JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

Credit Card Fraud Detection Using Random Forest Algorithm

Mrs. B. Rama Devi,[1]Sk. Neha,[2]M. V Jyothsna,[3]M. Srilekha,[4]N. Mounika,[5]N. Maruthi Priya[6]

[1]Asst. Professor, Department of Computer Science and Engineering

[2,3,4,5,6]Student, Department of Computer Science and Engineering

[1,2,3,4,5,6]QIS College of Engineering and Technology, vengamukkalapalem,ongole-523272

Abstract-Cashless purchases may be made using a credit card, which is widely accepted both online and offline. Making money and other kinds of transactions is a simple, convenient, and routine occurrence these days. As technology advances, so do the number of people who commit credit card theft.In the global statement enhancement, financial deceit has a significant compounding effect. These scams have cost the economy billions of dollars. These transactions are carried out with such finesse that they resemble real ones. Because of this, basic design techniques and other less composite ways will not be able to work. All banks now need a well-organized technique of fraud detection in order to reduce chaos and establish order. To detect Master Card fraud, we applied machine learning in this research. Random Forest Algorithm and OD techniques are also used to improve the best solution for fraud detection concerns. Efforts to reduce false alarms and increase fraud detection are still proven. Since European cardholders have had 284,807 communications, a data collection of card transactions has been collected. Slightly of these tactics may be used to the bank's credit card scam detection system to identify and prevent the scam.

Keywords— Random Forest Algorithm, Credit card fraud, Local Outlier Factor, Machine Learning, Logistic Regression.

I. INTRODUCTION

Financial fraud is on the rise, putting the banking system, large corporations, and the government at risk of huge losses [1]. When a fraudster uses a credit card without the owner's knowledge, they are committing credit card fraud. Credit card fraud may be committed in two ways: physically taking the card or utilising the card's subtle information, such as the number, CVV, expiration year, and name, without the cardholder's permission [2]. Criminals might use this information to begin significant transactions or purchases before the cardholder is aware of them. The objective is to detect all false transactions with a high degree of precision, while minimising the inaccurate scam setups.. The system identifies trends in the payment method based on the user's prior transactions (minimum 10- 15 transactions) [3]. Similar transactions that were later found to be fraudulent are part of the credit card scam detection issue. Maximize the possibilities of accuracy by using techniques like random forest, isolated forest, logistic regression, etc. The access mechanism is one of the most important tools for verifying the security of data. Authorized users will have access to the data and the system, as promised. It is possible to identify users who are making an attempt to abuse a system in an illegal manner. System [4] relies heavily on this strategy for safety. The hybrid classifiers, where the DCNN and NN are combined, continue to be used for classification. The MS-SL model, in addition, ensures that the NN's unseen neurons are set to the ideal level [5].Machine learning and data science are covered in this paper. It also compares credit card fraud detection methods utilising logistic regression, isolated forest, SVM and Local Outlier Factor algorithm. It also shows how these two

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

professions may be combined to tackle difficult challenges..Finally, the accuracy, compassion, specificity, and stability of cataloguing degrees are used to evaluate the presenting judgements of these four methodologies.

II. RELATEDWORKS

Credit Card Scam Recognition Based on Operation Behavior was suggested by John Richard D. Kho and Larry A. Vea et al. With the widespread use of EMV chip cards, the formerly chaotic behaviour that had been modelled using long-standing Magnetic stripe card tools has been mostly tamed.Despite this, a slew of papers stand ready to raise questions about the initiative and the need for EMV. Despite the article's warning that the discovery prototype must be accessible in the event of a failure of the skill, it is nevertheless important to keep this in mind. [6].Suman et al. Credit Card Fraud Detection Survey Proposal. In today's world, banks have more power than ever to protect their customers' money against fraud. This paper's primary goal is to describe methods that can be used to detect credit card scams [7]. The use of these technologies will aid in the detection of credit card fraud and provide a compliant conclusion.Using Machine Learning and Data Science, credit card scams may be discovered, according to S. P. Maniraj, Aditya Saini, and others.As long as recognition card companies are able to identify false acclaim card transactions, customers will not be charged for drugs they did not purchase. Machine Learning assumes that such a malfunctioning container exists [8]. Finding Problematic with Credit Card Scams includes displaying past credit card communications via the statistics of those who have been subjected to scams available. [9].University of Louisiana at Lafayette (ULB) and Kaggle have launched a machine learning group to combat credit card fraud. The data needed to build the implementation model was accessible. The data was skewed (fewer fraud incidents were reported) [10].

