

Kenneth Woodard

Lab 1

COSC 425

Question 1.5

My heat map is not what it is supposed to be, and I apologize. I wish I had more time. I couldn't figure out how to get it to just show the diseases and not the countries. Obviously, seeing the countries on both axes but not the diseases is not ideal.

Question 2.1

The accuracy score of the tree classifier is 0.8807692307692307. The body mass index was the first feature it used to split the data. The value it split it on was 52.45

Question 2.2

The tree with max depth of 15 performed the best. Basically, if the depth is significantly more than 15, overfitting occurs. This just means that the tree becomes too stuck on the training data; therefore it loses accuracy on test data.

Question 2.5

The tree with a min sample split of 30 performance better than the other tree. The minimum sample split being higher basically forces the tree to be more refined. It won't allow a split in the decision tree until the minimum number of samples is met.