Detecting Credit Card Frauds Using ML Algorithm

GUNUPATI VENKATESWRLU1, KANAMATHARED DY RESHMA REDDY2

#1Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

#2Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

Abstract_ In this paper the main focal point on savings credit card fraud identification in real world. Here the savings credit card fraud detection is primarily based on transactions. Generally savings card fraud things to do can take place in each on line and offline. But in present day world on-line fraud transaction things to do are growing day by way of day. So in order to locate the on line fraud transactions quite a number techniques have been used in current gadget.In this paper we have proposed different machine learning algorithms or finding the fraudulent transactions and the accuracy of those transactions.

1.INTRODUCTION

There are quite a number fraudulent things to do detection strategies has applied in credit score card transactions have been stored in researcher minds to strategies to strengthen fashions primarily based on synthetic talent , records mining, fuzzy good judgment and computer learning. Credit card fraud detection is drastically difficult, however additionally famous trouble to solve. In our proposed device we constructed the savings card fraud detection the use of Machine learning. With the development of computing device gaining knowledge of techniques. Machine getting to know has been recognized as a profitable measure for fraud detection. A giant quantity of information is transferred for the duration of on-line transaction processes, ensuing in a binary result: true or fraudulent. Within the pattern fraudulent datasets, facets are constructed. These are information factors particularly the age and fee of the client account, as properly as the starting place of the savings card.

There are heaps of elements and every contributes, to various extents, toward the fraud probability. Note, the degree in which every characteristic contributes to the fraud rating is generated via the synthetic talent of the computing device which is pushed by means of the education set, however is no longer decided by way of a fraud analyst. So, in regards to the card fraud, if the use of playing cards to commit fraud is tested to be high, the fraud weighting of a transaction that makes use of a savings card will be equally so. However, if this had been to shrink, the contribution degree would parallel. Simply make, these fashions self-learn except express programming such as with guide review. Credit card fraud detection the usage of Machine studying is finished by means of deploying the classification and regression algorithms.

2.LITERATURE SURVEY

[1] The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada. "Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Zavarsky."

This look up paper focuses on the introduction of a scorecard from relevant contrast criteria, features, and talents of predictive analytics supplier choices in modern times being used to word credit score score card fraud. The scorecard gives a side-byside distinction of 5 financial savings card predictive analytics supplier choices adopted in Canada. From the ensuing look up findings, a checklist of credit card fraud PAT vendor reply challenges, risks, and obstacles was once as soon as outlined.

[2] BLAST-SSAHA Hybridization for Credit Card Fraud Detection. "Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar"

This paper recommend to use two-stage sequence alignment in which a profile Analyser (PA) first determines the similarity of an incoming sequence of transactions on a given financial savings card with the proper cardholder's preceding spending sequences. The exclusive transactions traced by means of skill of the profile analyser are subsequent surpassed on to a deviation analyser (DA) for potential alignment with preceding fraudulent behaviour. The ultimate resolution about the nature of a transaction is taken on the basis of the observations by using the usage of these two analysers. In order to gain on line response time for every PA and DA,

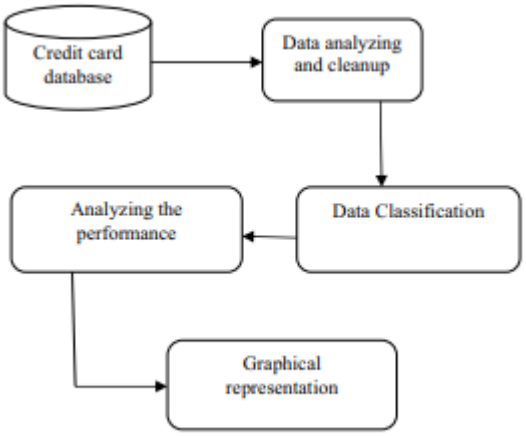
we advocate a new method for combining two sequence alignment algorithms BLAST and SSAHA.

[3] Fraudulent Detection in Credit Card System Using SVM & Decision Tree. "Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree P. Deshpande". With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplate system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by implementation of this hybrid approach, financial losses can be reduced to greater extend

3.PROPOSED WORK

We propose a Machine learning model to detect fraudulent credit card activities in online financial transactions. Analyzing fake transactions manually is impracticable due to vast amounts of data and its complexity. However, adequately given informative features, could make it is possible using Machine Learning. This hypothesis will be explored in the project. To classify fraudulent and legitimate credit card transaction by supervised learning Algorithm such as Random forest. To help us to get awareness about the fraudulent and without loss of any financially.

