Machine Learning Vision Board
(Project 1 - ITI AI)

Prepared By:
Mohamed Younis Mohamed - King Salman Internationl University
Youssef Helmy Abdelaty - King Salman Internationl University
Ahmed Abd ElSalam ElSayed - Sinia University

Supervised By:
Dr. Salwa Osama

Teams Members

Name	Department	Task Distribution
Mohamed Younis Mohamed	Computer Engineering	Feature Engineering + Data Preprocessing
Youssef Helmy Abdelat	Computer science	UI + Streamlit Integration
Ahmed Abd ElSalam	Computer science	ML Models Implementatic

Project Description

This project is a Streamlit-powered AI Dashboard that allows users to explore datasets and apply different Machine Learning techniques in an interactive way. The system integrates data preprocessing, model training, evaluation, and visualization into one unified platform. It is designed to make working with datasets easier, faster, and more intuitive, especially for students, researchers, and anyone interested in learning Machine Learning.

The dashboard supports both Supervised Learning (e.g., classification and regression) and Unsupervised Learning (e.g., clustering), providing users with the ability to experiment, compare results, and gain insights from their data without the need for complex coding.

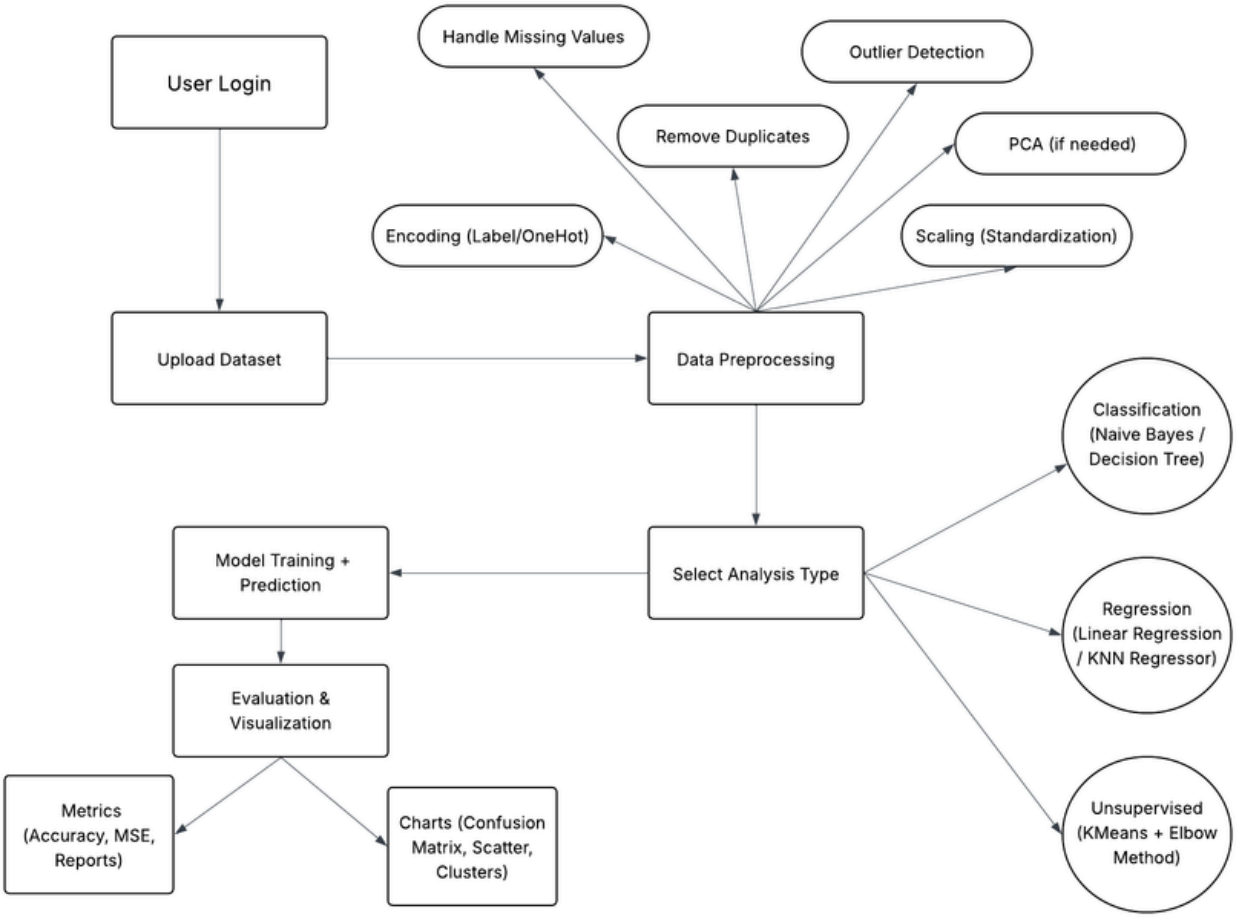
Main Features:

