International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Secure Scan -Analysis of Credit Card Fraud Detection Using Machine Learning

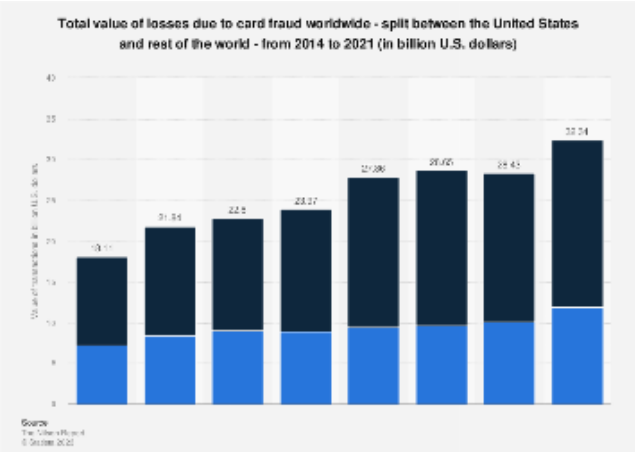
[1]Dr.N.Ramarao
[2]Aditya Kumar Singh, [3]G.Aditya Reddy, [4]D.Adarsh
[5]U.Aishwarya,[6]S.Ajay Ray,[7]O.Akanksha
[234567]Student,[1] Assistant Professor
Artificial Intelligence & Machine Learning
Department Of Artificial Intelligence & Machine Learning
Malla Reddy University, Hyderabad, Telangana, India

Abstract: Credit Card usage has been drastically increased across the world , now people believe in going cashless and are completely dependent on online transactions .The credit card has made the digital transaction easier and nowadays credit card frauds are drastically increasing in numbers compared to earlier times . A powerful fraud detection system is required to stop these frauds . Fraud detection is the process of monitoring the transaction behaviour of a cardholder to detect whether an incoming transaction is authentic and authorised or not otherwise it will be detected as illicit .Machine learning -based fraud detection systems rely on machine learning algorithms that can be trained with historical data and can perform classification , regression or finding the patterns among the data .

Keywords: Credit card machine learning , fraud detection ,historical data, algorithms.

I.INTRODUCTION

Credit card frauds are increasing day to day across the country which leads to demand for robust and trustable models which can work using the previous historical data . Some estimates say less than 1% of credit card fraud is actually detected while others say it could be higher but it is impossible to know. The truth is that most credit card fraud does go undetected , which is a major reason why it's become a favourite among crime rings and fraudsters .



The graph represents the increase in the count of credit card frauds globally where the modern problems can be solved using modern technologies such as machine learning . Machine learning can deal with this historical data to find the solutions to the problem .

II.LITERATURE REVIEW

Machine learning plays a major role in detecting fraud where the algorithms can be used according to the task such as classification or regression .There are many existing approaches present for the fraud detection. These existing systems can be improved and can become more robust in detecting the fraud by making the modifications using the current machine learning algorithms. The datasets used for the detection are mostly biased because there is only 2 to 3 percent of data

is fraud labelled among the total .This problem can also be simplified by machine learning modules .

III.PROBLEM STATEMENT

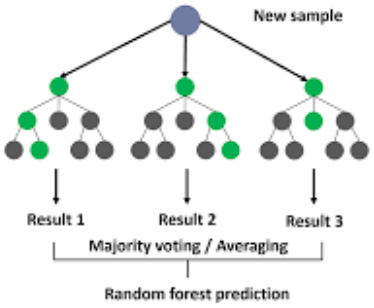
The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the data of the ones that turned out to be fraud. Using feasible machine learning algorithms on an imbalanced dataset such as logistic regression, correlation matrix the dataset is pre-processed and analysed and classified as training and testing dataset. The performance of the techniques is gauged based on accuracy, sensitivity, specificity, and precision. Machine learning also reduces operational costs ,as the amount of data and experience increases the results of machine learning becomes more accurate.

About the data

Out of 284,807 there are 492 frauds .The dataset is highly unbalanced and the positive class(frauds) account for 0.172% of all transactions. The input values are PCA transformations .The only features which have not been transformed with PCA are time and amount . feature "CLASS" is the variable and it takes 1 in case of fraud or 0 otherwise. time: the time of transactions. V1-V28: various attributes of credit card transactions .amount: amount involved in credit card transactions. class: whether the given credit card transaction is fraudulent (1) or not (0).

IV.METHODOLOGY

The process includes a machine learning algorithm called the random forest classifier which is a type of decision tree classification technique .



The random forest is a collection of decision trees where the prediction is done by considering the average of all the decision trees included in the classification

Steps included

1.Importing the required modules

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
```

2.loading dataset

```
df=pd.read_csv('creditcard.csv')
df.head()
```

3.Exploratory data analysis

Finding the null values