3.1 MODULES



1: Exploratory Data Analysis: In this module we will first collect all the credit card dataset and store it in a database. Then we will perform some descriptive analysis about the dataset.

2: Data Cleaning In the next step, after analyzing the dataset then we have to clean the data. In this cleaning process all the

duplicate values and null values that are present in the dataset will be removed and a new dataset will be obtained.

3: Preprocessing of dataset In this module the cleaned dataset will be preprocessed where the dataset will bedivided based on amount and transaction time.

4 : Dataset Partition In this module first the dataset will be divided into two partitions as trained dataset and testing dataset. After the data partitions the Random Forest Algorithm is applied. After applying Different machine learning algorithms finally a confusion matrix is obtained.

5:Evaluation Now the resultant data obtained in the form of confusion matrix can be evaluated by using graphical representation which gives better accuracy

4.ALGORTITHMS

4.1 Random Forest Algorithm

Random forests is a supervised studying algorithm. It can be used each for classification and regression. It is additionally the most bendy and handy to use algorithm. A wooded area is comprised of trees. It is stated that the extra timber it has, the greater sturdy a wooded area is. Random forests creates choice timber on randomly chosen facts samples, receives prediction from every tree and selects the fantastic answer by way of capacity of voting. It additionally gives a exceedingly desirable indicator of the characteristic importance. Python SKLEARN built in incorporates guide for CART with all selection timber and random wooded area classifier.

Random forests has a range of applications, such as suggestion engines, photo classification and function selection. It can be used to classify loyal mortgage applicants, become aware of fraudulent recreation and predict diseases. It lies at the base of the Boruta algorithm, which selects necessary elements in a dataset.

5.RESULTS AND DISCUSSIONS

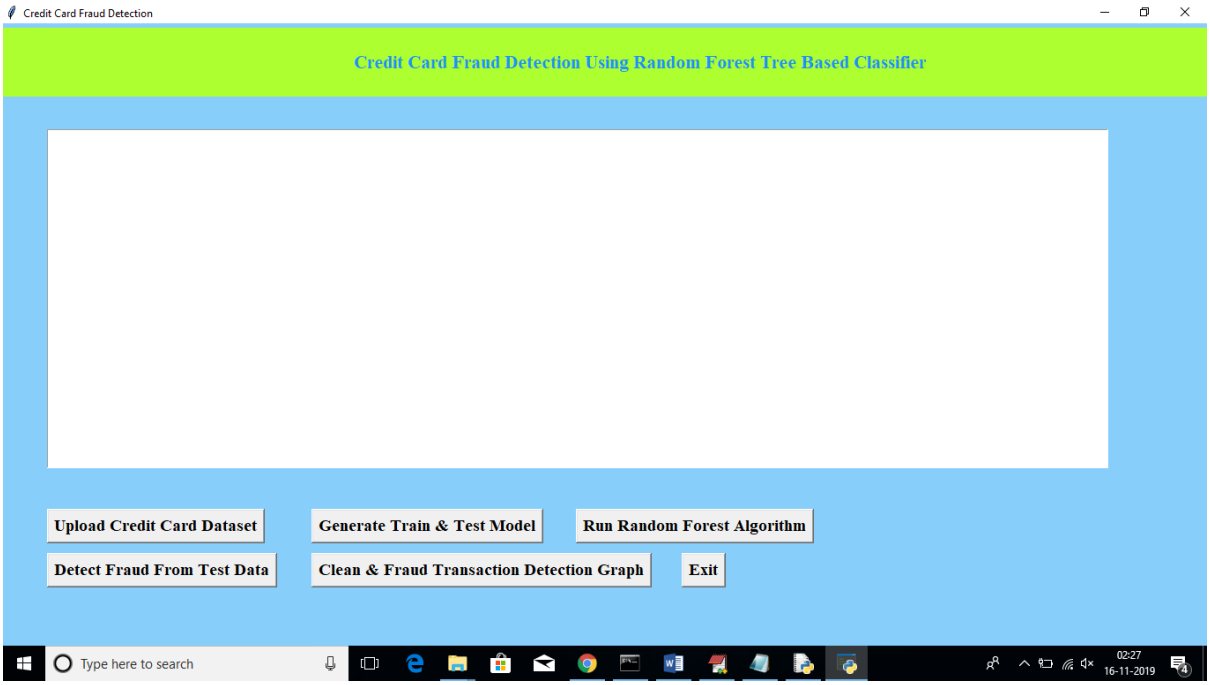


Fig 1:In above screen click on 'Upload Credit Card Dataset' button to upload dataset

