# Machine learning algorithms for Credit card fraud prevention and detection

**[1]Ramakant Ganjeshwar, [2]Dr. Partha Roy, [3]Prof. D.P.Mishra**

[1]M.Tech Scholar, [2]Associate Professor, [3]Associate Professor
Computer Science and Engineering,
Bhilai institute of Technology, Durg, India

***Abstract*: Now a days, the problem of credit card fraud is constantly increasing. All these problems are constantly happening due to rapid increase in the payment process. Credit card fraud occurs when someone's credit card is lost or in the hands of an unknown person. And they use the found credit card in a fake way .Nowadays many people are struggling with these kinds of problems. So this project is designed to avoid credit card frauds. This is the application of data sciences. The main focus of this project is to use machine learning algorithms. In this we will use Isolation Forest and Local Outlier Factor Algorithms. The results are in these algorithms is based on accuracy, precision, recall and F1 score.**

**Index Terms: Credit card fraudulent, Application of data science, isolation forest, Outlier Factor**

## I. INTRODUCTION

Credit card fraud is increasing day by day with the fraud increasing administrative department, agencies, banking industries and many other offices. Because the utility of the internet is constantly increasing these days For all these reasons, credit card fraud is increasing day by day. But the reason for credit card fraud is not just online payment but also offline payment process. Results are not accurate due to using data mining method of credit card fraud detection Machine learning algorithms have to be used to avoid losses from credit card fraud .This is a card issued by the bank, it means that you can borrow money, after some time, that money has to be returned with interest. Credit card fraud means that another person can use the card without permission. Using this fraud detection you will find out the false and correct payment. It is not necessary that only credit card fraud will be fraud. In banking sector, debit card can also be fraudulent. Only the card is the medium for fraud of fraud transaction It is a crime because misuse of one's funds is done without permission. It is easy to fraud with credit card, you earn more money in less time, this project is being made to avoid these antics. Nowadays a lot of credit card frauds are coming up using the internet. There are a lot of frauds in this commerce platform, people are not even able to know.

## II. RELATED WORKS

.In 2020, Ruttalasilusha,V. Gnaneshwar,R.Ramesh, G. Ramakoteshwara Rao have Researched the techniques of credit card fraud detection they use the Random forest algorithms and adaboost algorithms[1]. In Random forest algorithms randomly select sample data and create the decision tree for classifying fraud or non-fraud dataset majority vote[1] is performed and the decision tree for classifying fraud or non-fraud dataset and majority vote is performed and the decision trees may Result fraud and non-fraud case and finally finding the Accuracy, Precision, recall, and F1 Score for the dataset and prepare the comparison result and proof the random forest algorithms is best for the fraud detection.

In 2020 Darshankaur and Shubpreetkaur implement the Machine learning approach their main Aim of the work is merit and constraints of these approach here all merit and models are combined for constructing a strong model this implement recommends all the necessary input and output they are used various steps for card fraud detection dataset from UCI repository and implement the cross validation technique for dividing the data set into train set and test set they implement tae KNN classification[2] model for predicting the test set and implement the naïve-bayes classification there all parameter are use for Analyze the performance of hybrid model.

In 2020 S.Abhnyaa, H. Sangeetha, R.A. Kartikayan, K.Saran, shriram, D.piyush they used the Random forest algorithm [3] for classifying the the credit card fraud dataset they find the advantage over the choice tree they adjust the propensity to over fit to there set of preparation these are evaluated Randomly each node at that point parts of element that are chosen Random subset of the full list of capabilities to prepare each individual trees and choice tree is constructed in case it is incredibly fast prepare huge information the information freely prepared for other in here all credit card transaction they applies various technique for detecting fraud the researcher develop the models based on artificial intelligence, data mining fuzzy logic and machine learning they use Supervised machine learning like Random forest Algorithm such as Random Forest Algorithm to detect online or offline fraud card transaction.

