

Detecting the Credit Card Fraud by applying the Random Forest Algorithm

Dr. J. Sasi Kiran¹, Dr. K.G.S Venkatesan², D. Venkaiah³, K. Narayana Rao⁴, G. Siva Prasad⁵

^{1, 2, 3, 4, 5} Department of Computer Science and Engineering,

^{1, 2, 3, 5} QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

⁴AUCE(A), Andhra University

¹sasikiran.j@qiscet.edu.in, ²venkatesan.kgs@qiscet.edu.in, ³venkaiah.d@qiscet.edu.in

⁴narayanaraok@andhrauniversity.edu.in, ⁵g.sivaprasad@qiscet.edu.in

Corresponding Author Mail: qispublications@qiscet.edu.in

Article Info

Page Number: 236-246

Publication Issue:

Vol. 69 No. 1 (2020)

Article History

Article Received:

12 September 2020

Revised: 16 October 2020

Accepted: 20 November 2020

Publication: 25 December 2020

Abstract

Credit Card is universally accepted both offline and online. With credit card one can make cashless purchases. Making transactions for money and other things is now easy, convenient, and commonplace. The number of credit card thefts increases along with technological advancements. Financial fraud has a huge compounding effect in the worldwide statement enhancement. The economy has suffered billion-dollar losses due to these frauds. These transactions are carried out so deftly that they appear authentic. Fundamental design methods and other simple approaches won't be able to function as a result. In order to decrease confusion and promote order, all banks now require a fraud detection method which is well-organized. In this study, machine learning is used to identify Master Card fraud. In order to enhance the most effective response to fraud detection issues, the Random Forest Algorithm and OD approaches are also utilized. It is still shown to reduce false alarms and improve fraud detection. The collection of data of credit card transactions has been made since 284,807 communications have been made by European cardholders. To recognize and stop the scam, some of these techniques might be applied to the credit card scam detection system of the bank.

Keywords: Local outlier factor, Random Forest algorithm, detection of Credit card fraud, Logistic Regression, Machine Learning,

I. INTRODUCTION

The banking sector, major firms, and the government are at risk of suffering significant losses due to the growth in financial fraud. [1]. Credit card fraud occurs when a perpetrator uses a card without the permission of owner. Using the card physically or using the card's sensitive data, like the number of card, CVV number, valid till year, and card holder name, without the cardholder's consent, are two ways that credit card fraud can be conducted. [2]. This information could be used by thieves to start large purchases or transactions before the cardholder is aware of them. The goal is to accurately identify every fraudulent transaction while minimizing faulty scam setups. Based on the user's previous transactions, the system determines patterns in the payment method.(at least 10 to 15 transactions) [3]. The problem with detecting credit card scams includes the transactions that were similar and later decided to be scam. Utilize the methods like random forest algorithm and logistic regression, etc. to

increase the likelihood of correctness. The most crucial tool for ensuring the certainty of data is the access mechanism. As promised, the system and data will be accessible to authorized users. It is feasible to identify users who seek to illegally exploit a system. This strategy is used by the System for security [4]. The hybrid classifiers, which mix NN and DCNN, are still in use for classification. Additionally, the model MS-SL makes sure that the level of the NN's invisible neurons is set to the perfect level. [5]. In this paper, ML and concepts of data science are discussed. Additionally, it contrasts techniques for detecting credit card fraud that use the local outlier factor algorithm, isolated forest, SVM, and logistic regression. It demonstrates how these two professions can work together to address complex problems. In the end, the evaluation of the presenting judgments of these four methodologies is done using the correctness, compassion, specificity, and stability of cataloguing degrees.

