

Detecting Credit Card Frauds Using Different Machine Learning Algorithms

P.Visalakshi¹ K.V.Madhuvani², Sunilraja, S.R³

¹Asst.Professor, Dept. of Computer Science and Engineering,
SRM Institute of Science and Technology(Formerly SRM University),
Kattankulathur Campus, Chennai, South India.
email: visalakp@srmist.edu.in

^{2,3}Final year B.Tech, Dept. of Computer Science and Engineering,
SRM Institute of Science and Technology(Formerly SRM University),
Kattankulathur Campus, Chennai, South India.

²email:kk6872@srmist.edu.in

³email :sr8856@srmist.edu.in

ABSTRACT

The main focus of this research paper is to analyze the savings fraud credit card's identification in current era. Based on transaction the savings account fraud credit cards' detection is identified. Both online and offline savings fraud things are happening in online and offline transactions of accounts in the real world. But day by day the rate of happening of frauds are increased in an exponential manner. In order to detect the frauds in an online transaction a quite a number of techniques have been used in current era. Based on this, an exhaustive survey has been done and conclusion made. By considering the survey, we have proposed different machine learning algorithms for detecting the transactions fraudulent and the accuracy trusted records or transactions. We have used random forest, decision tree, SVM, Gaussian NBLogistic algorithms to find credit card fruds. The consequences acquired from processing the dataset provides the maximum accuracy as amble of above 90%.

Keywords

Support Vector Machine(SVM), NBLogistic algorithms(Naïve Bayes Logistic algorithms).

Introduction

There are various fraudulent activities detection techniques has implemented in credit card transactions have been kept in researchers' minds to develop new methods and models based on artificial intelligence, data mining, fuzzy logic and machine learning. Credit card fraud detection is significantly difficult, but also popular method to solve. In this proposed methodology, we developed fraud detection method for the credit card using Machine Learning(ML). There are many more advancements in "Machine Learning techniques. Machine learning has shown and act as eye opener for a successful measure for identification of frauds in online banking transactions. A large volume of data is transferred during online transaction processes, which results in a binary result such as genuine or fraudulent. In the sample of fraudulent datasets, the features are being constructed. These are called as data points namely the age and value of the customer account, as well as called as an initiative points of the credit card system.

There are millions of features and each contributes, to varying extents, towards the fraud probability. Particularly, the level of each feature contributes to the fraud score is generated by the artificial intelligence of the machine which is driven by the training set, but is not determined by a fraud analyst. So, in regards to the card fraud, if the use of cards

to commit fraud is proven to be high, the fraud weighting of a transaction that uses a credit card will be equally so. However, if this were to shrink, the contribution level would parallel. Simply make, these models self-learn without explicit programming such as with manual review. Credit card fraud detection using Machine learning is done by deploying the classification and regression algorithms”.

Literature Review

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

This paper focuses on the introduction of a scorecard from relevant contrast criterias, features, and talents of predictive analytics supplier choices in modern times being used to word credit score card fraud. The scorecard gives a side-by side 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]**Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar, “BLAST-SSAHA Hybridization for Credit Card Fraud Detection.**

This paper recommends to use two-stage sequence alignment in which a profile Analyzer(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 analyzer are subsequent surpassed on to a deviation analyzer (DA) for potential alignment with preceding fraudulent behavior. The ultimate resolution about the nature of a transaction is taken on the basis of the observations by using the usage of these two analyzers. 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]**Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree”, “Fraudulent Detection in Credit Card System Using SVM & Decision Tree 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.

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”.

