**PROJECT OVERVIEW**

**PROJECT NAME:** Credit Card Fraud Detection

**PROBLEM DEFINITION:** The problem aims to develop a machine learning-based system that analyses transaction data in real-time,effectively detecting credit card fraud while minimizing false positives.This solution will help financial institutions protect against fraudulent transactions,reducing financial losses and ensuring customer trust.

**DESIGN:**

To develop a machine learning model we need to install some packages and they are:

* Pandas
* Seaborn
* Matplotlib
* RobustScaler from sklearn.preprocessing
* train\_test\_split from sklearn.model\_selection
* classification\_report,confusion\_matrix and accuracy\_matrix from sklearn.metrics

Dataset is taken from <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/>

Since the data in the dataset is the raw data,it needs to undergo the following stages:

* **DATA COLLECTION:**Collect historical credit card transaction data from various sources, including legitimate and fraudulent transactions.
* **DATA PREPROCESSING:**
* Clean and preprocess the data to handle missing values,outliers, and inconsistencies.
* Normalize or scale features as needed.
* Create relevant features, such as transaction frequency, geographic patterns, and behavioral features.
* **FEATURE SELECTION:**We need to choose relevant features from the dataset, such as transaction amount, location, time, and more.
* **MODEL SELECTION:**
* We need to choose appropriate machine learning algorithms for fraud detection, considering both supervised and unsupervised techniques
* Consider using a combination of models to enhance accuracy e.g., ensemble methods.
* **MODEL TRAINING:**Train models using historical data, and periodically retrain them to adapt to evolving fraud patterns.
* **MODEL EVALUATION:** We need to continuously monitor the performance of the system using metrics like accuracy, precision, recall, F1-score, and AUC.
* **DATA PRIVACY AND SECURITY:**
* We need to ensure that sensitive data, both at rest and in transit, is encrypted.
* Need to implement strict access controls to limit who can view and manipulate the data.
* Need to ensure compliance with relevant data privacy regulations (e.g., GDPR, HIPAA) and industry standards.
* **SCALABILITY:**
* We will ensure that the system can handle a large volume of transactions and scale horizontally as needed.
* Also consider cloud-based solutions for scalability and reliability.