II. RELATEDWORKS

John Richard D. Kho and Larry A. Veal et al proposed the Credit Card Scam Recognition Based on Operation Behavior. The previously chaotic behavior that was modeled using long-standing Magnetic stripe card tools has mostly been controlled with the widespread introduction of EMV chip cards. In spite of this, a number of papers are prepared to cast doubt on the plan and the necessity of EMV. It is crucial to remember this even if the essay cautions that the prototype of discovery needs to be available if the skill fails. [6]. Credit Card Fraud Detection Survey was a proposal by Suman et al. Banks has greater authority than ever in the modern world to safeguard the money of their clients from fraud. The main objective of this essay is to describe techniques for spotting credit card scams [7]. By utilising these technologies, credit card theft can be identified and a compliance result can be reached. According to S. P. Maniraj, Aditya Saini, and others, credit card frauds may be identified using ML algorithms and data science. Clients won't be billed for medications they didn't buy as long as recognition card firms can spot fraudulent acclaim card transactions. Machine learning makes the assumption that such a broken container exists. [8]. The statistics of those who have fallen victim to scams are available when searching for Credit Card Scams problems, which includes displaying previous credit card conversations. [9]. A machine learning team has been established by the USB(University of Louisiana at Lafayette) and Kaggle to fight credit card fraud. The information required to create the model of implementation was available. Data were biased (lesser fraud instances were given) [10].

III. PROPOSED SYSTEM ARCHITECTURE

Random forest comes under the category of supervised learning. The problems which consist of classification or regression can be solved with it. Moreover, it can be viewed as the most convenient and adjustable algorithm. The strength of the forest depends on number of trees in it. If there are more trees the forest is said to be strong. On arbitrarily chosen data samples, random forests creates decision trees and collect projections from each tree, then choose on the best option. Additionally, it offers a fairly accurate indicator of the feature's relevance. The built-in CART support in Python SKLEARN works with all decision trees and random forest classifiers. Applications for random forests include recommendation systems such as restaurant recommendation, classification of pictures, and feature selection problems. It can

be employed to categorize dependable loan candidates, spot fraud, and forecast sickness. The Boruta algorithm, that selects important features in a dataset, is built around it. Throughout this project, we use the built-in Random Forest technique in Python to identify credit card fraud transactions. We collected this dataset from the 'kaggle' website at the URL below

URL for the dataset(from Kaggle): <https://www.kaggle.com/mlg-ulb/creditcardfraud>

Transaction data has been transformed to a numerical representation using the PCA algorithm to protect consumers' privacy. Instances from the dataset are listed below.

"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"

4,0.22965763450793,1.141003507049326,0.0453707735899449,0.20261273673594,1.19188
0988597645,0.272708122899098,0.00515900288250983,0.0812129398830894,0.464959994
783886,-0.0992543211289237, -1.41690724314928,