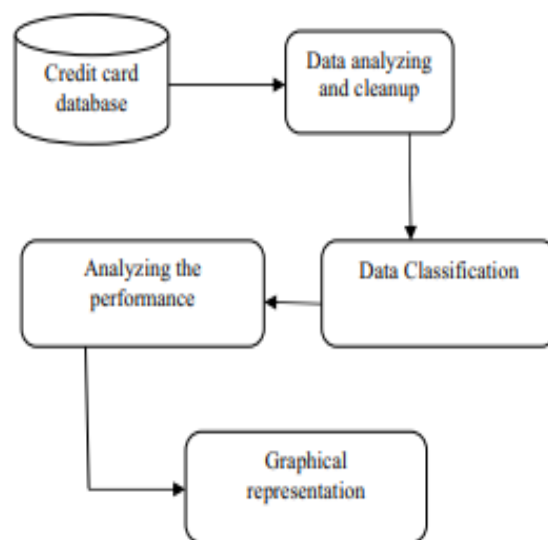


Figure 1.“Machine learning model to detect fraudulent credit card activities

To classify fraudulent and legitimate credit card transaction by supervised learning Algorithm. To help us to get awareness about the fraudulent and without loss of any financially”.

Modules

- **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.
- **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.
- **Preprocessing of dataset:** In this module the cleaned dataset will be preprocessed where the dataset will be divided based on amount and transaction time.
- **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.
- **Evaluation:** The resultant data obtained in the form of confusion matrix can be evaluated by using graphical representation which gives better accuracy.

ALGORITHMS

1) Random Forest Algorithm

Random forest 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 create 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 an 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 forest 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.

2) SVM

Input: “set of (input, output) training pair samples; call the input sample features $x_1, x_2 \dots x_n$ and the output result y . Typically, there can be lots of input features x_i .

Output: set of weights w (or w_i), one for each feature, whose linear combination predicts the value of y . (So far, just like neural nets...)

Difference: we use the optimization of maximizing the margin (street width) to reduce the number of weights that are nonzero to just a few that correspond to the important features that ‘matter’ in deciding the separating line(hyperplane)...these nonzero weights correspond to the support vectors (because they ‘support’ the separating hyperplane)

3) DECISION TREE

The decision tree Algorithm belongs to the family of supervised machine learning algorithms. It can be used for both a classification problem as well as for regression problem.

The goal of this algorithm is to create a model that predicts the value of a target variable, for which the decision tree uses the tree representation to solve the problem in which the leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

4) GAUSSIAN NB

We are going to implement the Naive Bayes classifier in Python using my favorite machine learning library scikit-learn. Next, we are going to use the trained Naive Bayes (supervised classification), model to predict the Census Income.

As we discussed the Bayes theorem in Naive Bayes classifier post. We hope you know the basics of the Bayes theorem. If not let’s quickly look at the basics of Bayes theorem once. Bayes’ theorem is based on conditional probability.

The conditional probability helps us to calculate the probability that is related to something will be happening at present.

5) LOGISTIC

Logistic Regression is a Machine Learning algorithm which is used for the classification problems. It is a predictive analysis algorithm and based on the concept of probability. We can call a Logistic Regression a Linear Regression model but the Logistic Regression uses a more complex cost function, this cost function can be defined as the ‘**Sigmoid function**’ or also known as the ‘logistic function’ instead of a linear function”

