IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Fraud Identification of Credit card using Machine Learning

Nayana B M [1], Namratha M N [2], K Vishnuvardhan [3], Ch S V Gani Prabhakar [4], Rekha Jayaram [5]

[1,2,3,4] Students, Department of Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India

[5]Assistant Professor, Department of Information Science and Engineering, Dayananda Sagar College of Engineering Bangalore, Karnataka, India

Abstract: Online Transaction is a popular mode of payment and most of the payments done by using credit card but credit card fraud happens continuously and it leads to huge lose. So all the banks and other financial industries supports to the progress of Fraud identification of credit card. The fraudulent transactions can takes place in different ways, identification of credit card fraud is very important to the credit card companies so that customers will not be charged for the items those are not purchased. This paper explained the concepts related to credit card and used different machine learning algorithms like logistic regression and random forest method .2,84,808 credit card transactions of dataset is collected from kaggle and these transactions are from European bank. Identification of credit card is an example of classification and in this process includes analyzing and preprocessing the data set along with the deployment of Machine learning algorithms.

Index Terms **-** Credit card fraud, logistic regression, random forest

1.INTRODUCTION

Credit card fraud is refers to misuse of someone's account without the owner being aware of it. Analyzing and studying these fraud transactions, prevention methods needs to be taken against these fraud transactions so that it'll helps to avoid similar situations in future.

The scenario of credit card fraud, the fraudster make use of other person's credit card for personal uses without taking permission from owner of the card. In 2018 Credit card fraud losses in London estimates 844.8 million US Dollar. To stop these fraudulent transaction either prevention or detection of credit card must be done.

These credit card fraud transactions can be takes place in two types: one is robbing a physical card and the other one is stealing some important information of the credit card like CVV, Card Number, Expire Date, Type of card and other.

The main aim of this paper is to analyze and preprocess the data by using Machine learning algorithms and find out which algorithm is best to identify credit card fraud.

The rest of the paper describes as follows: section 2 consists of related work regarding credit card system, section 3 consists of system framework, section 4 consists of Methodology, section 5 consists of results and discussions and in the next section it consists of conclusion and then references.

2.RELATED WORK

A survey of various papers which discussed about many Machine learning algorithms to find out the best one to detect credit card fraud were analyzed as below:

[1] Lakshmi S V S S, Selvani Deepthi Kavila used three Machine learning algorithms i.e., Logistic regression, decision tree, Random forest on dataset and the work is implemented using R language and the performance was measured using sensitivity, specificity, accuracy, error rate. Conclude that random forest is better compared to rest of techniques.

[2] S P Maniraj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed make use of local outlier factor and isolation forest algorithm on PCA transformed data set.

[3] Kaithekuzhical Leena Kurien & Dr. Ajeet Chikkamannur has did the research on credit card ,where SMOTE technique is used to balance the dataset and later they used logistic regression and random forest method, considering Precision ,F1 score, recall and ROC curve as performance metrics. Concluded the limitation that when there are multiple trees in the forest the algorithm works slowly for random forest method.

[4] Andhavarapu Bhanusri, K.Ratna Sree Valli , P.Jyothi , G.Varun Sai , R.Rohith Sai Subash proposed a system in which author performs sampling on dataset and test against logistic regression, Random forest, naïve bayes, AdaBoost method and used support, accuracy, recall as performance measures.

[5] Varun Kumar K S, Vijaya Kumar V G, Vijay Shankar A, Pratibha K developed a system where author applies a SMOTE technique, data sets were tested against different supervised machine learning algorithms like Decision Trees, Naive Bayes

Classification, Least Squares Regression, Logistic Regression and SVM are used to detect fraudulent transactions in real-time datasets.

[6] Navanshu Khare and Saad Yunus Sait used random forest, decision tree, SVM, Logistic regression. These techniques directly applied on raw data set and results were evaluated using accuracy, sensitivity, specificity, precision. Concluded that random forest is best.