-0.153825826253651,0.75106271556262,0.16737196252175

0.0501435942254188,-0.443586797916727,0.00282051247234708,-0.61198733994012,

-0.0455750446637976,-0.21963255278686,-0.167716265815783,-0.270709726172363,

-0.154103786809305,-0.780055415004671,0.75013693580659,-0.257236845917139,

0.0345074297438413,0.00516776890624916,4.99,0

9,-0.33826175242575,1.11959337641566,1.04436655157316,-0.222187276738296,

0.49936080649727, -0.24676110061991,0.651583206489972,0.0695385865186387,

-0.736727316364109, -0.366845639206541,1.01761446783262,0.836389570307029

1.00684351373408,-0.443522816876142,0.150219101422635,0.739452777052119

-0.540979921943059,0.47667726004282,0.451772964394125,0.203711454727929

-0.246913936910008,-0.633752642406113,-0.12079408408185,-0.385049925313426

-0.0697330460416923,0.0941988339514961,0.246219304619926,0.0830756493473326

3.68,0

16,0.694884775607337,-1.36181910308009,1.02922103956032,0.834159299216716

-1.19120879445965,1.30910881872952,-0.878585911450457,0.4452901278385

-0.446195831557423,0.568520735086962,1.01915061274695,1.29832870056251

0.420480265280796,-0.372650997239682,-0.807979512809369,-2.04455748288968

0.515663469043577,0.625847298442513,-1.30040816880609,-0.138333940419021

-0.295582931552029,-0.571955006812512,-0.0508807005036795,-0.304214501020644

0.0720010061385359,-0.422234430367677,0.0865533980909977,0.0634986493439305

231.71,0

The column names highlighted above are the titles of the other decimal columns of the dataset. The last column in the above three rows has a class label, in which 0 indicates that transaction values are normal and 1 indicates that fraudulent values are present. The Random Forest algorithm can be used to train the "CreditCardFraud.csv" file. Test data is uploaded. The Random Forest train model will be tested with the test data to decide whether the test data contains basic or fraudulent transaction signatures. When test data is uploaded, it will simply consist transaction data; there won't be a class label. The application will then forecast and provide the outcome. Figure 1 shows the test data file.

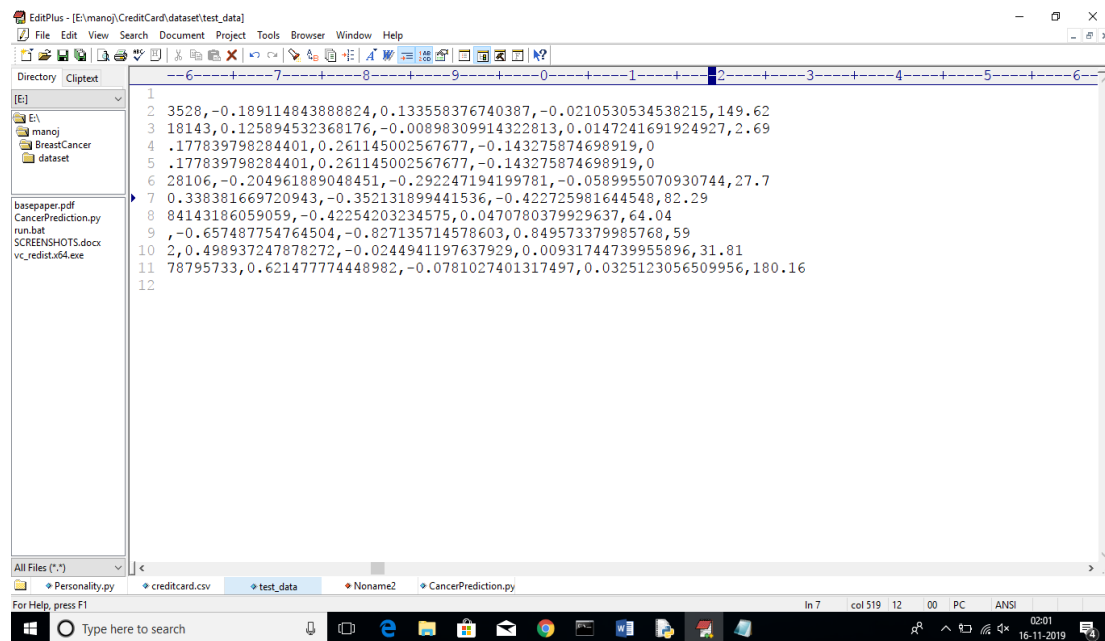


Fig.1 Test data file

There are no 0 or 1 values in test data file shown in above screen, random forest algorithm will be used by the application to perform prediction from the given test data and give the result.

IV. RESULTS AND DISCUSSION

fig.2 to fig. 9. Shows the execution process and screenshots of the work

Double click on 'run.bat' file to run the project and below screen will be visible.



Fig.2 Credit Card Dataset upload

Select the 'Upload Credit Card Dataset' button.

