A Mini Project Synopsis on

Credit Card Fraud Detection

T.E. - I.T Engineering

Submitted By

Sakshi Shinde   19104055

Shubhangi Lanke   19104063

Divya Shepal   19104053

Under The Guidance Of

Prof. Manjusha Kashilkar

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year : 2021-22

CERTIFICATE

This to certify that the Mini Project report on **Credit Card Fraud Detection** has been submitted by Sakshi Ravindra Shinde (19104055), Divya Sanjiv Shepal (19104053) and Shubhangi Babaji Lanke (19104064) who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2021-2022** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Manjusha Kashilkar
Guide

Prof. Kiran Deshpande   Dr. Uttam D.Kolekar
Head Department of Information Technology   Principal

External Examiner(s)
1.
2.

Place:A.P.Shah Institute of Technology, Thane
Date:

TABLE OF CONTENTS

References

Acknowledgement

Chapter 1

1. Introduction

Credit card is a small thin plastic or fiber card that contains information about the person such as a picture or signature and person named on it to charge purchases and service to his linked account charges for which will be debited regularly. Nowadays card information is read by ATMs, swiping machines, store readers, bank and online transactions. Each card has a unique card number which is very important, its security mainly relies on physical security of the card and also privacy of the credit card number.

There is a rapid increase in the credit card transaction which has led to substantial growth in fraudulent cases. Many data mining and statistical methods are used to detect fraud. Many fraud detection techniques are implemented using artificial intelligence, pattern matching. Detection of fraud using efficient and secure methods are very important.

Credit card frauds are increasing heavily because of fraud financial loss is increasing drastically. Nowadays Internet or online transactions are growing as new technology is coming day by day. In these transactions the Credit card holds the maximum share. In 2018 Credit card fraud losses in London estimated US dollar 844.8 million. To reduce these losses prevention or detection of fraud must be done. There are different types of frauds occurring as technology is growing rapidly. So there are many machine algorithms used to detect fraud nowadays hybrid algorithms, artificial neural networks are used as it gives better performance.

In a planned system, we are applying the random forest algorithm for classifying the credit card dataset. Random Forest is an associate in the nursing algorithmic program for classification and regression. Hence, it is a collection of decision tree classifiers. The random forest has an advantage over the decision tree as it corrects the habit of over fitting to their training set. A subset

of the training set is sampled randomly so that to train each individual tree and then a decision tree is built, each node then splits on a feature designated from a random subset of the complete feature set. Even for large data sets with many features and data instances, training is extremely fast in the random forest and because each tree is trained independently of the others. The Random Forest algorithm has been found to provide a good estimate of the generalization error and to be resistant to overfitting.

1.1 Purpose

- Fraud detection and prevention refers to the strategies undertaken to detect and prevent attempts to obtain money or property through deception

- Credit card fraud costs consumers and the financial company billions of dollars annually, and fraudsters continuously try to find new rules and tactics to commit illegal actions. Thus, fraud detection systems have become essential for banks and financial institutions, to minimize their losses.

1.2 Objectives

- To Identify Fraudulent Credit card transaction
- To detect 100% of the Fraudulent transactions while minimizing the incorrect fraud classifications.
- To estimate, perceive or avoid objectionable behaviour , which consist of fraud , intrusion and defaulting .

1.4 Scope

- Can be applied in banking organizations.
- Can be used to decrease fraud attempts.
- Can be used to enhance the process and guarantee the security of the account.

2.Problem Definition

- Unauthorized transactions and unwanted usage of an account by someone other than the owner of the account.
- Massive stream of payment requests were   generated on a daily basis.

3. Proposed System

- To help us classify fraudulent and legitimate credit card transactions by supervised learning algorithms such as Random forest.
- To help us to get awareness about the fraudulent and without loss of any financially.

4. Project Outcomes

- To classify fraudulent and legitimate credit card transactions by supervised learning algorithms such as Random forest.
- To help us to get awareness about the fraudulent and without loss of any financially.

5. Technology stack

- VS Code

    Intuitive keyboard shortcuts, easy customization and community-contributed keyboard shortcut mappings let you navigate your code with ease.

6. Project Design(Flow Of Modules)

- Importing Dataset
- Conversion of data into frames format
- Data Sampling
- Data Separation For Training and Testing
- Assigning train datasets to the models
- Applying algorithm and Creating the model
- Making Predictions for test dataset
- Calculating Accuracy

7. Project Scheduling:

1. Sakshi Shinde (19104055):

    i) Work to train and test the model

2.Shubhangi Lanke (19104064):

    i) Work to apply different algorithms on model

3.Divya Shepal (19104053):

    i) Work to calculating accuracy

8.Screenshot of Application


```
PROBLEMS  9    OUTPUT    TERMINAL    DEBUG CONSOLE

Data columns (total 31 columns):
 #    Column  Non-Null Count   Dtype
---   ------  --------------   -----
 0    Time    277968 non-null  float64
 1    V1      277968 non-null  float64
 2    V2      277968 non-null  float64
 3    V3      277968 non-null  float64
 4    V4      277968 non-null  float64
 5    V5      277968 non-null  float64
 6    V6      277968 non-null  float64
 7    V7      277968 non-null  float64
 8    V8      277968 non-null  float64
 9    V9      277968 non-null  float64
 10   V10     277968 non-null  float64
 11   V11     277968 non-null  float64
 12   V12     277968 non-null  float64
 13   V13     277968 non-null  float64
 14   V14     277968 non-null  float64
 15   V15     277968 non-null  float64
 16   V16     277968 non-null  float64
 17   V17     277968 non-null  float64
 18   V18     277968 non-null  float64
 19   V19     277968 non-null  float64
 20   V20     277968 non-null  float64
 21   V21     277968 non-null  float64
 22   V22     277968 non-null  float64
 23   V23     277968 non-null  float64
 24   V24     277968 non-null  float64
 25   V25     277968 non-null  float64
 26   V26     277968 non-null  float64
 27   V27     277968 non-null  float64
 28   V28     277968 non-null  float64
 29   Amount  277968 non-null  float64
 30   Class   277968 non-null  int64
dtypes: float64(30), int64(1)
memory usage: 65.7 MB
Number of Genuine transactions:  277502
Number of Fraud transactions:  466
Percentage of Fraud transactions: 0.1676
Shape of train_X:  (194577, 29)
Shape of test_X:  (83391, 29)
```

