

### DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

### SHARDA SCHOOL OF ENGINEERING AND TECHNOLOGY

**SHARDA UNIVERSITY, GREATER NOIDA**

**Credit Card Fraud Detection**

***A proposal was presented***

***a portion of the prerequisites for the Bachelor's degree, of Computer Applications***

### by

**Raghav Sukhwal (2020554158)**

### Dhruv Kumar (2020562984)

**Shivam Shrama (2020001649)**

### Supervised by:

**Dr. Amit Kumar Assistant professor**

### School of Engineering and Technology

**NOVEMEBR, 2022**

## 

## CERTIFICATE

This is to certify that the "Credit Card Fraud Detection" report that Raghav Sukhwal (2020554158) , Shivam sharma (2020001649)and Dhruv kumar(2020562984) submitted to Sharda University is a true record ofthe final year project work that they completed in the "Department of Computer Scienc & Applications, Sharda School of Engineering and Technology, Sharda University" in order to fulfil the requirements for the degree of "Bachelor of Computer A degree or certificate has not been granted by any other University or Institute using the findings from this study, either in full or in part. For the purpose of conferring a degree or certificate, no other university or institute has recognised the project's findings in full or in part..

#### Signature of the Guide Name: Mr. Amit Kumar

**Rank: Assistant Professor**

**Name of Department Head's Signature:** Prof.(Dr.) Arun Prakash Aggarwal

#### University of Sharda

#### Date:

**The External Examiner's Signature**

**Date:**

## ACKNOWLEDGEMENT

A significant project offers a unique chance for growth and development. We feel extremely fortunate and humbled to have had the guidance of so many excellent people while we completed this project.

First and first, we want to express our gratitude to **Dr. Arun Prakash Aggarwal, HOD, CSA**, for giving us the chance to work on this project.

Our sincere appreciation to Sir **Mr. Amit kumar** for his assistance with our project. Despite being quite busy with academic work, **Mr. Amit kumar** made some time to listen to us, lead the way, and keep us on the right track. Without her assistance, we are unsure of where we would be now.

The CSA department kept track of our development and set up all the facilities to make things simpler. We select this time to recognise theis contribution gretefully .

Raghav Sukhwal

(2020554158)

Dhruv Kumar

(2020562984)

Shivam Shrama

(2020001649)

## ABSTRACT

Credit card thieves have easy access to basic targets. Due to the growth of e-commerce and other websites, there are now more choices for making payments online, which increases the possibility of internet fraud As a result of an increase in fraud rates, academics are now applying a range to detect and analyse online transaction fraud, machine learning algorithms. The project's main objective is to develop and put into use a special fraud detection algorithm for streaming. analysing and extracting behavioural patterns from historical transaction data customer transaction data. whereby cardholders are categorised based on the amount of their transactional activity. Then, merge the transactions made by cards from various groups using the sliding window approach to get each group's distinctive behaviour. In the future, different classifiers are trained separately using the groups. After choosing the classifier with the highest rating score as one of the most effective techniques to detect frauds, a feedback mechanism is then employed to solve the problem of idea drift. For this investigation, we made use of a European credit card fraud dataset. The proliferation of online payment methods brought about by e-commerce and other The risk of internet fraud has grown due to websites. As a result of an increase in fraud rates, academics are now applying a range of machine learning techniques to identify and analyse online transaction fraud. The primary goal of the project is to create and implement a unique fraud detection algorithm for streaming transaction data to: analyse and extract behavioural patterns from past customer transaction data. wherein cardholders are categorised based on the amount of their transactional activity. Then, merge the transactions made by cards from Various groups are examined using the sliding window method to determine each group's unique behaviour. Future training of various classifiers employing the groups will take place independently. Choosing the classifier with the highest rating score is thus one of the finest methods for detecting fraud. Consequently, idea drift is handled through a feedback mechanism. mechanism. For this report, we used statistics on European credit card fraud.