In 2020 Hassan Najadat , olaAltiti, Ayah Abu Aqoulen and Mutazyounes used different Machine learning models including models logistic regression, Decision Tree, Adaboosting, Naïve base, voting and Random forest Algorithm on the credit card fraud detection dataset they perform the three techniques for imbalanced data for SMOTE Technique, Random Over Sampling, Random under Sampling they also used deep learning algorithms on the .credit card fraud detection dataset they perform the three techniques for imbalance data for SMOTE technique random aver sampling and random under sampling they are used deep learning algorithm for the dataset they tested different deep learning model these experiment perform BiLSTM-Maxpooling-BiGRUMaxpooling is best one for deep learning algorithm they built the BiLSTM and BiGRU models above model is two input Model which are numerical and categorical feature the model embedding for categorical feature and 0.1 embedding for output dropped to avoid over fitting by using the spatial dropout as well as the output passed in BiLTM and BiGRU layers simultaneously where global max-pooling applied to both the model and most feature of the output Combined together and find the performance evolution based on are under ROC curve and calculate the result.[4]

2020 B.Nitin, RohitRavula, ShaikGangina sultana have implementing mastercard fraud by using Adaptive boosting Technique there are many machine learning models but they use Adaboosting[5] are a few things which increases the accuracy and other Parameter and performance for the model will be} Meta Algorithm it can be utilized in conjunction with many other sorts of learning Algorithm to enhance performance the output of other learning algorithm is combined into a waited sum the ultimate output if the boosted classifier. Adaboost is sensitive to noisy data and outlier final model may be proven to converge to a stronger model they visualize the information by using the Machine learning Technique.

In 2020 AndhavarapuBhansuri, K.Ratnasreevalli, P.Jyothi, G. VarunSai, R.RohitSaiSubash they take the dataset from kaggle is the credit card fraud dataset and perform the sampling on the dataset they divide the 70% data in training dataset and 30% data in testing dataset and perform the various machine learning algorithm like naïve bayes, logistic Regression, Random forest and adaboost Algorithms and finding the outcome of test dataset and calculate the performance like Accuracy, Precision, Recall, F1 score[6].

In 2019 S P Maniraj, Aditya Saini, Swarna Deep Sarkar they Obtained the Credit Card fraud detection they plot the different graph to check for inconsistencies in the dataset and to visually Comprehend it they visualize the number of fraud transaction is much lower than the legitimate once for counting the fraudulent Transaction is Much lower than the legitimate ones for counting the fraudulent transaction and non-fraud transaction and they represent the graph for Distribution time feature and show the graph the time of which transaction were done within two days. It can be seen that the least numb of transaction and show the graph the time of which transaction were done within two days. It can be seen that the least number of transaction and show the graph for Distribution of Monetary Value feature they plot the graph for amount class representation for the amount transaction for maximum time and plotting the heat Map for finding correlation Matrix and once anomalies are detected the system can be used for report them to the concerned Authorities for Testing purpose comparing the output for finding accuracy, Precision, Recall, F1-Score etc.[7]

In 2019 VaishnaviNathDornadula, Geetha s Research on credit card transaction are always unfamiliar with previous transaction made for the customer this unfamiliarity is very difficult problem in real world. Here they use the clustering method to divide the card holder into different different cluster based on theire transaction and using the sliding window method and finding the cardholder behavioral pattern of transaction and followed by the average amount of time.[8]

In 2018 KuldeepRandhawa,CHUKiongloo,Manjeevanseera, cheepenglim, ashoke k. nandi al planned a method mistreatment machine learning to sight mastercard fraud detection. Initially, commonplace models were used at the moment hybrid models came into image that created use of AdaBoost and majority balloting strategies. in public obtainable knowledge set had been accustomed appraise the model potency and another knowledge set used from the institution and analyzed the fraud. Then the noise was value-added to the information sample through that the hardiness of the algorithms can be measured. The experiments were conducted on the idea of the theoretical results that show that the bulk of balloting strategies smarts} good accuracy rates so as to sight the fraud within the credit cards. For more analysis of the hybrid models noise of regarding 100 percent and half-hour has been value-added to the sample knowledge. many balloting strategies have achieved a decent score of zero.942 for half-hour value-added noise[9]. Thus, it had been ended that the balloting methodology showed abundant stable performance within the presence of noise.