Results and Discussions

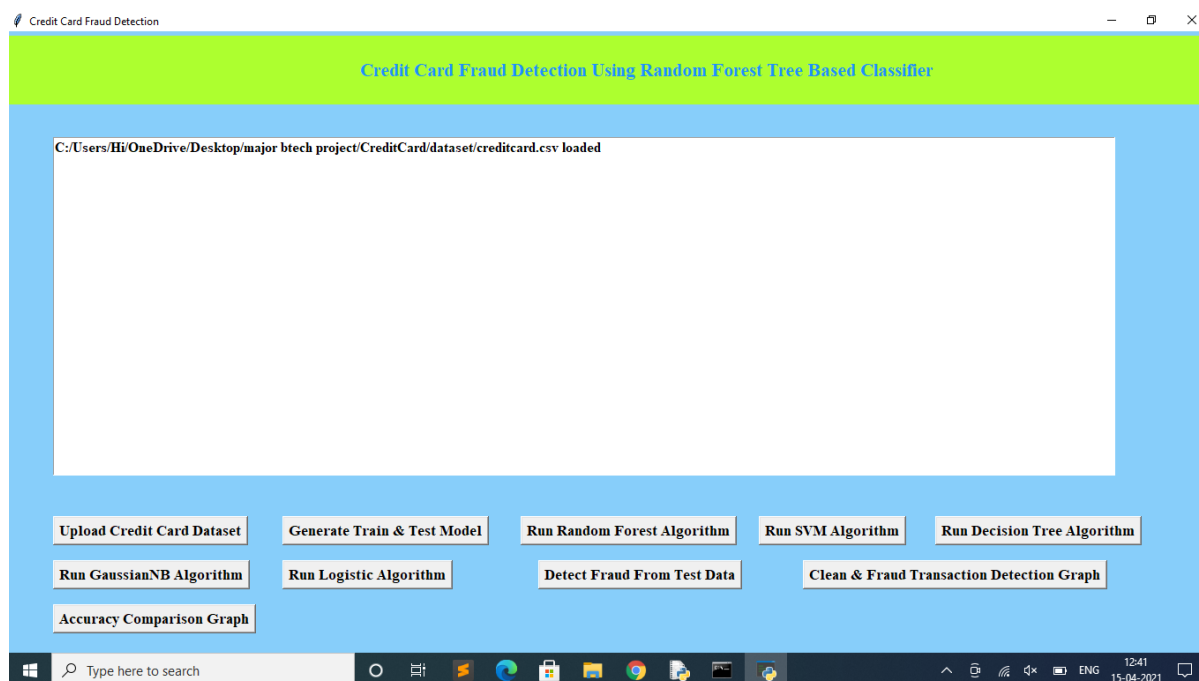


Fig 2. Home Screen

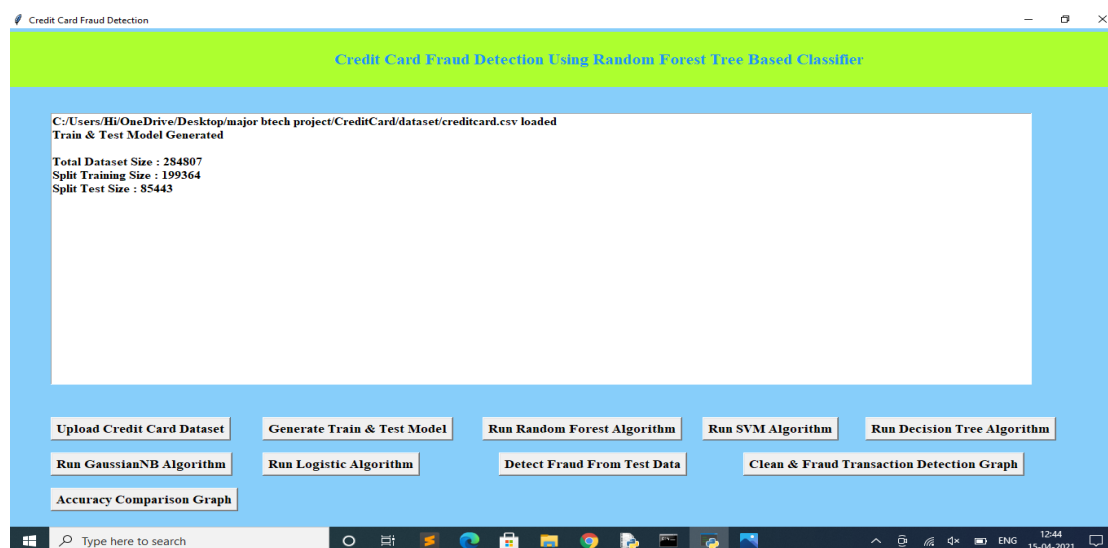


Fig 3. Generation of random forest model

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