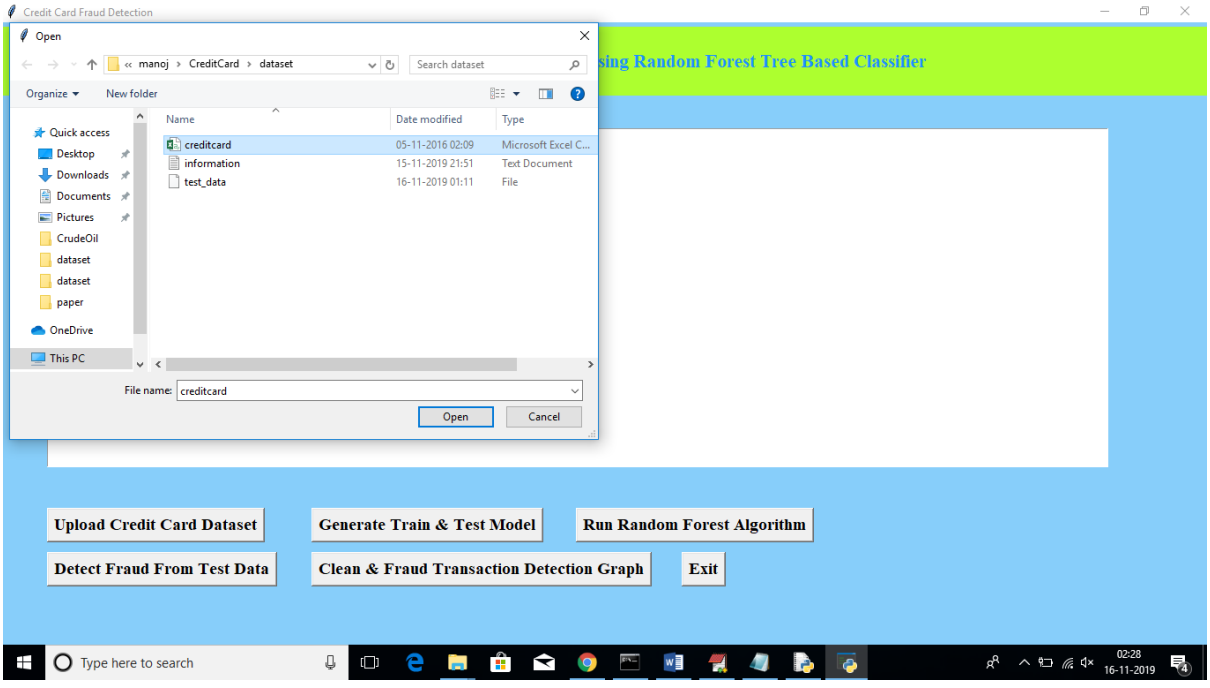


Fig 2: After uploading dataset will get below screen

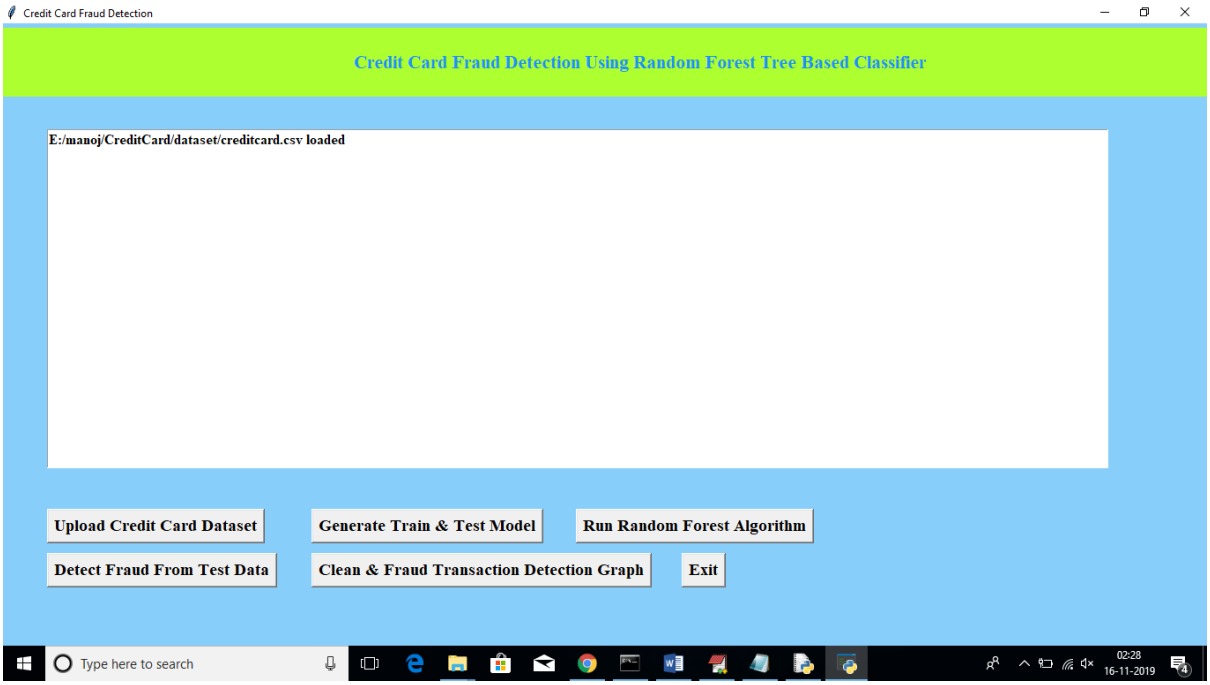


Fig 3:Now click on 'Generate Train & Test Model' to generate training model for Random Forest Classifier

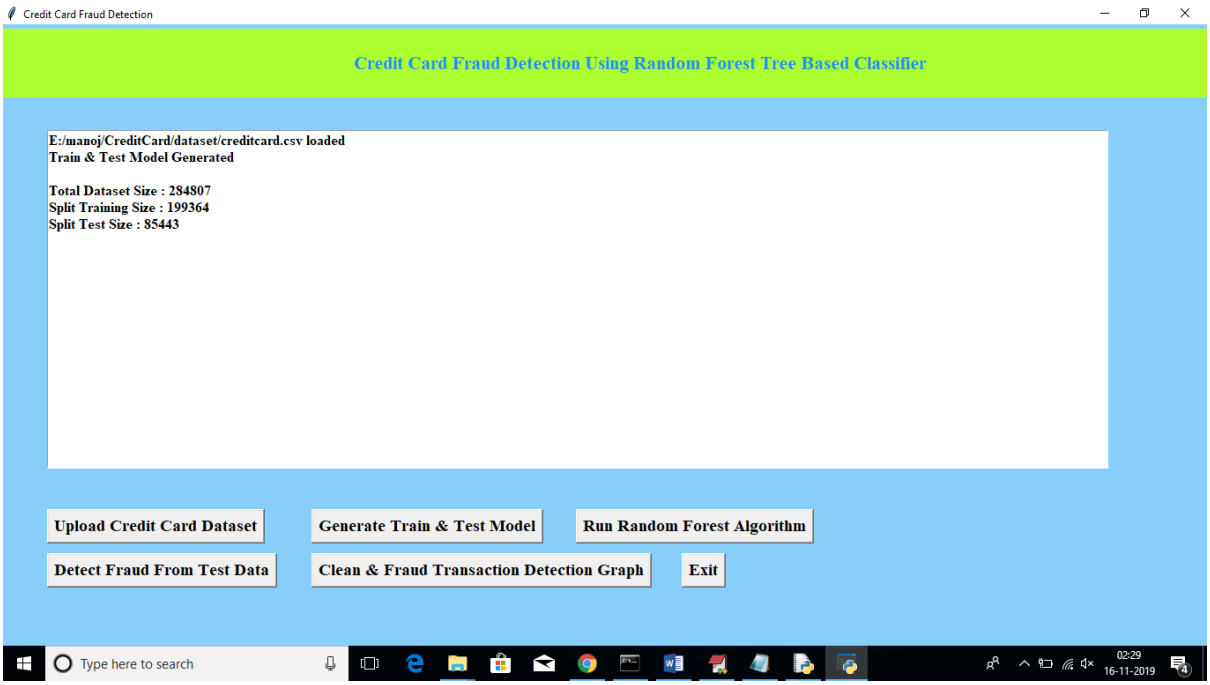


Fig 4: In above screen after generating model we can see total records available in dataset and then application using how many records for training and how many for testing. Now

click on "Run Random Forest Algorithm' button to generate Random Forest model on train and test data

