**Practical Technical Assessment AI**

Sarjitha P S

NSTI(W)Trivandrum 03-07-2024

**Tasks:**

1. **Data Preprocessing:**

o Load the dataset.

o Handle missing values.

o Encode categorical variables.

o Scale/normalize the features.

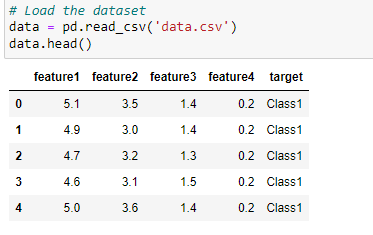
**1. Data Preprocessing**

**AIM :**

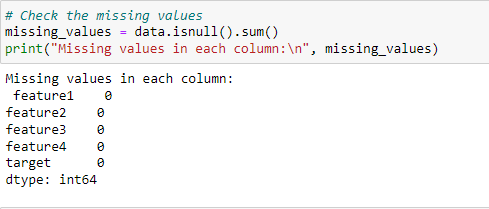
Prepare the dataset for analysis and modeling.

**PROCEDURE :-**

**Step 1:-Load the Dataset:** Use pandas to load the dataset from a CSV file.

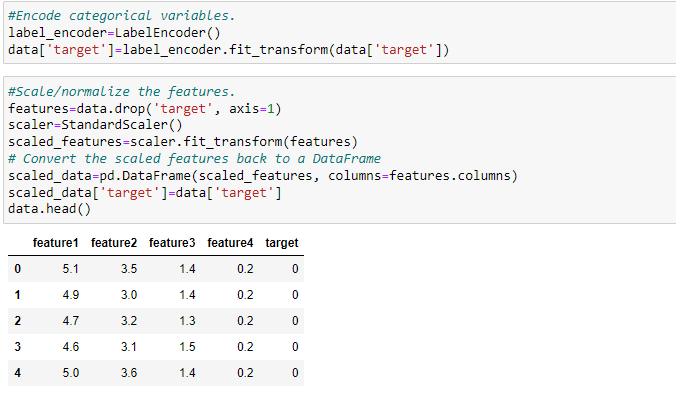


**Step 2:-Handle Missing Values:**  Address missing values using imputation or dropping strategies.



**Step 3:-Encode Categorical Variables:** Convert categorical variables into numerical values using one-hot encoding or label encoding.

**Step 4:-Scale/Normalize Features:** Apply standardization or normalization to features using techniques like StandardScaler or MinMaxScaler.



**2. Exploratory Data Analysis (EDA)**

**AIM :**

**Gain insights into the dataset's structure, distribution, and relationships.**

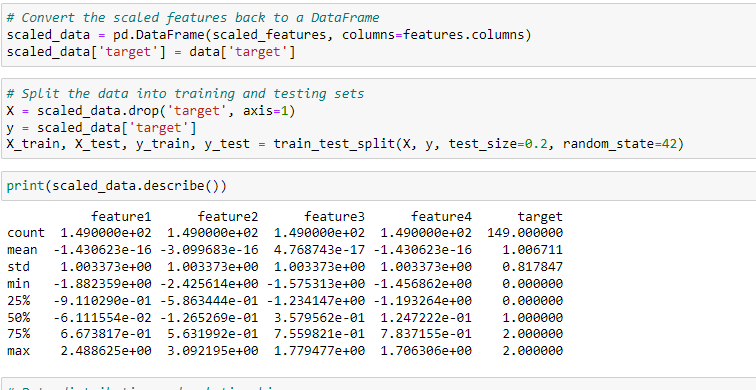
o Provide statistical summaries of the dataset.

