# Introduction

This file contains the results and process we operated a skewed Dataset. By skewed dataset we mean the data corresponding different classes were very skewed. So, the dataset selected for this purpose was ‘credit\_card.csv’. Where based on features we have to predict whether a transaction was Fraud or not.

The columns of the dataset are:

A screenshot of a computer

Description automatically generated

A screenshot of a calculator

Description automatically generated

Where, ‘Time’ represents the time it took for the transaction. ‘Amount’ represents the amount of transaction happened and ‘Class’ represents whether the transaction was ‘Fraud:1’ or not ‘Not Fraud: 1’.

# Dataset

Description of Dataset:

A table with numbers and symbols

Description automatically generated

A table with numbers and symbols

Description automatically generated

Null values:

A close-up of a computer code

Description automatically generated

Total Columns:

A group of numbers and symbols

Description automatically generated

Total number associated with different classes:

A computer screen shot of a number

Description automatically generated

As could be seen here that the data is highly skewed as the percentage of distribution between classes is very different. What problem it brings is if our proposed algorithm predicts class 0 (No Frauds) every time then also it is giving an accuracy of 99.83 % here. So we need to sort that as well.

A black rectangular object with white text

Description automatically generated

A graph of a graph

Description automatically generatedPlotting continuous data:

As could be seen that the distribution of continuous features is not normal hence, we need to transform those to more continuous form. We would use transformations for that.