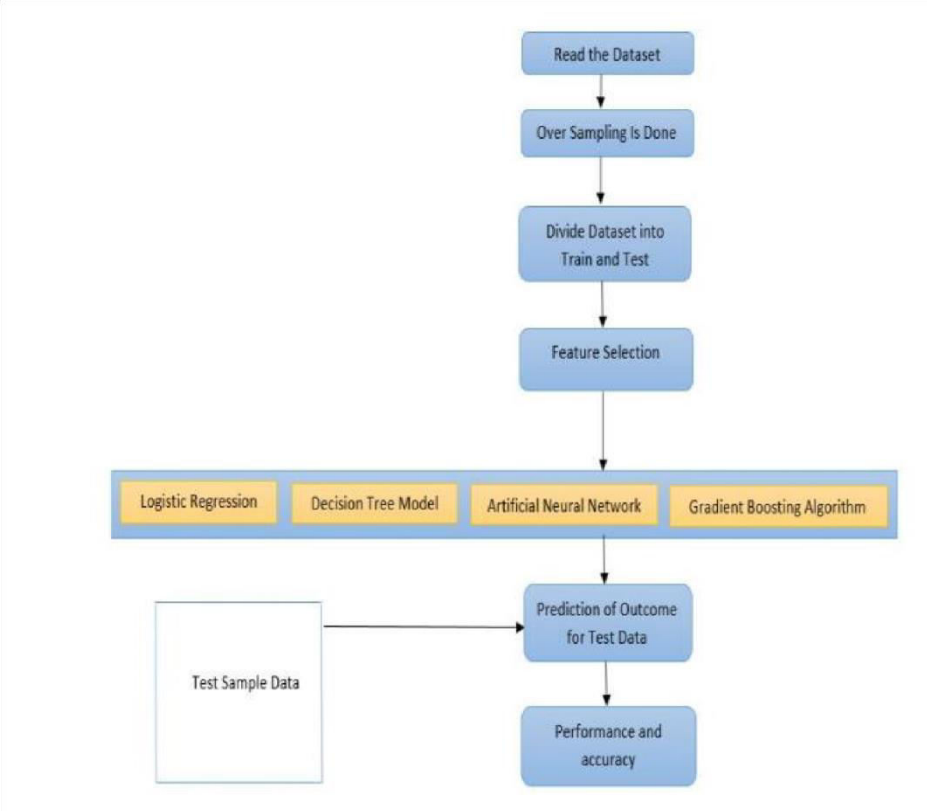
Credit card generators are computer emulation software that generates acceptable credit card numbers and expiration dates for more advanced thieves.

These generators are quite good at creating genuine credit card numbers and can be downloaded for free from the internet. Making them available to a large number of people who run fraudulent businesses

Some of the approaches to detection of such fraud are:

- Logistic Regression
- Artificial Neural Network
- Genetic Algorithm
- Gradient Boosting Algorithm
- Support Vector Machines
- Decision tree
- Fuzzy Logic
- Hidden Markov Model

II. SYSTEM ARCHITECTURE



PROBLEM STATEMENT AND OBJECTIVE:
Credit card fraud is on the rise these days, and as a result, financial losses are skyrocketing. Every year, billions of dollars are lost due to fraud. There is a dearth of study to analyze the fraud. To detect credit card fraud, many machine learning techniques are used. The following algorithms are used: Logistic Regression, Decision Tree Model, Artificial Neural Network, Gradient Boosting Algorithm, and Hybrid methods. The project's goal is to detect credit card fraud using a machine learning algorithm that takes into account the timing and amount of the transaction.

III. LITERATURE REVIEW

This document provides an overview of the many strategies for detecting credit card fraud. Creditcards have been the most popular method of payment for online transactions in recent years, but incidences of fraud related with them are also on the rise. Despite the current circumstances of the different measures created for its detection, credit card fraud is on the rise. Fraudsters are so skilled that they come up with new ways to commit fraud every day, necessitating the ongoing deployment of fresh ideas for fraud detection systems. Most of the techniques developed in detecting various credit card fraudulent transactions are based on Artificial Intelligence, Fuzzy logic, neural network, logistic regression, Machine learning, logistic regression, decision tree, Bayesian network, meta learning, Genetic Programming, and so on.

On the public platform, there are numerous literature or research papers on credit card fraud detection. Data mining applications, fraud detection, and adversarial detection are among the approaches utilized in credit card fraud, according to a survey research done by Clifton Phua and his interns. Suman discussed supervised and unsupervised learning strategies for fraud detection in another research study.