III. PROPOSED SYSTEM ARCHITECTURE

Random forests is a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance. Python SKLEARN inbuilt contains support for CART with all decision trees and random forest classifier. Random forests has a variety of applications, such as recommendation engines, image classification and feature selection. It can be used to classify loyal loan applicants, identify fraudulent activity and predict diseases. It lies at the base of the Boruta algorithm, which selects important features in a dataset. In this project we are using python Random Forest inbuilt algorithm to detect fraud transaction from credit card dataset, we downloaded this dataset from ‘kaggles’ web site from below URL

Dataset URL: https://www.kaggle.com/mlg-ulb/creditcardfraud

To provide privacy to users transaction data kaggles peoples have converted transaction data to numerical format using PCA Algorithm. Below are some example from dataset.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

"Time","V1","V2","V3","V4","V5","V6","V7","V8","V9","V10","V11","V12","V13","V14","V15","V16","V17","V18","V19","V20","V21","V22","V23","V24","V25","V26","V27","V28","Amount","Class"

0,-1.3598071336738,-0.0727811733098497,2.53634673796914,1.37815522427443,-0.338320769942518,0.462387777762292,0.239598554061257,0.0986979012610507,0.363786969611213,0.0907941719789316,-0.551599533260813,-0.617800855762348,-0.991389847235408,-0.311169353699879,1.46817697209427,-0.470400525259478,0.207971241929242,0.0257905801985591,0.403992960255733,0.251412098239705,-0.018306777944153,0.277837575558899,-0.110473910188767,0.0669280749146731,0.128539358273528,-0.189114843888824,0.133558376740387,-0.0210530534538215,149.62,"0"

0,1.19185711131486,0.26615071205963,0.16648011335321,0.448154078460911,0.0600176492822243,-0.0823608088155687,-0.0788029833323113,0.0851016549148104,-0.255425128109186,-0.166974414004614,1.61272666105479,1.06523531137287,0.48909501589608,-0.143772296441519,0.635558093258208,0.463917041022171,-0.114804663102346,-0.183361270123994,-0.145783041325259,-0.0690831352230203,-0.225775248033138,-0.638671952771851,0.101288021253234,-0.339846475529127,0.167170404418143,0.125894532368176,-0.00898309914322813,0.0147241691924927,2.69,"0"

406,-2.3122265423263,1.95199201064158,-1.60985073229769,3.9979055875468,-0.522187864667764,-1.42654531920595,-2.53738730624579,1.39165724829804,-2.77008927719433,-2.77227214465915,3.20203320709635,-2.89990738849473,-0.595221881324605,-4.28925378244217,0.389724120274487,-1.14074717980657,-2.83005567450437,-0.0168224681808257,0.416955705037907,0.126910559061474,0.517232370861764,-0.0350493686052974,-0.465211076182388,0.320198198514526,0.0445191674731724,0.177839798284401,0.261145002567677,-0.143275874698919,0,"1"

Above bold names are the column names of this dataset and others decimal values are the content of dataset and in above 3 rows last column contains class label where 0 means transaction values are normal and 1 means contains fraud values. Using above 'CreditCardFraud.csv' file we will train Random Forest algorithm and then we will upload test data file and this test data will be applied on Random Forest train model to predict whether test data contains normal or fraud transaction signatures. When we upload test data then it will contains only transaction data no class label will be there application will predict and give the result. Test data file is shown in the figure 1.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