1. Secure Login (simple authentication with email and password).
2. CSV Upload to allow flexible dataset input.
3. Automatic Data Preprocessing:

- Missing values handling.
- Removing duplicates.
- Outlier detection & removal (Z-score method).
- Automatic Encoding (OneHot / Label).
- Scaling numeric features.
- PCA (if number of features > 10).

4. Analysis Modes:
   - Classification: Naive Bayes, Decision Tree.
   - Regression: Linear Regression, KNN Regressor.
   - Unsupervised: KMeans clustering with optimal k detection (KneeLocator).
   - Visualization: Confusion Matrix, Accuracy/MSE charts, Clustering plots.

System Workflow Diagram

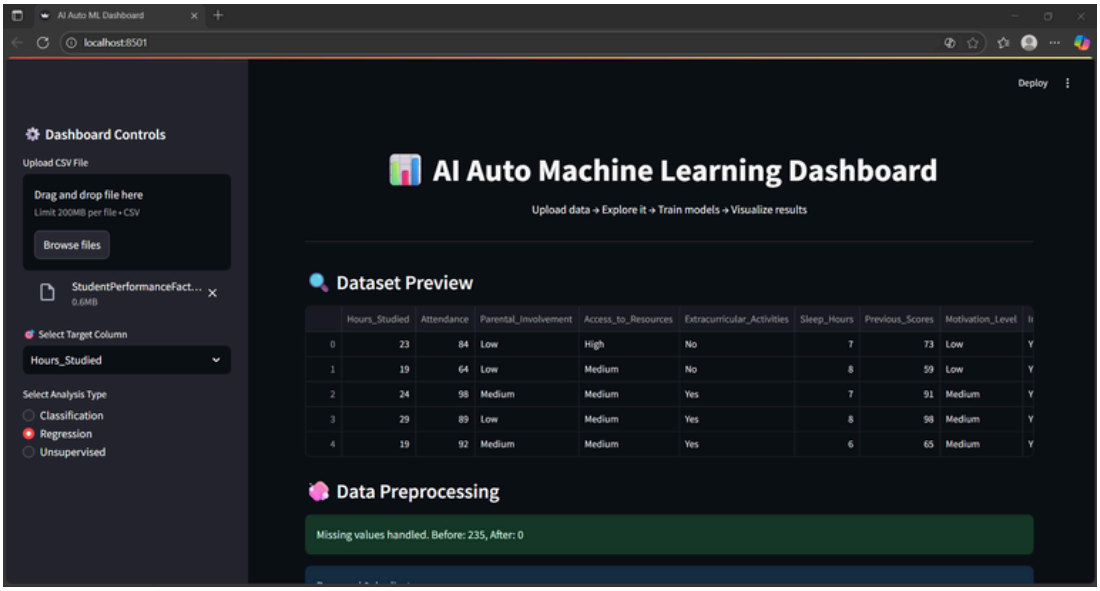


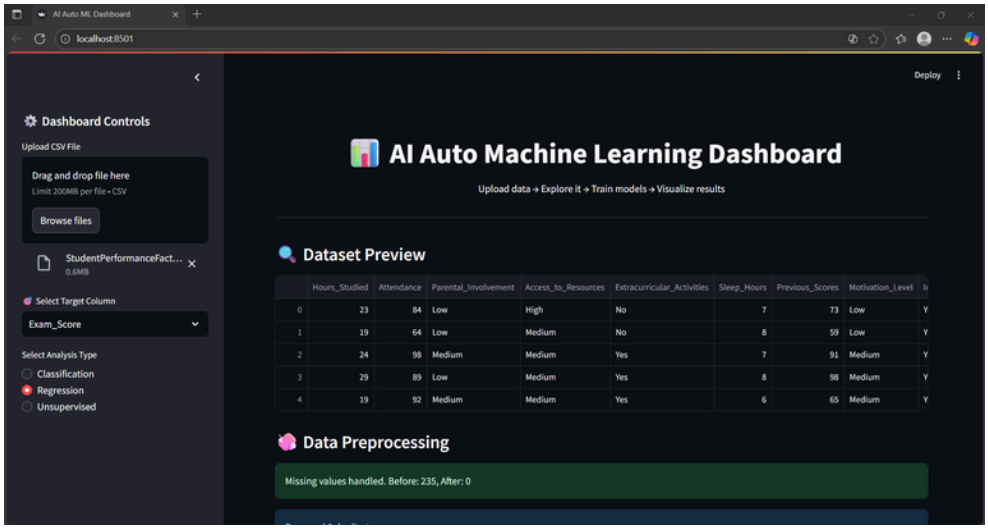
Depended libraries

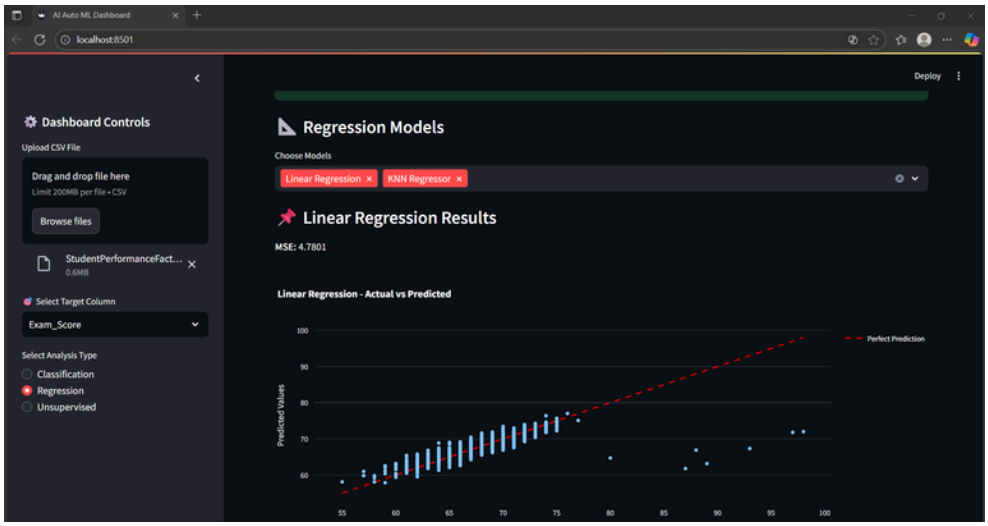
Library Name	Command in Terminal	File (.py)	Why?
streamlit	pip install streamlit	app.py	Web-based interactive dashboard
pandas	pip install pandas	app.py	Data manipulation

Library Name	Command in Terminal	File (.py)	Why?
numpy	pip install numpy	app.py	Numerical operations
matplotlib	pip install matplotlib	app.py	Visualizations
plotly	pip install plotly	app.py	interactive visualization
seaborn	pip install seaborn	app.py	Statistical plottinges (heatmaps, confusion matrices)