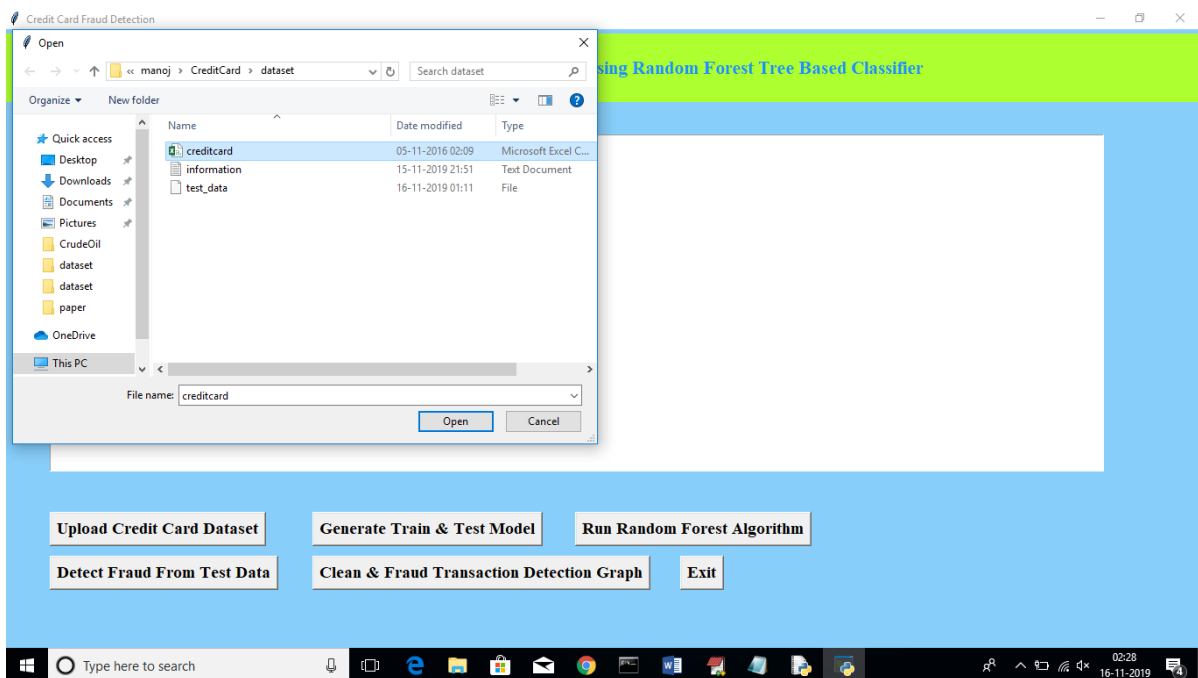


Figure.3 Browse Credit Card Dataset

Below screen will be visible after uploading dataset



Fig.4 Train and Test Model generated

Now select 'Generate Train and Test Model'. It will produce the model of training for Random Forest Algorithm

