CMTH 642 DATA ANALYTICS: ADVANCED METHODS

ASSIGNMENT 2

Preparation:

The dataset is related to white Portuguese "Vinho Verde" wine. For more info: https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Import the following file:
http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv

This assignment will require writing R codes.
Please submit your RMD file together with an output file (PDF, Word or HTML) format in the same assignment link in D2L.

QUESTIONS

1. Check the datatypes of the attributes. (3 points)
2. Are there any missing values in the dataset? (4 points)
3. What is the correlation between the attributes other than Quality? (10 points)
4. Graph the frequency distribution of wine quality by using Quality. (10 points)
5. Reduce the levels of rating for quality to two levels as Pass and Fail. Assign the levels of 3, 4 and 5 to level Fail; and 6, 7, 8 and 9 to level Pass. (10 points)
6. Normalize the data set. (12 points)
7. Divide the dataset to training and test sets. (12 points)
8. Use the Logistic Regression algorithm to predict the quality of wine using its attributes. (12 points)
9. Display the confusion matrix to evaluate the model performance. (12 points)
10. Evaluate the model performance by computing Accuracy, Sensitivity and Specificity. (15 points)

This is the end of Assignment 2

Ceni Babaoglu, PhD