Lab Assignment – Decision Trees

Problem Statement:

Build a decision tree classifier to predict the type of wine based on its chemical characteristics such as alcohol content, pH, etc.

Dataset:

The dataset used for this assignment is the Wine Quality dataset, which can be downloaded from the UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/Wine+Quality

The dataset consists of 1599 records with 11 features (fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol) and a target variable (quality). The data is stored in a CSV file named 'winequality-red.csv'.

Requirements:

- Load the dataset using the pandas library.
- Preprocess the data if necessary (e.g., handle missing values, scale the features, etc.).
- Split the dataset into training and testing sets.
- Train the decision tree classifier using the training data.
- Predict the target variable for the test data using the trained classifier.
- Evaluate the performance of the classifier using accuracy, precision, recall, and F1-score metrics.

Useful Hints:

- You can use the DecisionTreeClassifier class from the sklearn library to build the decision tree classifier.
- You can use the train_test_split function from the sklearn library to split the data into training and testing sets.
- You can use the accuracy_score, precision_score, recall_score, and f1_score functions from the sklearn library to evaluate the performance of the classifier.