DOCUMENTATION  
Machine Learning Practical

short line

*SUBMITTED BY:*

*SHUHAIB*

*NSTI CALICUT*

# Activity

AIM :

Instruction:

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.

Using a given dataset, apply various machine learning techniques to classify and predict

outcomes. Evaluate the performance of your models using different statistical methods,

confusion matrix, and cross-validation.

Dataset:

Assume you have a CSV file named data.csv which contains the following columns:

• feature1, feature2, ..., featureN: Numerical and categorical features

• target: The class label or continuous variable for prediction

You can use any dataset, such as the Titanic dataset, Iris dataset, or any other relevant dataset,

but for simplicity, let's assume this generic structure.

Tasks:

1. Data Preprocessing:

o Load the dataset.

o Handle missing values.

o Encode categorical variables.

o Scale/normalize the features.

2. Exploratory Data Analysis (EDA):

o Provide statistical summaries of the dataset.

o Visualize the data distribution and relationships between features using plots.

3. Classification:

o Apply Logistic Regression, Decision Tree, and Random Forest classifiers.

o Use a confusion matrix to evaluate the performance of each classifier.

o Perform cross-validation to assess the model stability.

4. Regression:

o Apply Linear Regression and Decision Tree Regressor.

o Evaluate the models using R-squared and Mean Squared Error (MSE).

o Perform cross-validation to assess the model stability.

5. Confusion Matrix:

o For classification tasks, plot the confusion matrix and compute the following

metrics:

▪ Accuracy

▪ Precision

▪ Recall

▪ F1 Score

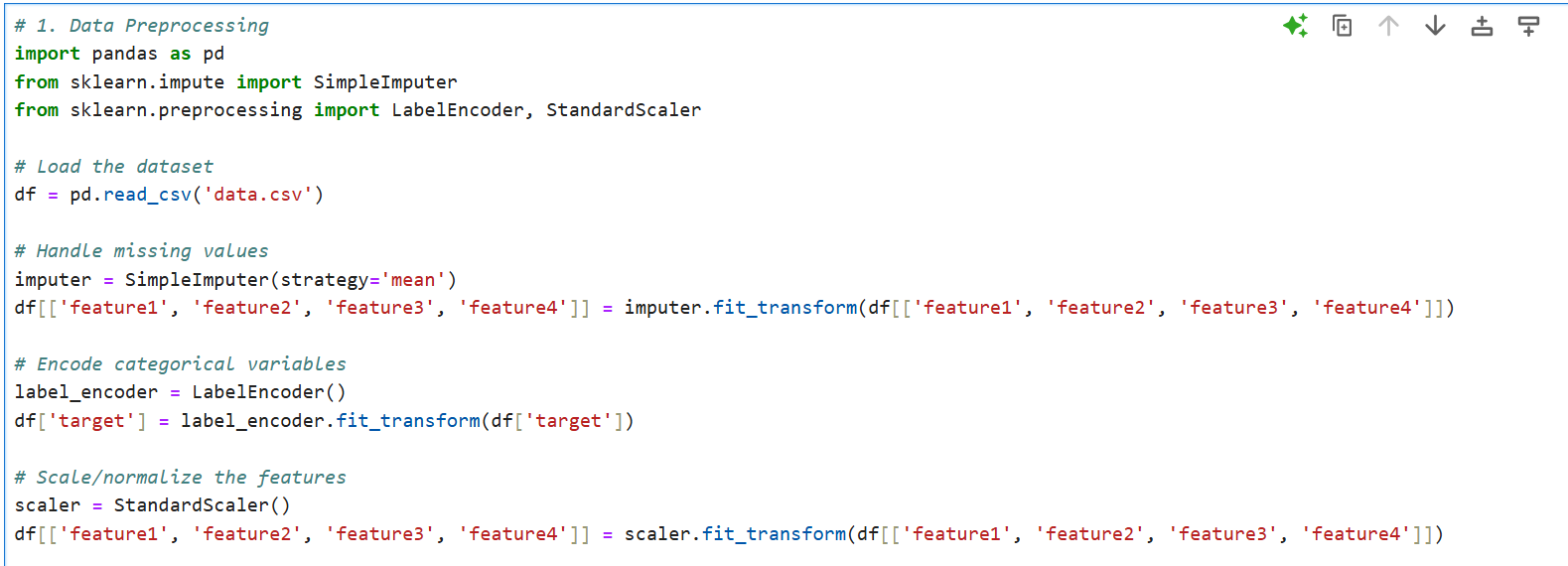
6. Cross-Validation:

o Implement k-fold cross-validation for both classification and regression

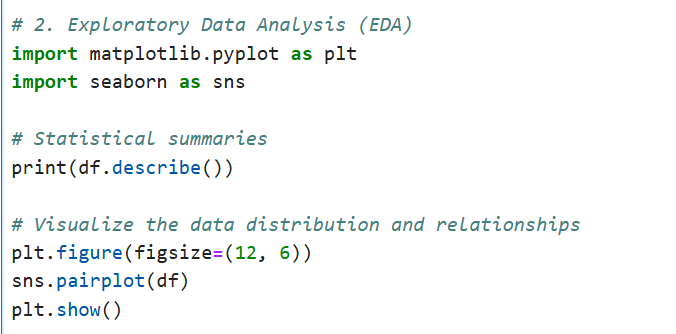
models.o Report the mean and standard deviation of the cross-validation scores

**Steps**

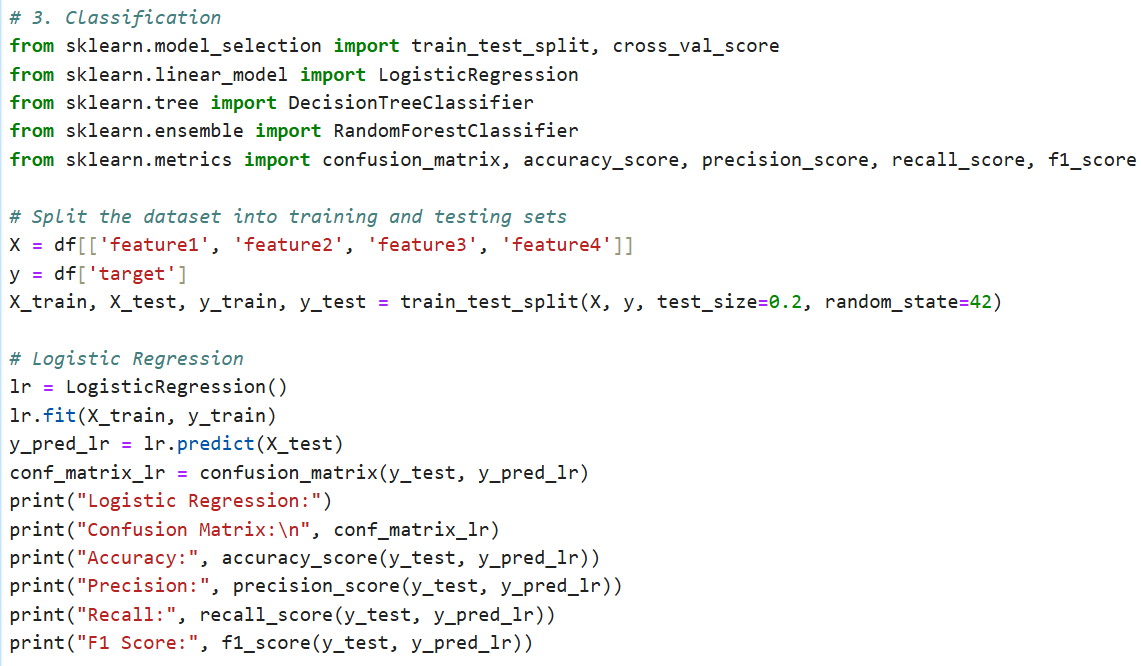
Step 1: Data Preprocessing

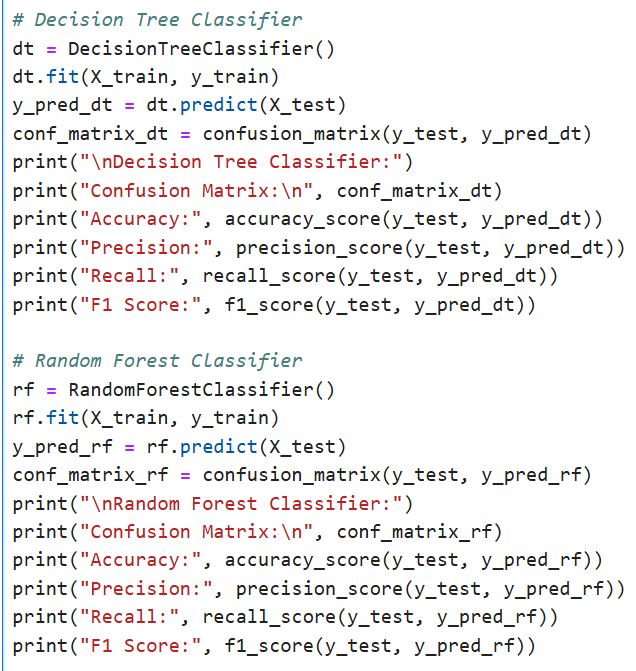


Step 2: Exploratory Data Analysis (EDA)

