

## **Project Proposal**

### **Motivation:**

Credit card fraud happens frequently and leads to massive financial losses. Online transaction has increased drastically significant no of online transaction are done by online credit cards. Therefore, banks and other financial institutions support the progress of credit card fraud detection applications. Fraudulent transactions can happen in different ways, and they can be placed into various categories. Identification of fraud credit card transactions is important to credit card companies for the prevention of being charged for items transaction of items which the customer did not purchase. Faced by credit card fraud transactions, I wanted to build a classifier so that I can put in all my transactions and predict if any transactions are a fraudulent or not.

### **Datasets used:**

Following is the link of the datasets that are used to train and test the classifier:  
[Credit Card Transactions - Train Set | Kaggle](#)

### **Existing System:**

This investigates and checks the performance of Decision tree, Random Forest, SVM and logistic regression on highly skewed credit card fraud data. The Random Forest algorithm will perform better with a larger number of training data, but speed during testing and application will suffer, imbalanced dataset problem and requires more pre-processing. The results indicate about the optimal accuracy for logistic regression, decision tree, Random Forest and SVM classifier.

Existing model with Principal Component Analysis selected features from data set and perform the machine learning algorithms with accuracy. They show Random Forest model reaches the best accuracy 99.7

### **Proposed System:**

Credit card fraud is a serious and ever-growing problem with billions of dollars lost every year due to fraudulent transactions. Fraud has always been present and will always be. It is also ever changing, as the technology and usage patterns change over time, which makes CCFD (credit card fraud detection) a particularly hard problem. Furthermore, a classification of mentioned techniques into two main fraud detection approaches, namely, misuses (supervised) and anomaly detection (unsupervised) is presented. Again, a classification of techniques is proposed based on capability to process the numerical and categorical datasets. Data will be processed by machine learning based on Decision tree, SVM, Linear regression etc., here we introduce the hybrid feature selection algorithm for fraud classification we better to achieve the best accuracy from the existing algorithms.