# CREDIT CARD FRAUD DETECTION WITH MACHINE LEARNING

## Phase 2: Innovation

### SHORT EXPLANATION:

Credit card fraud detection is the process of identifying and preventing unauthorized or fraudulent transactions made using credit or debit cards. This is typically done using machine learning and data analysis techniques. The goal is to distinguish between legitimate card transactions and fraudulent ones, protecting cardholders and financial institutions from financial losses. Key methods include monitoring transaction data, analyzing patterns, and using predictive models to flag and block suspicious transactions in real-time. The field of credit card fraud detection is crucial in today's digital economy, where the use of payment cards is widespread and fraudsters continually devise new tactics to exploit vulnerabilities.

### DATASET EXPLANATION:

The given dataset:

https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

**DETAILS ABOUT COLUMN:**

The columns that we are going to use are :

- Time
  Number of seconds elapsed between this transaction and the first transaction in the dataset
- Amount
  Transaction amount
- Class
  1 for fraudulent transactions, 0 otherwise

**LIBRARIES TO BE USED AND WAY TO DOWNLOAD:**

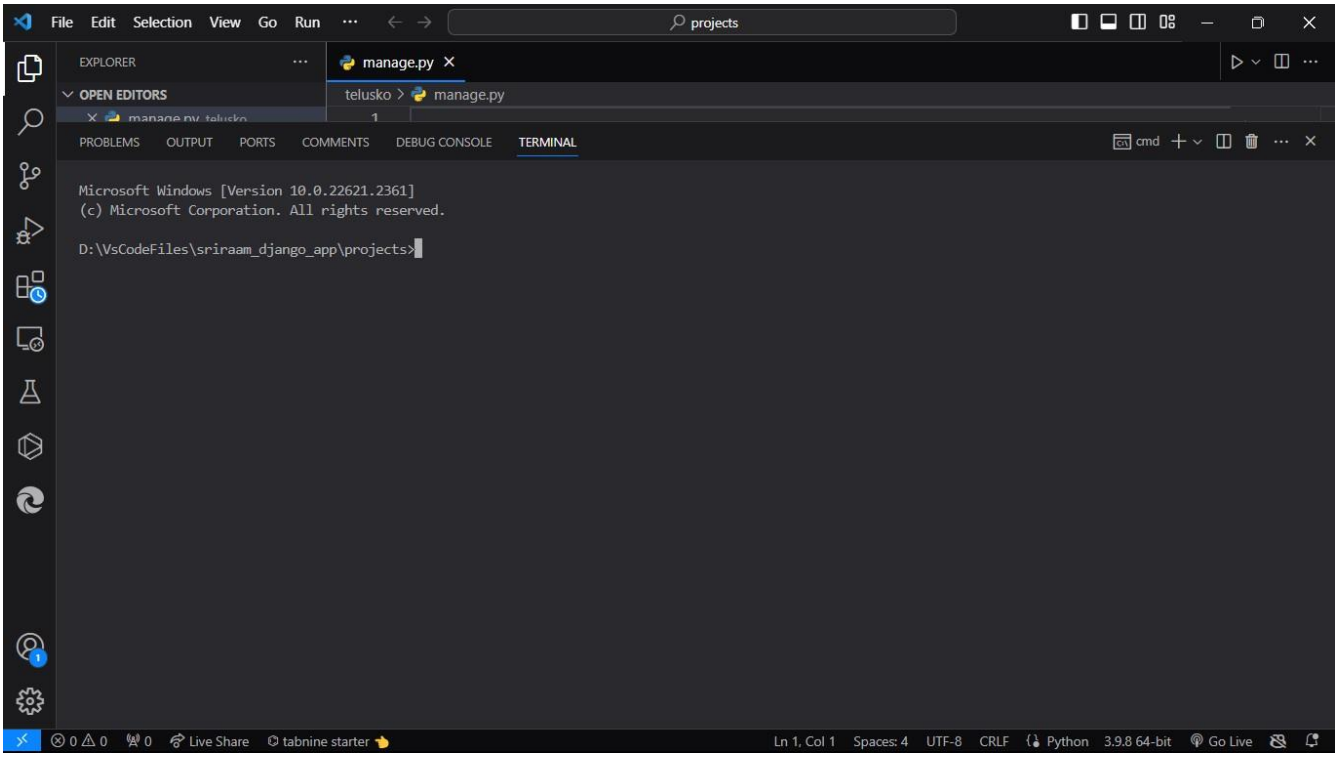
The libraries to be used :

- numpy
  - NumPy is an open-source library in Python that provides support in mathematical, scientific, engineering, and data science programming.
- Pandas
  - Pandas is an open-source library in Python that is made mainly for working with relational or labeled data both easily and intuitively
- Sklearn
  - A comprehensive library for machine learning that includes various algorithms for classification, model evaluation, and preprocessing techniques.
- Scipy
  - Offers a wide range of statistical and optimization functions for data analysis.
- Matplotlib
  - Used for data visualization, which is essential for data exploration and model performance evaluation.