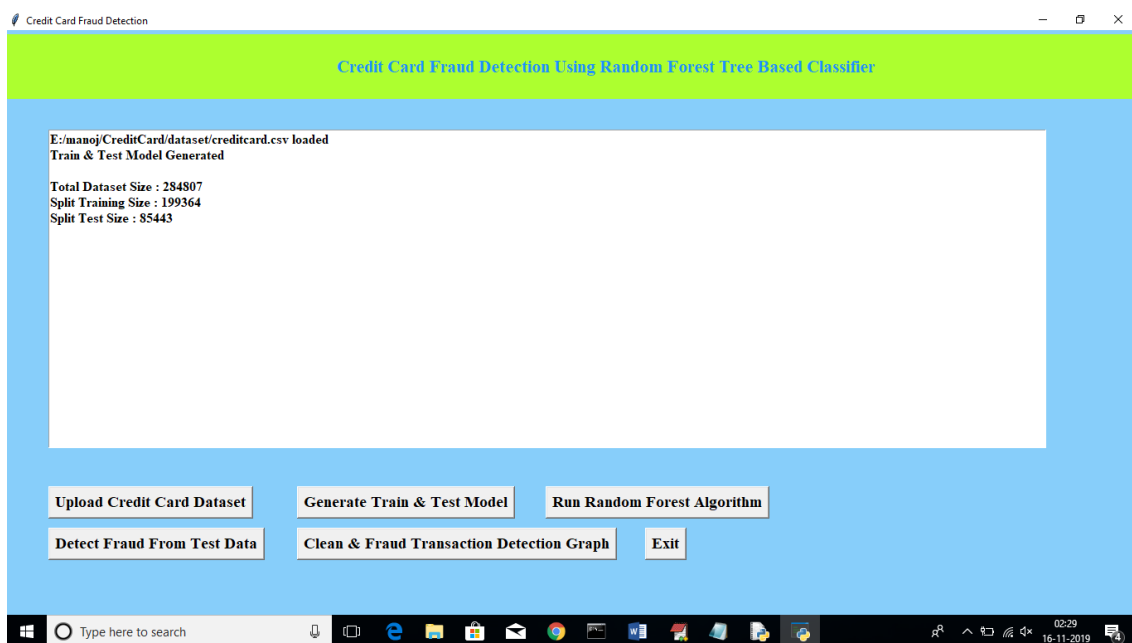


Figure 5 shows the total number of records in dataset

Total records available in dataset will be visible on the screen after generating training model and then the information of records that is how many records used for training and how many used 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

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.

