Question 6:

Dummy Dataset 1:

Treesize: 3, Classification Rate: 1.0

The decision tree is very accurate and the size is small for dummy set 1. This is because attribute 5 entirely determines the output for each example.

Dummy Dataset 2:

Treesize: 11, Classification Rate: 0.65

The accuracy was much lower because the training data set was too small. The tree splits on 5 attributes and only has 20 examples.

Connect4 Dataset:

Treesize: 41521, Average Classification Rate: 0.758800

Connect 4 had an extremely large data compared to the other datasets. This is because there are many different possibilities of future outcomes to the game, especially when the board is relatively empty.

Car Dataset:

Treesize: 408, Average Classification Rate: 0.941750

Compared to the connect4 dataset, there are much fewer attributes and relatively many more examples. This leads to a much higher classification rate and accuracy.

Question 7:

car:
The decision tree used for the car data set might be used for a car dealer website. The decision tree could classify a car's value based on make, model, year, mileage, etc.

This would be very useful for car dealerships to determine the price of a car to sell or for car owners to determine the price of their car.

connect4:
An algorithm that can be applied to the connect4 decision tree is the minimax algorithm. Connect4 is a two player game and this algorithm could be nicely applied to the game. The bot would look ahead and to future states and predict what the opponent will play. Then the bot will make its move based off of that.