In 2018 Abhimanyu Roy and J. Sun and R. Mahoney and L. Alonzi and S. Adams and P. Beling. projected deep learning topologies for the detection of fraud in on-line cash group action. This approach comes from the factitious neural network with in-built time and memory elements like future short term memory alternative and several other} other parameters. Consistent with the potency of those elements in fraud detection, virtually eighty million on-line transactions through mastercard are pre-labeled as dishonest and legal. they need used high performance distributed cloud computing surroundings.[10] The study projected by the researchers provides a good guide to the sensitivity analysis of the projected parameters as per the performance of the fraud detection. The researchers conjointly projected a framework for the parameter standardization of Deep Learning topologies for the detection of fraud. This permits the institution to decrease the losses by avoiding dishonest activities.

## III. METHODOLOGY

In this paper proposes the machine learning algorithms to detect the fraudulent activity .The basic flow diagram can be represented with the following Flow chart(figure 1.)

**Incoming transaction -** This is a live transaction, the main objective of our paper is to find out whether this transaction is a fraud transaction or a valid transaction.

**Customer transaction data-** This is our data set in which old transactions are kept, on the basis of which we can predict which transaction is fraud and which transaction is valid transaction. It helps to detect fraudsters.

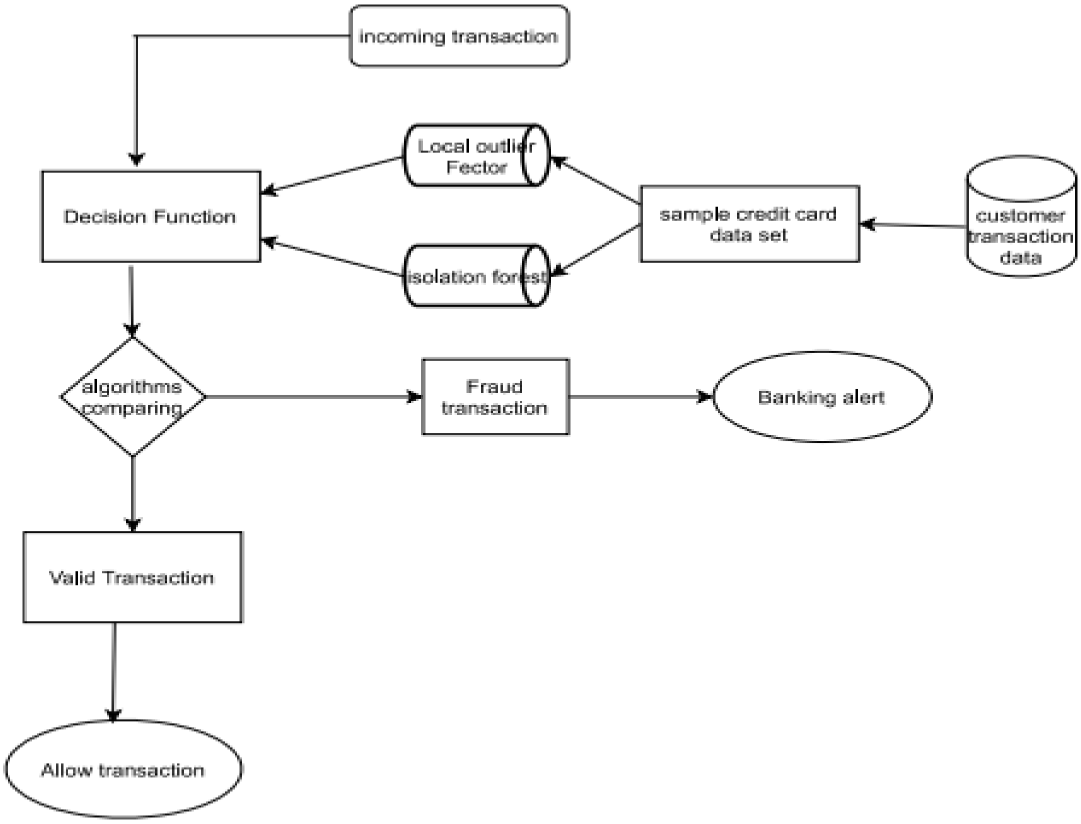


**Fig.1 Flow chart**

**Decision Function -** This function collects all the transactions and on the basis of the transaction compares the previous data by performing machine learning algorithm and detects and separates the fraud transactions and valid transactions.

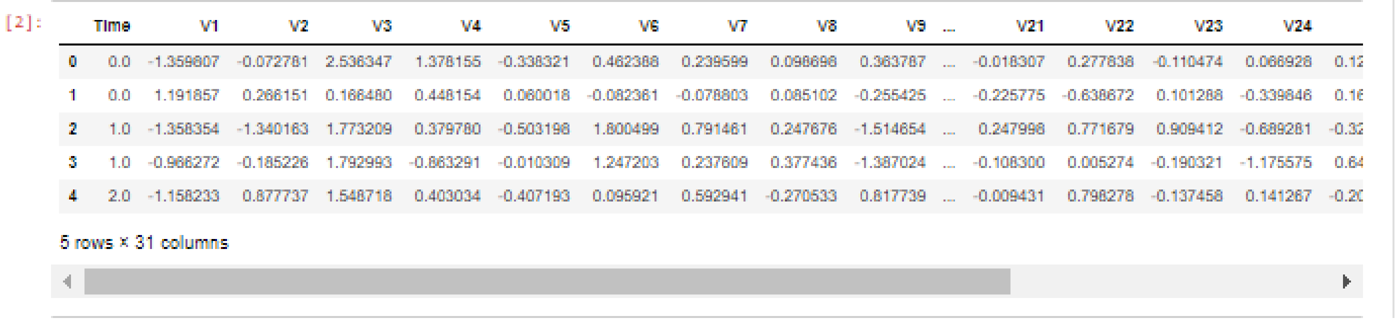