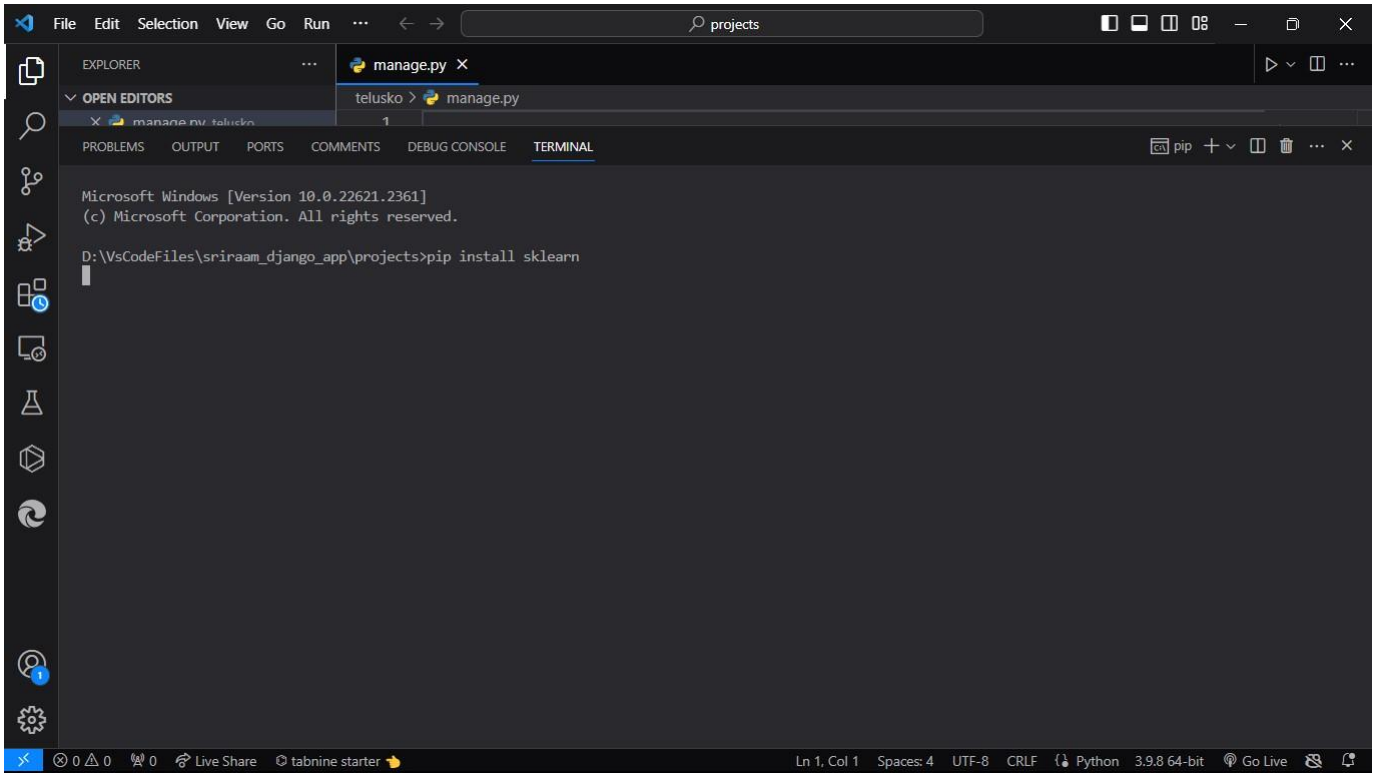
The above mentioned python libraries for developing the ML model can be downloaded  easily by pip installing in VS code .