Keywords: Credit card, Electronic, Fraud detection, Genetic algorithms

## CONTENTS Page No.

[TITLE i](#_bookmark0)

[CERTIFICATE ii](#_bookmark1)

[ACKNOWLEDGEMENT iii](#_bookmark2)

[ABSTRACT iv](#_bookmark3)

[LISTOF FIGURES v](#_bookmark4)

[LISTOF TABLES vi](#_bookmark5)

[CHAPTER1: INTRODUCTION 1](#_bookmark6)

* 1. Problem Statement 1
  2. Project Overview 2
  3. Expected Outcome 2
  4. Hardware & Software Specifications 3
  5. Other Non-Functional Requirements 3

CHAPTER2: LITERATURESURVEY 4

* 1. Existing Work 4
  2. Proposed System 6
  3. Feasibility Study 7

CHAPTER3: SYSTEM DESIGN& ANALYSIS 8

* 1. Project Perspective 8
  2. System features 9
  3. Methodology 10
  4. Implementation 11
  5. Testing Process 14

CHAPTER4: RESULTS AND OUTPUTS 16

4.1 Proposed Model Outputs 16

[CHAPTER5: CONCLUSION 17](#_bookmark7)

* 1. Future Scope 17

CHAPTER6: REFERENCES 18

## LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| Fig1.1 | Fraud Type Tree | 1 |
| Fig1.2 | Application workflow | 2 |
| Fig2.1.1 | Credit Card Fraud Detection | 5 |
| Fig3.2.1 | Matrix of Correlation for Attributes (both the X and Y axis show different attributes present in dataset) | 8 |

### CHAPTER 1 INTRODUCTION

A customer's (the cardholder's) credit card is typically a card that enables users to routinely take out cash in advance or make purchases up to their credit limit. Since credit card users may roll over balances, they profit from time. their debt to the following payment cycle, delaying its repayment to a later period. Credit card theft may happen to easy targets. Unknown to the owner, a big sum of money might be quickly and safely transferred. removed. Fraud detection is quite challenging since fraudsters frequently try to pass off fake transactions as real ones. to get a job. Credit card fraud was the most commonly reported type of data breach in 2017 (133,015 reports), According to information published by the FTC [10], bank fraud was reported 55,045 times, phone fraud 55,045 times, and job or tax-related fraud 82,051 times. Nearly 179 million people were impacted by these data breaches. records.



Fg.1.1

Fraud type tree

Due to scams frequently appearing in the news over the past several years, particularly credit card thefts, the majority of people worldwide are most aware of them. The credit card dataset is quite asymmetrical since there will be more legitimate transactions than fraudulent ones. Even while card-not-present fraud rates are still higher, EMV cards, which are smart cards that save their data on integrated circuits rather than magnetic stripes as technology progresses, have made certain on-card payments safer.A 2017 US Payments Forum study [10] found that as chip card security increased, fraudsters focused their attention on CNP purchase-related activity. Fig. displays the total number of CNP fraud incidents recorded throughout the pertinent years. 2.

Requirement

* + Software requirement
    1. Windows 10, 11.
    2. Testnet Goerli
    3. Alchemy Deploy-er
    4. Mozilla, Chrome, Opera
    5. Visual Studio Code
    6. Cranq IDE
    7. Metamask Wallet
  + Hardware Requirement

1. High Core Processor i5 10gen, Ryzen 7 and above
2. Disk Space 20 GB (including IDE/tools)
3. Ram 8 GB
4. 15’’ Colour monitor
5. Keyboard 122keys
   * Non Functional Requirements

1.UX - Designing the customer experience is a challenge in the creation of Blockchain-based web applications. The issue is how much of the technology involved, Blockchain, should be made available to your use

2.Necessity of feasibility analysis - By analysing a project's potential, feasibility studies can determine its logistical, financial, and commercial problems.

3.Development Operations - The technology to create a blockchain-based platform is continually developing as blockchain is still a relatively new concept. To enhance the DevOps experience and implement new capabilities into the system, time is needed.

### LITERATURE SURVEY

To detect fraud, a variety of Techniques for supervised and semi-supervised machine learning are employed. [8]. Three significant issues with the card fraud dataset that we want to address are the severe class imbalance, the availability of labelled and unlabeled samples, and the requirement to coordinate several transactions

