Name: Dipen Prajapati

NetID: dp1435

RUID: 237004764

Project Proposal:

Dataset 1: **Credit Card Fraud Detection Dataset**

This dataset provides financial transaction records collected to detect fraudulent activities in credit card usage. Fraudulent activities are a critical challenge for financial institutions and this dataset allows for the development of machine learning models to identify fraud in real-time. Each transaction is represented by set of anonymized features (V1–V28) along with the transaction amount, time and binary target variable indicating whether the transaction was fraudulent (1) or legitimate (0). The anonymization of features ensures the confidentiality of sensitive information while still enabling robust analysis.

Task: **Classification**
The task involves identifying whether a transaction is fraudulent or legitimate. This binary classification task is crucial for improving the security of financial systems and reducing fraud related losses.

Total number of samples: 550000

Total number of columns: 31

Link: https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023

Dataset 2: **Nifty Bank Index Minute Data (2015 to 2022)**

This dataset contains high-frequency stock market data for the Nifty Bank index which recorded at one-minute intervals. It includes stock market metrics such as the open, high, low, close prices of trades. The dataset incorporates various technical indicators such as Moving Averages, Relative Strength Index (RSI) and Bollinger Bands which are essential for predictive modeling in financial markets. Such minute-level data is valuable for developing machine learning models to forecast stock trends and understand market behavior.

Task: **Regression**
The task involves predicting a continuous variable such as the future price of the Nifty Bank index or its return by using past OHLC values and technical indicators. This helps in developing investment strategies and decision-making tools for stock market traders.

Total number of samples: 550000

Total number of columns: 60

Link: https://www.kaggle.com/datasets/debashis74017/nifty-bank-minute-data?select=NIFTY+BANK+-+1+minute_with_indicators_.csv

Dataset 3: **Fruit Recognition Dataset**

This image dataset consists of a diverse collection of fruit images including multiple classes such as apples, bananas, grapes, oranges and many more. Each image is labeled with the corresponding fruit type which is making it ideal for image recognition tasks. The dataset contains images with variations in lighting, orientation and background. It provides a realistic scenario for machine learning applications. Such datasets are commonly used in computer vision tasks to develop systems that can automate food identification, support dietary tracking or even assist in agricultural practices.

Task: **Classification**
The task is to classify images into their respective fruit categories. This is a **multi-class classification** problem where the goal is to predict one out of 33 fruit classes for each image.

Number of Images: 22495

Number of Classes: 33

Link: https://www.kaggle.com/datasets/sshikamaru/fruit-recognition?select=test