Question 6:

The dummyDataSet 1 and 2 have size 3 and 11 with 1.0 and 0.65 classification rate, and both of them have input size 20. As result, it seems like dummy1 is more accurate. The reason of high classification rate for set 1 is the index 5 entries in the first dataset have a perfect correlation with the labels. On the other hand, the reason of low classification rate for set 2 is that the training data set has a small size which does not give the best attribute for the testing set.

The connect4DataSet has size 41521 with 0.763 classification rate, and it has input size 67557. There has 42 attributes and each attribute has 3 different choices which has 3^42 combinations. Therefore, it needs a large amount of attributes to classify and makes that the correlate pretty low with the classification. It generates a large tree, and does not give a high classification rate.

However, the carDataSet has size 408 with 0.94 classification rate, and it has input size 1728. There has 6 attributes, and the first three has 4 different choices and last three has 3 different choices which has (4^3) \* (3^3) = 1728 combinations. Therefore, each attributes contributes a lot without useless attributes. It generates a small tree, and give a high classification rate.

Question 7:

The decision tree can be used for (secondhand car) dealer website. The customer can searches by choosing the attribute they want such as engine, price, etc. and each search corresponds to a class. After collecting a large amount of data, the agent can build a decision tree to give the recommendations to users. When a user makes his/her decision, the system will collect it and apply to the decision tree, and give the best recommendation.

The connect4 can be applied minimax algorithm on it. We can limit the depth by using a heuristics function since the search space is huge. For example, we can check all consecutive pieces. We add 1 if its piece is ours, and subtract 1 if its piece is opponent. Collect all the result and we can use the minimax to decide which move to take by our decision tree.