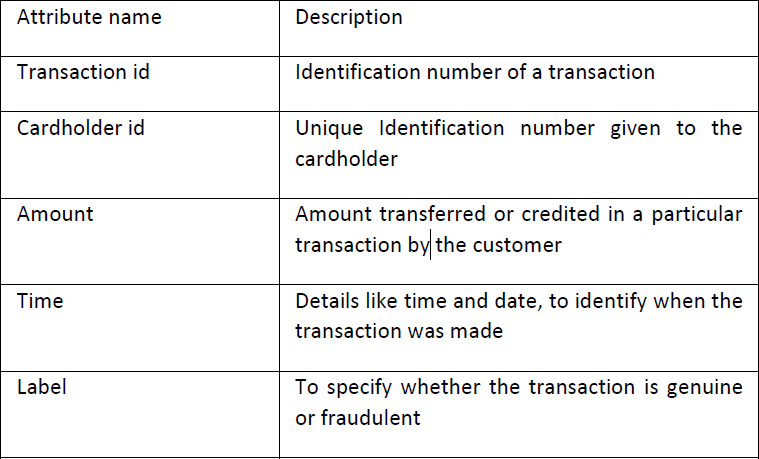
Several Supervised machine learning techniques [3Real-time datasets [including Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression, and SVM] are used to identify fraudulent transactions. To educate the behavioural characteristics of typical and abnormal transactions, two random forest models were utilised.algorithms [6] are utilised. Based on CART and Random Trees, these are random forests. Random forest produces acceptable results for small data sets, but there are still significant drawbacks for uneven data. The upcoming project will try to solve the previously mentioned issue..

### The random forest algorithm itself need improvement. Research is being done in managing highly skewed credit card fraud data using meta-classifiers and meta-learning approaches to evaluate the effectiveness of Logistic Regression, K-Nearest Neighbor, and Naive Bayes. It's possible that fraud incidents won't always be correctly identified when using supervised learning approaches. using a limited Boltzmann machine and a deep auto-encoder (RBM)model [2], it may be possible to identify abnormalities in typical patterns by generating frequent transactions. Additionally, by combining the majority voting and adaboost approaches, a hybrid strategy is produced [4].

**Suggested System**

Comparing card transactions to a customer's prior purchases is always an alien concept. When they are referred to as idea drift difficulties [1], this unfamiliarity presents a very serious challenge in the actual world. Concept drift may be thought of as an unpredictable variable that evolves over time. These factors significantly throw the data out of balance. Our study's primary goal is to identify a workable solution to the Concept Drift issue. The essential characteristics that are noted throughout each transaction are listed in Table 1 [1].

Table 1 shows raw credit card transaction data.



* 1. *Dataset Description*

There are transactions done by a cardholder in the dataset [11].r over a two-day period or in September 2013. 492 transactions out of the 284,807 total transactions—or 0.172%—are fake. This dataset has a severe imbalance. The majority of the dataset's properties are transformed using principal component analysis since it is thought that disclosing a customer's transactional information would compromise their privacy (PCA). Table 2 demonstrates that while the features "time," "amount," and "class" are not PCA-applied features, V1, V2, V3,..., and V28 are.

Table 2: European dataset attributes

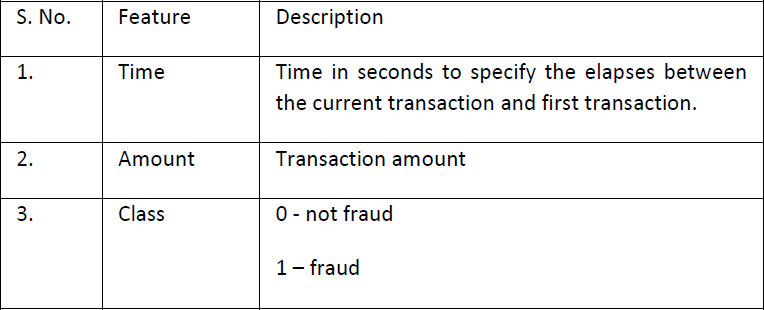
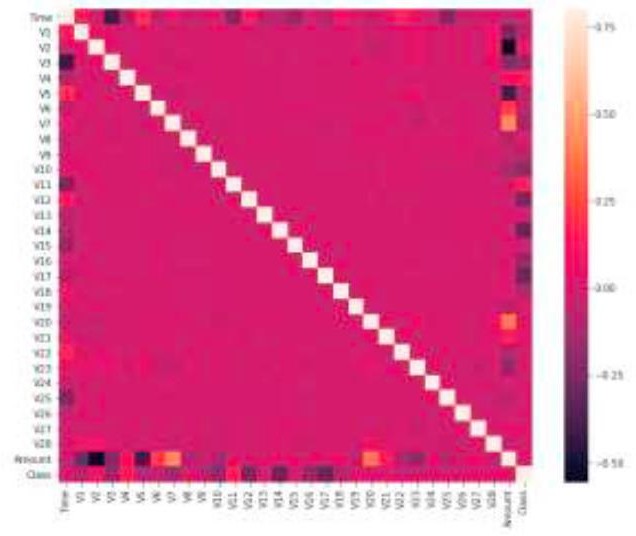


