**Project Report**

**Credit Card Fraud Detection**.

**Aim of the Study**

With businesses moving online, fraud and abuse in online systems is constantly increasing as well. Traditionally, rule-based fraud detection systems are used to combat online fraud, but these rely on a static set of rules created by human experts. This project uses machine learning to create models for fraud detection that are dynamic, self-improving and maintainable. Importantly, they can scale with the online business.

The goal of this project is to help a bank to detect fraudulent credit card transactions. We think this is a major issue for banks for four main reasons:

**The loss of money:** credit card replacement, refund, administrative costs... Banks are losing huge amounts of money each year because of undetected frauds

**The customer experience:** having your money stolen sucks, and this is why banks need to detect frauds as soon as possible, before the transaction is confirmed.

**Justice:** as part of the economical and the societal environment of a country, we think that banks are willing to fight against fraud and theft.

**Political credibility:** banks' reputation with regulation authorities is at stake here.

**Selection of the Response Variable**

First of all, response variable refers to the dependent variable whose variation depends on another variable. For example, in my project of credit card fraud detection the response variable is the number of transactions.

**Choice of Factors and Levels**

Selecting factors is important in Quantitative Finance. Factor is a special case of vector that is solely used for representing nominal variables. In the credit card fraud detection dataset, we are building, we might use a Factor to represent transactions, because it uses two Categories/levels. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependent cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

**Choice of Experimental Design**

In general, designs that are true experiments contain three key features. So, while selecting ant project in machine learning these three key features are much more important.

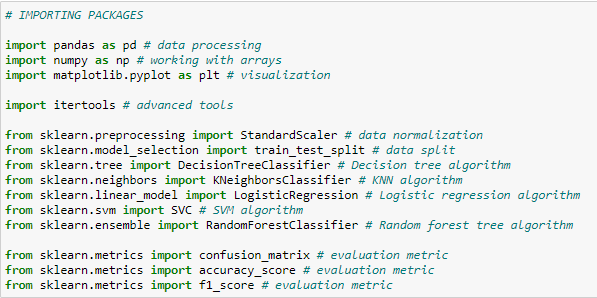
Their names are following down:

* Independent &dependent (response) variables.
* Pretesting (Before) and post testing.
* Experimental and control groups.

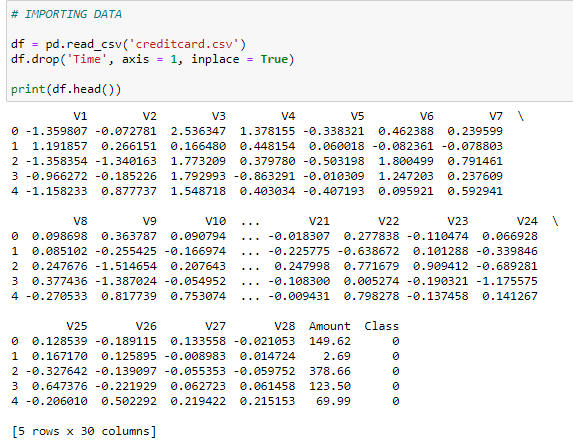
**Performing the Experiment**

The following steps I took for my project

**Step 1: Importing the required packages into our python environment.**

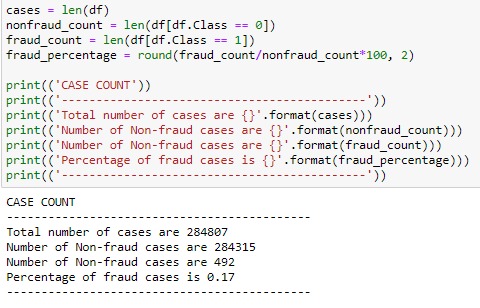


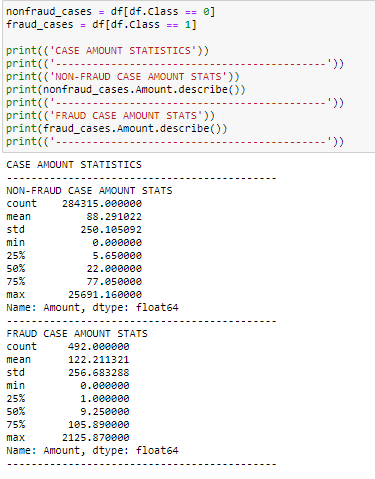
**Step 2: Importing the data**



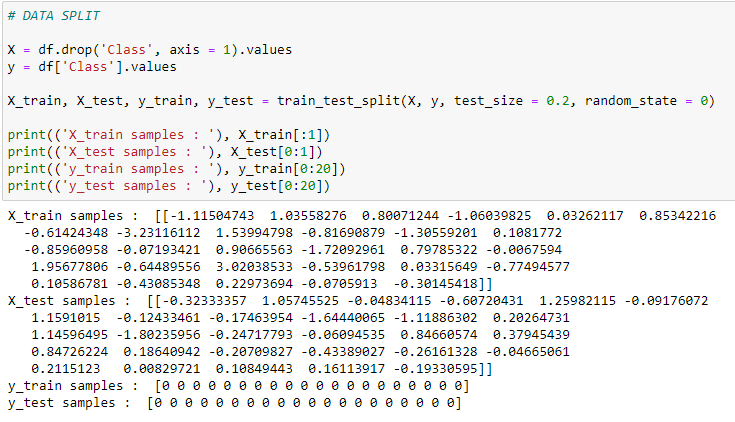
**3.Processing the data to our needs and Exploratory Data Analysis**

