Question 6:

The classification rate of dummy dataset 1 is 100% and the tree size 3. From the decision tree, the classification only depends on attribute 6 of the example. If attribute 6 is true, the result is false and if attribute 6 is false, the result is true. So the accuracy is very high with a small tree size.

As for dummy dataset 2, the classification rate was 65% while the tree size was 11. This was probably due to the fact that the classification depended on more than one attribute in dataset 2 compared to dataset 1, which makes the decision process more complicated, leading to the bigger tree size and lower classification rate. Because of this as well as the data size being the same, the decision tree did not perform as well for dummy dataset 2.

The classification rate for the Connect4 dataset was 74.5% while the tree size was 41521. This is probably due to the fact that there are 42 attributes with 3 possible values, which would lead to a huge number of combinations (3^{42}) despite there only being 67,557 examples and 3 classes. This means that each attribute contributes little to the data distribution, which means that a large amount of attributes are necessary to classify the examples. This would explain the large tree size.

The classification rate for the Car dataset was 94.65% while the tree size was 408. This is probably due to the fact that there are only 6 attributes with the first 3 having 4 possible values and the last 3 having 3 possible values, leading to 1728 combinations. Because the data set also has 1728 examples, one attribute can contribute a lot to the data distribution.

Question 7:

The Car dataset can be used for car dealer websites to help users make decisions using a data set that contains searches that all users make with attributes such as the car's mileage, safety, style, mpg, or make and model. The data collected can be used to build a decision tree to give recommendations to the user.

We can apply the mini max algorithm for the Connect4 dataset to minimize the possible loss for putting down a disk. This would inform a bot's decision and allow it to make better decisions during a game.