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

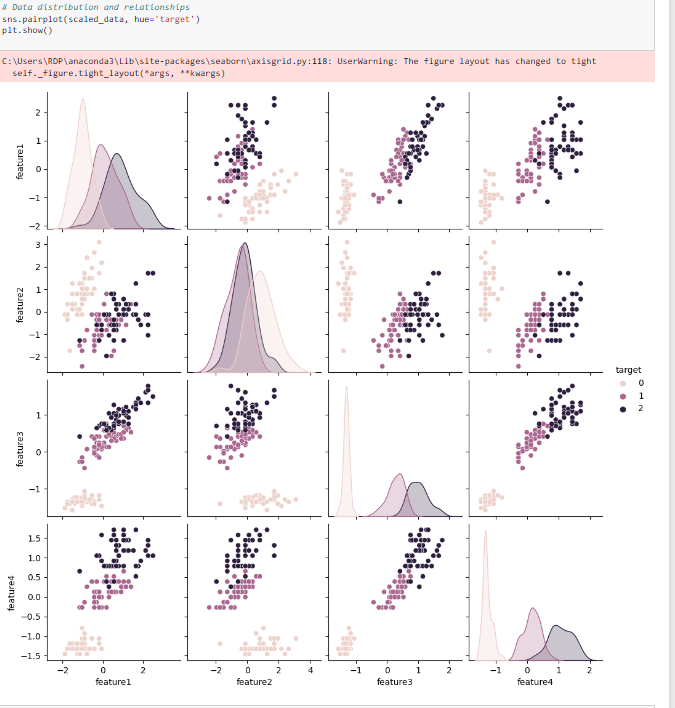
**PROCEDURE :-**

**Step1 :-Statistical Summaries:** Generate statistical summaries of the dataset using pandas' describe() function.

**Step 2 :-Visualize Data Distribution:** Use histograms, box plots, and density plots to visualize the distribution of features.



**Relationships Between Features:** Explore relationships using scatter plots, pair plots, and correlation matrices.



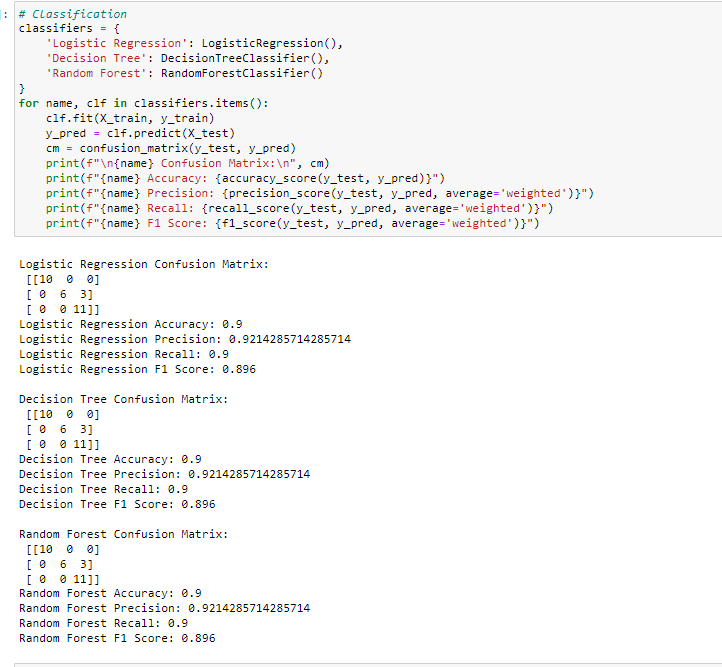
**3. Classification**

**AIM :**

**Build and evaluate classification models.**

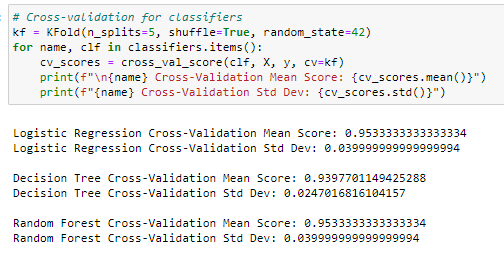
**PROCEDURE**

**Step1 :-Apply Classifiers:** Implement Logistic Regression, Decision Tree, and Random Forest classifiers using scikit-learn.



**Evaluation:** Use confusion matrices to assess the performance of each classifier. Calculate accuracy, precision, recall, and F1 score.