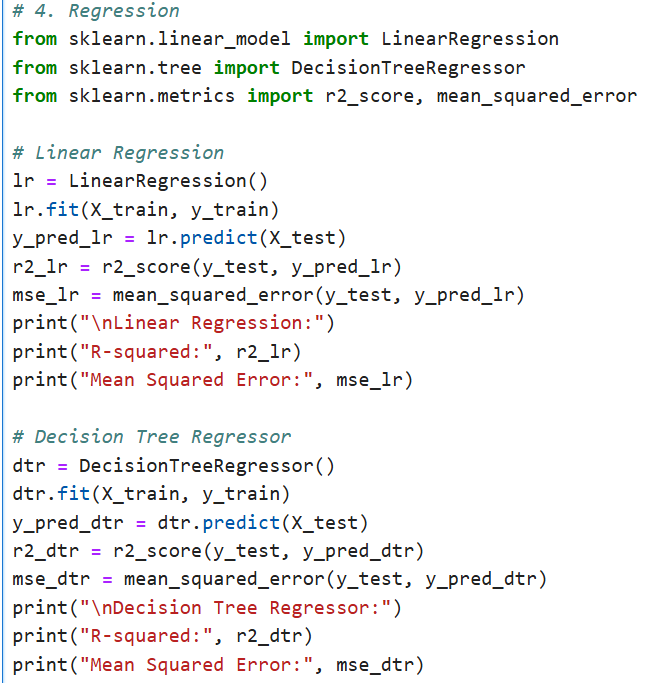


Step 3: Classification

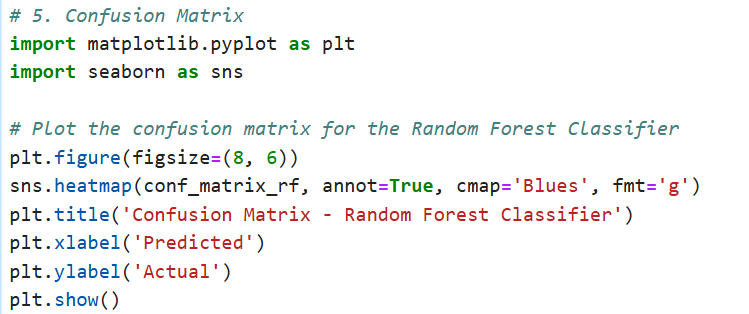




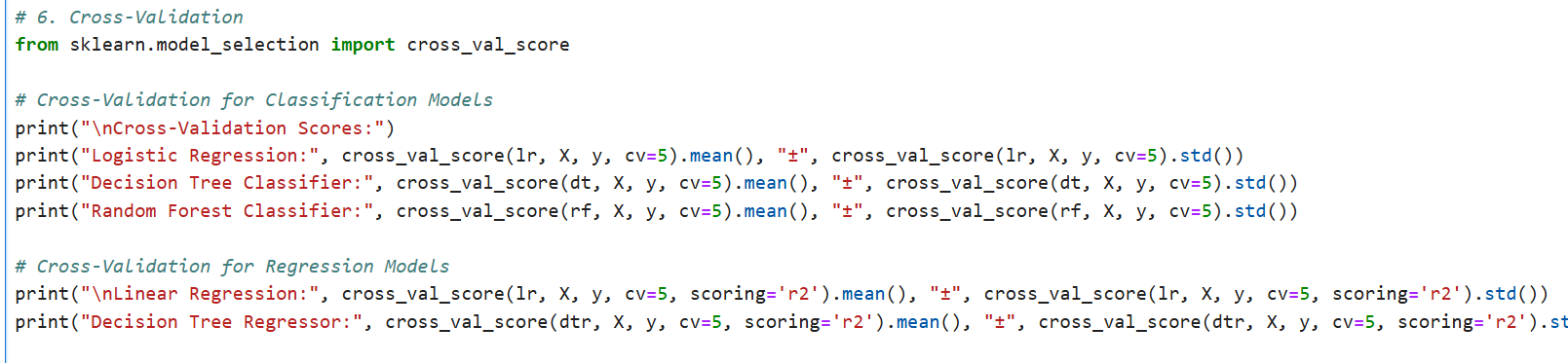
Step 4: Regression



