**Team Members:** Oyindamola Obisesan, Michael Olubode, Judah Odiachi, Adesoji Akanji

**Project Title:** Credit Card Fraud Detection Using Machine Learning Algorithms.

**Description of the problem:** Fraud detection is a series of activities that hinders money from being obtained by illegal means which can include identity theft or stolen credit card. Fraud is a wide spread activity that can occur in a large number of industry but the concentration in this work is the credit card companies.

The target in this work would be to determine if a transaction is fraudulent or not by considering certain variables from the dataset provided such as the location the transaction is being initiated, the IP address of the device, the historical transaction pattern or trend of the credit card owner, and the actual information of the transaction (how much?, what was the purchase?). These factors would be analyzed to be able to make the decision if the transaction should be initiated or declined.

The detection would be done by machine learning which would involve the analysis and exploration of the available data, development of a model that will provide the best classification for the target (1- fraud, 0-no fraud). The user information and pattern is run through the model to train it and some fraudulent transactions has be available to train the model to find patterns and behavior to classify the data as fraudulent or not.

The classifier in this case has to be very accurate in prediction true positive as false negatives can be very costly. False positives have a lower penalty/cost but it might be upsetting to customers if all legitimate transactions are labeled as fraudulent and being declined.

The data would most likely contain a lot of no fraud transactions which makes it an imbalanced data set and care would be taken during the analyses and performance prediction of the model by using appropriate model evaluation techniques.

**Type of the Problem:** This is a supervised learning. The training data set has transactions already labelled as fraudulent or non-fraudulent. The machine learning algorithms would be used to assign the data to the appropriate classifier label. Some of the algorithms that would be analyzed are support vector machine (SVM), logistic regression, random forest, KNN, XGBoost.

**Work Process Summary:** Data mining, feature selection, data visualization and exploration, developing training model, validate model using performance metrics such as confusion metrics, F1-score, precision, recall, log loss etc. to appropriately classify the data as accurately as possible. The aim is to detect the fraudulent activities while reducing the false negatives as much as possible and detecting as many true positives as possible.

**Data Sourcing:** The data from this project would be sourced from kaggle from a similar problem statement.