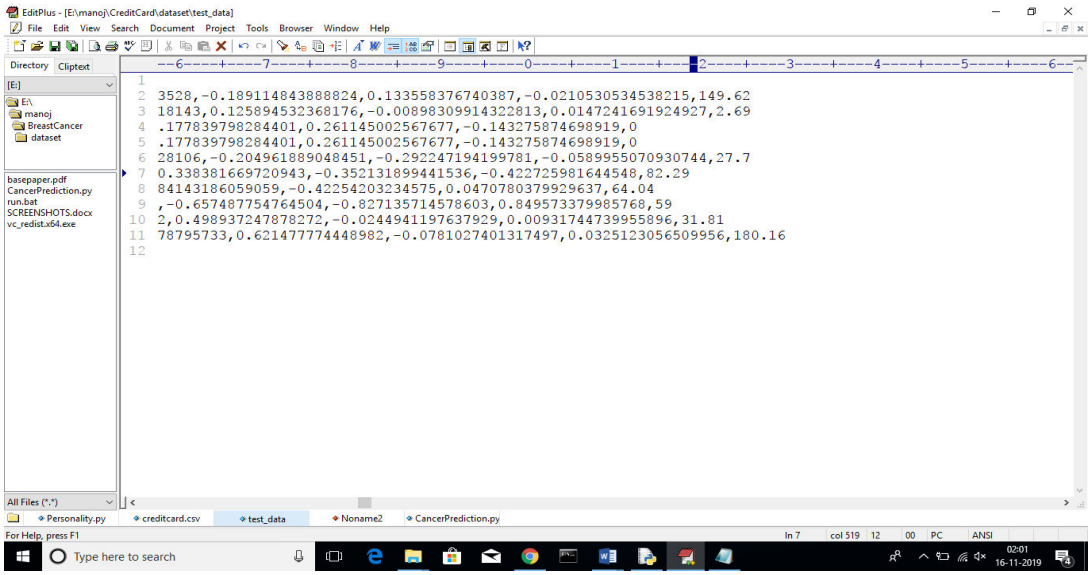


Fig.1 Test data file

In above screen in test data file there are no 0 or 1 values, application will predict from this test data using random forest and give the result.

IV. RESULTS AND DISCUSSION

The execution process and screen shots of our work is shown from fig.2 to fig. 9.
To run project double click on 'run.bat' file to get below screen

Fig.2 Upload Credit Card Dataset'

