## ONLINE PAYMENTS FRAUD DETECTION



**UNIVERSITY OF ENGINEERING**

**&**

**MANAGEMENT, JAIPUR**

## Online Payments Fraud Detection

Submitted in the partial fulfillment of the degree of

**BACHELOR OF TECHNOLOGY**

In

**COMPUTER SCIENCE & ENGINEERING**

Under

**UNIVERSITY OF ENGINEERING & MANAGEMENT, JAIPUR**

BY

**Pritam Rajak**

**University Roll no: 12020002001056**

**University Registration no: 204202000200044**

UNDER THE GUIDANCE OF

**Prof. Sagarika Ghosh**

COMPUTER SCIENCE & ENGINEERING



UNIVERSITY OF ENGINEERING & MANAGEMENT, JAIPUR

**Approval Certificate**

This is to certify that the project report entitled “**Online Payments Fraud Detection**” submitted by **Pritam Rajak** (Roll:**12020002001056**) in partial fulfillment of the requirements of the degree of **Bachelor of Technology** in **Computer Science & Engineering** from **University of Engineering & Management, Jaipur** was carried out in a systematic and procedural manner to the best of our knowledge. It is a bonafide work of the candidate and was carried out under our supervision and guidance during the academic session of 2011-2015.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Prof. Sagarika Ghosh**

Project Guide, Assistant Professor (CSE)

UEM, JAIPUR

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Prof. Mrinal Kanti Sarkar Prof. A Mukherjee**

HOD (CSE) Dean

UEM, JAIPUR UEM, JAIPUR

**ACKNOWLEDGEMENT**

The endless thanks goes to Lord Almighty for all the blessings he has showered onto me, which has enabled me to write this last note in my research work. During the period of my research, as in the rest of my life, I have been blessed by Almighty with some extraordinary people who have spun a web of support around me. Words can never be enough in expressing how grateful I am to those incredible people in my life who made this thesis possible. I would like an attempt to thank them for making my time during my research in the Institute a period I will treasure. I am deeply indebted to my research supervisor, Professor Sagarika Ghosh me such an interesting thesis topic. Each meeting with her added in valuable aspects to the implementation and broadened my perspective. She has guided me with her invaluable suggestions, lightened up the way in my darkest times and encouraged me a lot in the academic life.

Pritam Rajak

**ABSTRACT**

The technique of sensor fusion addresses the issues relating to the optimality of decision-making in the multiple-sensor framework. The advances in sensor fusion enable to perform intrusion detection for both rare and new attacks. This thesis discusses this assertion in detail, and describes the theoretical and experimental work done to show its validity. The attack-detector relationship is initially modeled and validated to understand the detection scenario. The different metrics available for the evaluation of intrusion detection systems are also introduced. The usefulness of the data set used for experimental evaluation has been demonstrated. The issues connected with intrusion detection systems are analyzed and the need for incorporating multiple detectors and their fusion is established in this work. Sensor fusion provides advantages with respect to reliability and completeness, in addition to intuitive and meaningful results. The goal for this work is to investigate how to combine data from diverse intrusion detection systems in order to improve the detection rate and reduce the false-alarm rate. The primary objective of the proposed thesis work is to develop a theoretical and practical basis for enhancing the performance of intrusion detection systems using advances in sensor fusion with easily available intrusion detection systems. This thesis introduces the mathematical basis for sensor fusion in order to provide enough support for the acceptability of sensor fusion in performance enhancement of intrusion detection systems.

# Table of Contents

[Table of Contents](#_3znysh7) **1**

[Figure list](#_6dqe2bk3iqnk) **2**

[**1. CHAPTER**](#_tyjcwt) **3**

[INTRODUCTION](#_3dy6vkm) 3

[1.1 online payment system.](#_1t3h5sf) 3

[1.2 online payment fraud](#_4d34og8) 3

[1.3 Types of payment :](#_26in1rg) 3

[1.4 Online Payments Fraud Detection with Machine Learning](#_jj1rqn1f4q1) 4

[**2. CHAPTER**](#_txjyhc6pf1t2) **4**

[LITERATURE REVIEW](#_ymaf63lio7yw) 4

[2.1 Online Payments Fraud Detection with Machine Learning](#_abtmoh9t5myr) 5

[**3. CHAPTER**](#_8rwuhtjqb9fo) **6**

[METHODOLOGY](#_3ozkfh9vhhzg) 6

[3.1 Algorithm used in this project :](#_rufva1d2gfyu) 6

[3.2 flow chart of my project](#_otcr64ml5g3e) 7

[3.3 Decision tree classifier](#_g85xgqhnmwq2) 7

[3.3.1 Use of decision tree](#_z337ya) 8

[3.3.2 Decision tree terminologies](#_3j2qqm3) 8

[3.4 logistic regression](#_o99hhu6j0xcw) 9

[3.4.1 logistic regression algorithm](#_aspspzdumsy0) 9

[3.4.2 use of logistic regression](#_4skrqma4p8qe) 9

[3.4.3 logic of logistic regression](#_sjk7l9nhga3h) 9

[3.4.4 logistic regression equation](#_81n8n4lktv9d) 10

[**RESULTS & DISCUSSION**](#_1za7mt4gsa9z) **12**

[**CONCLUSION & FUTURE SCOPE**](#_7f8gcqd7fyeg) **14**

[**APPENDIX**](#_qsva5q9h0yst) **15**

[**BIBLIOGRAPHY**](#_f0cr2d5ddq40) **17**

# Figure list

Figure title Figure no.

Figure 1: flow chart 6

Figure 2: decision tree flow diagram 7

Figure 3: Logistic regression 8

Figure 4: Decision tree accuracy 10

Figure 5: Logistic regression accuracy 10

Figure 6: confusion matrix 10

Figure 7: result output 11

Figure 8: code 14

# CHAPTER

## INTRODUCTION

### online payment system.

Electronic Payments entail the transfer of funds through electronic or digital mediums. You can choose from different e-payment methods like mobile wallets, bank cards, mobile banking, etc. E-payments are quick and efficient, and the fund transfer typically takes place instantly.

