CREDIT CARD FRAUD DETECTION

Rachitha E[1], Rani H E[2], Prathiksha[3], Swathi B V[4],

[1] *Student, Information Science & Engineering, SJB Institute of Technology, Karnataka, India*
[2] *Student, Information Science & Engineering, SJB Institute of Technology, Karnataka, India*
[3] *Student, Information Science & Engineering, SJB Institute of Technology, Karnataka, India*
[4] *Assistant Professor, Information Science & Engineering, SJB Institute of Technology, Karnataka, India*

ABSTRACT

In recent days the use of credit card for all types of transactions has been increased drastically all over the globe which indicates the increase in fraudulent transaction. A large number of fraudulent transactions are made each and every second across the global where transactions can be determined using various modern techniques like data mining, machine learning and few others which have been used recently. This paper uses various genetic algorithms for finding accurate solutions for the modern day problems like detecting the unusual transactions. As the number of online transactions are increasing day by day the responsibilities of the banks to increase the security to these transactions. Our proposed system focuses on series of machine learning models where we select the best method among these available algorithms.

Keywords*: machine learning, Data mining*

1. INTRODUCTION

A credit card is a thin card which is made of plastic and is handy, it contains information such as signature are picture for the identification purpose. It gives the authorization to the person to whom it belongs, for the purpose of purchasing and charge services to his /her account (charges for which he will be billed periodically). The information on the credit card can be read by automated teller machines, store readers ,bank and can also be used in online banking systems. The security of the credit card depends on the physical security of the plastic card as well as the privacy of the credit card number.

A fraud is detected when one individual uses the other individual’s card for their own use without intimating the owner about the transaction. During this kind of transactions, the fraudsters can use the card until its available time limit is depleted. Therefore, a solution which reduces the total available limit on the credit card should be found.

This type of fraud is done when any person with pure intensions to defraud uses card of unknown which has been lost, stolen, cancelled or revoked and misuses which results in fraud. Using number of credit card without having actual card can be also known as credit card fraud. Identity theft also have been increasing and have contributed to fraudulent transaction. This effects consumer credit industry as it has become one of the fastest growing type and which is most complicated and difficult in solving.

2. PROPOSED SYSTEM

This proposed system uses the PCA which means principal component analysis where these principal components are computed which are further used to perform the change of the basis on the data available. In our paper, we have opposed PCA on the dataset for cleaning the data and for reducing the data.

During the data pre-processing, data will be cleaned, integrated, transformed and reduced to the required form. The missing data is handled by applying data handling techniques. The number of fraudulent transactions is always less when compared to the number of genuine transactions. To overcome this, we have used Random over-sampling techniques where the minority occurrences are raised. This pre-processed data set will be trained first and then tested.

