**Credit Card Fraud Detection:**

**Problem Statement:**

For many banks, retaining high profitable customers is first goal, however credit card fraud posses significant threat to this goal. This will affect their financial losses, trust and credibility of both banks and customers. With the rise in digital payments, the rise in fraudulent transactions is also increasing in different ways.

It has become necessity for the banks to mitigate the Credit Card fraud detection using Machine Learning models to monitor proactively and to have fraud mechanism in place. This helps the banks to reduce manual reviews, charges and fees.

Our goal in the project is to apply different Machine Learning algorithms in order to classify credit card fraud from a dataset of European transactions.

**Project Pipeline:**

Below is the brief summary of project pipeline

**Data Understanding:**

Given dataset contains 4932 frauds out of 284807 transactions occurred in 2 days of September 2013, Dataset contains only numerical inputs which are the results of a PCA Transformation.

Usually in any dataset we would perform EDA for the available features to identify the relationship and hidden patterns. Here, in this case the data set is modified using PCA and all the features other than time and money are Principle Components and hence the distribution of features is Gaussian. We need to check for skewness in the data and try to mitigate it as it might cause problems during model-building.

Given dataset is highly imbalanced hence need to rescale the same using suitable techniques like SMOTE and ADASYN.

**Test/Train Split:**

We need to perform train/test split, in order to check the performance of the models with unseen data. Here, for validation, we can use the k-fold cross-validation method. We need to choose an appropriate k value so that the minority class is correctly represented in the test folds

**Model Building/Hyper Parameter Tuning:**

This is the final step at which we can try different models and fine-tune their hyper parameters until you get the desired level of performance. Through various sampling techniques we can try and see if we can get better models. We will build these models and evaluate them

* Logistic Regression
* KNN
* Random Forest and
* XGBoost

**Model Evaluation:**

We need to use appropriate evaluation metrics to find the performance of the model. Since it is an imbalanced data set, we need to make sure we which are fraudulent transactions than non-fraudulent. Accuracy is not always correct matrix for solving this kind of classification problems having imbalanced data hence we can evaluate the model with ROC Curve, Precision, Recall etc