**Cross-Validation:** Perform k-fold cross-validation to evaluate the stability and generalizability of the models.



**4. Regression**

**AIM :**

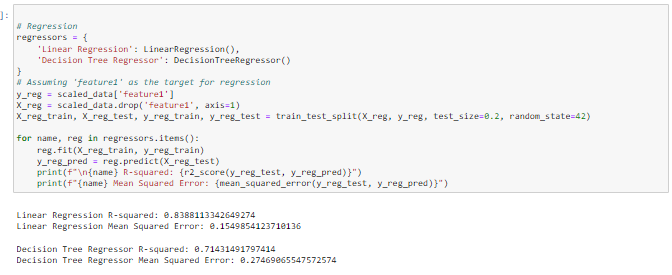
**Build and evaluate regression models.**

**PROCEDURE :-**

**Step 1:-Apply Regressors:** Implement Linear Regression and Decision Tree Regressor using scikit-learn.

**Evaluation:** Evaluate the models using R-squared and Mean Squared Error (MSE).

**Cross-Validation:** Perform k-fold cross-validation to assess the stability and generalizability of the models.



**5. Confusion Matrix**

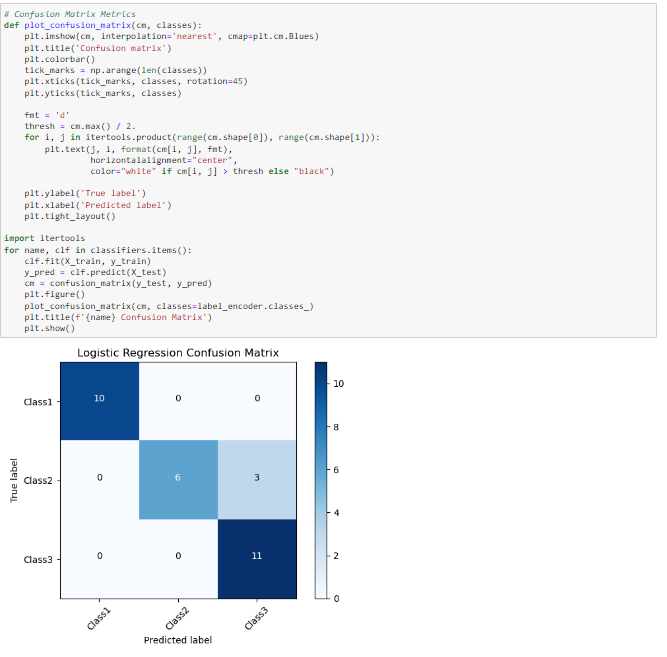
**AIM :**

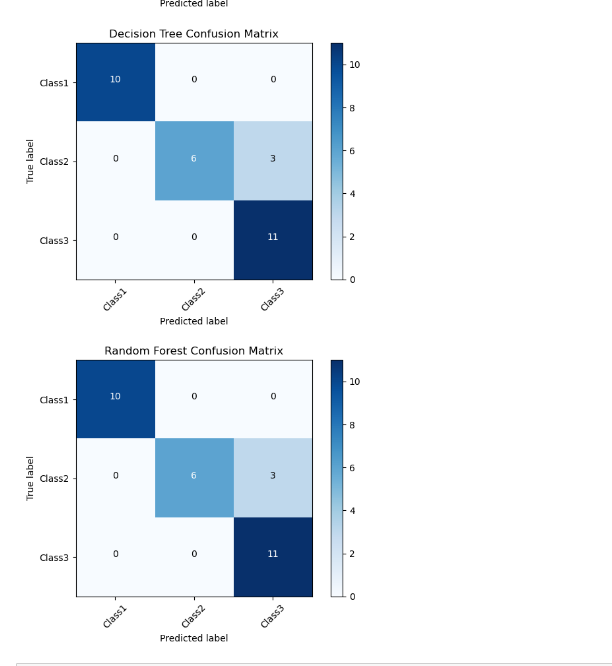
**Visualize and analyze the performance of classification models.**

**PROCEDURE :-**

**Step 1:-Plot Confusion Matrix:** Use heatmaps to plot the confusion matrix.

**Compute Metrics:** Calculate accuracy, precision, recall, and F1 score based on the confusion matrix.





**6. Cross-Validation**

**AIM:**

**Assess the performance of models across different subsets of the data.**

**PROCEDURE :-**

**Step 1:-Implement k-fold Cross-Validation:** Use scikit-learn's KFold or cross\_val\_score for both classification and regression models.

**Report Scores:** Report the mean and standard deviation of the cross-validation scores to understand the model's stability and variance in performance.