[7] Mr.Manohar.s, Arvind Bedi, Shashank kumar, Shounak kr Singh, author states that after preprocessing the data tested against random forest, decision tree and SVM methods and concludes that SVM is best in terms of accuracy but with low precision values.

3.PROPOSED TECHNIQUE

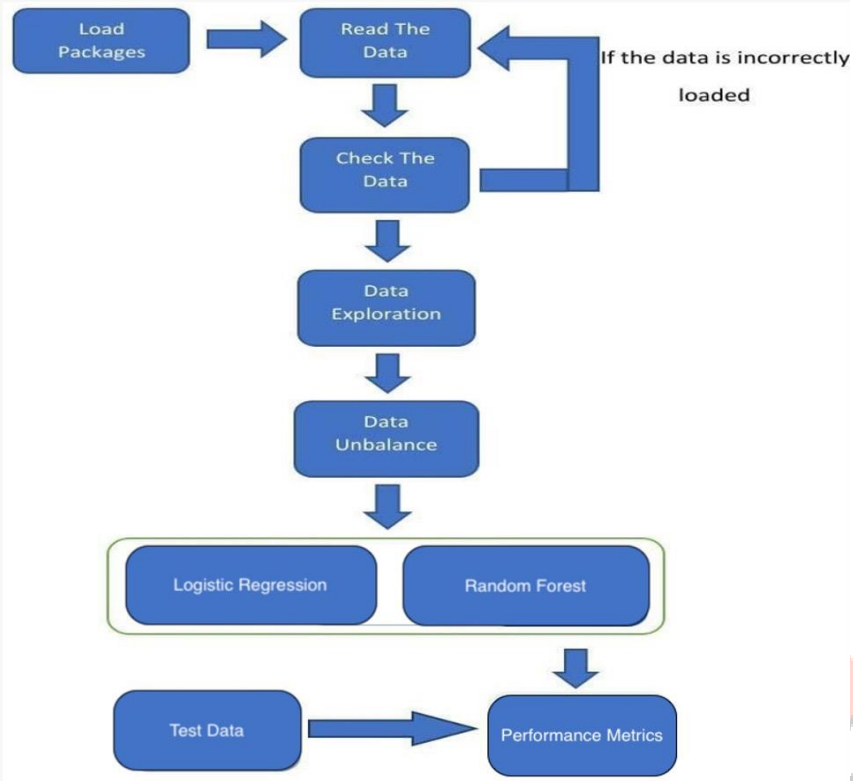

Fig-1:System architecture

Above figure(1) demonstrates the key steps that has to be followed to develop a proposed model. After balancing the data set using SMOTE technique the feature extraction will be performed later the output of the training data will be tested against the test data.

```
In [3]: data.head(10)
Out[3]:
```

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.167
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.327
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.647
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.206
5	2.0	-0.425966	0.960523	1.141109	-0.168252	0.420987	-0.029728	0.476201	0.260314	-0.568671	...	-0.208254	-0.559825	-0.026398	-0.371427	-0.232
6	4.0	1.229658	0.141004	0.045371	1.202613	0.191881	0.272708	-0.005159	0.081213	0.464960	...	-0.167716	-0.270710	-0.154104	-0.780055	0.750
7	7.0	-0.644269	1.417964	1.074380	-0.492199	0.948934	0.428118	1.120631	-3.807864	0.615375	...	1.943465	-1.015455	0.057504	-0.649709	-0.415
8	7.0	-0.894286	0.286157	-0.113192	-0.271526	2.669599	3.721818	0.370145	0.851084	-0.392048	...	-0.073425	-0.268092	-0.204233	1.011592	0.373
9	9.0	-0.338262	1.119593	1.044367	-0.222187	0.499361	-0.246761	0.651583	0.069539	-0.736727	...	-0.246914	-0.633753	-0.120794	-0.385050	-0.069