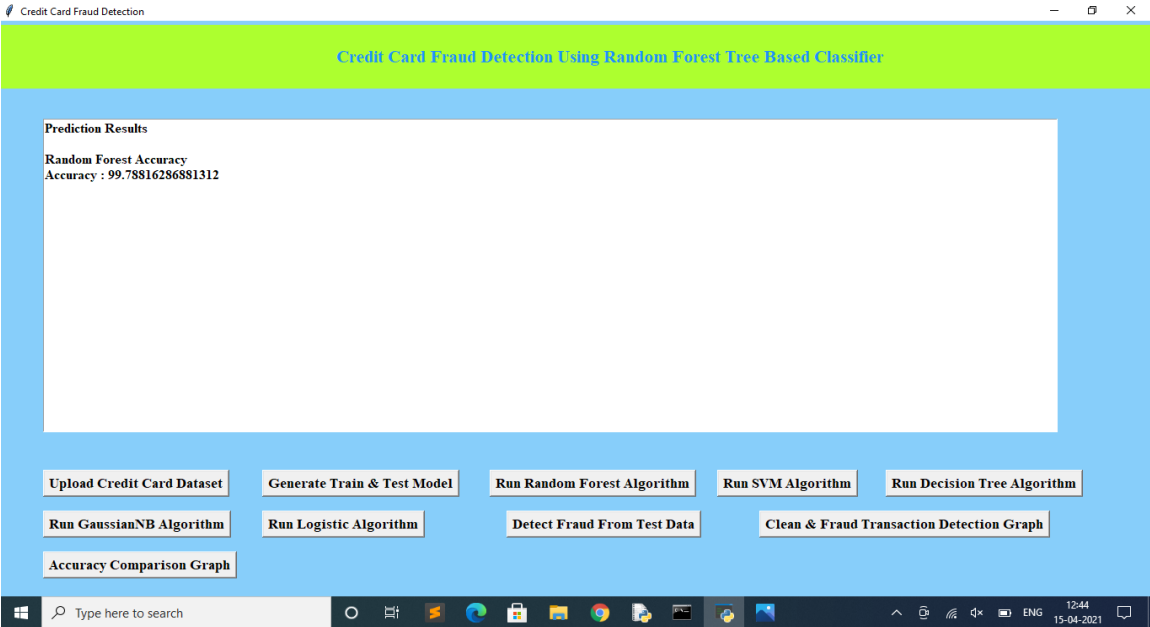


Fig 4. Accuracy building model

In above screen we can see Random Forest generate 99.78% percent accuracy while building model on train and test data.

