PRACTICAL TECHNICAL ASSESMENT

 |

MUHAMMED HIDASH

adit22AP02601

**Activity**

1. Load the dataset and apply necessary preprocessing steps.

2. Perform exploratory data analysis (EDA) to understand the dataset.

3. Implement classification models and evaluate them using a confusion matrix and

cross-validation.

4. Implement regression models and evaluate them using R-squared, MSE, and crossvalidation.

5. Visualize the confusion matrix for at least one classification model.

6. Report and interpret the results of each model.

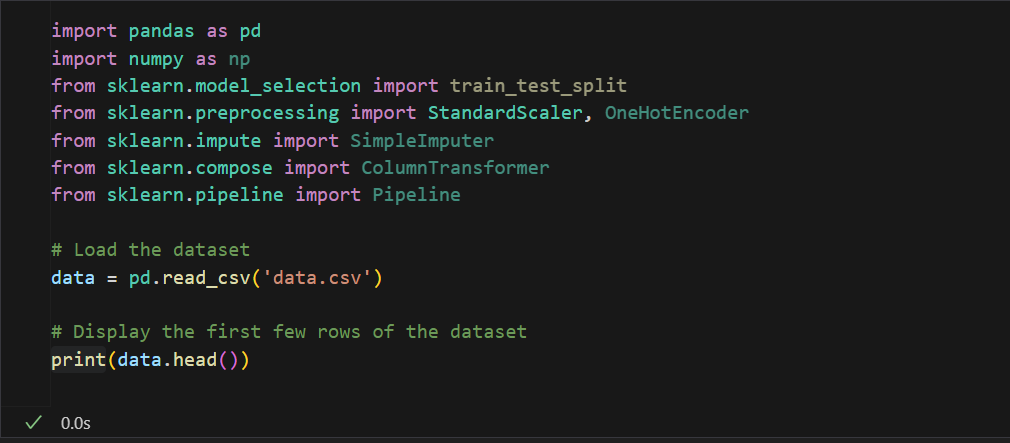
**Requirements**

* PC / Laptop
* VS code with python installed
* Dataset (data.csv)

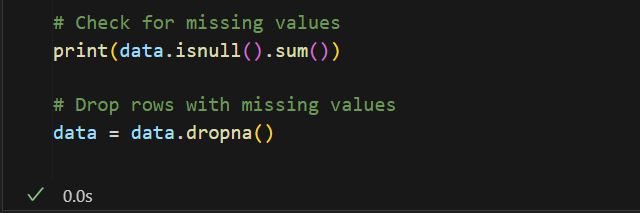
**Procedure**

**Data Preprocessing**

* Load the dataset using pd.read\_csv('data.csv')



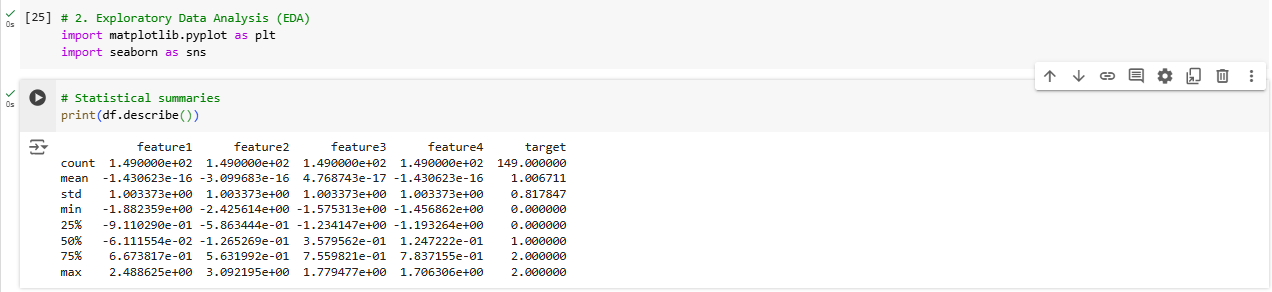
* Handle missing values.