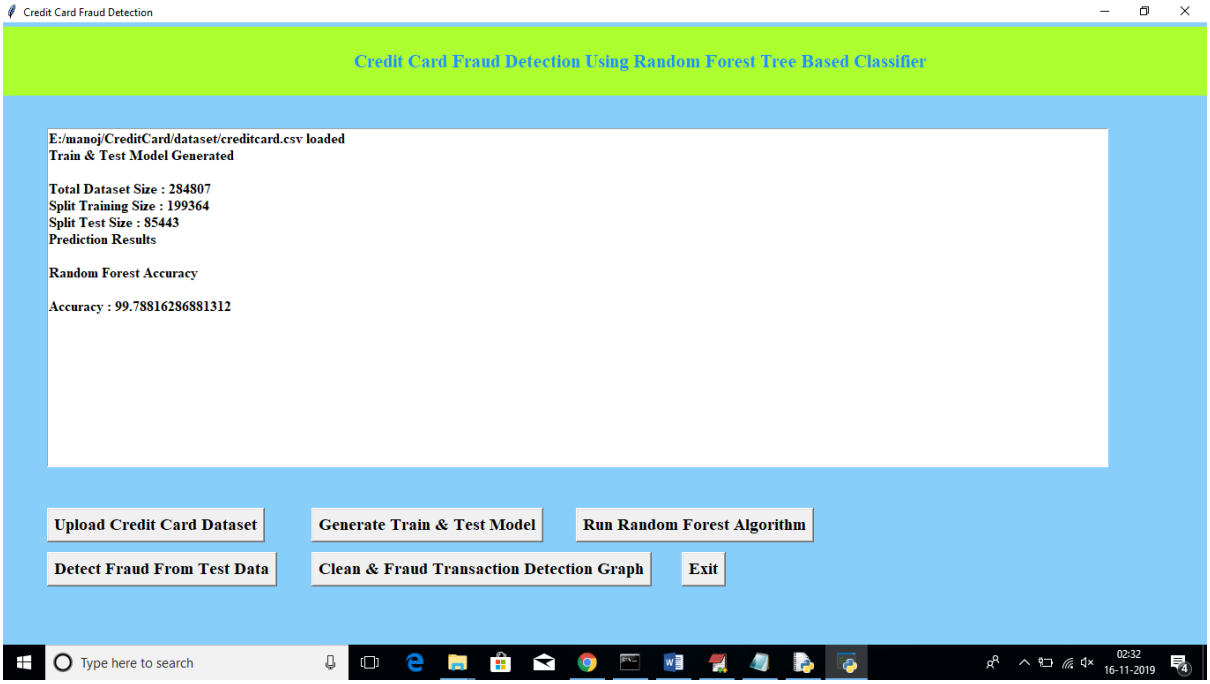


Fig 5: In above screen we can see Random Forest generate 99.78% percent accuracy while building model on train and test data. Now click on 'Detect Fraud From Test Data' button to upload test data and to predict whether test data contains normal or fraud transaction

