Question 6:

Dummy Data Set 1

Training Time	0.00493813
Classification Rate	1
Tree Size	3

The first dummy data set had a classification score of 100% and a tree size of 3 (a root node and two leaf nodes). This is because the algorithm used the gain function to split based on attribute 5 first. Splitting based on attribute 5 creates homogeneous classification values in either node. Because attribute 5 determines the classification, the score was 100%.

Dummy Data Set 2

Training Time	0.01106215
Classification Rate	0.65
Tree Size	11

Even though the second dummy set size was the same size as the first (20), its training time was over twice as large. This is because the data had to be split by many more attributes (hence the larger tree size), meaning more recursive calls in the subtree method. The tree only split on attributes 2, 0, 4, 5, and 6. Perhaps if it had split the data based on more attributes, the classification score would be higher. A data set as complex as this one probably required more than 20 training data samples. A larger training data set would have allowed the algorithm to become more fitted to the data and bigger by making more splits.

Connect 4 Data Set

Training Time	1071.87177
Classification Rate	0.7555
Tree Size	41521

This data set had an extremely high training time. This is because, the training data set had 67577 samples and 42 attributes. Also the data set is not like data set 1 where a single split can classify the data. Determining the winner based on the configuration of the board is very complex. This means many recursive calls to the subtree method before a tree is made. One reason the tree is not as accurate as one would expect could be overfitting. Having a training data set with that many features could cause the tree to fit the training data too perfectly (even fitting the noise of the data set), making it hard to generalize the tree to data it has yet to see. The extremely large tree size (41521) is often a consequence of the tree being too narrowly fit to the data. This could be fixed with pruning, which would have minimized useless splits in the tree.

Car Data Set

Training Time	0.32299519
Classification Rate	0.94525
Tree Size	408

This data set had a much lower training time and tree size and a much higher classification rate than the previous data set. This is because the problem of classifying a car based on six

attributes is much simpler than the previous problem and the number of training samples (1728) was perfect for a data set with 6 attributes. A very small data set could have caused the decision tree to interpret the data's noise too "seriously" and over fit. It did not, as evidence by the small tree sixe

Question 7:

This dataset (https://www.kaggle.com/CooperUnion/cardataset) determines a car's MSRP based on features like make, model, year, and transmission. A regression-based decision tree built off of this data set can be useful for websites like Car Fax that monitor car sales and determine whether or not a sale is legit. For each sale posted on the website, the website may also put the projected price of the car so that buyers can determine whether or not to buy the car.

If one were building an AI for connect four, they could use Q-Learning based off of a Markov Decision Process model of Connect 4 to train an algorithm to play Connect 4 optimally. The decision tree we generated could be used to create a reward value for endgame states based on whether or not the decision tree classifies the configuration as winning. This way the algorithm will learn early-on to avoid moves that lead to a losing configuration classified by the decision tree and make moves that lead to winning configurations.