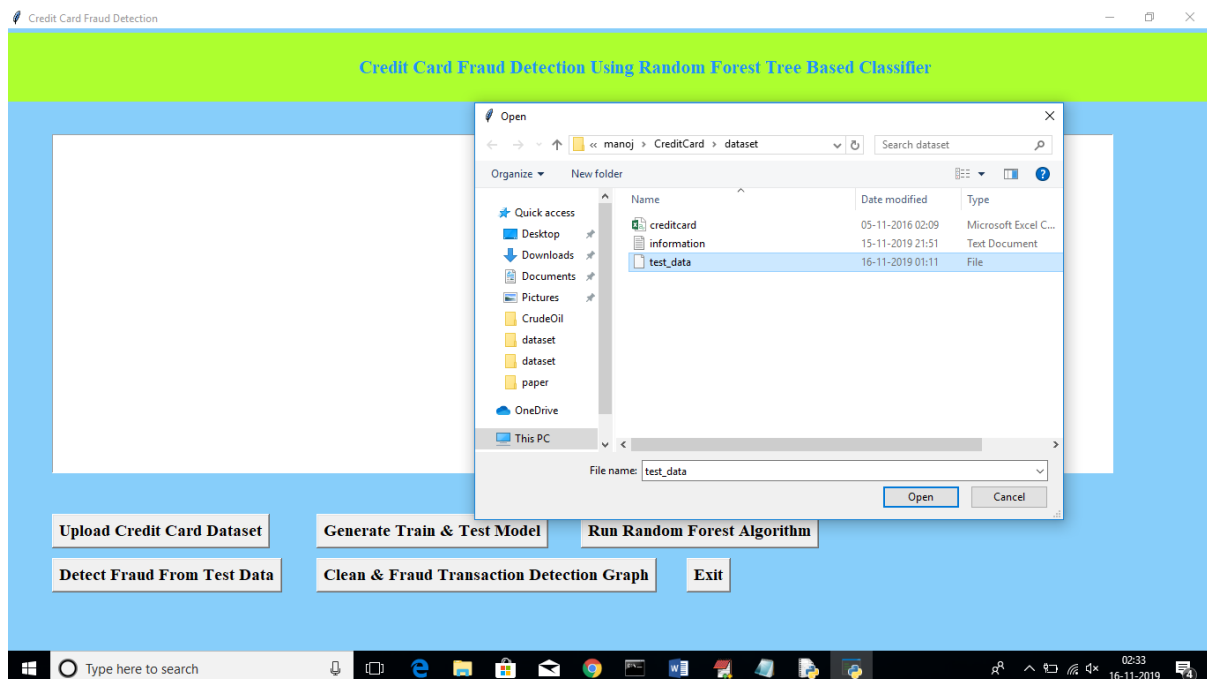


Fig.7 Browse Test data file

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

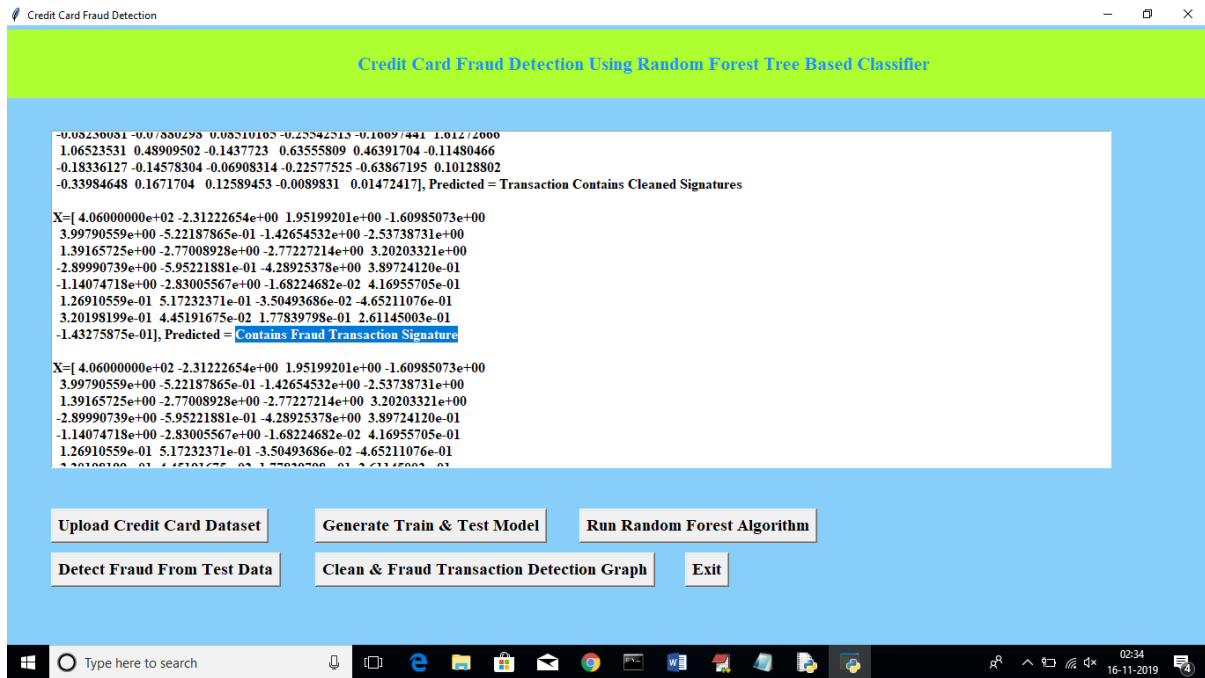


Fig.8 Prediction Details

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

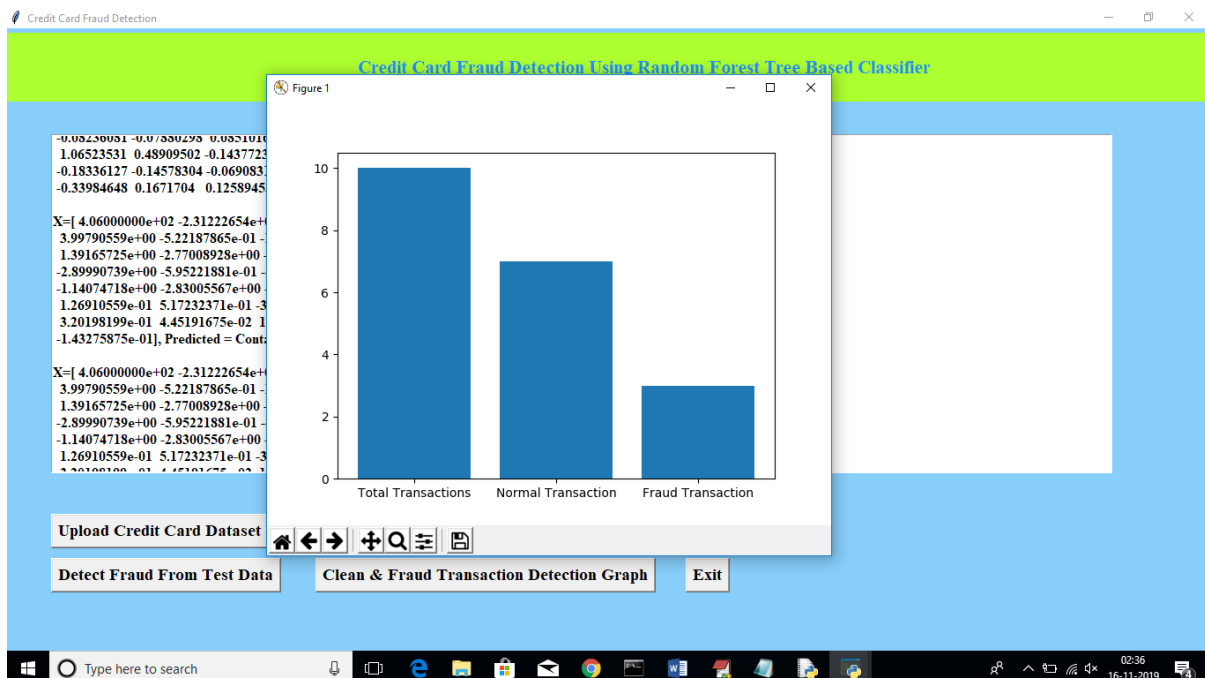


Fig.9 Clean & Fraud Transaction Detection Graph

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

When it comes to referring to this criminal deception as "Credit card fraud, there is no reluctance. In this study, the Local Outlier Factor and Random Forest techniques are compared. Along with this, a section on credit card fraud was included. The findings of the study show that Isolation Forest is an important presenting tool despite the uneven data and the compressed timeframe. Because it is built on machine learning techniques, the database can only determine its own existence by creating more records over time. The main goal of scheme authorizations was to keep sections of often occurring procedures connected so that their outcomes may be combined to improve the correctness of final result. This prototype might be improved by including more steps. As the dataset is increased, the accuracy of procedure improves. More data will improve the prototype's accuracy in identifying frauds and lower the amount of false positives.

REFERENCES

- [1] John Richard, D. Kho, Larry A. Veal, "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, Sonapat, "Survey Paper on Credit Card Fraud Detection", International Journal of Advanced Research in Computer 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, Regino Criado, and Santiago Moral, "Credit Card Fraud Detection through Parental 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.
- [11] Sudhamathy G: Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, vol. 8, no-5, pp. 1954-1966.
- [12] LI Changjian, HU Peng: Credit Risk Assessment for ural Credit Cooperatives based on Improved Neural Network, International Conference on Smart Grid and Electrical Automation vol. 60, no. - 3, pp 227-230, 2017.
- [13] Wei Sun, Chen-Guang Yang, Jian-Xun Qi: Credit Risk Assessment in Commercial Banks Based On Support Vector Machines, vol.6, pp 2430-2433, 2006.