10 rows × 31 columns

Fig-2 : Credit card transaction dataset

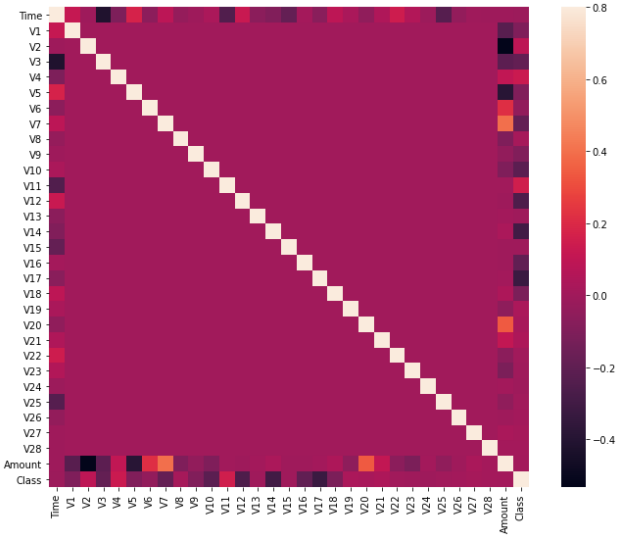

Fig-3: Heat map

In the way to generating the heatmap, first train the data set using different Machine learning algorithms. In this process extract the dataset and find the patterns between the fraudulent transactions. So using these patterns between the normal and fraudulent transactions, it'll be easy to identify the future fraudulent transactions.

Below algorithms are used in this paper to identify the fraudulent transactions of credit card:

Random Forest:

It is one of the tree based algorithm and also it is an algorithm for both regression and classification problems. Random forest consisting of no. of trees and produces the output by combining the output of all the trees. The process of combining the trees is called as ensemble method. The output of each decision trees will be merged to get accurate results.

It also performs row sampling.

Logistic Regression:

It is one of the supervised classification algorithm. The outcome of logistic regression probability has two values either yes or no, 1 or 0, true or false.

The Logistic Regression probability predict and returns the dependent variable, that is predicted from the independent variable.

Logistic regression is very much similar to linear regression but the output of linear regression will be straight line where as logistic regression produces the curve.

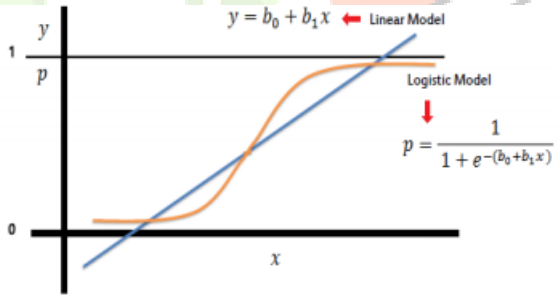

Fig-4: Logistic curve

5.RESULTS AND DISCUSSIONS

Figure 5 is the confusion matrix for random forest, where it shows the no. of genuine transactions and fraud transactions. This Confusion matrix plot between true class and predicted class.

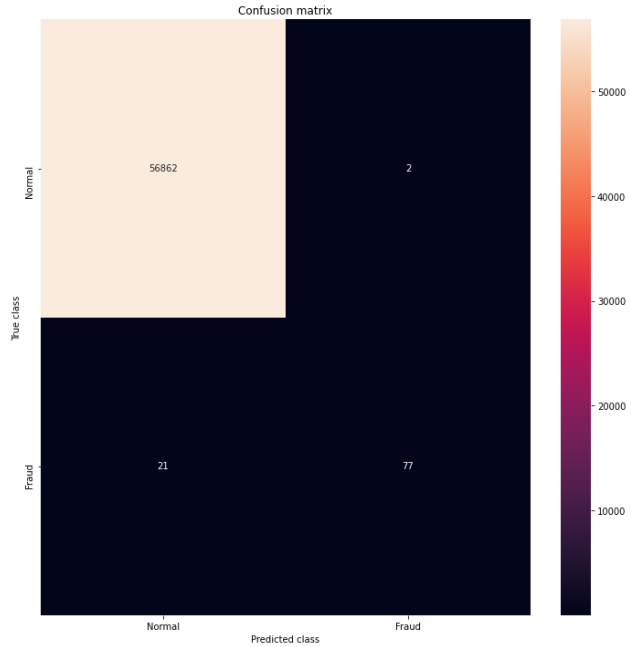

Fig-5: Confusion Matrix for random forest

The below figure represents Confusion Matrix for logistic regression. It produces the correct transactions and the incorrect ones also produces the true positives, true negatives, false positives and false negatives.