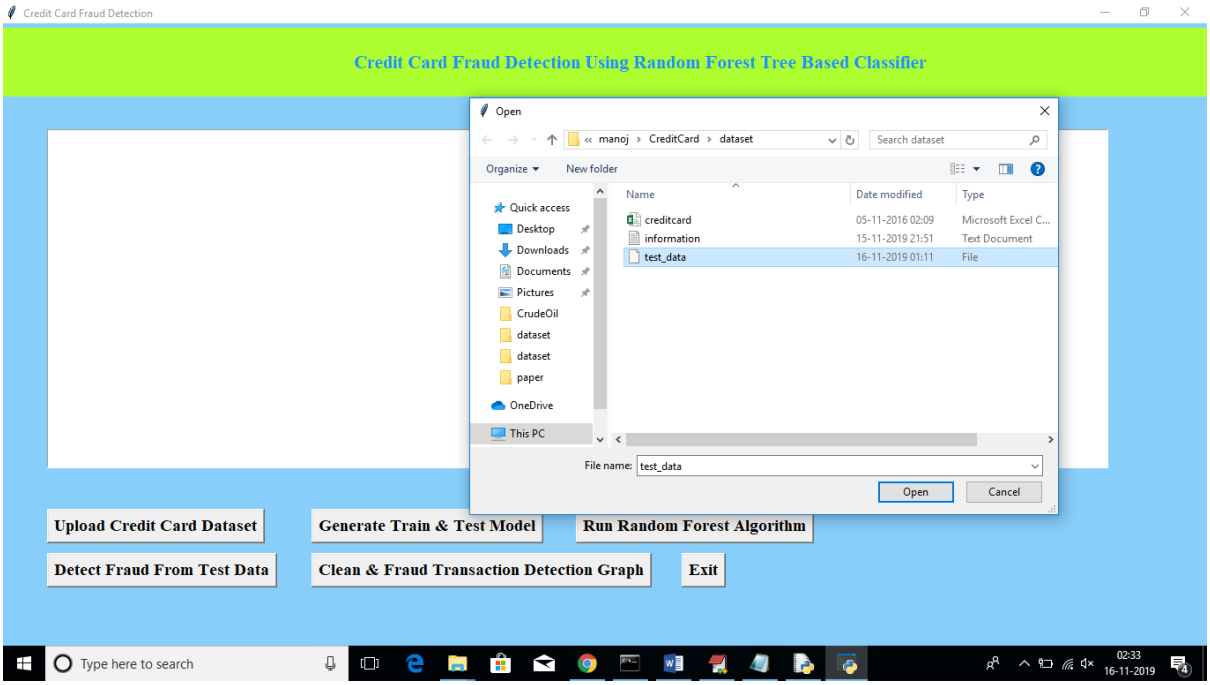


Fig 6:In above screen I am uploading test dataset and after uploading test data will get below prediction details

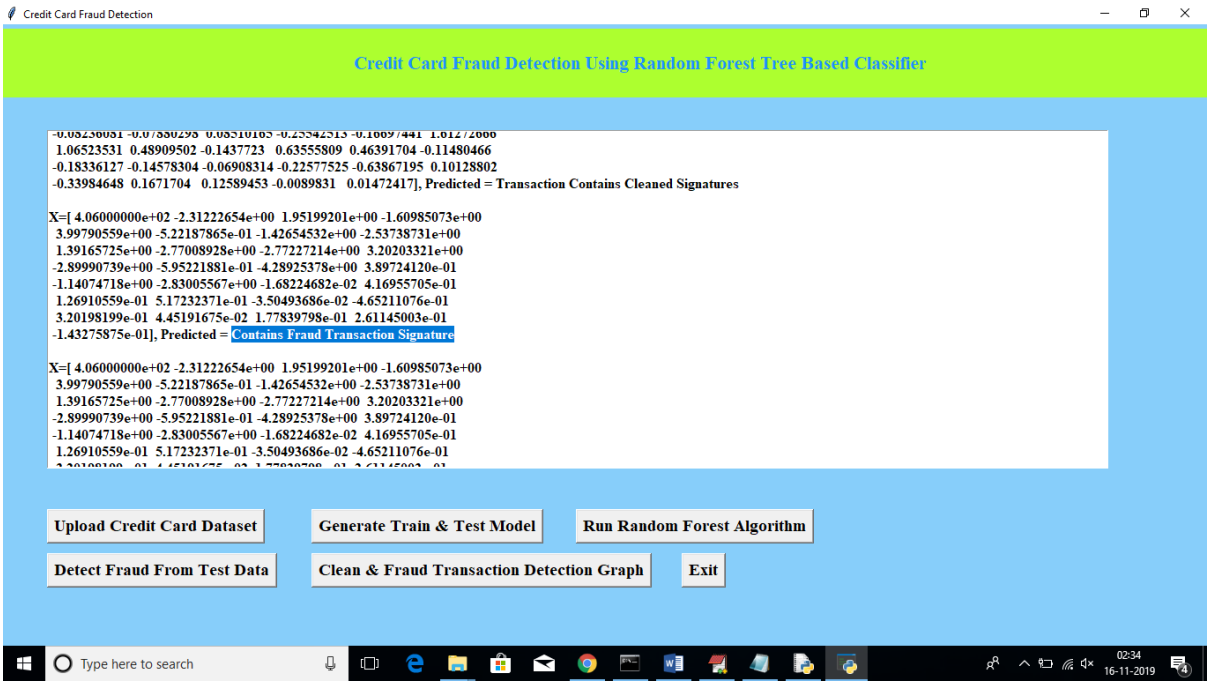


Fig 7:In above screen beside each test data application will display output as whether transaction contains cleaned or fraud signatures. Now click on 'Clean & Fraud Transaction