We use credit card fraud data set and find the sample of credit card fraud and deploying the model for finding the performance. The decision function are classifying using the machine learning algorithms .the isolation forest algorithms and local outlier fector. Comparing both algorithms which one is fraud transaction and valid transaction .machine learning algorithms worked to explore all types of sanction transaction and report the dubious one.

### 3.1 Dataset Description

Table 1. Attribute of dataset

| S.no | Feature | Description |
|---|---|---|
| 1. | Time | Time in second to elapses between current transaction & first transaction |
| 2 | Amount | Transaction amount |
| 3 | **Class** | 0-valid,1-fraud |

First we use the dataset for the kaggle websites they provide the credit card fraud dataset. In this dataset is two days transaction made in September 2013 by European cardholder this dataset contain 31 Numerical Feature. Namely V1-V28 ,Time & Amount. V1-V28 is the protected sensitive data.


[2]:

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | ... | V21 | V22 | V23 | V24 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | 0.12 |
| 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 | 0.16 |
| 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 | -0.32 |
| 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 | 0.64 |
| 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 | -0.20 |

5 rows × 31 columns

**Fig .2 dataset sample**

Time features describe the gapping of time between first transaction and the following one. The amount feature is the amount of rupees transacted. This dataset there are two classes class 0 and class 1 . class 0 represents the valid transaction. And class 1 represents the fraud transaction.