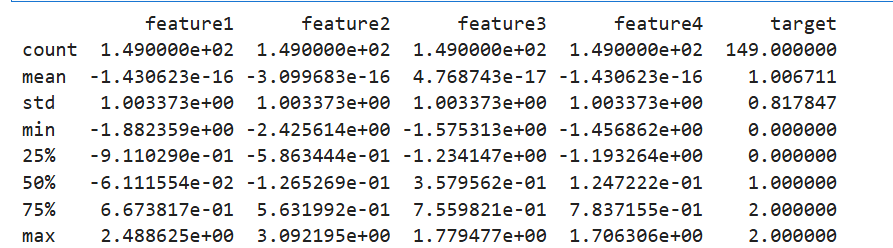
Step 5: Confusion Matrix Visualization

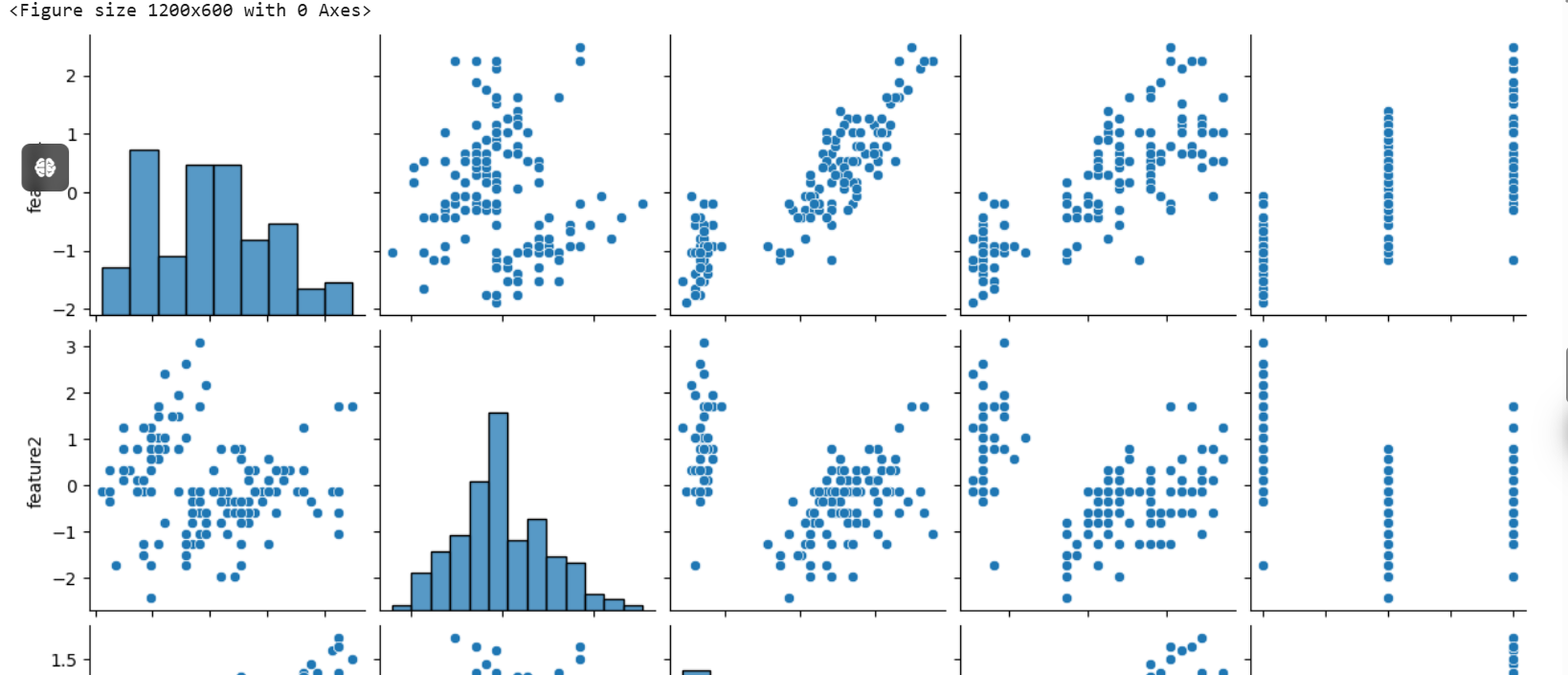


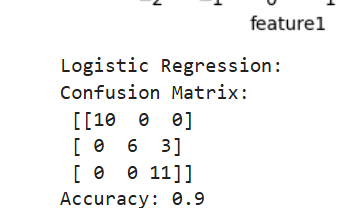
Step 6: Report and Interpret Results