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

In above screen click on 'Upload Credit Card Dataset' button to upload dataset

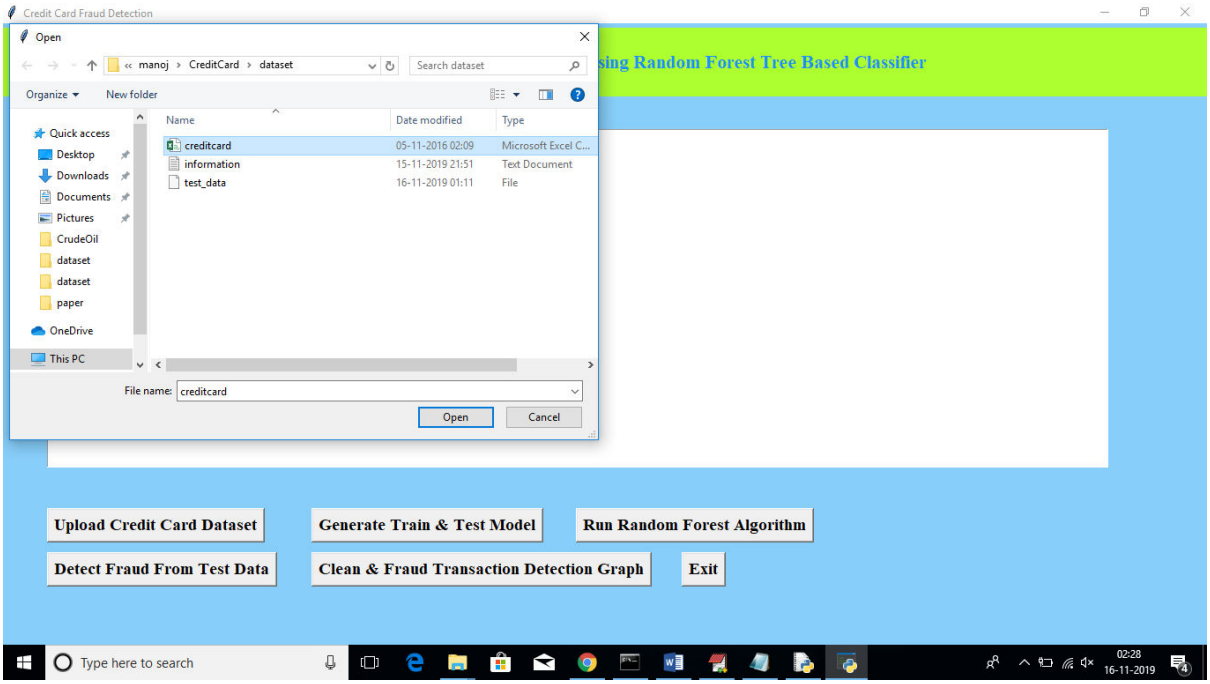
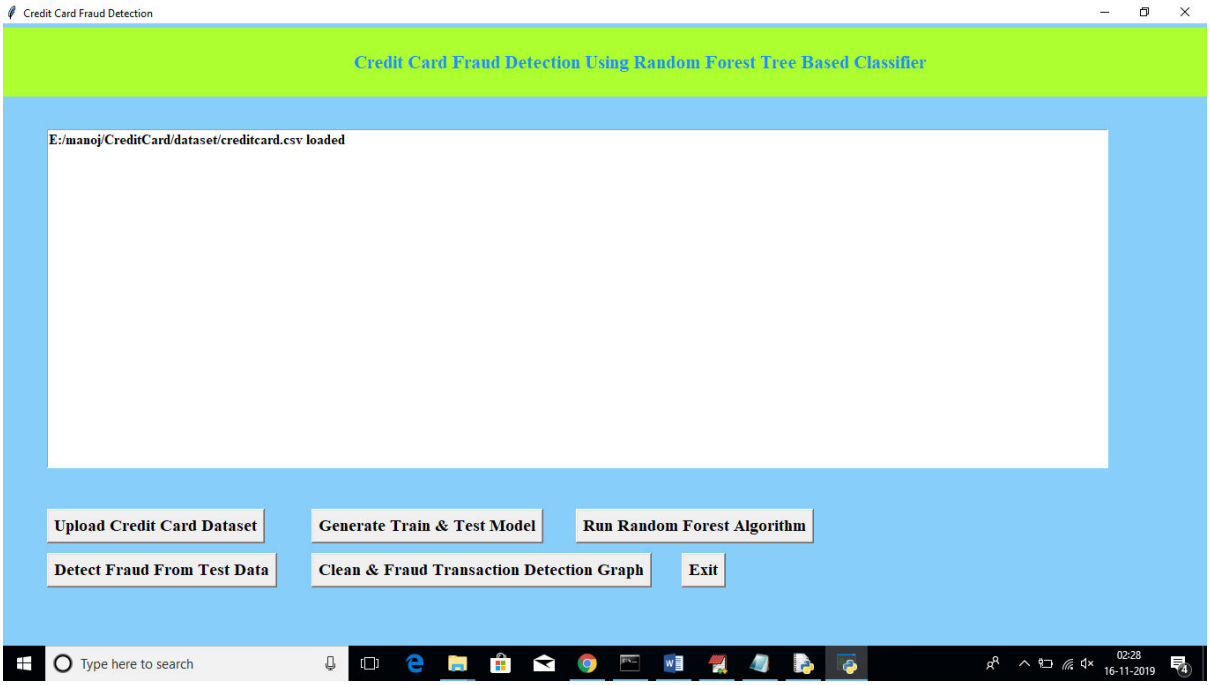