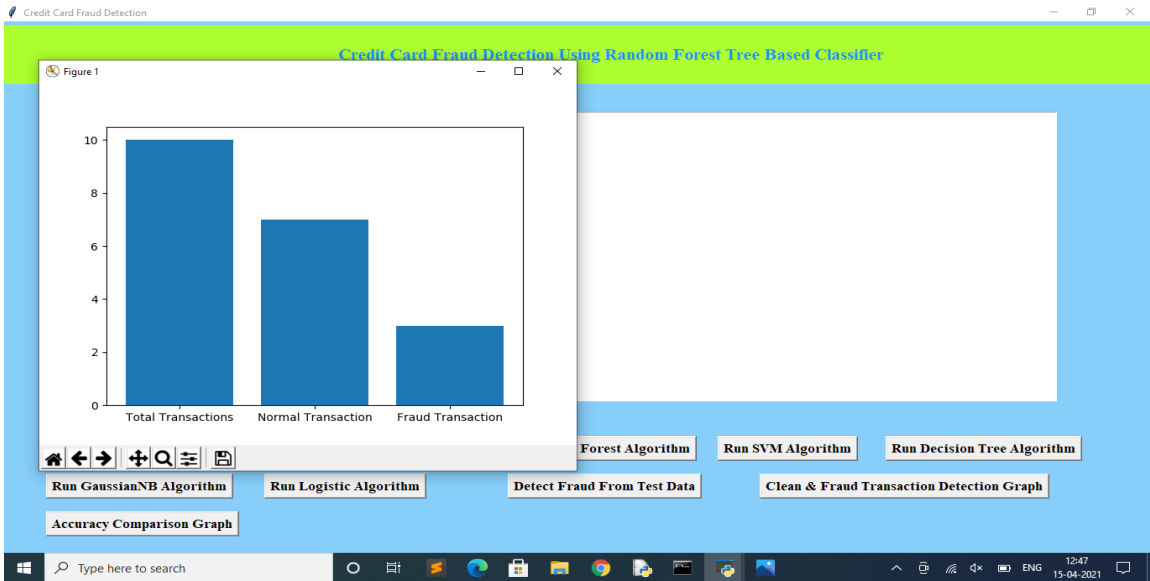


Fig 5. Detection of fraud transactions

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

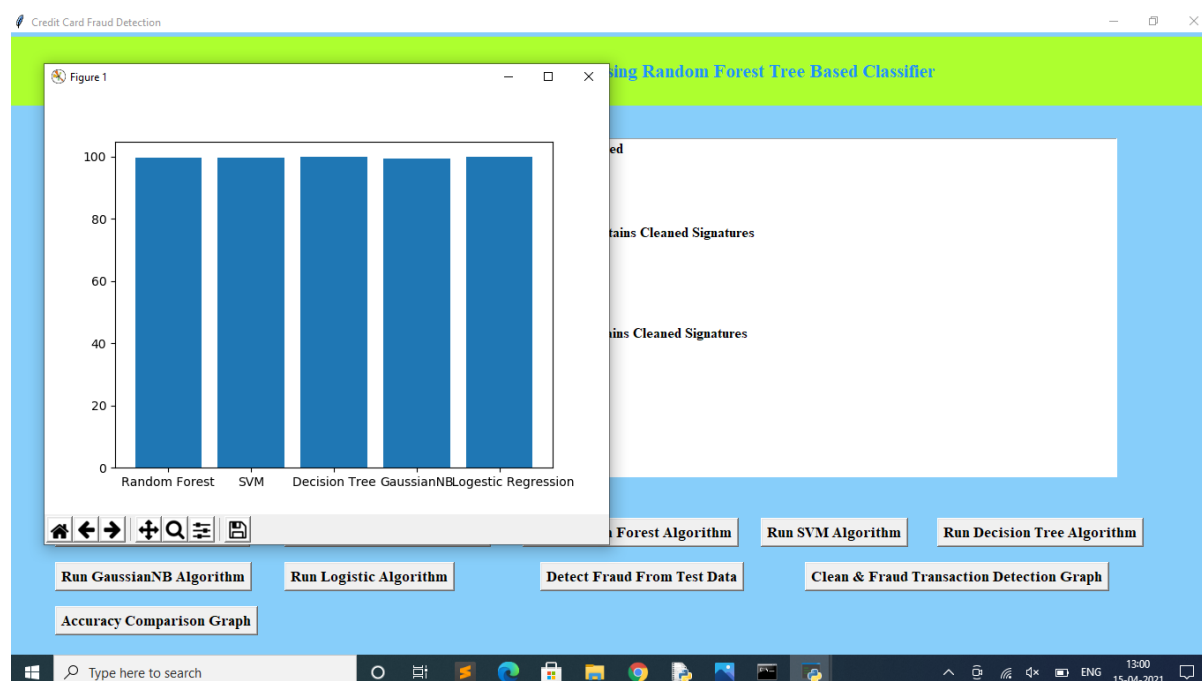


Fig 6. “Comparison of performances of algorithms in graph

Conclusion

In this project we have applied different machine learning algorithms like random forest, decision tree, SVM, Gaussian NB, Logistic for detecting credit cards frauds. As per our analysis each algorithm is 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 rural 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 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, pp. 12-24.