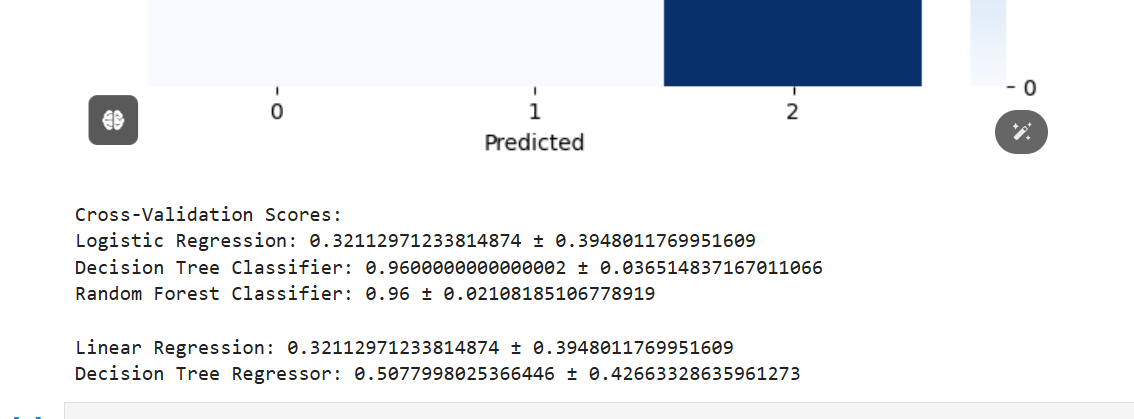


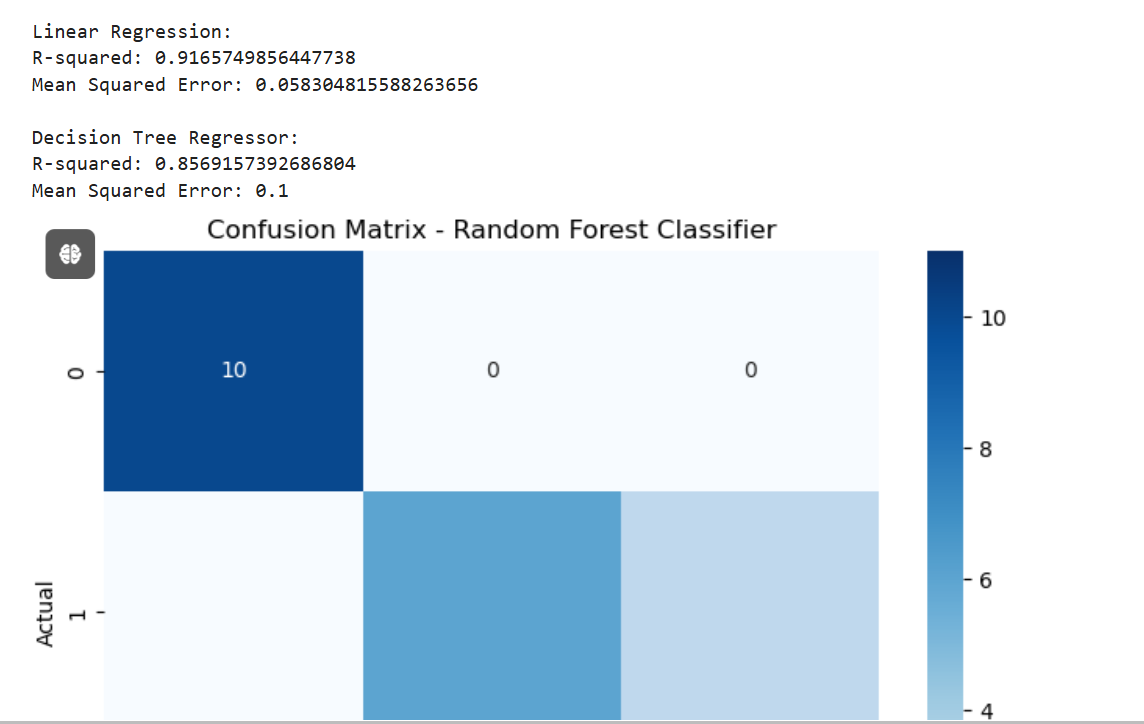
Output :

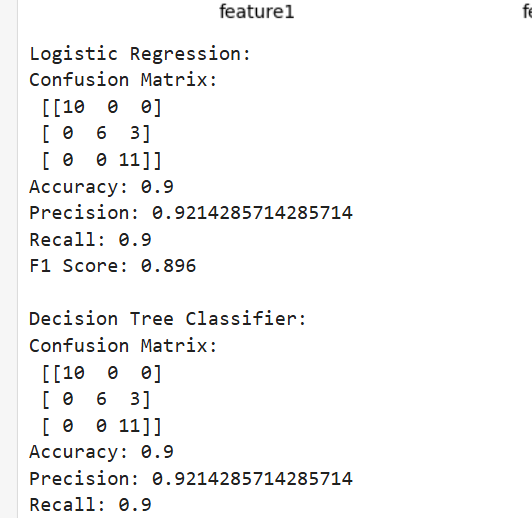












Result

Compile the results of each model, including accuracy, precision, recall, F1 score, MSE, and R-squared values. Interpret the performance of each model and discuss their strengths and weaknesses based on the evaluation metrics.