Fig.3 Browse Credit Card Dataset'

After uploading dataset will get below screen



ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

Fig.4 Generate Train & Test Model'

Now click on 'Generate Train & Test Model' to generate training model for Random Forest Classifier

Fig.5 Total records available in dataset

In above screen after generating model we can see total records available in dataset and then application using how many records for training and how many for testing. Now click on "Run Random Forest Algorithm' button to generate Random Forest model on train and test data.

Fig. 6 Run Random Forest Algorithm

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

In above screen we can see Random Forest generate 99.78% percent accuracy while building model on train and test data. Now click on 'Detect Fraud from Test Data' button to upload test data and to predict whether test data contains normal or fraud transaction.

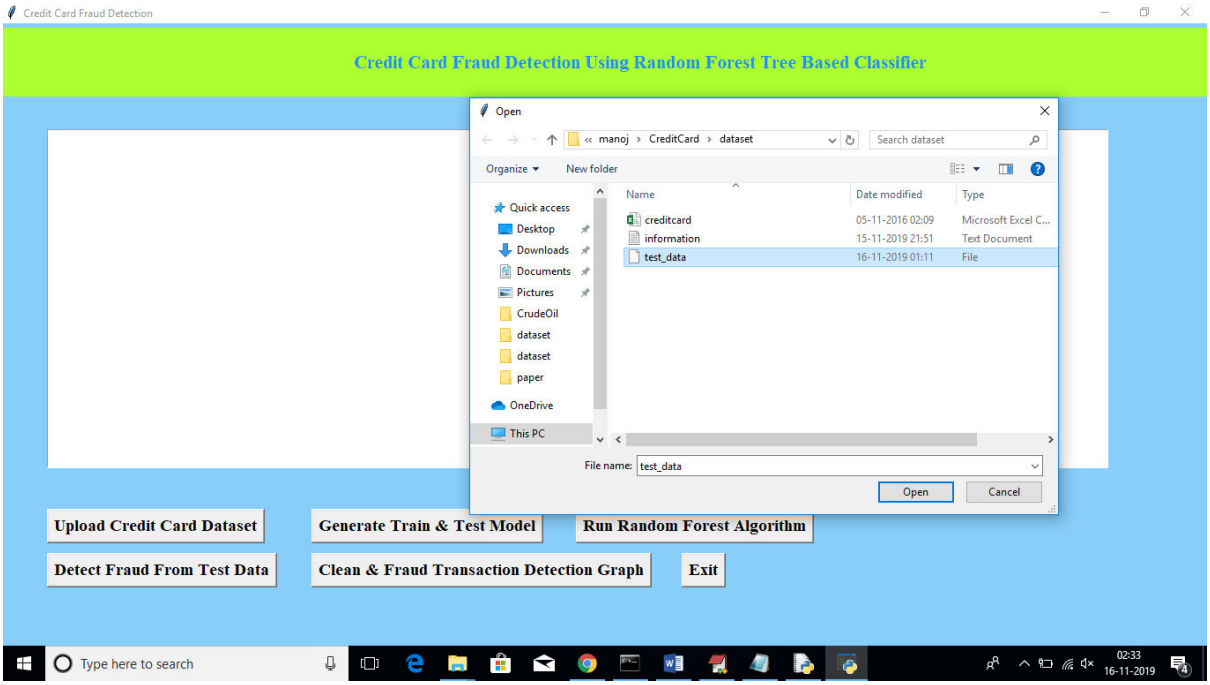


Fig.7 Browse Test data file

In above screen I am uploading test dataset and after uploading test data will get below prediction details.

