Project Brief: Analysing Chemical Components of Wines and Predicting their Quality.

We consider the dataset winequality-red.csv. The dataset gives information about the chemical components and quality of wine. It has 10 variables, 9 of which gives different levels of the composition of a wine and the 10th giving the quality of the wine.

The variables are:

fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality

The features then are the first nine variables and the target variable is quality.

Tasks:

1) Load this data set into your Python workspace using the required library.
2) Check for any null values and inconsistent values in the dataset.
3) Perform any required cleaning.
4) Do some exploratory data analysis. In particular, create a heat map representing the correlations between any pair of the variables. Create scatter plots between variables with strong correlation (correlation coefficient greater than 0.7 or less than -0.7). Use a bar chart to visualise the distribution of the target variable.
5) Using the three machine learning models explained during the course, train a model that predicts the target variable. Use the ratio 80:20 for your train test split. Display your accuracy score, confusion matrix and classification report.
6) Use any of the ensemble algorithms, repeat task (5).