Question 6:

The cars dataset generated a classification rate of about 94.1%, while the connect 4 dataset generated a classification rate of about 75.7%. The tree size of the connect 4 was 41521 while the tree size of the cars dataset was 408. This is because the car dataset had better organization in terms of classification and didn’t posses a lot of random values that were provided by the dataset of connect4. Unlike the connect-4 dataset: the cars dataset possessed a classification at the end such as “vgood” and “good”. These classifications on the quality of the car allowed the tests to be more efficient and avoid being clogged up in large sub-trees. The attributes also contributed to determining the classification so all in all, the dataset was more organized than the connect-4 dataset. The tree size for dummy-1 was 3 while the dummy-2 tree size was 11. The classification rate of the dummy-1 dataset was 100% while the classification rate of the dummy-2 was 65%. This follows the same concept and idea that the datasets for cars and connect-4 did.

Question 7:

For the cars data set, this tree can be used to make advances in selling products on the market. So when consumers buy products and stuff, a dataset that contains the product’s statistics could be used to ultimately draw the conclusion on whether that product was good or not. It doesn’t necessarily have to be automobiles. It could be anything that ranges from clothing to video game products. The dataset for selling products can still contain the “vgood” and “good” classifications (or maybe more elaborate grading systems). For clothing, there could be a value for price, using rating, and other patterns that the company has the resources to monitor. For the connect-4 data, this could be used to help create a more effective game bot/agent in order to challenge human users. Even though a bit disorganized, the dataset could begin to draw conclusions make predictions on when and where the player may think about moving his/her pieces.