We plot the various graph to visualize the data set

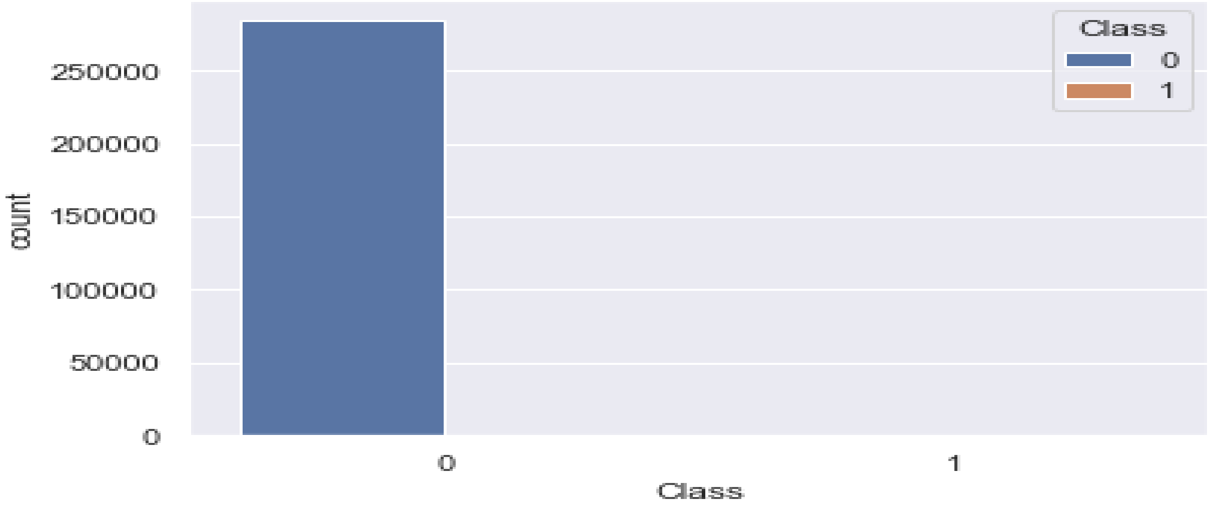


**Fig. 3 Count Fraudulent vs. non fraudulent transaction**

This graph shows the more than 25000 data are valid transaction

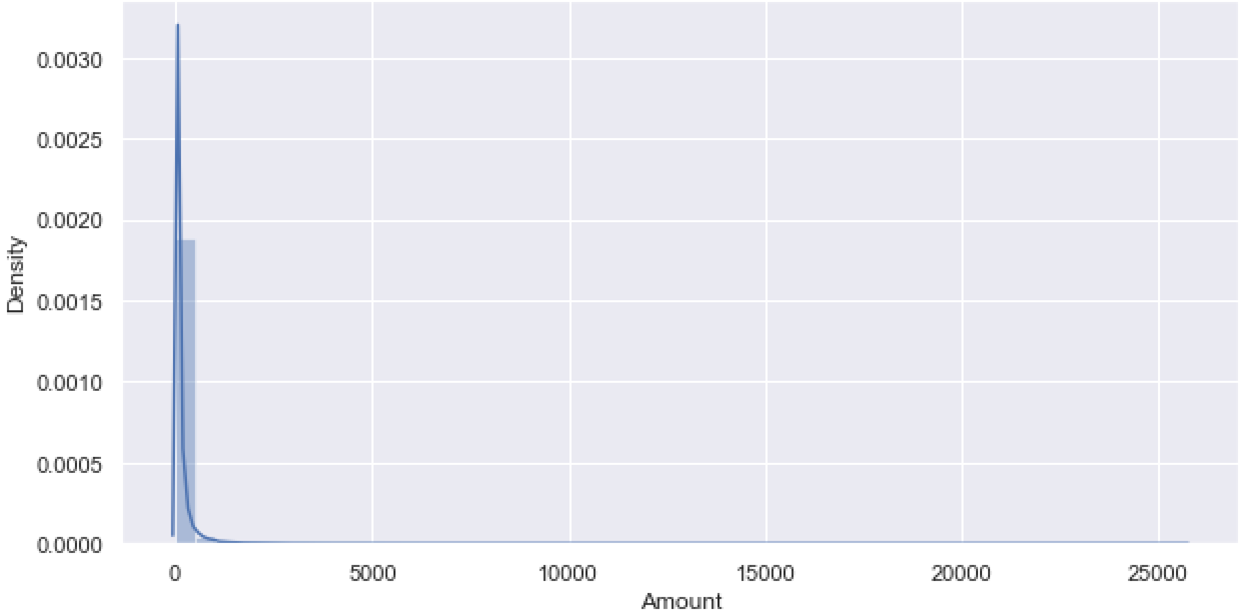


**Fig.4 Distribution of Amount Feature**

This graph shows the distributions of amount.