- [14] Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE,“BLAST-SSAHA Hybridization for Credit Card Fraud Detection”, vol. 6, no. 4 pp. 309-315, 2009.
- [15] Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines, Proceedings of International Multi Conference of Engineers and Computer Scientists, vol. I, 2011.
- [16] Adi Saputra¹, Suharjito²L: Fraud Detection using Machine Learning in e-Commerce, International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.
- [17] Dart Consulting,Growth Of Internet UsersIn India And Impact On Country’s Economy: [https:// www. Dartconsulting.co.in/market - news/ growth- ofinternet-users-in-india-and-impact -on-countryseconomy/](https://www.Dartconsulting.co.in/market-news/growth-ofinternet-users-in-india-and-impact-on-countryseconomy/)
- [18] Ganga Rama Koteswara Rao and R.Satya Prasad, “ ShieldingThe Networks Depending On Linux Servers Against Arp Spoofing, Int. Journal of Engi-neering and Technology(UAE),Vol.7, PP.75-79, 2018, ,
- [19] Heta Naik , Prashasti Kanikar: Credit card Fraud Detection based on Machine Learning Algorithms, International Journal of Computer Applications (0975 – 8887) Volume 182 , March 2019.
- [20] Randula Koralage, , Faculty of Information Technology, University of Moratuwa,Data Mining Techniques for Credit Card Fraud Detection.
- [21] Roy, Abhimanyu, et al:Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.
- [22] Credit Card Fraud Detection Anonymized credit card transactions labeled as fraudulent or genuine www.kaggle.com/mlg-ulb/creditcardfraud
- [23] Working on scikit-learn library in Python to classify - Anonymized credit card transactions labeled as fraudulent or genuine <https://github.com/mohitguptaomg/Kaggle-Credit-Card-Fraud-Detection>
- [24] Credit Card Fraud Detection using Neural Networks <https://github.com/SimarjotKaur/Credit-Card-FraudDetection>
- [25] C. Yue, B. Qu, and J. Liang, “A multiobjective particle swarm optimizer using ring topology for solving multimodal multiobjective problems,” IEEE Trans. Evol. Comput., vol. 22, no. 5, pp. 805–817, Oct. 2018.