scikit-learn	pip install scikit-learn	app.py	ML models, preprocessing, metrics
kneed	pip install kneed	app.py	Elbow method (finding best k for KMeans clustering)
scipy	pip install scipy	app.py	Outlier detection (Z-score method)

GUI







Datasets description

Dataset 1 (Student Performance Factors) - Regression :
link : https://www.kaggle.com/datasets/drewbamberger/student-performance-factors

Dataset 2 (Heart Disease Dataset)- Classfication :
This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer-valued from 0 (no presence) to 4.
link : https://www.kaggle.com/datasets/yasserh/heart-disease-dataset/data

Dataset 3 (Breast Cancer Prediction) - Classfication :
Worldwide, breast cancer is the most common type of cancer in women and the second highest in terms of mortality rates.Diagnosis of breast cancer is performed when an abnormal lump is found (from self-examination or x-ray) or a tiny speck of calcium is seen (on an x-ray). After a suspicious lump is found, the doctor will conduct a diagnosis to determine whether it is cancerous and, if so, whether it has spread to other parts of the body.
link : https://www.kaggle.com/datasets/merishnasuwal/breast-cancer-prediction-dataset

Dataset 4 (Mall_Customers) - Unsupervised :
link : https://www.kaggle.com/datasets/shwetabh123/mall-customers