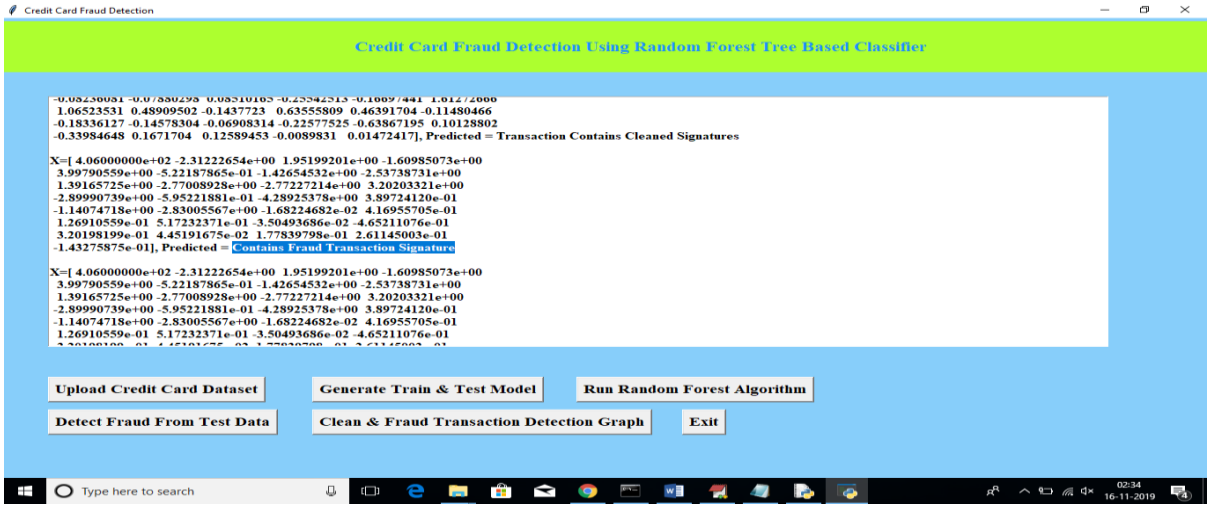


Fig.8 Prediction Details

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

In above screen beside each test data application will display output as whether transaction contains cleaned or fraud signatures. Now click on 'Clean & Fraud Transaction Detection Graph' button to see total test transaction with clean and fraud signature in graphical format. See below screen

