

Practical Technical Assessment

Total Marks: 50 marks

Hours: 2hrs 30min

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 cross-validation.
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:**
 - Load the dataset.
 - Handle missing values.
 - Encode categorical variables.
 - Scale/normalize the features.
2. **Exploratory Data Analysis (EDA):**
 - Provide statistical summaries of the dataset.
 - Visualize the data distribution and relationships between features using plots.
3. **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.
4. **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.

5. **Confusion Matrix:**

- For classification tasks, plot the confusion matrix and compute the following metrics:
 - Accuracy
 - Precision
 - Recall
 - F1 Score

6. **Cross-Validation:**

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