Information of data

```
Number of Genuine transactions:  277502
Number of Fraud transactions:  466
Percentage of Fraud transactions: 0.1676
Shape of train_X:  (194577, 29)
Shape of test_X:  (83391, 29)
Random Forest Score:  99.95443153337891
Decision Tree Score:  99.91845642815171
Confusion Matrix - Decision Tree
[[83213    41]
 [   27   110]]
Confusion matrix, without normalization
Confusion Matrix - Random Forest
[[83248     6]
 [   32   105]]
Confusion matrix, without normalization
Evaluation of Decision Tree Model

Accuracy: 0.99918
Precision: 0.72848
Recall: 0.80292
F1-score: 0.76389
Evaluation of Random Forest Model

Accuracy: 0.99954
Precision: 0.94595
Recall: 0.76642
F1-score: 0.84677
```

Accuracy and Precision Prediction
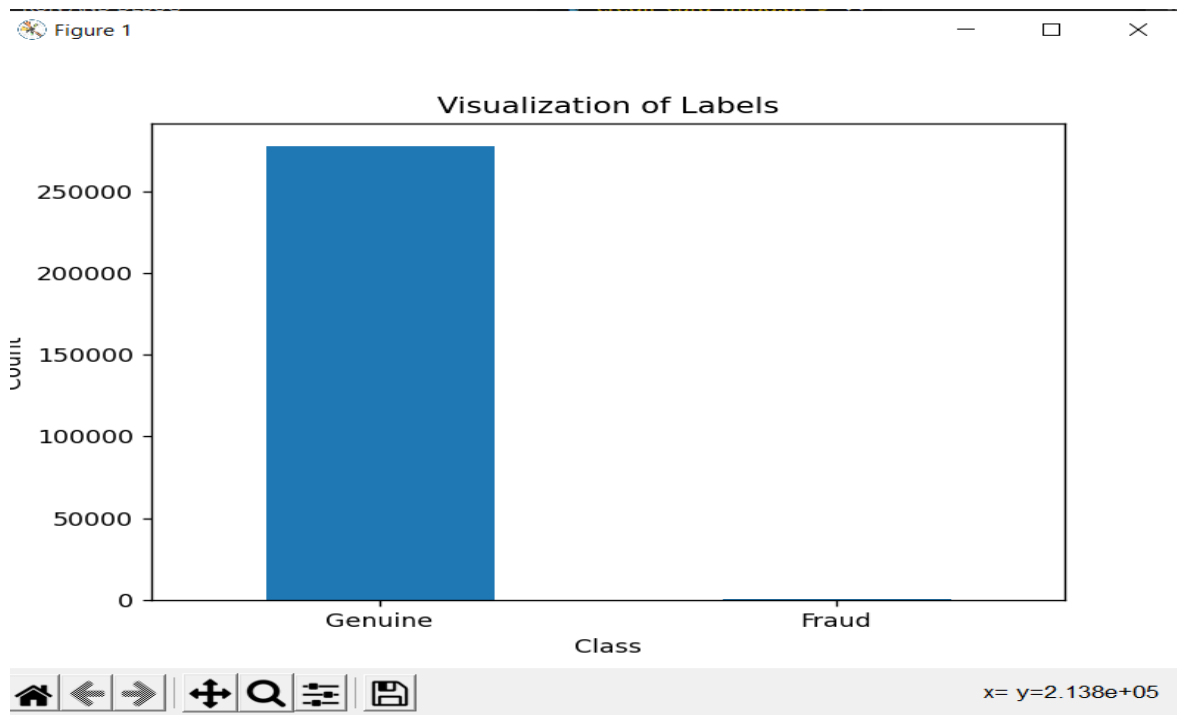
Visualization of Labels

9. Conclusion

Hence, we have acquired the result of an accurate value of credit card fraud detection i.e. 0.9994802867383512 (99.93%) using a random forest algorithm with new enhancements. In comparison to existing modules, this proposed module is applicable for the larger dataset and provides more accurate results. The Random forest algorithm 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.

In this python machine learning project, we built a binary classifier using the Random Forest algorithm to detect credit card fraud transactions. Through this project, we understood and applied techniques to address the class imbalance issues and achieved an accuracy of more than 99%.

REFERENCES

[1] "Credit Card Fraud Detection Based on Transaction Behaviour -by John Richard D. Kho, Larry A. Vea" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[2] CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 " A Comprehensive Survey of Data Mining-based Fraud Detection Research" published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800, Australia

[3] "Survey Paper on Credit Card Fraud Detection by Suman" , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

[4] "Research on Credit Card Fraud Detection Model Based on Distance Sum – by Wen-Fang YU and Na Wang" published by 2009 International Joint Conference on Artificial Intelligence