Fig. Figure 3 displays the dataset's correlation matrix. The attribute class is independent of the transaction's value and time, according to this matrix. Even the matrix clearly illustrates how the PCA applied attributes affect the transaction class.Fig. 3: Correlation Matrix for Attributes (both the X and Y axis show different attributes present in dataset)



### Methodology

First, we divide the cards into high, medium, and low transaction amount clusters/groups using range partitioning and the clustering technique.

Using the sliding-window method, we classify the transactions into the relevant categories, and then we extract certain traits from the window to identify patterns in cardholder behaviour. There are details given, like the highest and lowest transaction amounts, the typical quantity in the window, including the duration that has gone.

# Running code

* Import Data Set of Credit card fraud detection

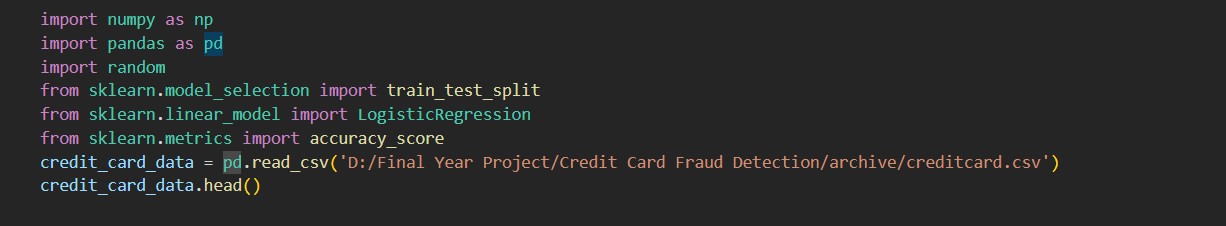


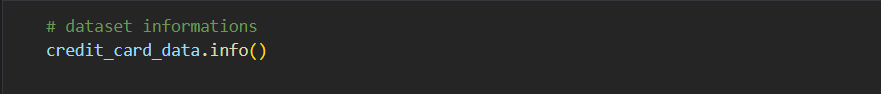
Fig 4

* Result from running code



Fig 5

* Dataset informations



* It shows column and rows



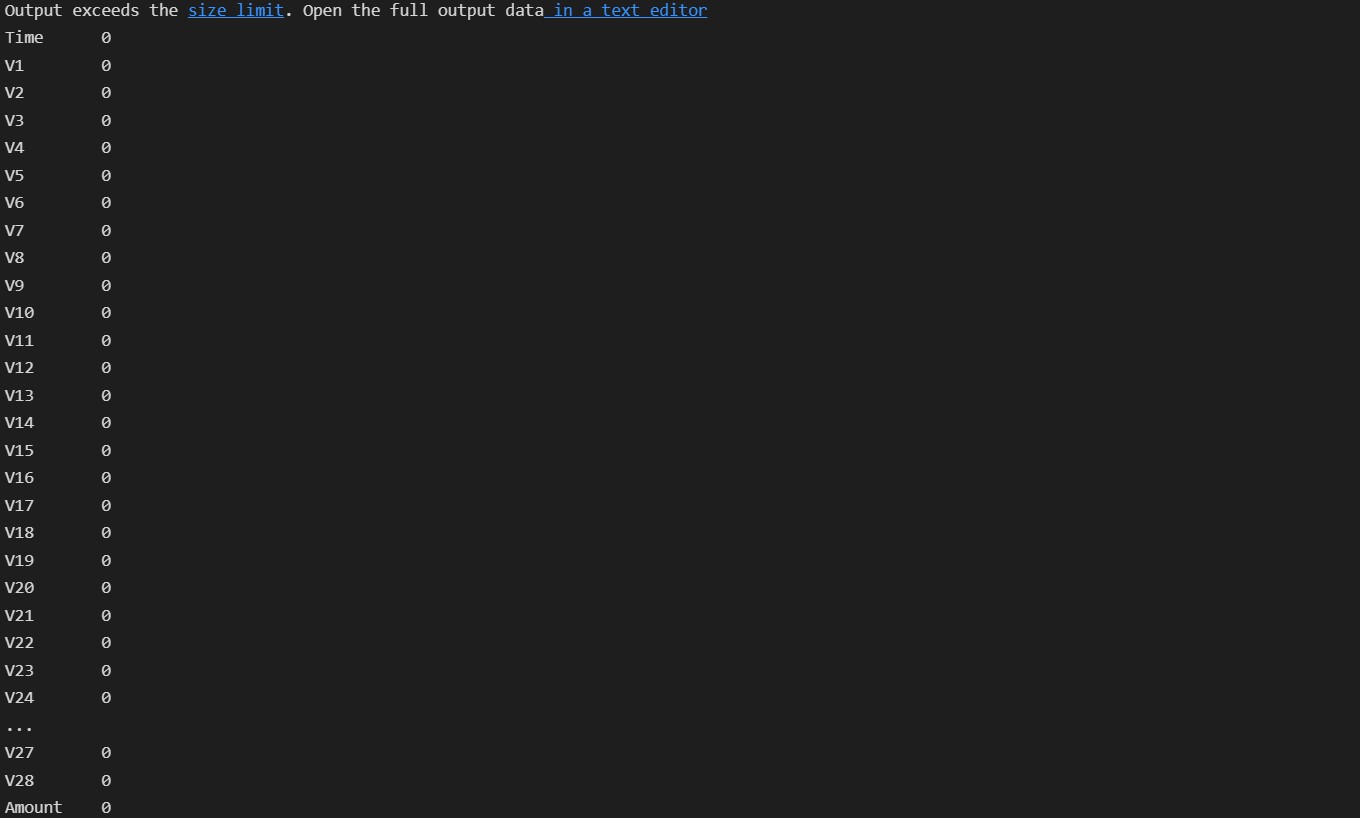
Fig 6

* checking the numbers of missing values in each column



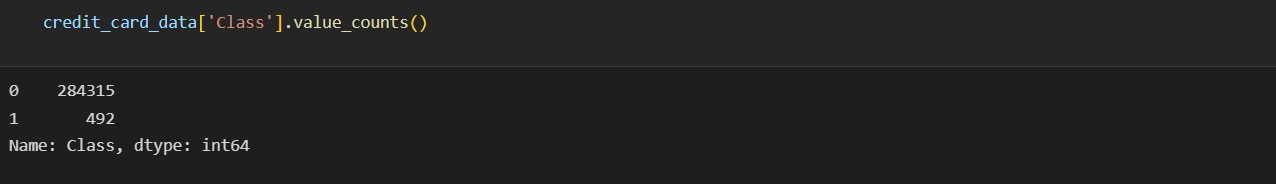
Fig 7

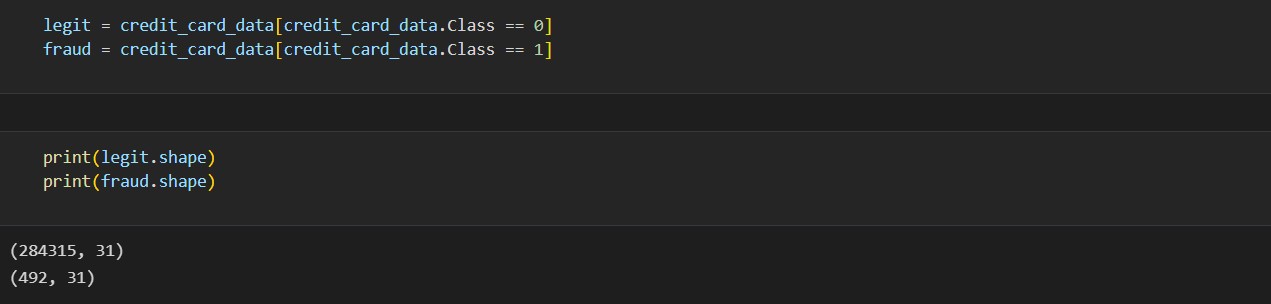
* Some values

 Fig 8

* fraud or non fraud values

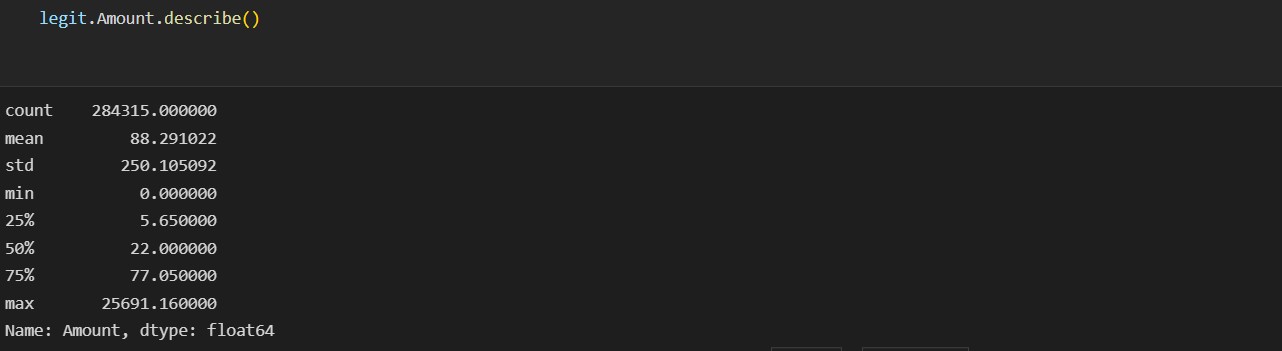
Fig 9

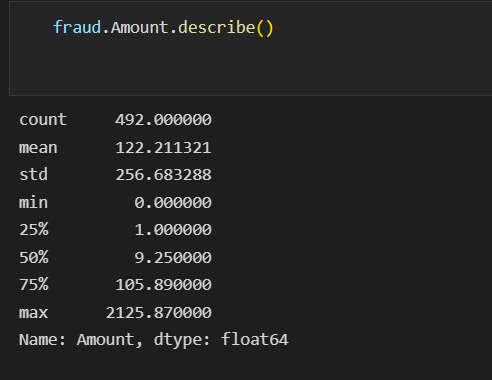


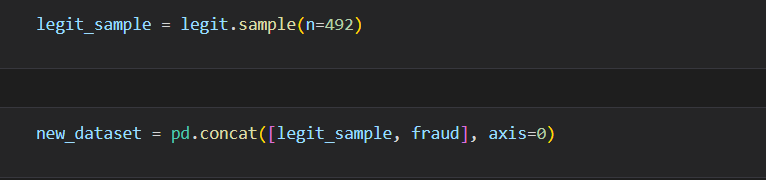
 Fig 10

* Minimum and maximum amount

9





Fig 12

* New\_dataset.head()