```
df.isnull().sum()
```

The null method helps to find if there are any null values present in the data and sum is used to count all the null values present in it . The presence of null values leads to error while training the data so it is highly important to treat them .

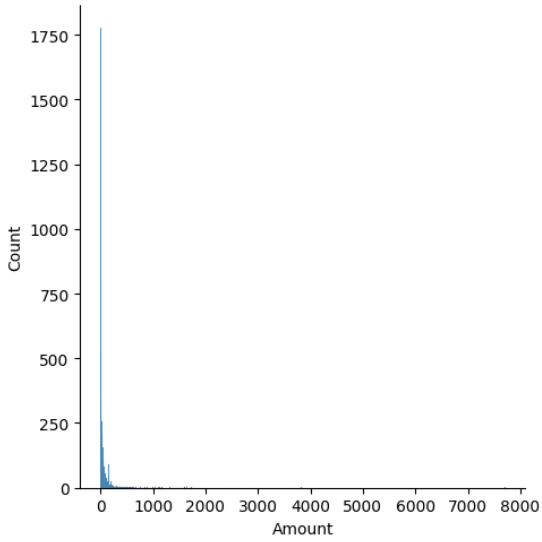
```
Time       0
V1         0
V2         0
V3         0
V4         0
V5         0
V6         0
V7         0
V8         0
V9         0
V10        0
V11        0
V12        0
V13        0
V14        0
V15        0
V16        0
V17        0
V18        1
V19        1
V20        1
V21        1
V22        1
V23        1
V24        1
V25        1
V26        1
V27        1
V28        1
Amount     1
Class      1
dtype: int64
```

After dropping the null values using the dropn_a function from the python modules the output is

```
Time                0
V1                  0
V2                  0
V3                  0
V4                  0
V5                  0
V6                  0
V7                  0
V8                  0
V9                  0
V10                 0
V11                 0
V12                 0
V13                 0
V14                 0
V15                 0
V16                 0
V17                 0
V18                 0
V19                 0
V20                 0
V21                 0
V22                 0
V23                 0
V24                 0
V25                 0
V26                 0
V27                 0
V28                 0
Amount              0
normalized_amount   0
dtype: int64
```

The displot of the amount

```
sns.distplot(df['Amount'])
```



.

Finding the fraud data percentage

```
fraud_pct=(fraud/(fraud+non_fraud))*100
print("The fraud transactions percentage:",fraud_pct)
```

Output is

```
The fraud transactions percentage: 0.05022601707684581
```

Splitting the training and testing data

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

Creating the random forest classifier

```
from sklearn.ensemble import
RandomForestClassifier
rf1=RandomForestClassifier()
rf1.fit(x_train, y_train)
```

Training the model

```
rf1.fit(x_train, y_train)
```

```
▾ RandomForestClassifier
RandomForestClassifier()
```

V EXPERIMENTAL RESULTS

The model is trained using the random forest classifier where its accuracy is high among all the models present .The accuracy is also calculated using the python module called as metrics .accuracy_score

```
accuracy=metrics.accuracy_score(actual
,predicted)
print(accuracy)
```

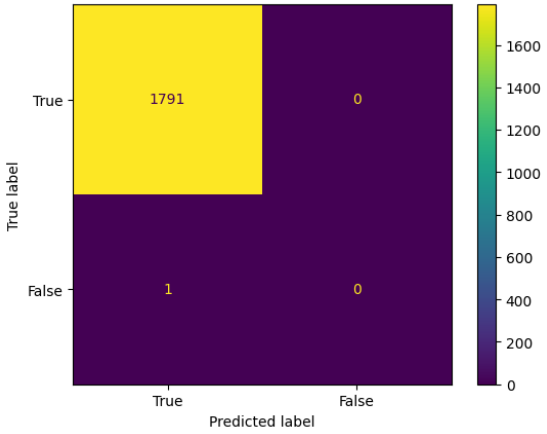
The output of the accuracy calculation

```
0.9994419642857143
```

The confusion matrix is used the visualise the accuracy of the model

```
from
sklearn.metrics._plot.confusion_matrix
import confusion_matrix
actual=y1_test
predicted=prediction_rf
confusion_matrix=metrics.confusion_mat
rix(actual,predicted)
cm_display=metrics.ConfusionMatrixDisp
lay(confusion_matrix=confusion_matrix,
display_labels=[True,False])
cm_display.plot()
plt.show()
from
sklearn.metrics._plot.confusion_matrix
import confusion_matrix
actual=y1_test
predicted=prediction_rf
confusion_matrix=metrics.confusion_mat
rix(actual,predicted)
cm_display=metrics.ConfusionMatrixDisp
lay(confusion_matrix=confusion_matrix,
display_labels=[True,False])
cm_display.plot()
plt.show()
```

The confusion matrix is provided after executing the code where it is shown as



1791 true values are predicted correctly among all the values where only one value is marked as false .

VI.CONCLUSION

```
accuracy=metrics.accuracy_score(actual
,predicted)
```

In this article we have developed the machine learning model to detect credit card frauds with highest accuracy. The main tasks also included dealing with the unbalanced data along with the model's robustness .

While the algorithm has reached 99 % of the accuracy which is best to use among all the methods included in the fraud detection methods.

VII . FUTURE WORK

In further process this machine learning model can be converted to the pickle file and it can be integrated with any static or dynamic websites where the inputs can be directly taken from the user and results can be evaluated and it can also be integrated with the alert systems where it can provide the alert message to the user before the fraud occurs . more data will surely make the model more accurate in detecting frauds and reduce the number of false positives.

VIII. REFERENCES

[1] Credit Card Fraud Detection through Parenclitic Network Analysis- By Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral" published by Hindawi Complexity Volume 2018, Article ID 5764370, 9 pages.

[2] Credit Card Fraud Detection using Machine Learning and Data Science International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181 http://www.ijert.org
IJERT V8I 090031
(This work is licensed under a Creative Commons Attribution 4.0 International License.)
Published by :
www.ijert.org
Vol. 8 Issue 09, September-2019

[3] INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING, ICRTAC 2019,Credit Card Fraud Detection using Machine Learning Algorithms.

[4] A machine learning based credit card fraud detection using the GA algorithm for feature selection .leberietal. Journal of Big Data (2022) 9:24 https://doi.org/10.1186/s40537-022-00573-8

[5] Google Colab [Online]. Available: https:// colab. resea rch. google. com/