* **Data Processing and EDA**





**4.Feature Selection and Data Split**



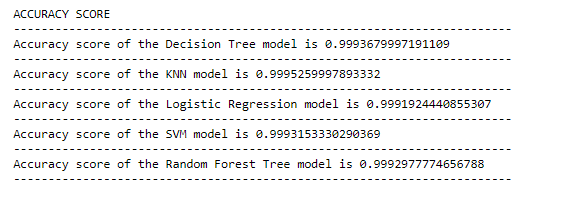
**5.Building 5 types of classification models**

* Decision Tree
* K-Nearest Neighbors (KNN)
* Logistic Regression
* Support Vector Machine (SVM)
* Random Forest



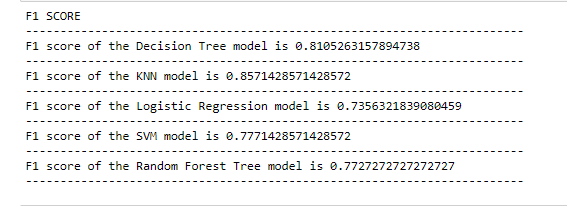
**6.Evaluating the created classification models using the evaluation metrics**

* Accuracy score
* **Accuracy score = No of correct predictions / Total no. of predictions**

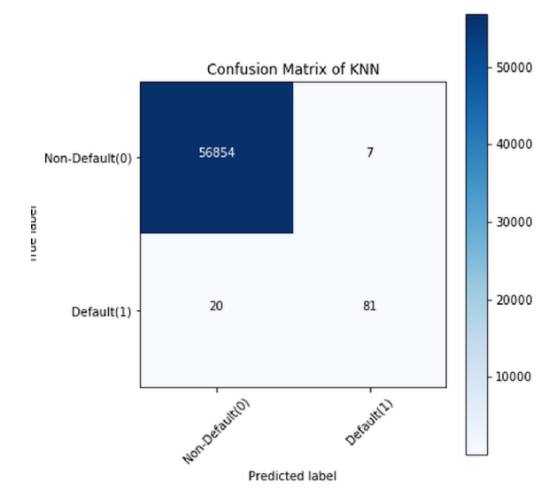
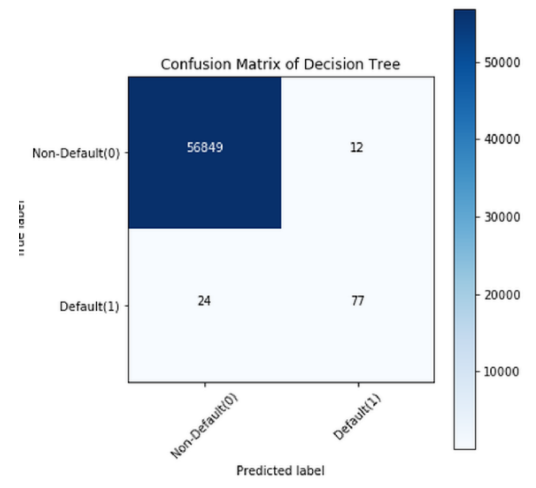


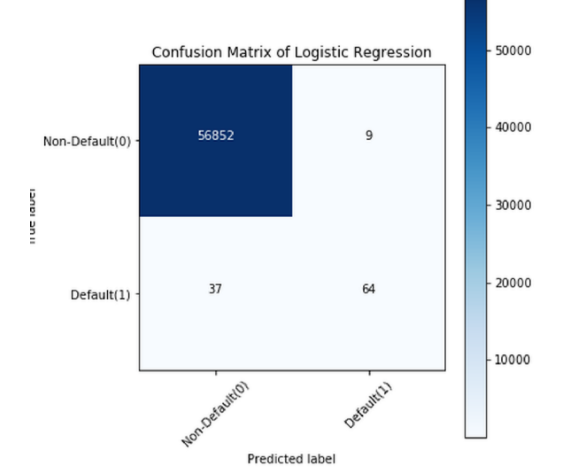
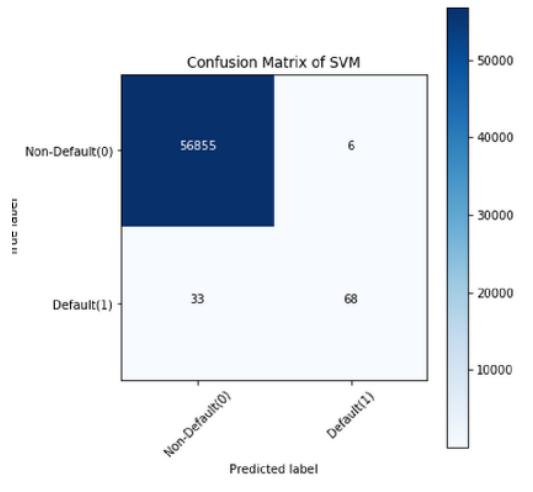
**F1 Score**

* **F1 score = 2((precision \* recall) / (precision + recall))**



**7.Confusion Matrix**



**DATASET**

For carrying out the credit card fraud detection, we will make use of the Card Transactions dataset that contains a mix of fraud as well as non-fraudulent transactions from Kaggle.

**Statistical Analysis of the Data**

First of all, we come to know that Statistics is the mathematical study of data. In context of machine Learning, Statistical Model is the use of statistics to build a representation of the data and then conduct analysis to infer any relationships between variables or discover insights. In view of my project statistical analysis refers to library implementation used for data classification and then analysis is on transactions.

**Conclusions and Recommendations**

We learnt how data can be analyzed and visualized to discern fraudulent transactions from other types of data. credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase. After a whole bunch of processes, we have successfully built five different types of classification models After that, we have evaluated each of the models using the evaluation metrics and chose which model is most suitable for the given case.