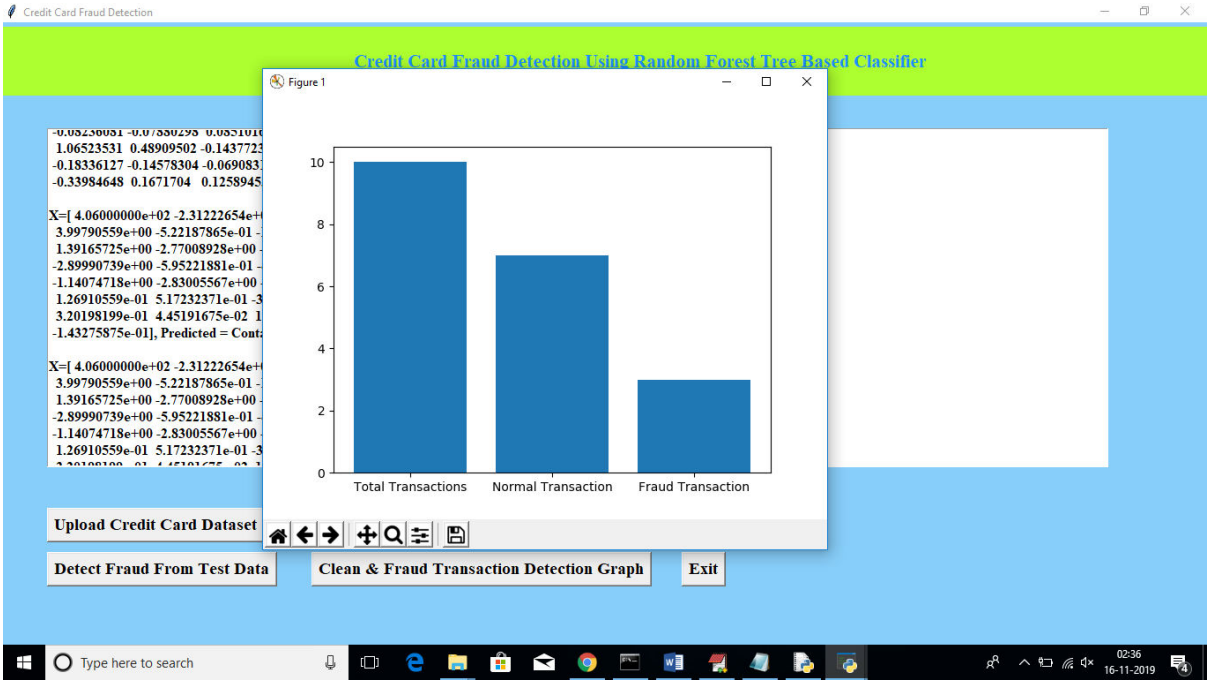


Fig.9 Clean & Fraud Transaction Detection Graph

In above graph we can see total test data and number of normal and fraud transaction detected. In above graph x-axis represents type and y-axis represents count of clean and fraud transaction

V. FUTURE SCOPE AND CONCLUSION

There is no hesitation when it comes to calling this act of unlawful deception, "Credit card fraud." Local Outlier Factor, Random Forest technique is examined in this research to see how they compare. In addition, there was a section devoted to credit card scams. The results of the study reveal that despite the unbalanced data and the compressed timeframe, Isolation Forest is a significant presenting tool. Because it's based on machine learning methods, the database's determination of existence only increases one ability with time, like the creation of more records. The fundamental objective of scheme authorizations targeted at frequent processes to stay integrated comprised by sections and their results may be merged to increase the accuracy of the final result. It's possible that adding more steps will make this prototype better. It was previously stated that the procedure's accuracy is enhanced as the dataset is expanded. Additional information will make the prototype more accurate in detecting frauds and reducing the number of false positives.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

REFERENCES

[1] John Richard, D. Kho, Larry A. Vea, "Credit Card Fraud Detection Based on Transaction Behaviour", 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017.

[2] Suman, GJUS&T Hisar HCE, Sonepat, "Survey Paper on Credit Card Fraud Detection", International Journal of Advanced Re- search inComputer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2016. Pages 237–243, https://doi.org/10.1093/ijlct/ctt041

[3] S P Maniraj and Aditya Saini, "Credit Card Fraud Detection using Machine Learning and Data Science", International Journal of Engineering Research & Technology (IJERT), Vol. 8 Issue 09, September-2019.

[4] ULB (2018), Kaggle, "Machine Learning Group-Credit Card Fraud Detection".

[5] Massimiliano Zanin, Miguel Romance, ReginoCriado, and SantiagoMoral, "Credit Card Fraud Detection through Parenclitic Network Analysis", Hindawi Complexity Volume 2018, Article ID 5764370.

[6] Steven J. Murdoch, Saar Drimer, Ross Anderson and Mike Bond, "Chip and PIN is Broken", IEEE Symposium on Security and Privacy, pp. 433-446.

[7] Ishu Trivedi, Monika, Mrigya,Mridushi, "Credit Card Fraud Detection-by" International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016.

[8] "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018.

[9] Yogesh M. Gajmal, R. Udayakumar,"Authentication based Data Access Control and sharing mechanism in Cloud using Blockchain technology" published by International Journal of Emerging Trends in Engineering Research, VOL. 8, NO. 9, September 2020.

[10] Arvind M Jagtap, Prof.Dr.Gomathi N, "Meta-Heuristic based Trained Deep Convolutional Neural Network for Crop Classification" ,International Journal of Emerging Trends in Engineering Research (IJETER) Volume 8. No. 7, July 2020.