Machine Learning Nanodegree Capstone Project

# Credit Card Fraud Detection using Supervised learning

Proposed by : Ahmed Roshdy

**Project overview:**

In this project I will build a credit card fraud detection solution for the credit card companies which shall be able to recognize fraudulent credit cards transactions .

**Domain Background**:

This project is proposed on the domain of supervised learning, supervised learning refers to a class of systems and algorithms that determine a predictive model using data points with known outcomes.  The model is learned by training through an appropriate learning algorithm (such as [linear regression](https://deepai.org/machine-learning-glossary-and-terms/linear-regression), [random forests](https://deepai.org/machine-learning-glossary-and-terms/random-forest), or [neural networks](https://deepai.org/machine-learning-glossary-and-terms/neural-network)) that typically works through some optimization routine to minimize a loss or error function

**Problem Statement :**

The main objective of this project is to detect the fraud credit cards , Credit card companies shall be able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase

**Datasets and Inputs:**

**Data Exploration:**

**The dataset used in this project is produced by kaggle website and available to download from** [**https://www.kaggle.com/mlg-ulb/creditcardfraud**](https://www.kaggle.com/mlg-ulb/creditcardfraud)

**the dataset is labeled and consisting of 284807 rows and 31 columns , and our target feature that we should predict is the Class column , and if the value of the class is 0 that means it’s transaction without fraud , and if it is equal 1 that means it’s transaction with fraud.**

**Solution statement , algorithms and technique:**

The proposed solution to this problem is to apply supervised learning algorithm to detect if the new transaction is fraud or not , and we will use the logistic regression model , after a lot of trying I found that the logistic regression model give the best result.

**Benchmark Model :**

For the benchmark model, we will use logistic regression model.

**Evaluation Metrics:**

The evaluation metric for this problem is simply precision, recall , and, f1-score

**Project Design:**

• Preprocessing

we will show the data and see the shape of it

• Data splitting

Split the data into a training set and validation set with an 80-20 split.

• Model training and evaluation

* I will start with the simple model architecture first before training and evaluating it. Then splitting the data , fit the model , and test the model by using predict function

Analysis

**Data Exploration:**

The dataset used in this project is produced by kaggle website and available to download from [**https://www.kaggle.com/mlg-ulb/creditcardfraud**](https://www.kaggle.com/mlg-ulb/creditcardfraud)

the dataset is labeled and consisting of 284807 rows and 31 columns , and our target feature that we should predict is the Class column , and if the value of the class is 0 that means it’s transaction without fraud , and if it is equal 1 that means it’s transaction with fraud.

**Exploratory Visualization:**

As I said our target feature that we should predict is the Class column

Time V1 V2 Amount Class

0 0.0 -1.359807 149.62 0

1 0.0 1.191857 2.69 0

2 1.0 -1.358354 378.66 0

3 1.0 -0.966272 123.50 0

4 2.0 -1.158233 69.99 0

That’s sample of the column and sample of the rows