Q6

Fake Set 1

Size: 3. Classification rate: 1.0.

Fake Set 1 has a 100% accuracy because of attribute #5. This attribute directly determined the label for the example. Since this defining value was so strong, it allowed for a definitive result, leading to 100% accuracy.

Fake Set 2

Size: 11. Classification rate: 0.65

Fake Set 2 has a much lower accuracy rate than Fake Set 1. This training data set was small. There was no defining attribute that could split the data like in Fake Set 1. The small sample size made it so that it could not find the best split to define the data, making the accuracy so low.

Car Set

Size: 408. Classification rate:0.939.

Connect4

Size > 40000. Classification rate: 0.764.