Assignment-3

Dataset: Red Wine Quality

The dataset is related to the red variant of "Vinho Verde" wine. It contains 1599 data points where features are the physicochemical properties and the target value is quality which is an integer score ranging from 0-10. Your task is to classify if the wine provided is good based on its physicochemical properties.

(i) Create a new column on the dataset with binary values (i.e, 0 or 1) telling whether the wine is of good quality or not. You can categorise wines with quality>=7 to be of good quality. Drop the original ‘quality’ column.

(ii) Perform the data pre-processing steps that you feel are important for the given dataset.

(iii) Apply following classification algorithms on the given dataset (you are allowed to use scikit-learn library until not specified ‘from scratch’):

(a) Logistic Regression
(b) K-Nearest Neighbors
(c) Decision Trees Classifier
(d) Random Forest Classifier
(e) Logistic Regression from Scratch

(iv) Evaluate all your models based on the accuracy score and f1 score obtained on the test dataset.