INNOVATION FOR CREDIT CARD FRAUD DETECTION

★ SHORT EXPLANATION ABOUT THE QUESTION

- ❖ Credit card fraud detection refers to the process of identifying and preventing unauthorized or fraudulent transactions made using credit or debit cards.

- ❖ The model used must be simple and fast enough to detect the anomaly and classify it as a fraudulent transaction as quickly as possible

- ❖ This involves the use of various algorithms, machine learning techniques, and data analysis to spot unusual patterns, behaviors, or transactions that may indicate fraudulent activity.

- ❖ The project aims to develop a machine learning-based system that analyzes transaction data in real-time, effectively detecting credit card fraud while minimizing false positives.

➢ DATA GATHERING

- ❖ Dataset is taken from kaggle competition and can be downloaded from

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

- The dataset contains transactions made by credit cards in September 2013 by European cardholders.

- 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.

★ DETAILS ABOUT COLUMNS

- **V1,V2, ……V28 :** It contains only numerical input variables which are the result of a PCA transformation due to confidential issues. Features V1, V2, … V28 are the principal components obtained with PCA
- **TIME :** It contains the seconds elapsed between each transaction and the first transaction in the dataset.
- **AMOUNT :** It is the transaction Amount, this feature can be used for example- dependant cost-sensitive learning.
- **CLASS :** It is the response variable and it takes value 1 in case of fraud and 0 otherwise.

★ DETAILS OF LIBRARIES TO BE USED

Numpy : For numerical computations and working with arrays.

```
Import numpy as np
```

Pandas : For data manipulation and analysis.

```
Import pandas as pd
```

Matplotlib : It enables users to generate visualizations like histograms, scatter plots, bar charts, pie charts.

```
Import matplotlib.pyplot as plt
From matplotlib import gridspec
```

Seaborn : It is a visualization library that is built on top of Matplotlib. It provides data visualizations that are typically more aesthetic and statistically sophisticated.

```
Import seaborn as sns
```

★ HOW TO DOWNLOAD LIBRARIES

To download libraries in Python, you typically use a package manager called “pip.”

- Open a Terminal or Command Prompt: On Windows open the Command Prompt. On macOS or Linux, open the Terminal.

- Install the Library. Use the following command to install a Python library using pip replacing Library_name with the name of the library you want to install:

```
Pip install library_name
```

->For example, if you want to install the popular library numpy, you would run:

```
Pip install numpy
```

- Wait for Installation: pip will download and install the library and its dependencies. You'll see progress in the terminal as it does this. Wait for Installation: pip will download and install the library

and its dependencies. You'll see progress in the terminal as it does this

- ❖ Verify Installation: You can verify that the library is installed correctly by opening a Python interpreter and trying to import it. For example:

  Import numpy

- ❖ If you don't get any error messages, the library is installed and ready to use.

Additionally, if you're using Python 3, you might use pip3 instead of pip (eg, pip3 install library_name) depending on your system configuration

★ TRAINING AND TESTING

- ❖ Use the training set to train selected machine learning model. During training, the model learns patterns in the data to distinguish between fraudulent and legitimate transactions

- ❖ Assess how well your model generalizes to new, unseen data. The goal is to detect fraudulent transactions accurately while minimizing false positives (legitimate transactions mistakenly flagged as fraudulent).

★ METRICS USED FOR CHECKING ACCURACY

- ❖ Metrics such as
- ❖ Accuracy
- ❖ Precision

- Recall
- F1-score
- AUC-ROC are commonly used for evaluation.