ML Workshop Project

Overview

In this project, you will work as a team to analyze **two datasets** and build machine learning models to estimate meaningful outcomes from each of them.

You will work on:

- **NYC Taxi dataset** to estimate expected amount of fare
- **Cyber Attacks dataset** to analyze and estimate patterns related to cyber security incidents.

By the end of the project, your team will **compare and present results from both datasets together in a single dashboard**, showcasing insights, model performance, and predictions.

Team Structure

- You will work in teams of **3 members**.
- All team members are expected to actively participate in:
    - Data exploration (EDA)
    - Data preprocessing
    - Model training and evaluation
    - Presentation and explanation of results

Datasets & Task Description

- **Dataset 1:** NYC Taxi Dataset
    - Build a machine learning model to estimate the Taxi fares
- **Dataset 2:** Cyber Attacks Dataset
    - Build a machine learning model to estimate or classify different attack types

For both datasets, you are required to:

- Understand the data and its features
- Apply EDA and preprocessing techniques
- Train and evaluate an appropriate machine learning model
- Interpret and explain your results

Deliverables

Each team must submit the following:

1. **Jupyter Notebook**
    - Includes EDA, preprocessing, model training, and evaluation for **both datasets**.
2. **User interface**
    - A single interface (using **Streamlit**) that presents:
        - Key EDA insights from the Taxi dataset
        - Key EDA insights from the Cyber Attacks dataset
        - Model results and estimations for both datasets
        - A clear comparison or separation between the two tasks

3. **Presentation**

Presentation Guidelines

During your presentation, your team should:

1. Explain the insights gained from EDA for both datasets.
2. Clearly describe preprocessing steps and their impact.
3. Explain the model training process.
4. Justify the choice of models used.
5. Discuss challenges faced and how you overcame them.
6. Demonstrate the dashboard and walk through the results for both datasets.

Note:
You will **not be judged on the design quality** of the Streamlit dashboard, as it was not part of the workshop content.
However, the dashboard **must be included** in the final deliverables and should correctly display results from **both datasets**.