Detection Graph' button to see total test transaction with clean and fraud signature in graphical format. See below screen

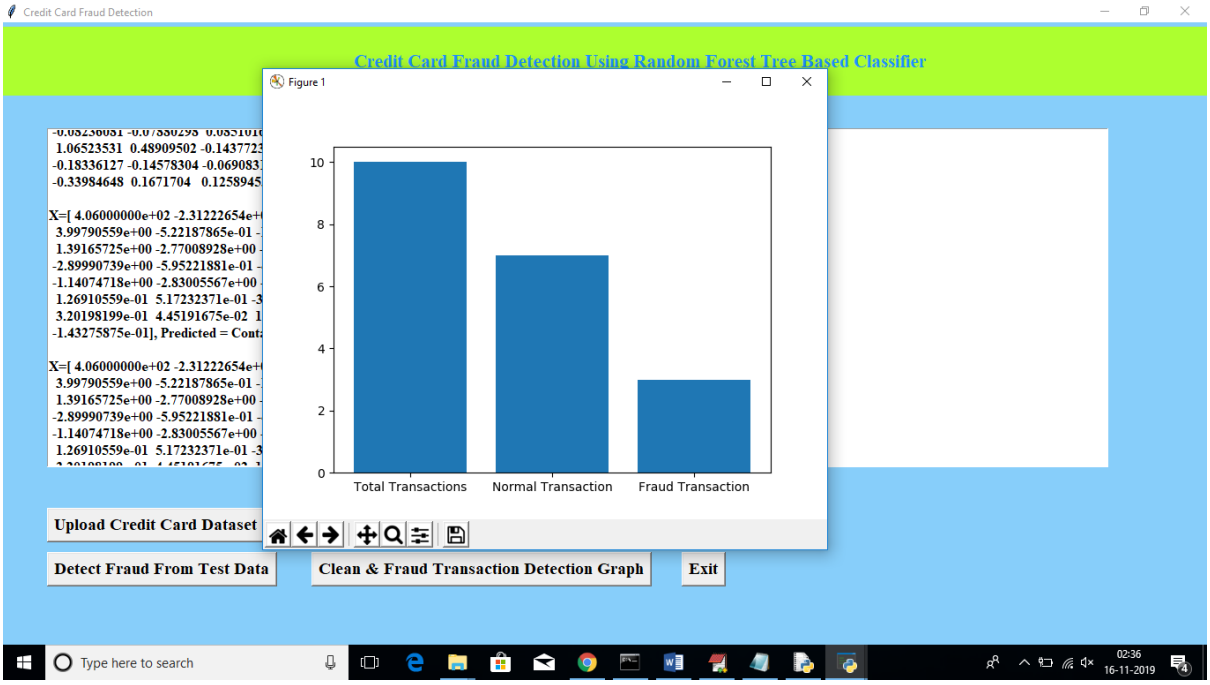


Fig 8:In above graph we can see total test data and number of normal and fraud transaction detected. In above graph x-axis represents type and y-axis represents count of clean and fraud transaction

6.CONCLUSION

In this project we have applied diff machine learning algorithms for detecting credit cards frauds.as per our analysis each algorithms working fine but accuracy wise there is slight variation. The decision tree will provide better performance with many training data, but speed during testing and application will still suffer. Usage of more pre-processing techniques would also assist. Our future work will try to represent this into a software application and provide a solution for credit card fraud using the new technologies like Machine Learning, Artificial Intelligence and Deep Learning.

REFERENCES

[1] Sudhamathy G: Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, vol. 8, no-5, pp. 1954-1966.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Sitaram patel, Sunita Gond , "Supervised Machine (SVM) Learning for Credit Card Fraud Detection, International of engineering trends and technology, vol. 8, no. -3, pp. 137-140, 2014. [7] Snehal Patil, Harshada Somavanshi, Jyoti Gaikwad, Amruta Deshmane, Rinku Badgujar," Credit Card Fraud Detection Using Decision Tree Induction Algorithm, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April- 2015, pg. 92-95

[8] Dahee Choi and Kyungho Lee, "Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System", vol. 5, no. - 4, December 2017, p. 12-24.