After this analysis we are ploting the heat map for colored representation of data and correlation between out predicting variable and the class variable.

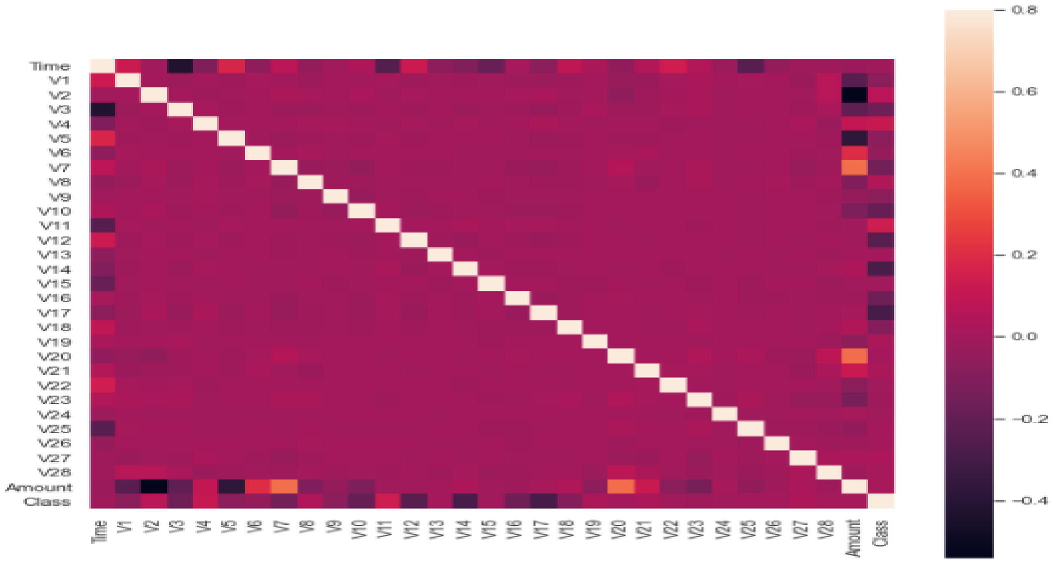


Fig.5 Heat map of Correlation

**A. Local outlier factor** – this method apply for Unsupervised outlier detection. It create an anomaly score that produce data points which are outliers in the data set.

**Pseudo code of Local outlier :**

```
Length based outlier detection
INPUT : A //Fraud data
            Max //Maximum hierarchy
            k // distance
            ג // threshold of local outlier factor
OUTPUT : LOF(A)
(1)  Outlier = len(A)
(2)  For each x in outlier do{
       Performing the k distance neighborhood of x
       Performing the rehabilitees distance of object x
       According to formula(3), performing the $p_k(x)$
       According to formula(4), computing the LOF(x)

    If (LOF(x) ≥ ג)

         Label x as Outlier.}

(3)  Return LOF(A)
```

**B. Isolation forest** - the technique of recognize rare events or recognize which can raise doubtful by being significant different from the rest of the observations.

