A circular emblem with a shield and gears on it

AI-generated content may be incorrect.

**AutoML Dashboard: A GUI-based Interactive Machine Learning Tool for CSV Data"**

**AI102L-F**

**Group Members:**

* Muhammad Ahmed Hammad 2024341
* Syed Abid Hussain Kazmi 2024603
* Awais Iftikhar 2024124

**1. Project Objectives**

The primary objective of this project is to design and develop a user-friendly, interactive machine learning application that simplifies the data analysis and model training process for users with little to no programming experience. Specifically, the project aims to:

* Allow users to upload and visualize CSV datasets via an intuitive GUI.
* Let users select independent (features) and dependent (target) variables.
* Provide options for handling missing data (filling, removing, or leaving as is).
* Automatically preprocess data and train a basic machine learning model.
* Present evaluation metrics and data insights in an interactive dashboard format.
* Demonstrate key machine learning concepts in a visual and hands-on manner.

**2. Methodology**

The project will be implemented in Python using the **Streamlit** framework to build the GUI. The system architecture and workflow are outlined below:

**Step 1: Data Upload and Display**

* GUI allows users to upload a .csv file.
* The dataset is displayed in tabular format with basic statistics (e.g., shape, column types, null values).

**Step 2: User Selection**

* Interface for selecting:
  + Independent variables (features)
  + Dependent variable (target)
  + A circular emblem with a shield and gears on it

    AI-generated content may be incorrect.
  + Null value handling strategy: Fill (with mean/median), Remove rows, or Leave as is

**Step 3: Data Preprocessing**

* Convert categorical features to numerical (label encoding or one-hot encoding).
* Handle missing values based on user selection.
* Normalize/scale data if needed.

**Step 4: Model Training**

* Train a basic model using Scikit-learn (e.g., Linear Regression, Decision Tree, or Random Forest depending on target variable type).
* Automatically detect regression or classification based on target variable type.

**Step 5: Output & Dashboard**

* Display model performance metrics (e.g., accuracy, precision, recall, R² score).
* Show data visualizations: feature importance, confusion matrix, distribution plots.
* Provide option to download trained model (as a .pkl file).

**3. Expected Outcomes**

By the end of the project, we aim to deliver:

* A fully functional **Streamlit-based app** that performs end-to-end machine learning tasks with user input.
* An **interactive dashboard** that simplifies understanding of machine learning workflows.
* **Clear documentation** and an intuitive UI suitable for beginners.
* Enhanced understanding of:
  + Data preprocessing techniques
  + Feature selection
  + Model training and evaluation
* A circular emblem with a shield and gears on it

  AI-generated content may be incorrect.
* **Bonus features** (if time permits):
  + Model comparison (train multiple models and compare)
  + Option to save and load previous sessions

This project will not only solidify our foundational knowledge of AI/ML concepts but also equip us with practical experience in building user-focused data applications.