

### Question 6:

#### DummySet1

Tree Size: 3

Classification Rates: (1.0, [])

This dummy set has a very small tree size so it's subset has a higher probability of being a pure subset and obtaining a 1.0 classification rate. This tree size is too small and is more likely to be incorrect and will become more accurate if more data is added to training. This set produces an empty confusion matrix, so we can not see the ways in which the classification model is confused when it is making predictions.

#### DummySet2

Tree Size: 11

Classification Rates: (0.65, [(0, 1), (0, 1), (1, 0), (0, 1), (0, 1), (0, 1), (0, 1), (1, 0)])

Compared to Dummy Set 1, the tree size is over 3x larger which leads to a lower classification rate. There is more opportunity to not have a pure subset as the tree gets larger and the subset for the test sets gets lower and lower in the tree.

#### Car

Tree Size: 408

Classification Rates: ([0.96, 0.95, 0.93, 0.92, 0.935, 0.965, 0.945, 0.945, 0.95, 0.95, 0.945, 0.955, 0.935, 0.93, 0.955, 0.94, 0.955, 0.945, 0.93, 0.935], 0.9437499999999999)

Avg. Classification Rate: 0.943750

Car testing attributes have a subset of attributes from the data set that leads to less uncertainty and thus is why the classification rate is very high. The smaller tree size leads to less variation compared to one with a much larger tree size like in the Connect 4 example.

#### Connect4

Tree Size: 41,521

Classification Rates: ([0.753, 0.767, 0.7635, 0.761, 0.7565, 0.7635, 0.749, 0.757, 0.7615, 0.78], 0.7611999999999999)

Avg. Classification Rate: 0.7612

The Connect 4 tree size is extremely large due to its very large set of training data. Therefore, there is more uncertainty due to the many different results from the training data and is why the classification rates are lower. The avg rate (0.7612) is still above 50% but there is still a great deal of uncertainty.

### Question 7:

For the cars dataset, a website or advertisement company could use data regarding which vehicles are being searched for, as well as color, make, model, etc and use this as part of a training set. This information could then be used to guess as to which car a particular consumer may be most interested in using their previous searches. The decision tree would provide smaller and smaller subsets based on the level of specificity that is provided for the test set.

For the connect4 dataset, we could incorporate an iterative deepening search tree to look several moves ahead. This would work by assigning a value to each leaf of the search tree, while also making the best choice for the other color, and then choosing the move with the highest heuristic value.