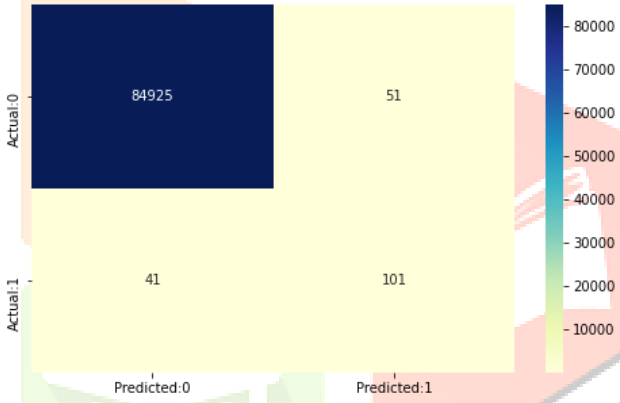

Fig-6: Confusion Matrix for logistic regression

Confusion matrix gives the visualization of the results obtained from methods.

Below table represents the results of different Machine learning algorithms. In this simulation process accuracy, precision, F1-score and recall were considered as performance measures.
Below table concludes that Random forest is the best Machine learning Algorithm to find the fraudulent Transactions.

Table-1: Accuracy, Precision, recall, F1-score comparison for different ML methods.

	Accuracy	Precision	recall	F1-score
Random Forest	99.9	97.4	78.6	87
Logistic Regression	99.8	75	68	71

6. CONCLUSION

Credit card fraud transactions are increasing day by day, because of this many peoples were losing their money so that companies are investing money to find the suitable method to find out the fraudulent transactions.

From the experiments which are demonstrated in this paper results that random forest produces the best accuracy based on the performance measures.

7. REFERENCES

[1] Lakshmi S V S S, Selvani Deepthi Kavila,"Machine Learning For Credit Card Fraud Detection System",International Journal of Applied Engineering Research, Vol 13, pp. 16819-16824, November 24 (2018).

[2] S P Maniraj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed "Credit Card Fraud Detection Using Machine Learning and Data Science",International Journal of Engineering Research & Technology,Vol. 8, September 09 (2019).
[3] Kaithekuzhical Leena Kurien & Dr. Ajeet Chikkamannur , "Detection And Prediction Of Credit Card Fraud Transactions Using Machine Learning", International Journal of Engineering Sciences &Research Technology, March (2019).
[4] Andhavarapu Bhanusri, K.Ratna Sree Valli , P.Jyothi , G.Varun Sai , R.Rohith Sai Subash "Credit card fraud detection using Machine learning algorithms", Quest Journals-Journal of Research in Humanities and Social Science, Volume 8, pp: 04-11, 2 (2020).
[5] Varun Kumar K S, Vijaya Kumar V G, Vijay Shankar A, Pratibha K, "Credit Card Fraud Detection using Machine Learning Algorithms", International Journal of Engineering Research & Technology, Vol. 9, July 7(2019).
[6] [1]Navanshu Khare and [2]Saad Yunus Sait , "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models", International Journal of Pure and Applied Mathematics,Vol 118, (2018).
[7] Mr.Manohar.s, Arvind Bedi, Shashank kumar, Shounak kr Singh, "Fraud Detection in Credit Card using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET),Vol 07, April (2020).
[8] Vaishnavi Nath Dornadula, Geetha S, "Credit Card Fraud Detection using Machine Learning Algorithms", International Conference on recent trends in advanced computing , (2019).
[9] Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, "Credit Card Fraud Detection-Machine Learning methods", 18th International Symposium INFOTEH-JAHORINA, 20-22 March (2019).