### online payment fraud

The introduction of online payment systems has helped a lot in the ease of payments. But, at the same time, it increased in payment frauds. Online payment frauds can happen with anyone using any payment system, especially while making payments using a credit card. That is why detecting online payment fraud is very important for credit card companies to ensure that the customers are not getting charged for the products and services they never paid for.

### Types of payment :

1. **Card present**

When the physical card is used to buy something in a shop, restaurant, bar or market.

**2. Card not present**

When the card details are used, the physical card itself isn’t passed from the buyer to seller. CNP payments can happen by mail or on the phone, but mainly happen online

# CHAPTER

## LITERATURE REVIEW

### Online Payments Fraud Detection with Machine Learning

To identify online payment fraud with machine learning, we need to train a machine learning model for classifying fraudulent and non-fraudulent payments. For this, we need a dataset containing information about online payment fraud, so that we can understand what type of transactions lead to fraud. For this task, I collected a dataset from Kaggle, which contains historical information about fraudulent transactions which can be used to detect fraud in online payments. Below are all the columns from the dataset I’m using here:

1. step: represents a unit of time where 1 step equals 1 hour
2. type: type of online transaction
3. amount: the amount of the transaction
4. nameOrig: customer starting the transaction
5. oldbalanceOrg: balance before the transaction
6. newbalanceOrig: balance after the transaction
7. nameDest: recipient of the transaction
8. oldbalanceDest: initial balance of recipient before the transaction
9. newbalanceDest: the new balance of recipient after the transaction
10. isFraud: fraud transaction

I hope you now know about the data I am using for the online payment fraud detection task.

# CHAPTER

## **METHODOLOGY**

### Algorithm used in this project :

For this project I am using a Decision tree algorithm and logistic regression to build the model.

The basic info about Decision tree algo:

The Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification problems too.

The basic info about logistic regression:

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.

The general motive of using Decision Tree and logistic regression is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data(training data).

### flow chart of my project

### 

figure 1: flow chart

The online payment fraud detection data is the dataset taken from kaggle. The data is separated into two part test and train data. Then the model is built with the test data. Two classifiers are used for this project. The prediction is done according to the train model. and the train model will predict whether the given data is fraud or not fraud.

### Decision tree classifier

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

### Use of decision tree

There are various algorithms in Machine learning, so choosing the best algorithm for the given data-set and problem is the main point to remember while creating a machine learning model. Below are the two reasons for using the Decision tree:

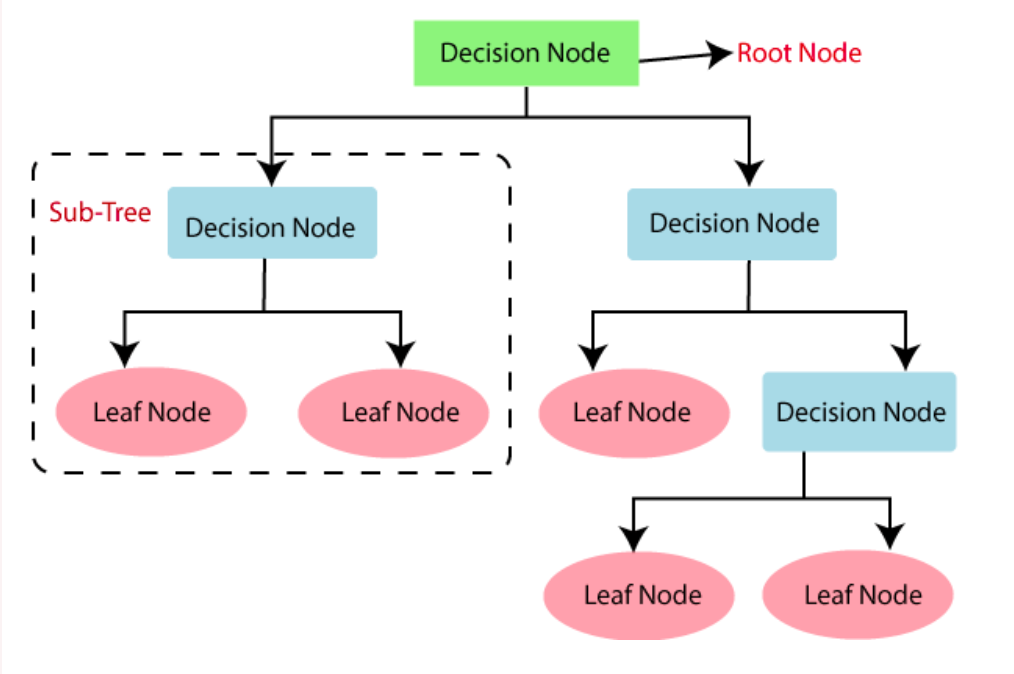
* a. Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
* b. The logic behind the decision tree can be easily understood because it shows a tree-like structure.

Figure 2: decision tree flow diagram

### Decision tree terminologies

* Root Node: Root node is from where the decision tree starts. It

represents the entire dataset, which further gets divided into two or more homogeneous sets.

* Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
* Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
* Branch/Sub Tree: A tree formed by splitting the tree.
* Pruning: Pruning is the process of removing the unwanted branches from the tree.
* Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

### logistic regression

### logistic regression algorithm

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.

### use of logistic regression

Logistic regression is commonly used for prediction and classification problems. Some of these use cases include: Fraud detection: Logistic regression models can help teams identify data anomalies, which are predictive of fraud.