- [26] P Ramprakash, M Sakthivadivel, N Krishnaraj, J Ramprasath. "Host-based Intrusion Detection System using Sequence of System Calls" International Journal of Engineering and Management Research, Vandana Publications, Volume 4, Issue 2, 241-247, 2014
- [27] N Krishnaraj, S Smys."A multihoming ACO-MDV routing for maximum power efficiency in an IoT environment" Wireless Personal Communications 109 (1), 243-256, 2019.
- [28] N Krishnaraj, R Bhuvanesh Kumar, D Rajeshwar, T Sanjay Kumar, Implementation of energy aware modified distance vector routing protocol for energy efficiency in wireless sensor networks, 2020 International Conference on Inventive Computation Technologies (ICICT),201-204
- [29] Ibrahim, S. Jafar Ali, and M. Thangamani. "Enhanced singular value decomposition for prediction of drugs and diseases with hepatocellular carcinoma based on multi-source bat algorithm based random walk." *Measurement* 141 (2019): 176-183. <https://doi.org/10.1016/j.measurement.2019.02.056>
- [30] Ibrahim, Jafar Ali S., S. Rajasekar, Varsha, M. Karunakaran, K. Kasirajan, Kalyan NS Chakravarthy, V. Kumar, and K. J. Kaur. "Recent advances in performance and effect of Zr doping with ZnO thin film sensor in ammonia vapour sensing." *GLOBAL NEST JOURNAL* 23, no. 4 (2021): 526-531. <https://doi.org/10.30955/gnj.004020> , https://journal.gnest.org/publication/gnest_04020
- [31] N.S. Kalyan Chakravarthy, B. Karthikeyan, K. Alhaf Malik, D.Bujji Babbu., K. Nithya S.Jafar Ali Ibrahim , Survey of Cooperative Routing Algorithms in Wireless Sensor Networks, *Journal of Annals of the Romanian Society for Cell Biology* ,5316-5320, 2021
- [32] Rajmohan, G, Chinnappan, CV, John William, AD, Chandrakrishan Balakrishnan, S, Anand Muthu, B, Manogaran, G. Revamping land coverage analysis using aerial satellite image mapping. *Trans Emerging Tel Tech.* 2021; 32:e3927. <https://doi.org/10.1002/ett.3927>
- [33] Vignesh, C.C., Sivaparthipan, C.B., Daniel, J.A. et al. Adjacent Node based Energetic Association Factor Routing Protocol in Wireless Sensor Networks. *Wireless Pers Commun* 119, 3255–3270 (2021). <https://doi.org/10.1007/s11277-021-08397-0>.
- [34] C Chandru Vignesh, S Karthik, Predicting the position of adjacent nodes with QoS in mobile ad hoc networks, *Journal of Multimedia Tools and Applications*, Springer US, Vol 79, 8445-8457,2020