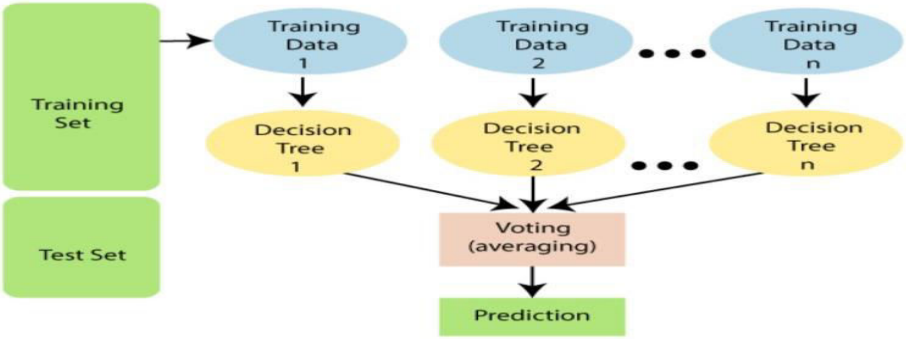
These methods were quite efficient and effective in particular sections of the domain, but they were unable to provide a long-term solution to the detection of credit card fraud. Wen-Fang Yu and Na Wang published a similar study in which they employed outlier mining, outlier detection mining, and the Distance sum algorithm to estimate fraud in an experiment using credit card data from a few commercial banks. Outlier mining is a data mining approach that is commonly used in sectors such as finance and the internet. It recognizes fields that aren't genuine. In this technique, we take fields of consumer behaviour and use them to calculate the distance between the field's observed value and a specified value.

There is a plethora of other literature that offers a completely different approach to detecting fraud. There has been some research on making the alert feedback interaction more efficient in the case of fraudulent transactions.
When a fraudulent transaction is detected, an alert is issued and transmitted to the authorized server system, which denies the transaction in turn. One of the features of Artificial Genetic Algorithm is the ability to approach the problem from a different perspective. This strategy led to more accurate fraud detection and fewer false alerts.

IV. MACHINE LEARNING ALGORITHM USED FOR CREDIT CARD FRAUD DETECTION

Random Forest Classifier:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.
As the name suggests, "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 of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.



The Working process can be explained in the below steps:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes

STEPS INVOLVED:

- ➢ Import the required packages into python environment.
- ➢ Import the data.
- ➢ Process the data and Exploratory Data Analysis.
- ➢ Selection of feature and Data splitting.
- ➢ Six types of classification models are build.
- ➢ using the evaluation metrics classification models are evaluated.

V. METHODOLOGY

The design of the flow of the system is as follows:

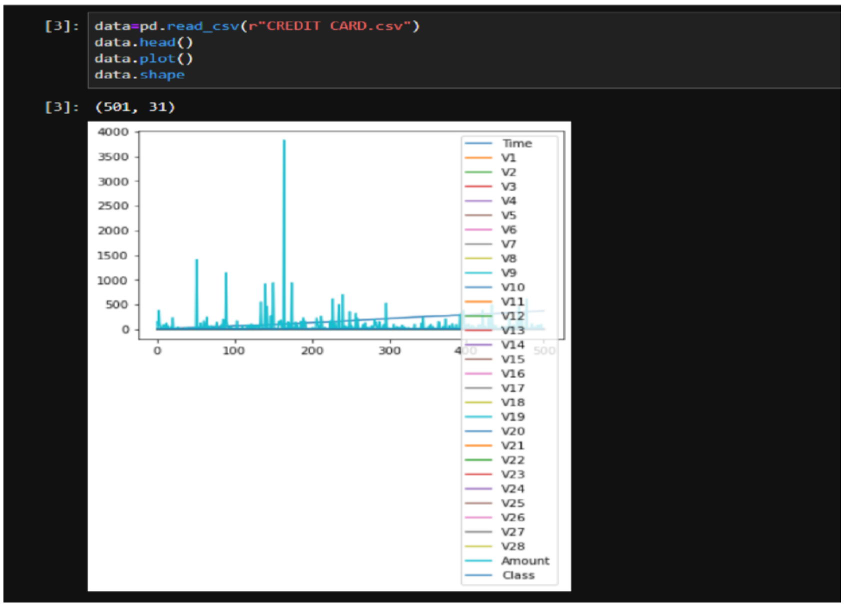
- ➢ Raw data is collected.
- ➢ Data is segregated and standardized.
- ➢ It is fed to the database; we will be using Excel as ours.
- ➢ Its sent for pre-processing where the data is fed next to train the system.
- ➢ The algorithms are used to predict results.
- ➢ Python scripts and libraries will be used to then visualize the obtained results for a better understanding of trends.