### logic of logistic regression

Figure 3: logistic regression diagram

### logistic regression equation

# The Logistic regression equation can be obtained from the Linear Regression equation. The mathematical steps to get Logistic Regression equations are given below:

# We know the equation of the straight line can be written as:

# 𝑦 = 𝛽0 + 𝛽1\* 𝑥

# In Logistic Regression y can be between 0 and 1 only, so for this let's divide the above equation by (1-y):

# 

# The equation of the sigmoid function is:

# The sigmoid curve obtained from the above equation is as follows:

# RESULTS & DISCUSSION

Accuracy output:

| MODEL | ACCURACY |
| --- | --- |
| Decision tree classification | 0.9997 |
| Logistic regression | 0.9995 |

Tabel 1: accuracy table

The above table gives the accuracy of the following model. By the analysis of the table we can say that decision tree calcification is more better than Logistic regression .

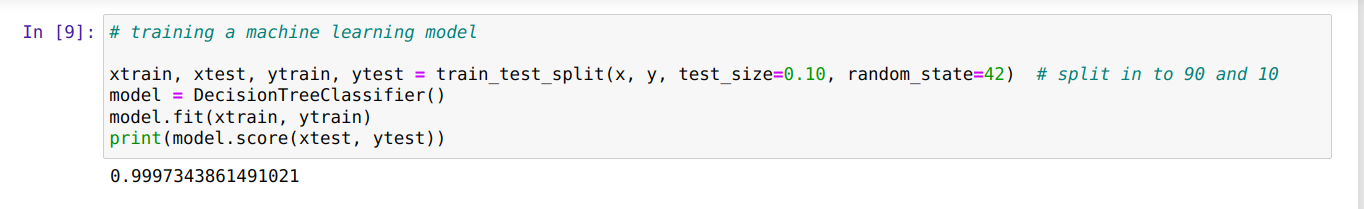


Figure 4: decision tree accuracy

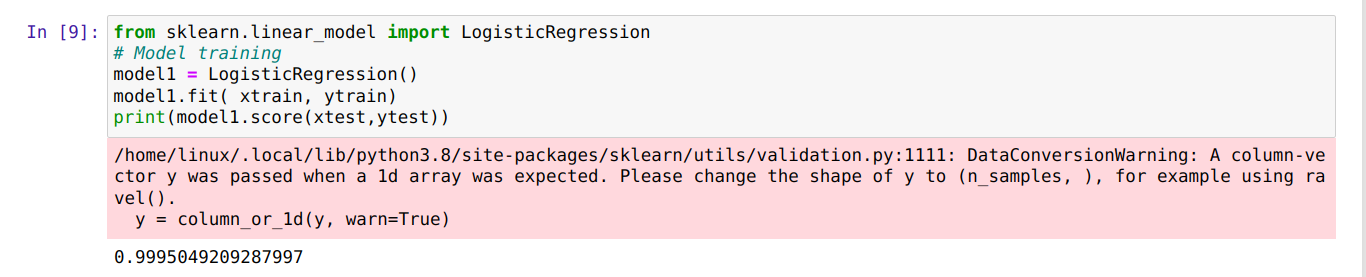


Figure 5: logistic regression accuracy

confusion matrix

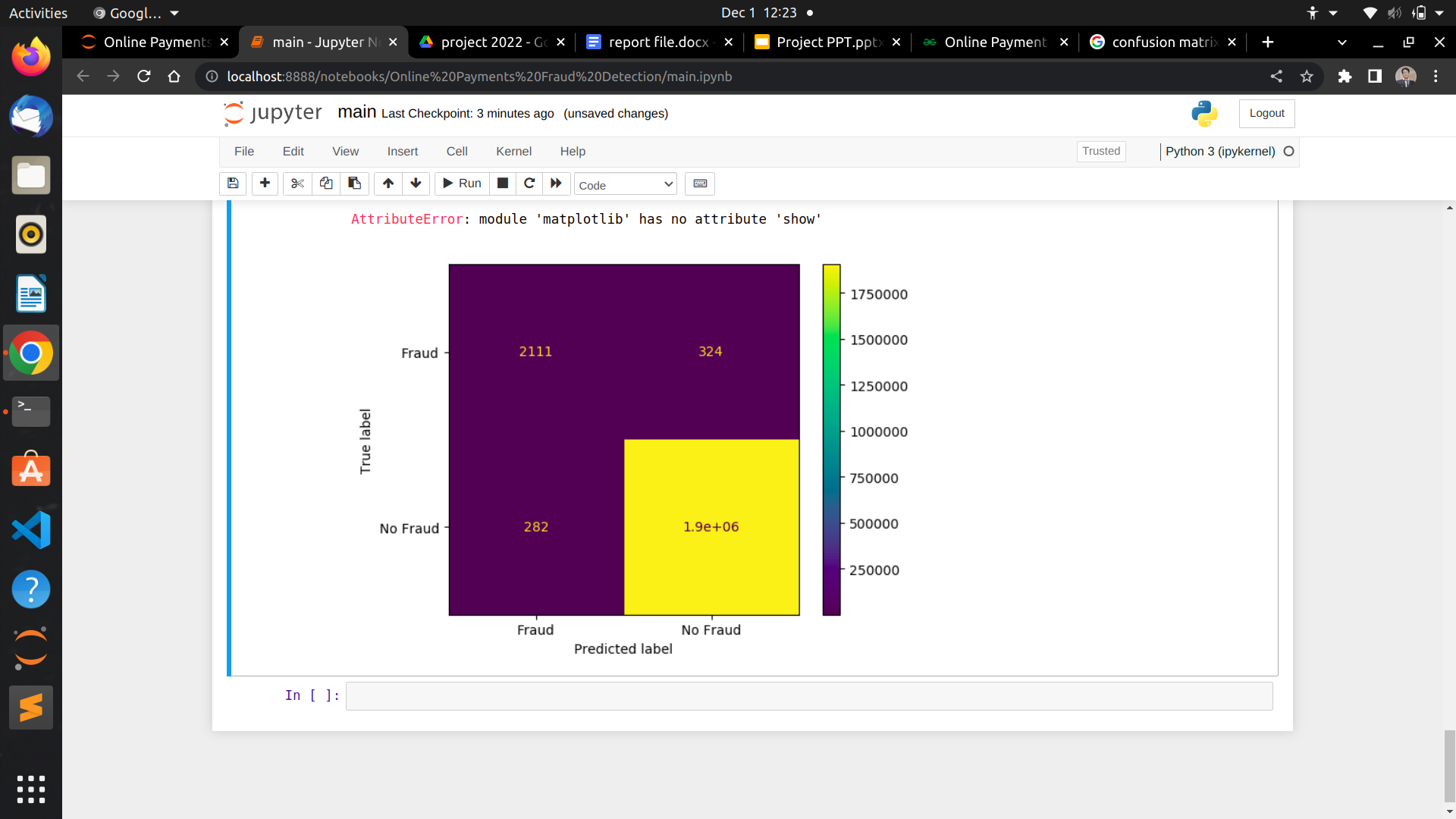
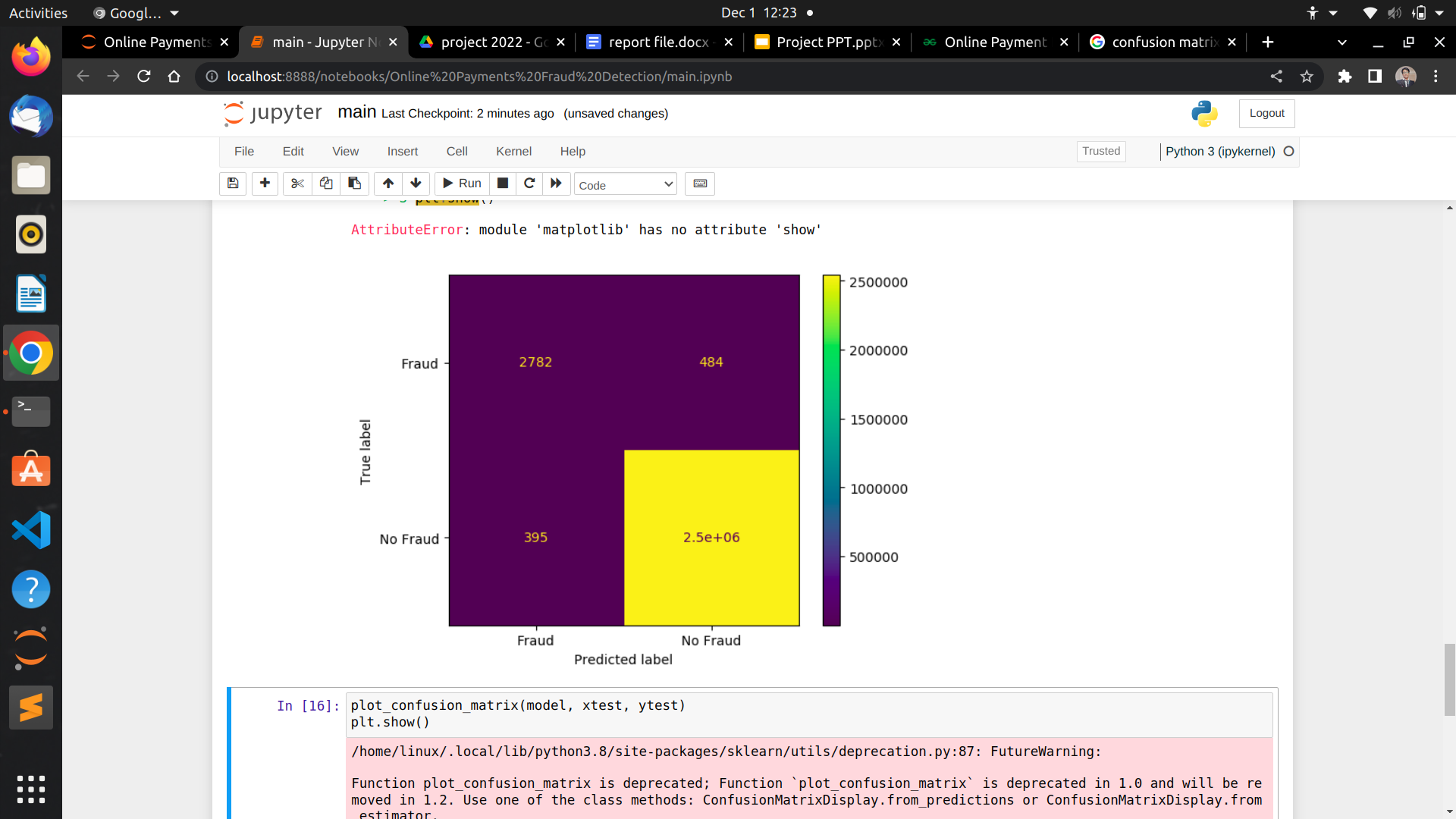
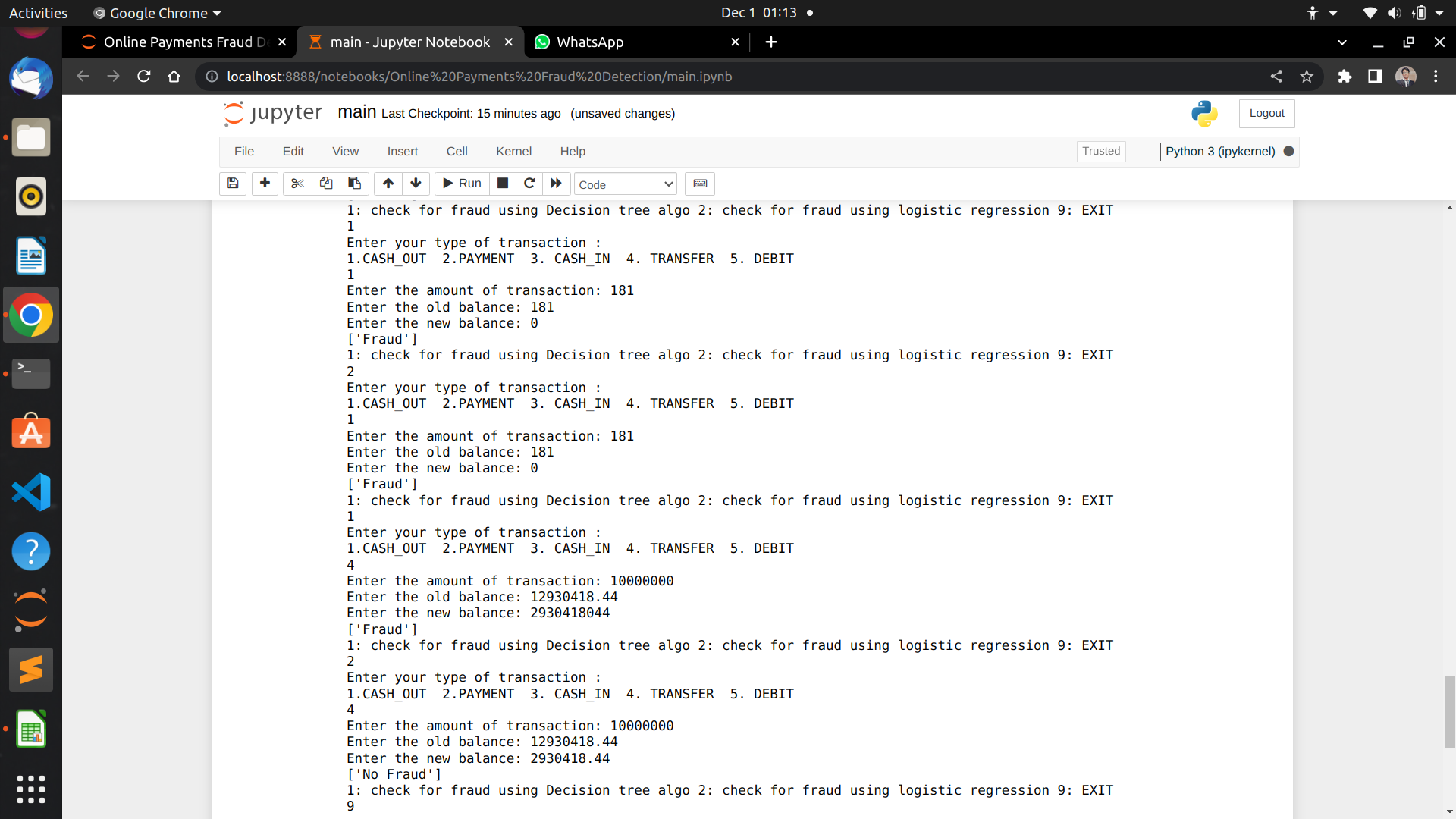


