

Description of the file directory and steps to run the script

Directory Tree:-

- Fraud-Detection is the main folder which contains the following:
 - main script 'fraud_detection.py' (the script itself contains detailed analysis and explanation)
 - script to tune hyperparameters 'fine_tune_model.py'
 - script which contains all the important helper functions 'utils.py' (all the helper functions are thoroughly explained in the script)
 - 'requirements.txt' file which contains all the libraries with the exact versions used in this project
 - the 2-slide presentation pitch 'Pitch.pdf'
 - a folder called **data** which contains the following:
 - 'labels_obf.csv' → fraud labels
 - 'transactions_obf.csv' → transaction data
 - a folder called **figs** which contains the following figures:
 - 'pipeline.png'
 - 'feature_engineering.png'
 - 'feature_importance.png'
 - 'XGBoostBase.png'
 - 'pr-best-model.png'
 - a folder called **processed_data** which contains:
 - 'processed_df.csv' → data after processing and feature engineering
 - 'precision-recall-best-model.csv'
 - a folder called **model** which contains:
 - 'best_model.pkl' → final trained model

Steps to run the script:-

- 1) Extract the zip file into any folder
- 2) Open a terminal or command prompt
- 3) Change the directory into this folder using the following command:
 - a. "cd location_of_the_folder" → this will take you inside the folder
- 4) Now we need to install the required libraries to run our script which can be done using the command below:
 - a. "pip install -r requirements.txt"
- 5) Now after successfully installing the required libraries, we can run our main script by the command below:
 - a. "python fraud_detection.py" → all windows users
 - b. "python3 fraud_detection.py" → some mac users who have multiple versions of Python installed