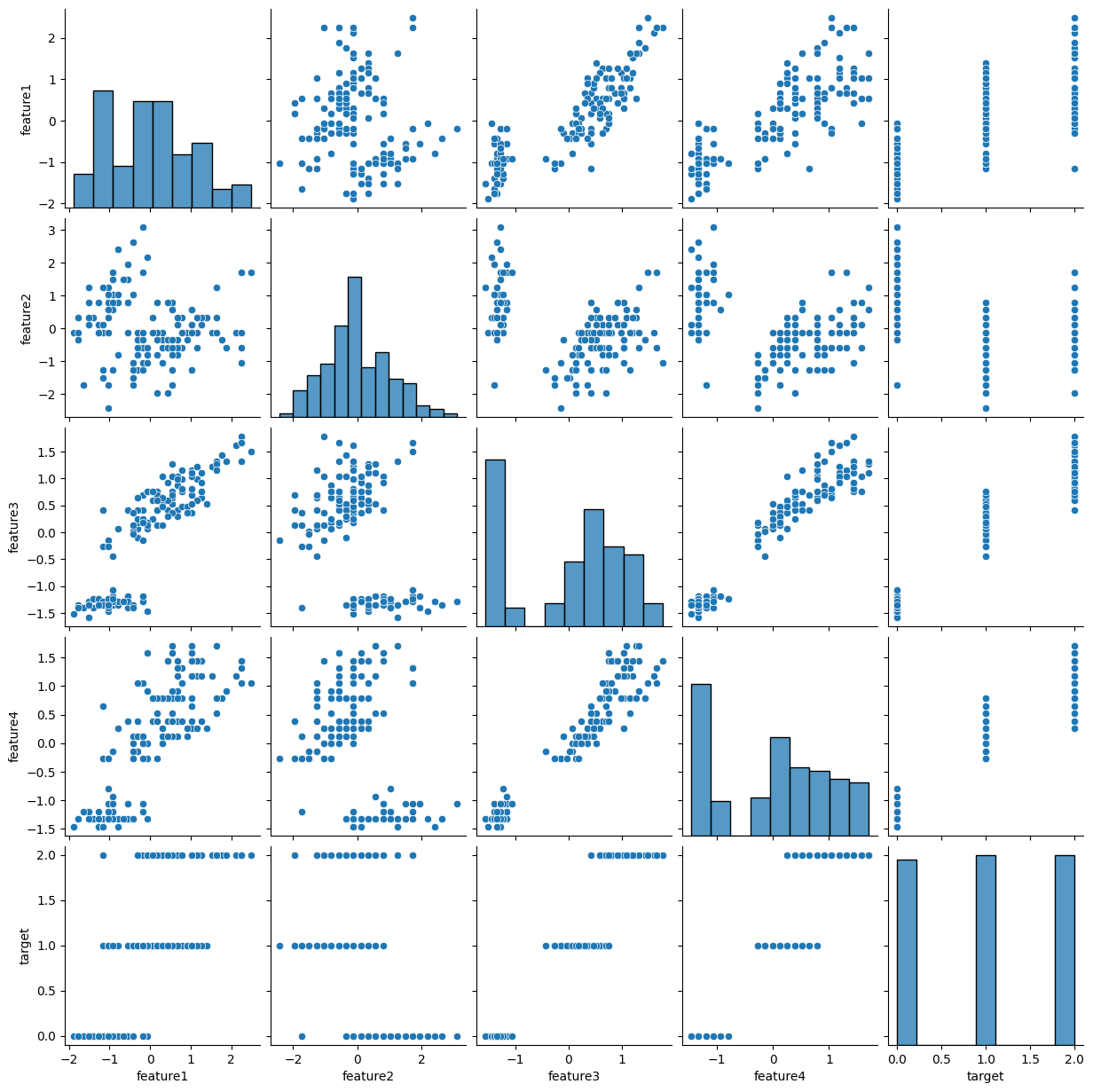


* Encode categorical variables.
* Scale/normalize the features.

**Exploratory Data Analysis (EDA)**

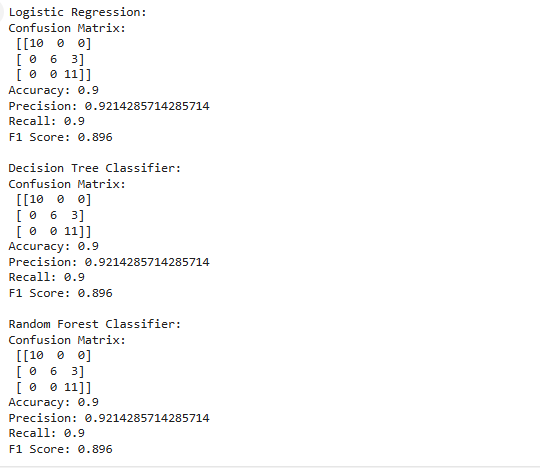
* Provide statistical summaries of the dataset.
* Visualize the data distribution and relationships between features using plots.