Figure 6: confusion matrix

output result:

Figure 7: result output.

It is the sample of output of this project, predicting the output using the train model.

We are getting output in both the algorithms but the estimate out we are getting from the decision tree algorithm .

The accuracy of the decision tree algorithm is 99% and providing the estimated output so we can say for this project the decision tree algorithm is best among this.

# 

# 

# CONCLUSION & FUTURE SCOPE

In this project I have learned many things such as Decision tree algorithm and Logistic algorithm, their concept , implementation of the algorithm about online payment frauds. In this project I got accuracy of 99% with the test data. The estimated value I have received from the Decision tree algorithm. The logistic regression gives output but in some of the cases it gives wrong output. but analyzing the outputs I can conclude that the decision tree algorithm is still now best for online payment fraud detection.

The future scope is to build a web extension for this model and an application for predicting the fraud and notify the user on time.

# 

# APPENDIX

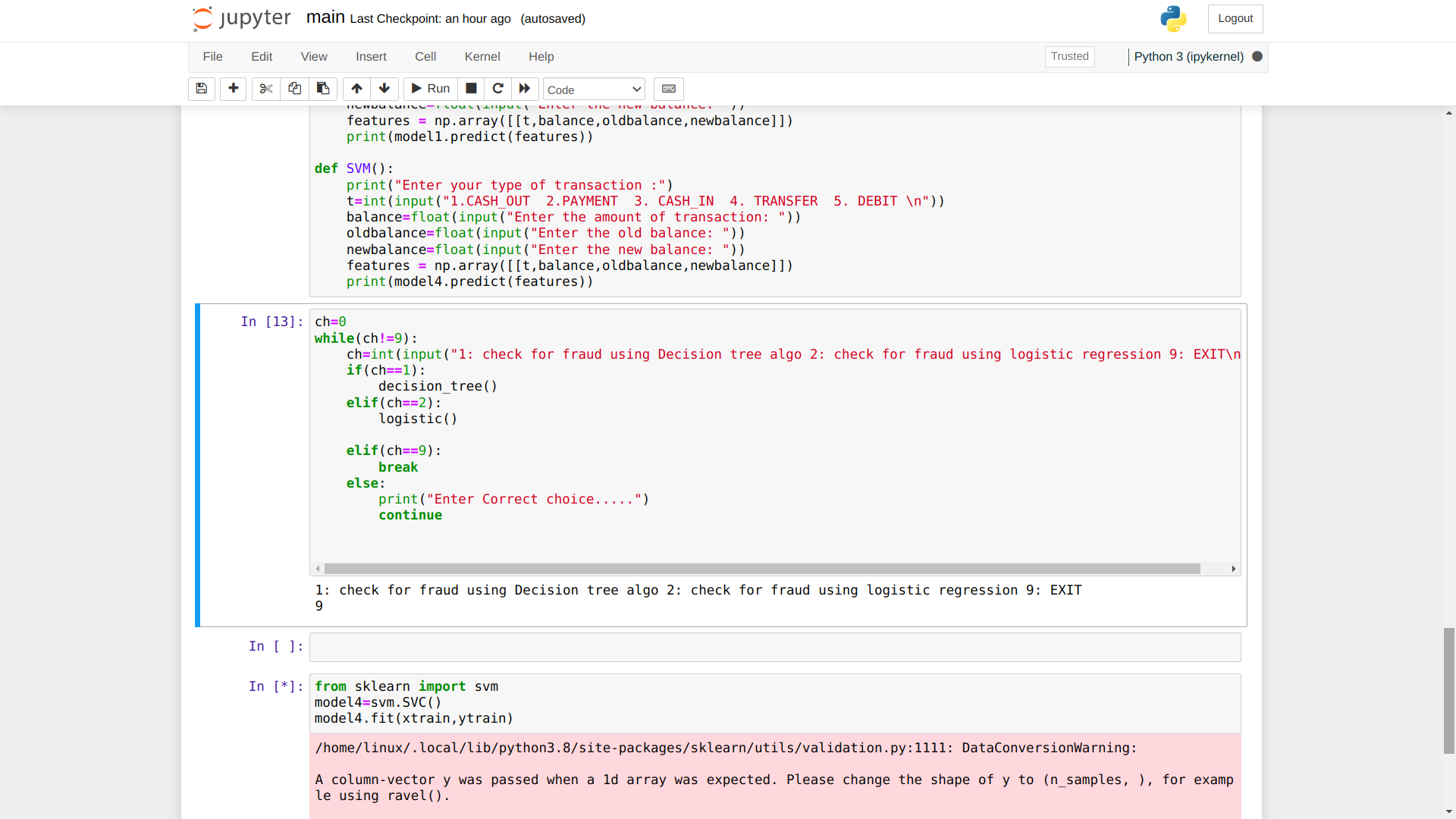
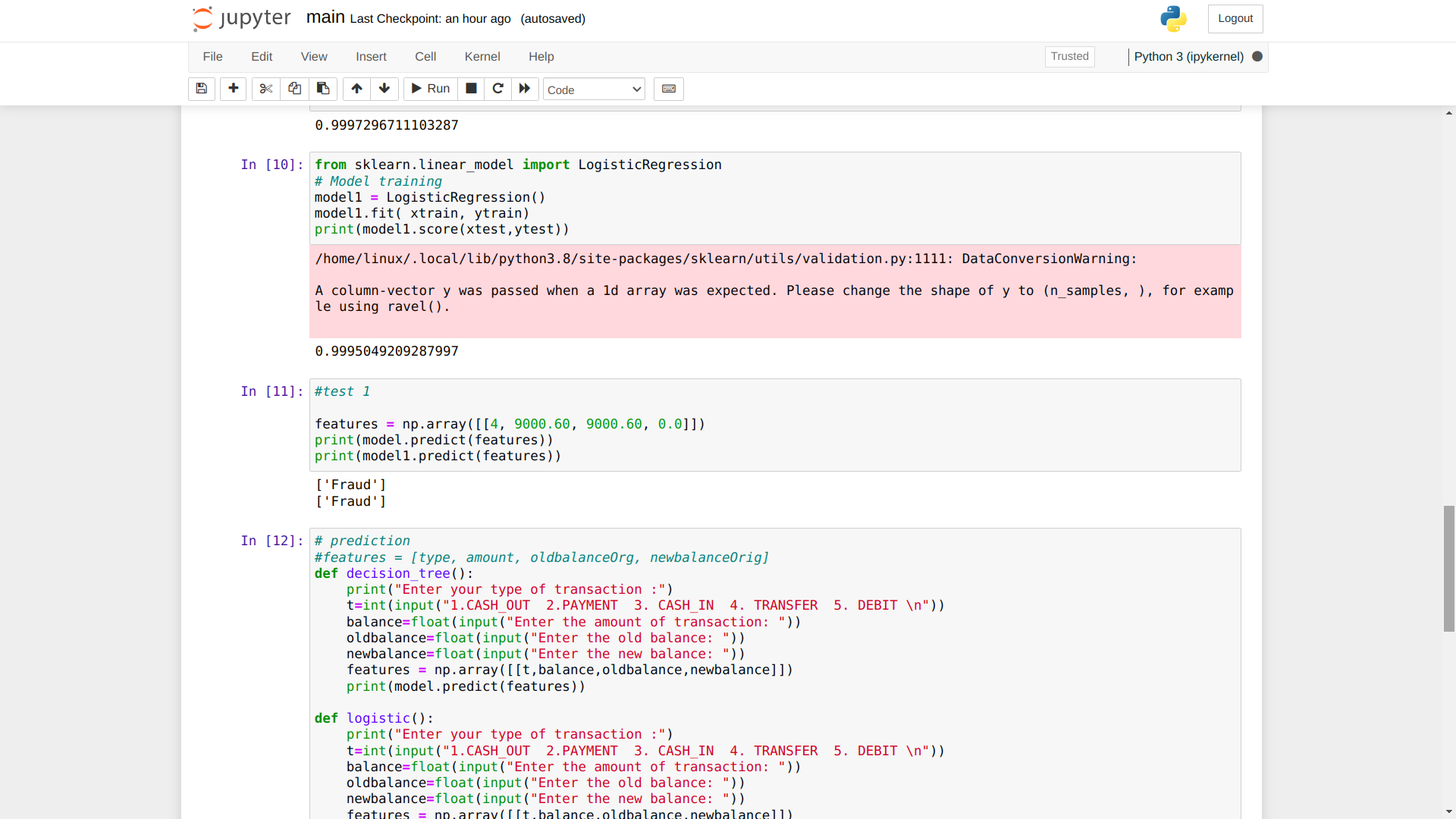
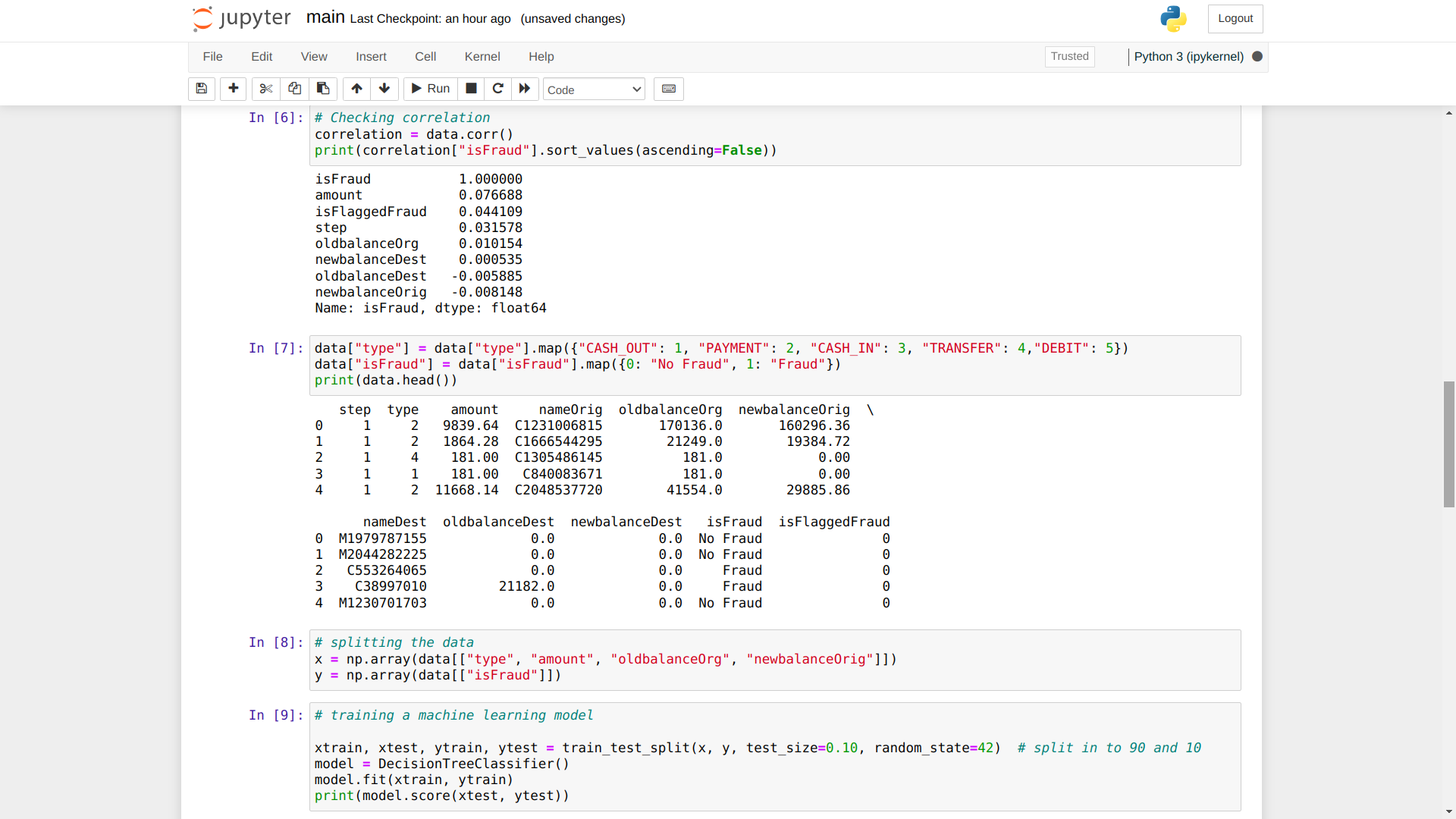
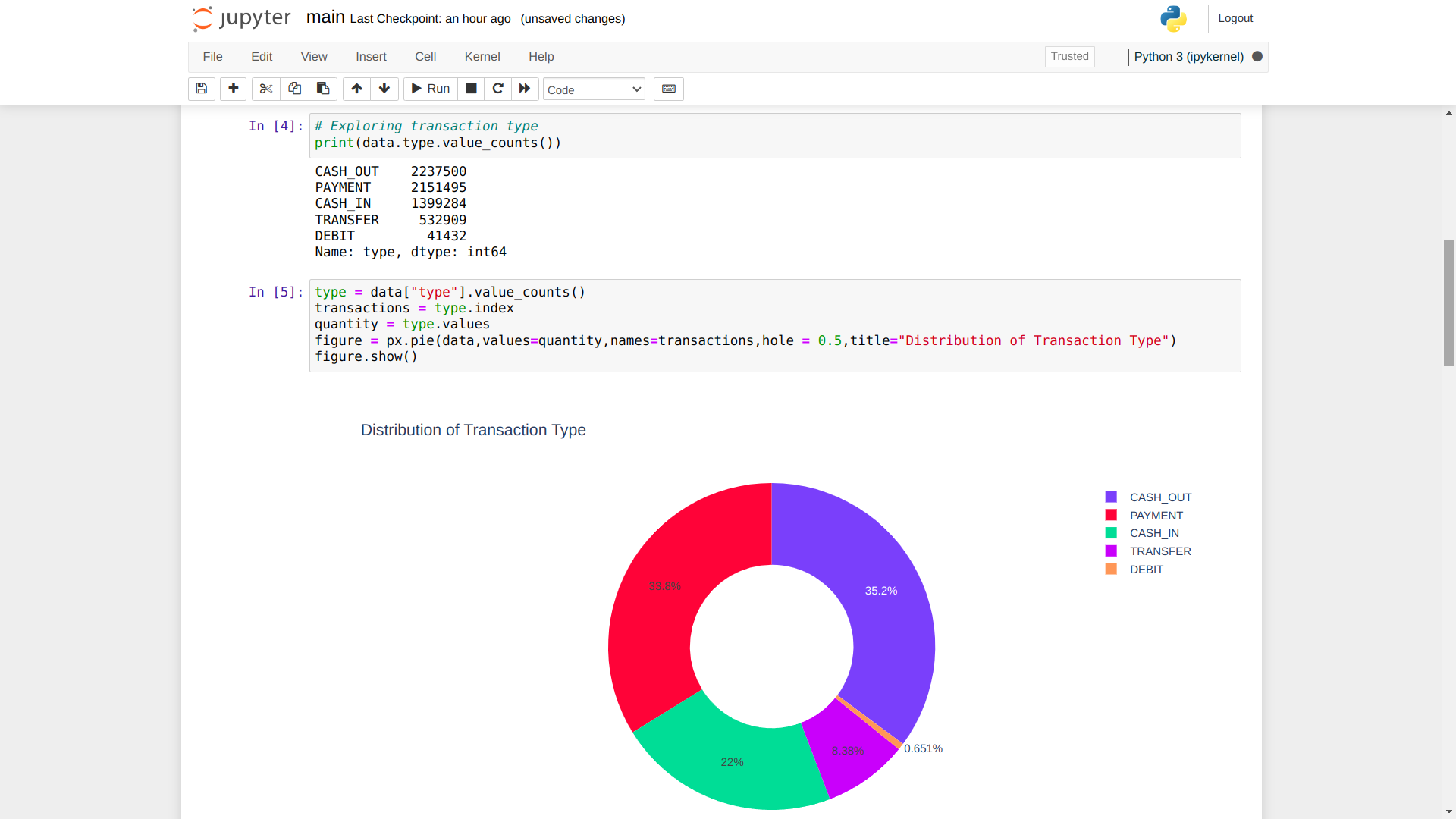