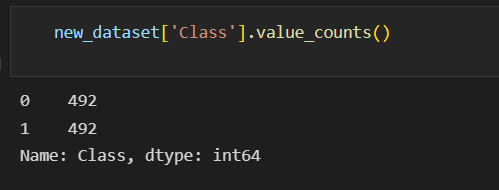


Fig 14

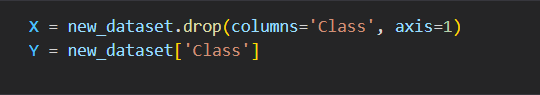
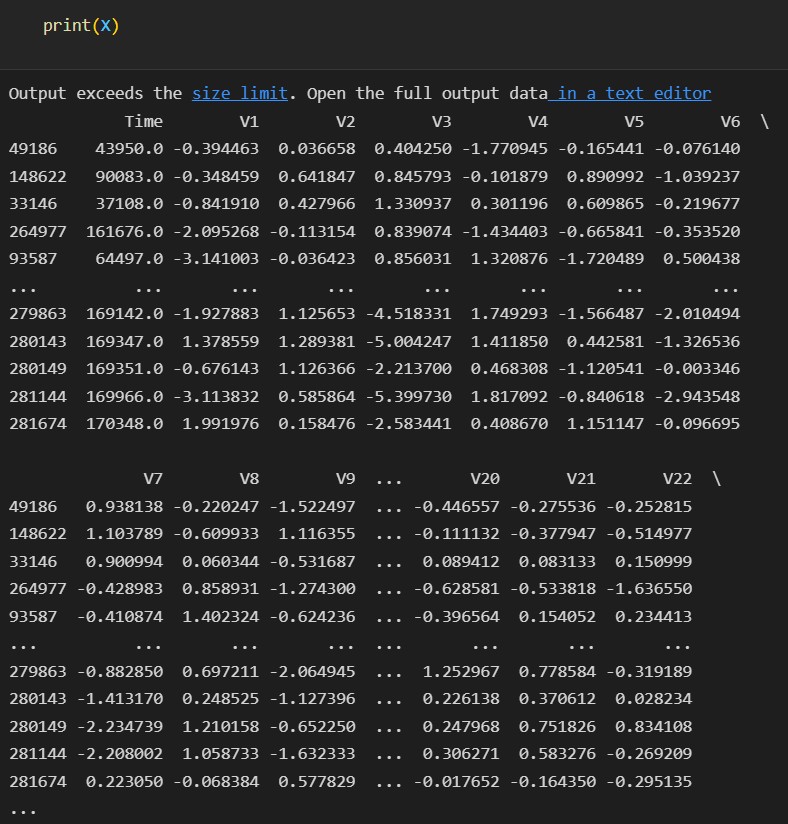