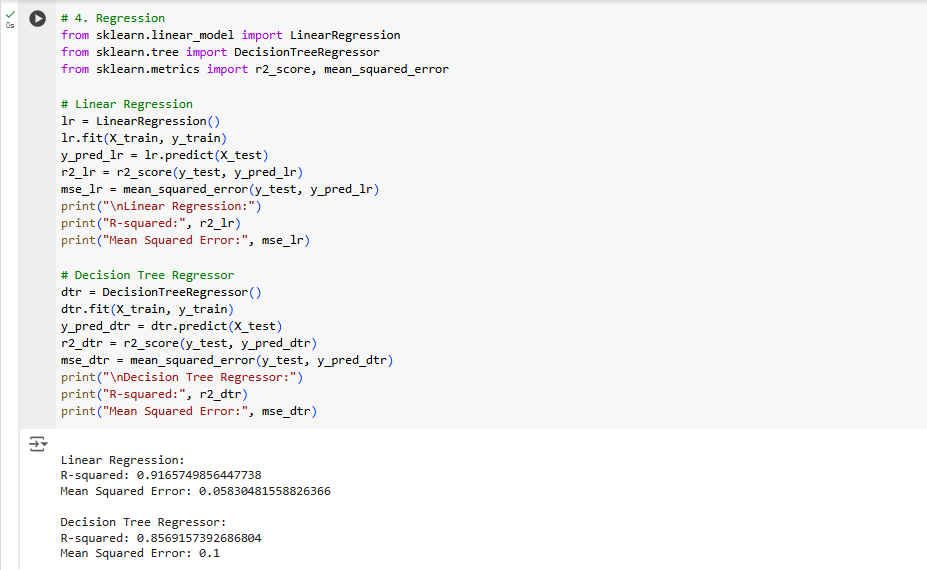
**Classification**

* Apply Logistic Regression, Decision Tree, and Random Forest classifiers.
* Use a confusion matrix to evaluate the performance of each classifier.
* Perform cross-validation to assess the model stability.



**Regression**

* Apply Linear Regression and Decision Tree Regressor.
* Evaluate the models using R-squared and Mean Squared Error (MSE).
* Perform cross-validation to assess the model stability.



**Confusion Matrix**

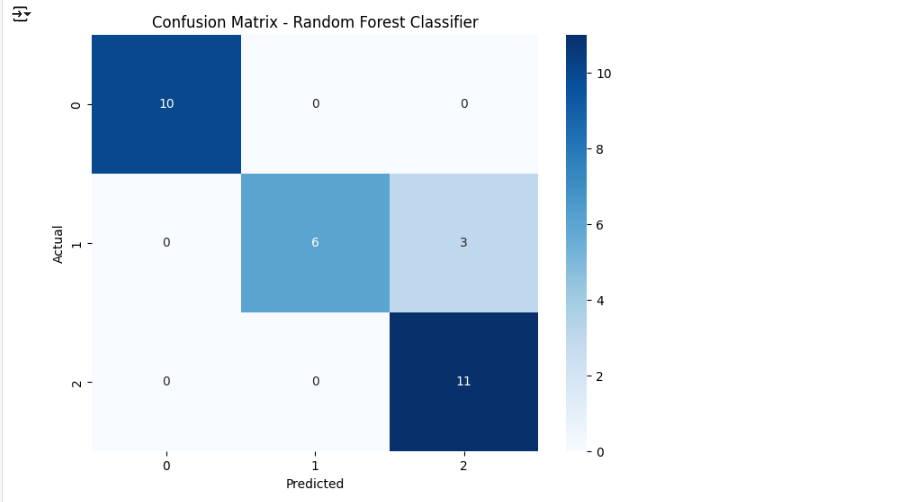
For classification tasks, plot the confusion matrix and compute the following metrics:

▪ Accuracy

▪ Precision

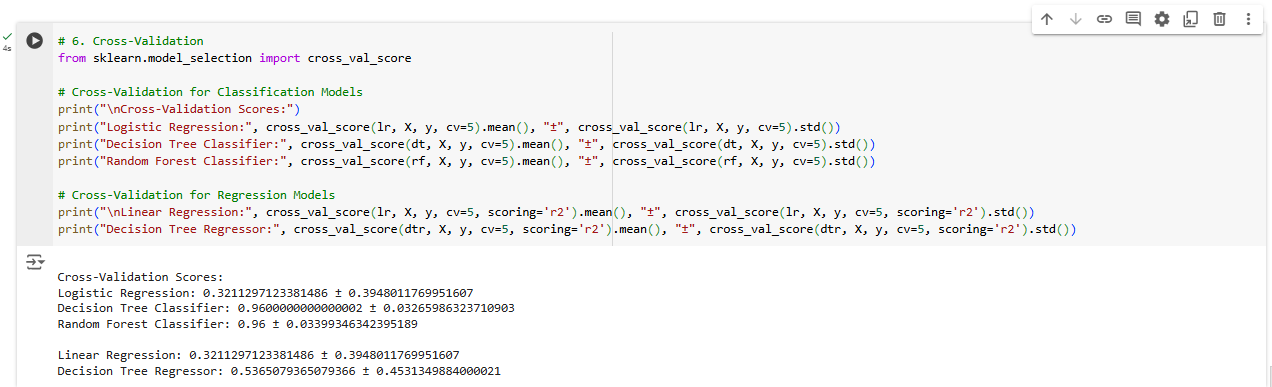
▪ Recall

▪ F1 Score



**Cross-Validation**

* Implement k-fold cross-validation for both classification and regression models.
* Report the mean and standard deviation of the cross-validation scores.



**Conclusion**

This documentation outlines the process of loading and preprocessing a dataset, performing exploratory data analysis (EDA), implementing classification and regression models, evaluating the models using various metrics, and visualizing results. The dataset used is assumed to have numerical and categorical features, with a target variable for prediction.