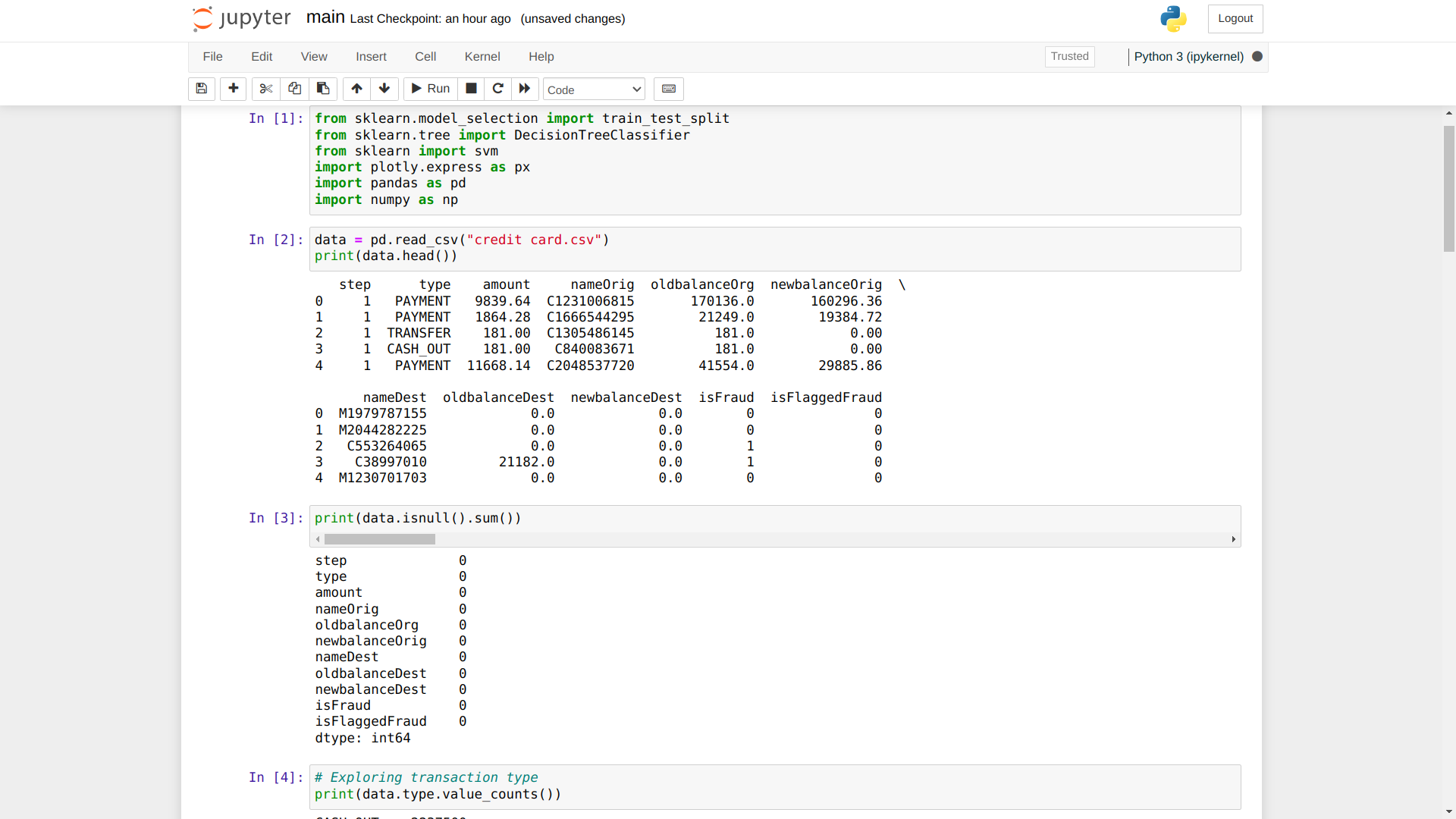


Figure 8 :code

# 

# BIBLIOGRAPHY

[1] <https://dataaspirant.com/credit-card-fraud-detection-classification-algorithms-python/>

[2] <https://youtu.be/coOTEc-0OGw>

[3] [https://www.sciencedirect.com/topics/computer-science/logistic-regression](https://www.)

[4]<https://www.researchgate.net/publication/354937786_Online_Transaction_Fraud_Detection_System_Based_on_Machine_Learning>