VI. PYTHON AND LIBRARY

- * **Python** is a high level and effective general use programming language. It supports multi-paradigms. Python has a large standard library which provides tools suited to perform various tasks. Python is a simple, less-clustered language with extensive features and libraries. Different programming abilities are utilized for performing the experiment in our work. In this thesis, the following python libraries were used
- * **Pandas** - It is a python package that provides expressive data structures designed to work with both relational and labelled data. It is an open source python library that allows reading and writing data between data structures
- * **Numpy** - It is an open source python package for scientific computing. Numpy also adds fast array processing capacities to python.
- * **Matplotlib** - It is an open source python package used for making plots and 2D representations. It integrates with python to give effective and interactive plots for visualization.
- * **Sklearn** - It is an open source python machine learning library designed to work alongside Numpy. It features various machine learning algorithms for classification, clustering and regression.

VII. RESULTS

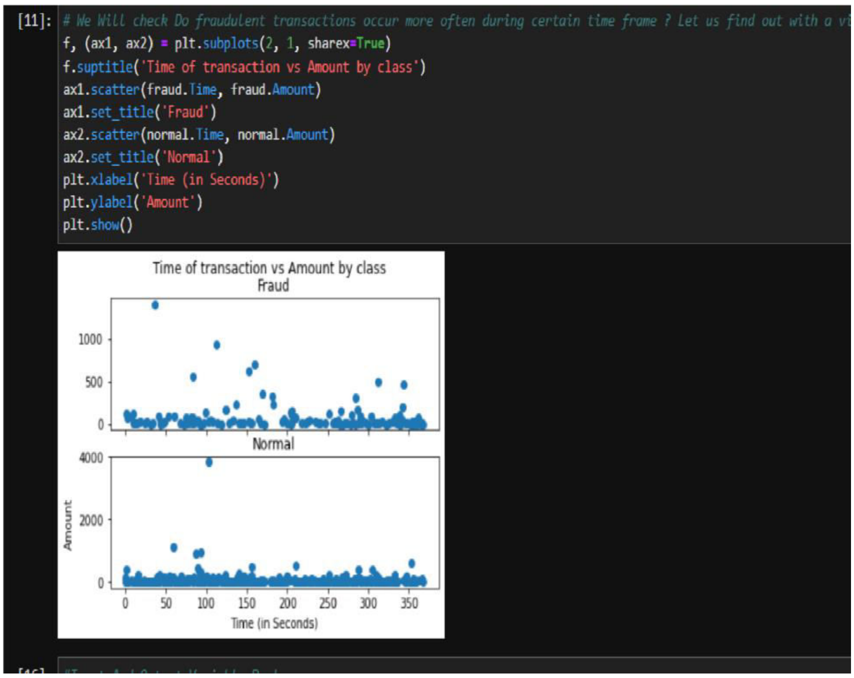
Import Data:



```
[3]: data=pd.read_csv(r"CREDIT CARD.csv")
     data.head()
     data.plot()
     data.shape
```

```
[3]: (501, 31)
```

Data Visualization:



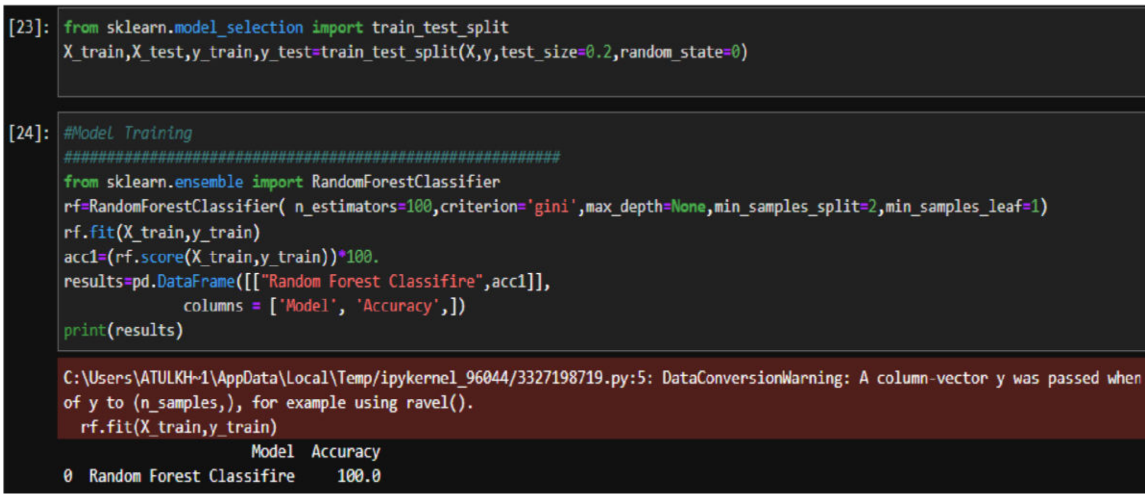
```
[11]: # We Will check Do fraudulent transactions occur more often during certain time frame ? Let us find out with a vi
      f, (ax1, ax2) = plt.subplots(2, 1, sharex=True)
      f.suptitle('Time of transaction vs Amount by class')
      ax1.scatter(fraud.Time, fraud.Amount)
      ax1.set_title('Fraud')
      ax2.scatter(normal.Time, normal.Amount)
      ax2.set_title('Normal')
      plt.xlabel('Time (in Seconds)')
      plt.ylabel('Amount')
      plt.show()
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005045 |

Data Splitting and Training with Random Forest Classifier:

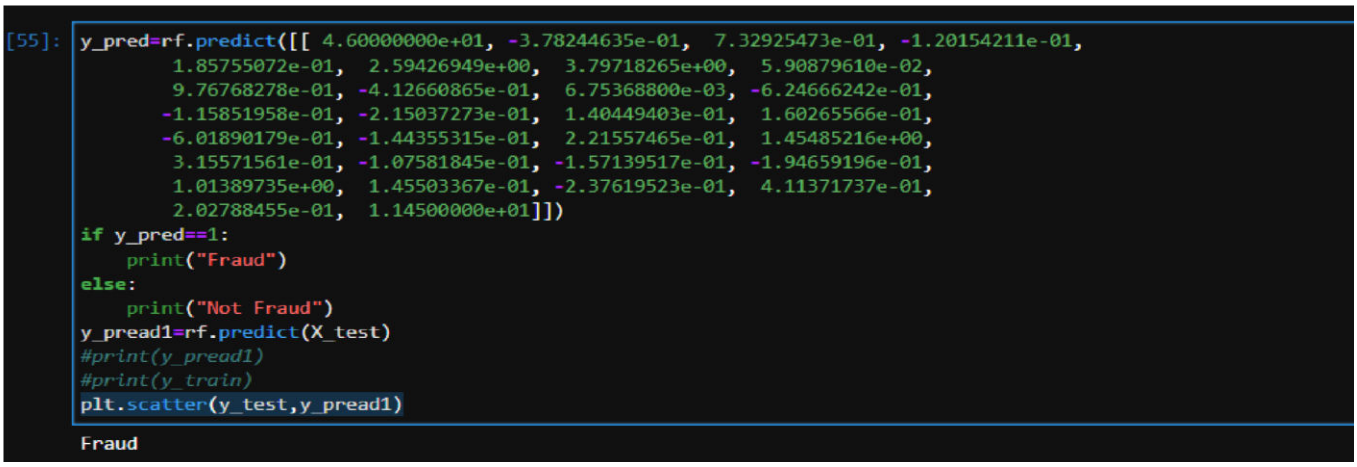


```
[23]: from sklearn.model_selection import train_test_split
      X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)

[24]: #Model Training
      ##########################################################
      from sklearn.ensemble import RandomForestClassifier
      rf=RandomForestClassifier( n_estimators=100,criterion='gini',max_depth=None,min_samples_split=2,min_samples_leaf=1)
      rf.fit(X_train,y_train)
      acc1=(rf.score(X_train,y_train))*100.
      results=pd.DataFrame([["Random Forest Classifire",acc1]],
                      columns = ['Model', 'Accuracy',])
      print(results)

      C:\Users\ATULKH~1\AppData\Local\Temp/ipykernel_96044/3327198719.py:5: DataConversionWarning: A column-vector y was passed when
      of y to (n_samples,), for example using ravel().
        rf.fit(X_train,y_train)
                            Model  Accuracy
      0  Random Forest Classifire     100.0
```

Predicted Output:



```
[55]: y_pred=rf.predict([[ 4.60000000e+01, -3.78244635e-01,  7.32925473e-01, -1.20154211e-01,
               1.85755072e-01,  2.59426949e+00,  3.79718265e+00,  5.90879610e-02,
               9.76768278e-01, -4.12660865e-01,  6.75368800e-03, -6.24666242e-01,
              -1.15851958e-01, -2.15037273e-01,  1.40449403e-01,  1.60265566e-01,
              -6.01890179e-01, -1.44355315e-01,  2.21557465e-01,  1.45485216e+00,
               3.15571561e-01, -1.07581845e-01, -1.57139517e-01, -1.94659196e-01,
               1.01389735e+00,  1.45503367e-01, -2.37619523e-01,  4.11371737e-01,
               2.02788455e-01,  1.14500000e+01]])
      if y_pred==1:
          print("Fraud")
      else:
          print("Not Fraud")
      y_pread1=rf.predict(X_test)
      #print(y_pread1)
      #print(y_train)
      plt.scatter(y_test,y_pread1)

      Fraud
```

VIII. CONCLUSION

To detect fraud in the credit card system, we applied machine learning techniques such as Logistic regression, Decision Tree, Gradient Boosting Algorithm, Artificial Neural Network, and Random Forest. Based on the differences in accuracy between the Logistic Regression Model and the Gradient Boosting Model, we conclude that the Gradient Boosting Model is the best model to apply for the given data. The Gradient Boosting Model approach is best for fitting this type of data. As a result of this research, we were able to learn how to use machine learning to construct our credit card fraud detection model. This model was built using a variety of Machine Learning algorithms. We discovered how data is frequently analyzed and visualized in order to detect fraudulent transactions in various types of data**.**

| DOI: 10.15680/IJIRCCE.2022.1005045 |

REFERENCES

[1].Amlan Kundu, Suvasini Panigrahi,Shamik Sural and Arun K. Majumdar, “BLAST-SSAHA Hybridization for Credit Card Fraud Detection,” IEEE Transactions On Dependable And Secure Computing, vol.6,Issue no. 4, pp.309-315, October-December 2009S.

[2].A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection,"June 2007.

[3].Dahl, J.: Card Fraud. In: Credit Union Magazine (2006).

[4].Ghosh and D.L. Reilly, “Credit Card Fraud Detection with a Neural-Network,” Proc. 27th Hawaii International Conference on System Sciences: Information Systems: Decision Support and KnowledgeBased Systems, vol. 3, pp. 621-630, 1994.

[5].Kaiyong Deng, Ru Zhang, Hong Guo,Kaiyong Deng,R Zhang, Dongfang Zhang,WenFeng Jiang,Xinxin Niu“Analysis and Study on Detection of Credit Fraud in E- commerce 2011.

[6].Krishna Modi; Reshma Dayma; 2017 International Conference on Intelligent Computing and Control (I2C2); 23-24 June 2017

[7].S P Maniraj;Aditya Saini;Shadab Ahmed ; International

[8].Journal of Engineering and Technical Research 08(09); September 2019;vol 08;page no. 110-115.

[9].S. Abinayaa, H. Sangeetha, R. A. Karthikeyan, K. Saran Sriram, D. Piyush; International Journal of Engineering and Advanced Technology (IJEAT) ;4, April, 2020;vol 09; page no. 1199-1201.

[10].K.Ratna Sree Valli , P.Jyothi , G.Varun Sai , R.Rohith Sai Subash; Quest Journals Journal of Research in Humanities and Social Science; 2 June 2019;Volume 8; page no: 04-11

[11].Lakshmi S V ; Selvani Deepthi Kavila; International Journal of Applied Engineering Research ISSN; 04 November 2018; Volume 13, pp. 16819-16824.

[12].Vaishnavi Nath Dornadulaa; Geetha S; INTERNATIONALCONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019; December 2019; vol 08; page no. 631–641.

[13].Ayushi Agrawal; Shiv Kumar; Amit Kumar Mishra; 2015 2 nd Internation al Conference on Computing for Sustainable Global Development (INDIACom); 11-13 March 2015; IEEE ;vol 09; page 231-242

[14].D. S. Sisodia, N. K. Reddy and S. Bhandari; IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI),2017; Chennai,pp.2747 2752.

[15].A. Roy and J. Sun and R. Mahoney and L. Alonzi and S. Adams and P. Beling, "Deep learning detecting fraud in credit card transactions," IEEE International Conference on in Systems and Information Engineering Design Symposium(SIEDS), pp. 129-134, 2018.

[16].K. Modi and R. Dayma, "Review on fraud detection methods in credit card transactions, "International Conference on Intelligent Computing and Control (I2C2), Coimbatore, pp. 1-5, 2017.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details