**Pseudo code of Isolation Forest**

**iforest(x,t,α)**

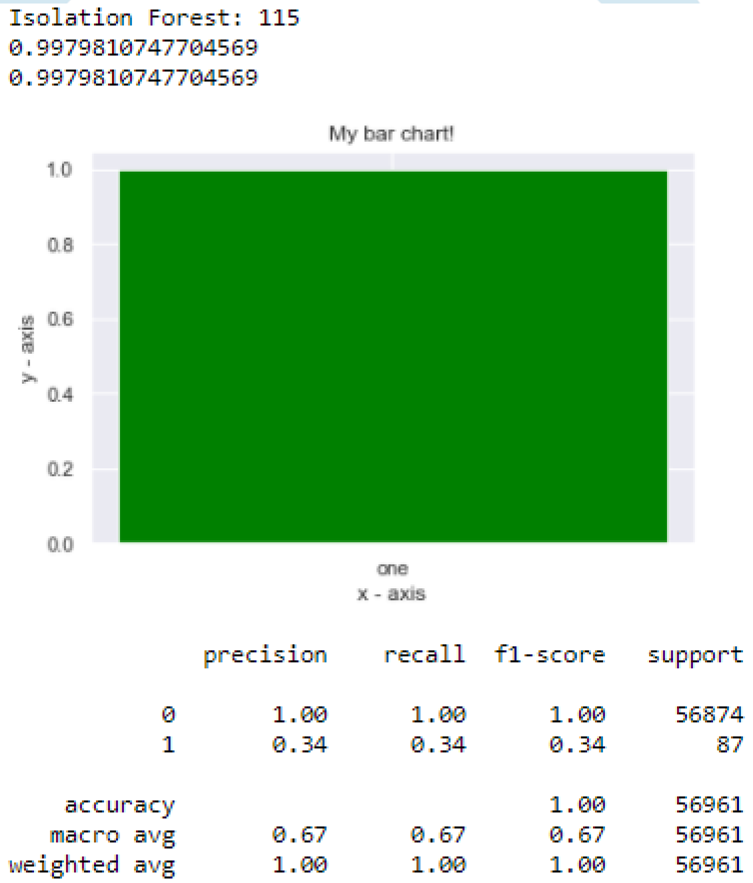
```
INPUT :  x -input data, t – number of trees, α – sub sampling size
OUTPUT: a set of t trees
1.   Initialize the forest
2.   Set height limit l = ceiling($\log_2 α$)
3.   for i=1 to t do
4.   $X^1$ = sample(X,α)
5.   Forest ← forest U trees($X^1$,0,1)
6.   End for
7.   Return forest
```

## IV. RESULTS

Table 2 .Accuracy Result

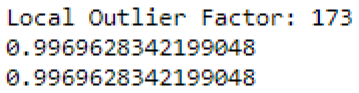
| MODEL | ACCURACY | PRECISION | RECALL | F1SCORE |
|---|---|---|---|---|
| Local outlier Factor | 99.6% | 100% | 100% | 100% |
| Isolation Forest | 99.7% | 100% | 100% | 100% |

This results In this result the classification report of both the algorithms is given in which class 0 means valid transaction and class 1 means in fraud transaction. In the number of false positive is detected and comparing the actual value this is use for calculate accuracy, precision, recall, F1 score. This results match the cross value for checking the false positives.