Fig 15

**Output**



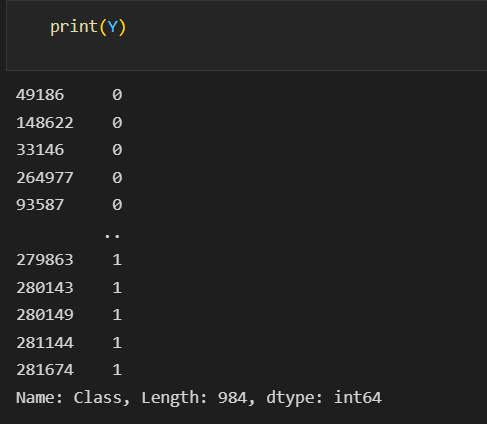
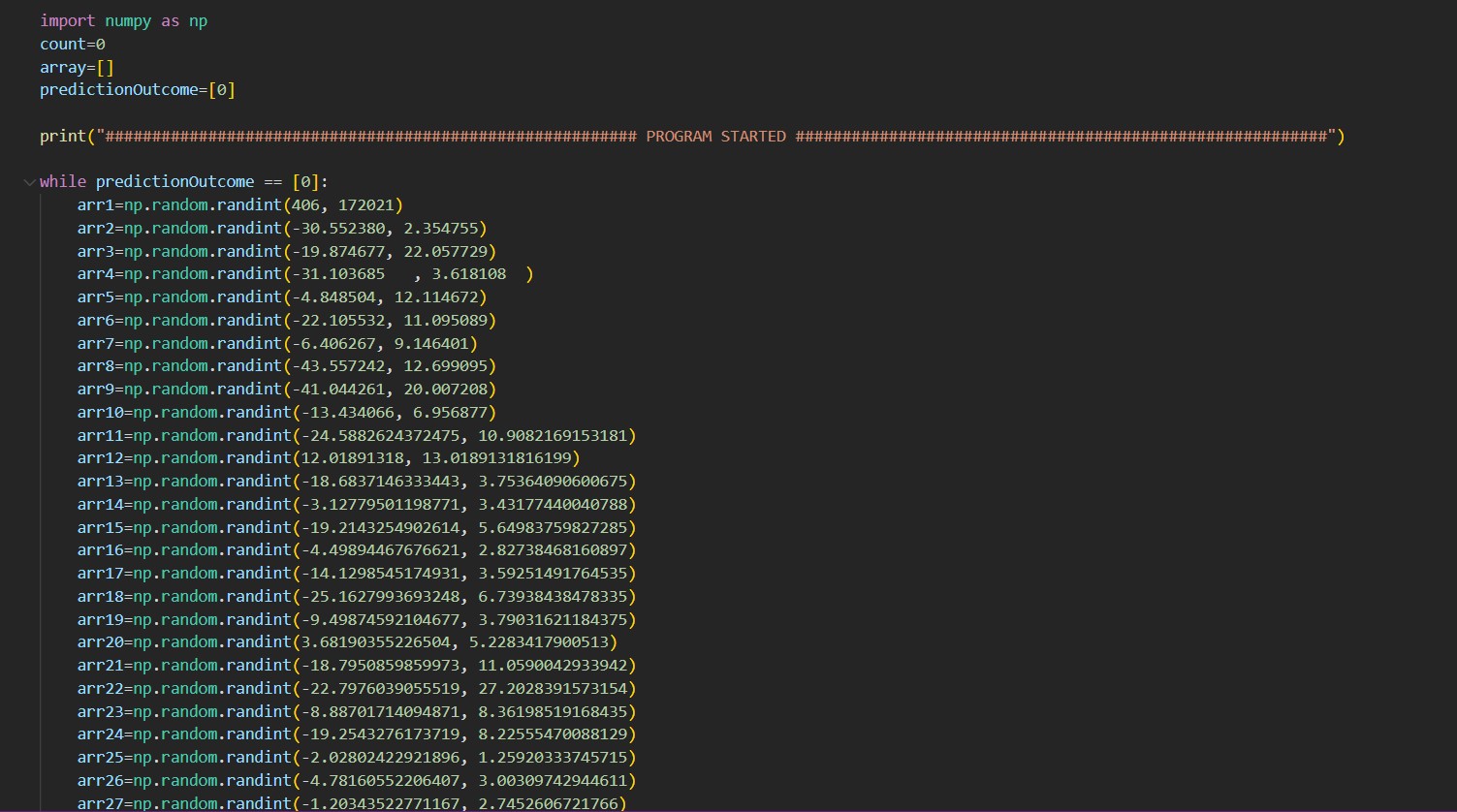
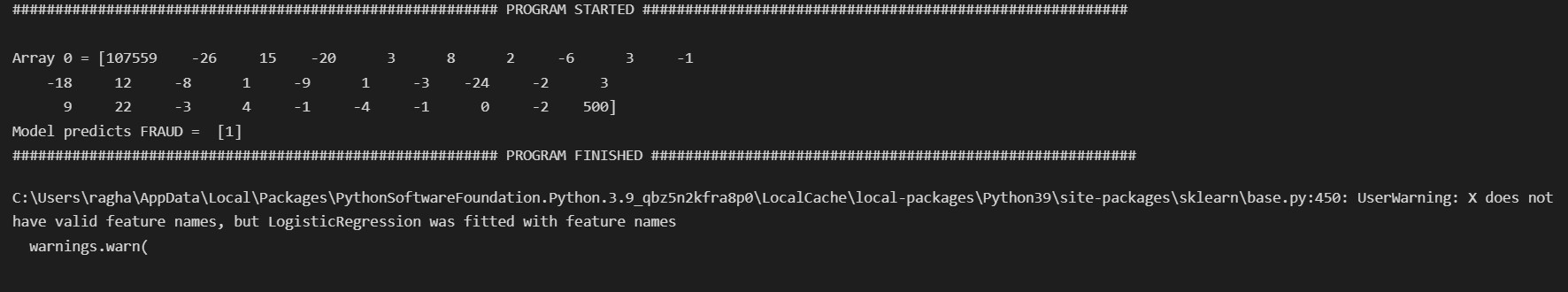


Fig 17

* Random data





**CHAPTER5 CONCLUSION**

We created a unique system for detecting fraud in this study, where consumers are classified based on Each cardholder has a profile created based on the transactions and behavioural patterns they make. Rating ratings are generated for each type of classifier after applying various classifiers to three different groups. The programme can change fast. to the transactional characteristics of new cards thanks to these dynamic parameter alterations. Then, a feedback system is used to solve the issue of concept drift. The Matthews Correlation Coefficient proved to be the most effective statistic for handling imbalance datasets. Other than MCC, there were choices. The classifiers performed better than previously when we attempted using the SMOTE to balance the dataset. The use of one-class classifiers, such as one-class SVM, is a different approach for addressing imbalance datasets. Finally, we identified the algorithms that delivered the best outcomes as being random forest, decision tree, and logistic regression..

**Reference**

* + 1. <https://www.gemini.com/cryptopedia/>
    2. <https://www.leewayhertz.com/>
    3. <https://cointelegraph.com/defi-101>
    4. <https://www.forbes.com/>

# https://[www.kaggle.com/ntnu-testimon/paysim1/home](http://www.kaggle.com/ntnu-testimon/paysim1/home)