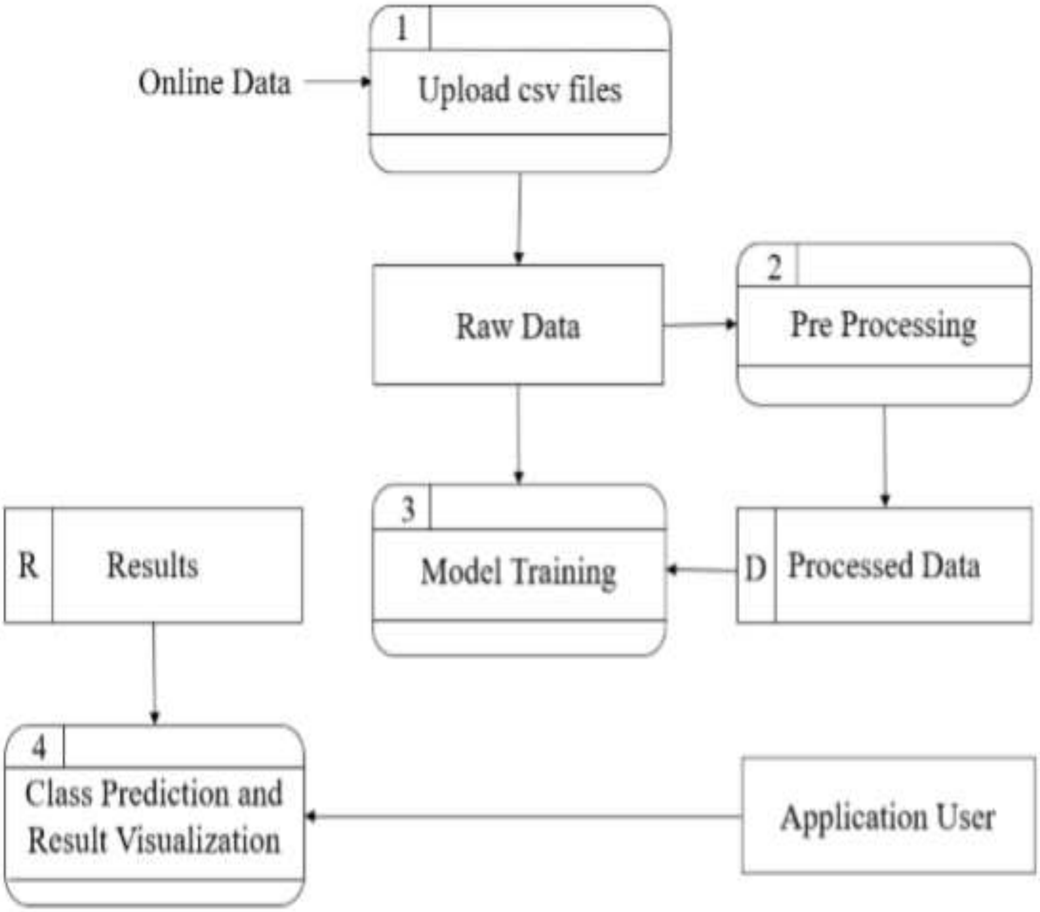


Fig -1: Flow diagram

2.1 Random Forest Algorithm

Random forest algorithm for finding the fraudulent transactions and the accuracy of these transaction and it is based on supervised learning algorithm where it uses decision trees for classification of the dataset and also handle the missing values.

2.2 Logistic Regression

It is a supervised learning used for visualization purpose also mainly used for classification of data and predict the target variable. Even though, it is named as regression this algorithm is used for classification purpose.

2.3 Naïve Bayes Algorithm

The naïve Bayes classifier is the probability-based classifier for the fraud detection. Here, the target classes probability and test case probability are calculated. Naïve **B**ayes classifier is based on applying the Bayes theorem with strong assumptions between the features and the different parameters that are computed such as, precession, recall and sensitivity.