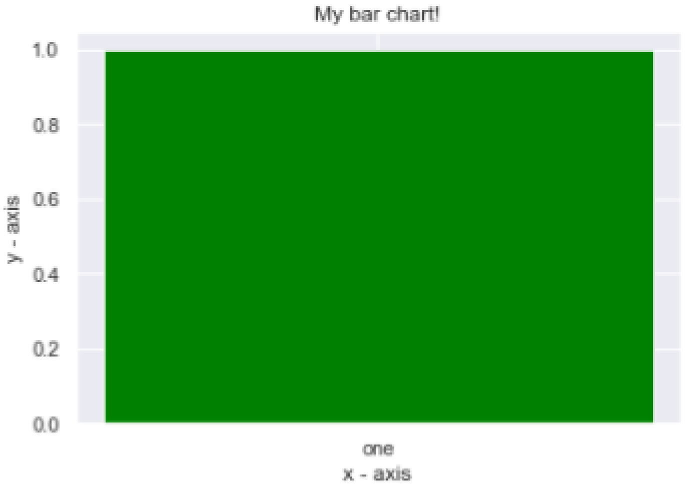


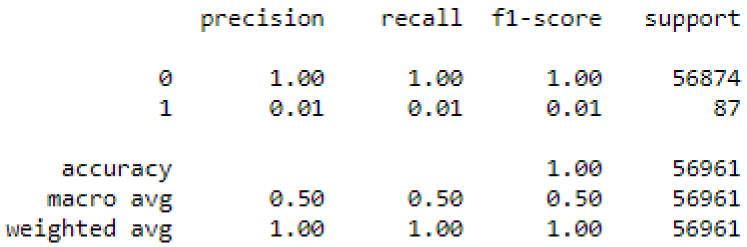
```
Isolation Forest: 115
0.9979810747704569
0.9979810747704569
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     56874
           1       0.34      0.34      0.34        87

    accuracy                           1.00     56961
   macro avg       0.67      0.67      0.67     56961
weighted avg       1.00      1.00      1.00     56961
```



```
Local Outlier Factor: 173
0.9969628342199048
0.9969628342199048
```





```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     56874
           1       0.01      0.01      0.01        87

    accuracy                           1.00     56961
   macro avg       0.50      0.50      0.50     56961
weighted avg       1.00      1.00      1.00     56961
```

## V. CONCLUSION

Without any doubt, credit card fraud is a criminal activity, in this paper, methods have been made to avoid fraud using common methods and many research papers have been studied in this paper, how using machine learning brought better results. It is explained in depth and how it is implemented through sudo code along with its experimental results

While the algorithms does reach over 99.7% accuracy and its precisions remains 34% this is high percentage of accuracy is to be expected due to the huge imbalance between the number of genuine transaction and number of valid transaction.

## REFERENCES

[1] RuttalaSailusha, V.Gnaneswar, R.Ramesh, G. Ramakoteswara Rao. **Credit Card Fraud Detection Using Machine Learning.**IEEE Xplore part Number CFP20K74-ART,Jun 2020

[2] DarshanKaur,Shubpreet Kaur. **Machine Learning Approach for Credit Card Fraud Detection(KNN & Naïve bayes)**. 1st international conference on intelligent communication and computational research(ICICCR-2020)

[3] S.Abhinayaa, H.Sangeetha, R.A.kartikayan, K.Saransriram, D.Piyush**Credit card fraud detection and prevention using Machine learning** International journal of Engineering and Advanced Technology(IJEAT)Vol-9 issue 4 ISSN 2249-8958,April 2020

[4] Hassan Najadat, Ola Altiti,Ayah Abu Aqoleh, MutazYounes**Credit card Fraud Detection Based on Machine and Deep Learning** International conference on Information and Communication System(ICICI) 978-1-7281-6227

[5] B.Nitin, RohitRavula, ShaikGangina Sultana **Credit Card Fraud Detection Using Adaboost**International Journal of Scientific Research & Engineering Trends Volume 6, Issue 2,ISSN(online):2395:566X

[6] AndhavarapuBhanusri, K.RatnaSreeValli, P.Jyothi, G.VarunSai,R.RohitSaiSubash**Credit card fraud detection by using Machine learning algorithms,** Quest Journal Journal of Research in Humanities and Social Science Volume 8 issue 2(2020) pp: 04-11 ISSN(online):2321-9467

[7] S P Maniraj, Aditya Saini, Swarna Deep Sarkar **Credit Card Fraud detection using Machine learning and Data Science,** International Journal of Engineering Research & Technology(IJERT) vol.8 Issue 09,ISSN 2278:0181, September 2019

[8] Vaishnavinath, Dornadula, Geetha s. **Credit Card Fraud Detection using Machine Learning Algorithms,** International conference on Recent Trends in Advanced Computing 2019.ICRTAC 2019

[9] Kuldeep Randhawa, chukiong loo, Manjeevalseera. **Credit card fraud Detection using Adaboost and Majority voting.** IEEE Access 2018.2806420

[10] A. Roy and J. Sun and R. Mahoney and L. Alonzi and S. Adams and P. Beling, **Deep learning detecting fraud in credit card transactions,** in Systems and Information Engineering Design Symposium(SIEDS),pp.129-134,2018.