The installation image is attached below

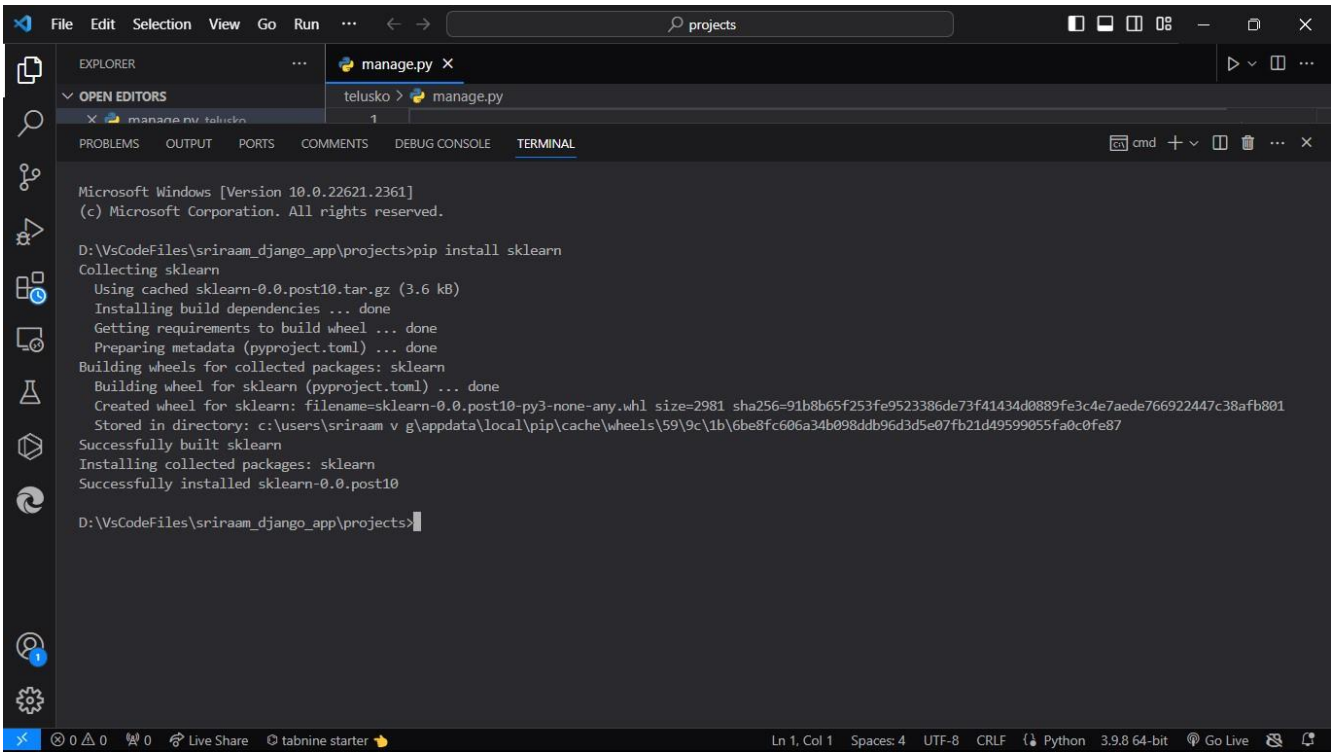
## Opening Terminal



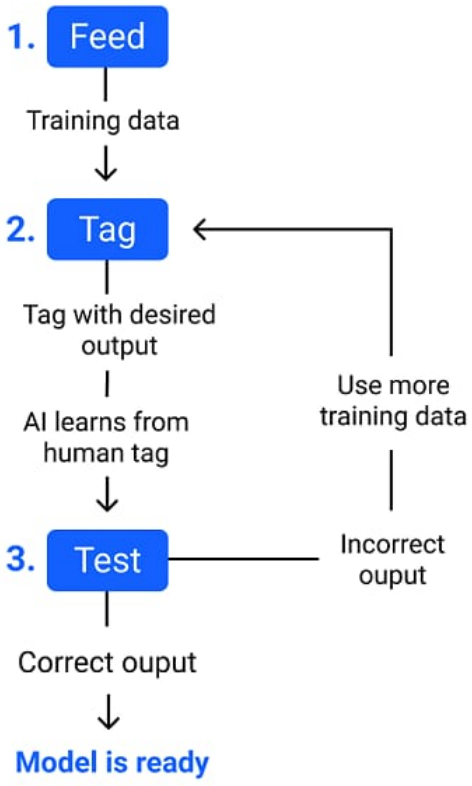

## Installing Dependencies



## Installed successfully




## TRAINING AND TESTING:



The phases available in training of ML model are:

- **Data Preparation**: Get your data ready, clean it up, and create features from it.
- **Data Splitting:** Divide your data into two parts - one for teaching the model (training data) and the other for testing how well it learned (testing data).
- **Model Building**: Choose a machine learning method and teach it using the training data. The model learns from the patterns in this data.
- **Model Training**: The model fine-tunes its settings to make the best predictions on the training data.

The phases available in testing of ML model are:

- **Testing Set:** The model is tested on the testing data, which it hasn't seen before.
- **Model Evaluation**: The model makes predictions on the testing data, and we compare these predictions to the actual answers (the ground truth).
- **Performance Metrics**: We measure how well the model did using metrics like accuracy, precision, recall, etc.

**METRICS:**

The metrics that we choose to execute the model preparation is F-1 score.

The F1-score is the harmonic mean of precision and recall. It provides a balance between the two metrics and is particularly useful when you want to consider both false positives and false negatives in your evaluation.