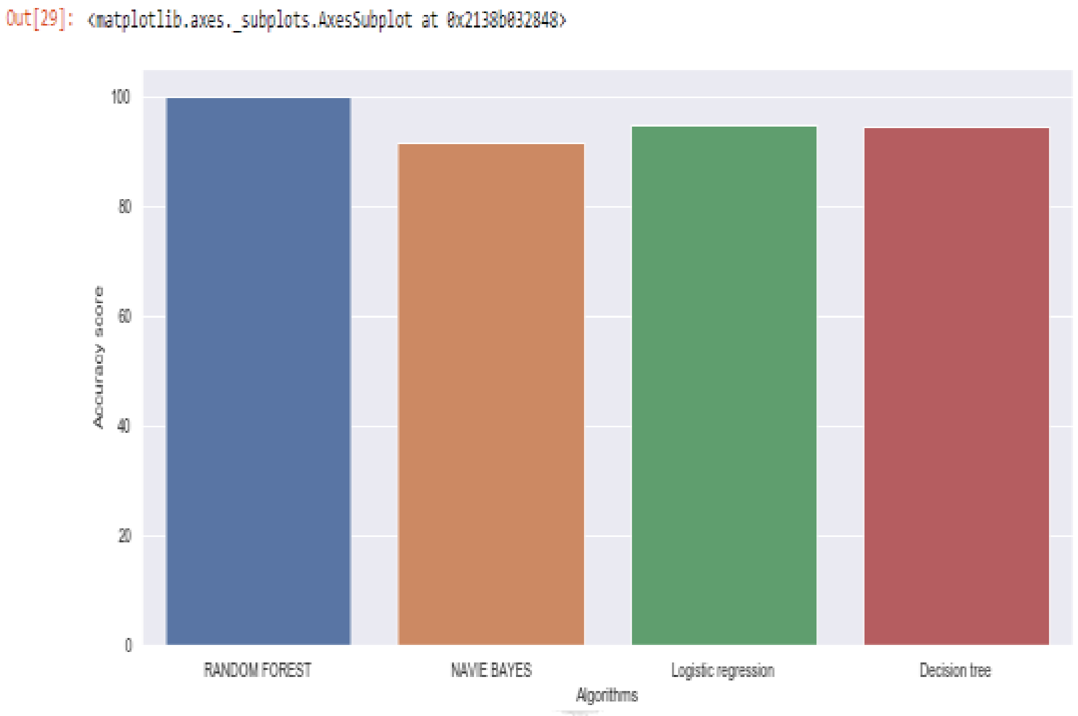
2.4 Decision Algorithm

Decision tree is a supervised learning technique. In this paper, it has been used for classifying of the data. It is used to create a training model, which is then used to predict the value or the class of target variable while learning simple decision rules which are taken from prior data.

3. RESULTS

In credit card fraud detection, after using random forest, naïve bayes, logistic regression, decision tree algorithm, the most accuracy seen is the random forest with 99.6% of accuracy which was the best one when compared to the accuracy of the rest followed by the least accuracy with 97.2% which is of naïve bayes.

Over all with the practical accuracy results found random forest suited the best with best accuracy compared to another algorithm**.**



```
Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0x2138b032848>
```

Fig - 2: Graphical Representation of Algorithms

Considering the obtained results, Random forest algorithm can be considered are the best algorithm which gives the best accuracy. Hence this algorithm can be used.

4. CONCLUSION AND FUTURE ENHANCEMENT

One of the common online criminal activities are the credit card fraudulent activities. This paper helps in the detection of fraudulent activities happening during online credit card transactions. here we have used random forest algorithm for the detection of fraudulent transactions. The random forest algorithm gives the accuracy around 99.6%, The below figure shows the graphical representation of the accuracies obtained when the following algorithms are applied to the dataset.

5. REFERENCES

1] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and G. N. Surname, **“**Random forest for credit card fraud detection**”,** IEEE 15th International Conference on Networking, Sensing and Control (ICNSC),2018.

[2] Dilip Singh Sisodia, Nerella Keerthana Reddy, Shivangi Bhandari, “Performance Evaluation of Class Balancing Techniques for Credit Card Fraud Detection” IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017).

[3] SamanehSorournejad , Zahra Zojaji , Reza Ebrahimi Atani , Amir Hassan Monadjemi, “A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective”,IEEE 2016

[4] E. Michael and S. Pedro, “A survey of signature-based methods for financial fraud detection,” *Computer and security*, vol. vol 28, no. 6, pp. 381–394.

[5] B. Adrian, “Detecting and Preventing Fraud with Data Analytics,” *Procedia Economics and Finance*, vol. 32, no. 15, pp. 1827–1836, 2015.

[6] H. He and E. A. Garcia, “Learning from Imbalanced Data,” *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[7] B. Zhu, B. Baesens, and K. L. M. Seppe, “An empirical comparison of techniques for the class imbalance problem in churn prediction,” *Information Sciences*, vol. 408, pp. 84–99, 2017.

[8]N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[9] G. E. a. P. a. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017)

[10] Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020). *Credit Card Fraud Detection Using Machine Learning. 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS).*