Question 6

Dummy Set 1
Classification Rate: 1.0
Tree Size: 3

Dummy Set 2
Classification Rate: 0.65
Tree Size: 11

Connect 4
Average Classification Rate: 0.755
Tree Size: 41521

Cars
Average Classification Rate: 0.94375
Tree Size: 408

Dummy Set 1 had those results because its data set had only one attribute to partition from with values: 0, 1. Hence why the tree had a size of 3 nodes. The training set is overfitting for the testing set. Dummy Set 2 had those results because the data is very scattered or "noisy". There were more attributes to partition on which affected the classification rate. Connect 4 had an extremely large tree size because there are many different children to a single node based on where to go next in Connect 4. In Cars, reference an exponentially decreasing graph of tree size vs training data. The training data isn't too much as the testing data set is not overfitting with this decision tree, but instead it is near the equilibrium point of the "perfect" size of training data set to not underfit or overfit as shown by the average classification rate.

Question 7

Related to the Cars dataset, another similar dataset to be used is number of doors on the car, color, interior fabric, where it has fog lights, trunk size, number of seats in the car, whether it has power seats or not, etc. The dataset contains attributes of a car that can be sold or researched online. When looking to buy a car online, a decision tree will be helpful in narrowing down search results by filtering these attributes to what the user is looking for in a car. Related to the Connect 4 dataset, a similar dataset to be used is not only possible open spots to place a chip, but an attribute to be added is to check if the opponent is one chip away from winning the game. If so, then prevent that from happening if possible. These attributes can be customized to the level of difficulty when playing a computer bot in Connect 4.