

There are two python files that answer their questions respectively

courseworkQuestion1.py -

source code that computes the answers for the questions in question 1.

The classifyTree function calculates the error rates and accuracy of the decision tree, the argument taken in is the dataframe and the function also encodes the values and creates a train and test set.

courseworkQuestion2.py -

This python file answers the questions relating to question 2, the python file drops the channel and region attributes. From there, the file calculates the mean, min and max values for every attribute and then completes the k-means algorithm and plots the clusters for every pair of attributes. Once completed the script finally calculates the BC, WC and BC/WC ratio values.