Implementing the ID3 Algorithm

Program prog.r contains the source for the implementation. To handle basic tree services, I used the data.tree library.

Figures 1 & 2 show runs of training, test, and the whole data set. I did a 75/25 split of the tennis set for training/test. You can see from the runs how wildly different the decisions trees are for different samples of data. It highlights that just slight changes to the Yes/No values for an attribute can affect greatly the entropy for that attribute, shifting the order in which it is chosen for entropy reduction.

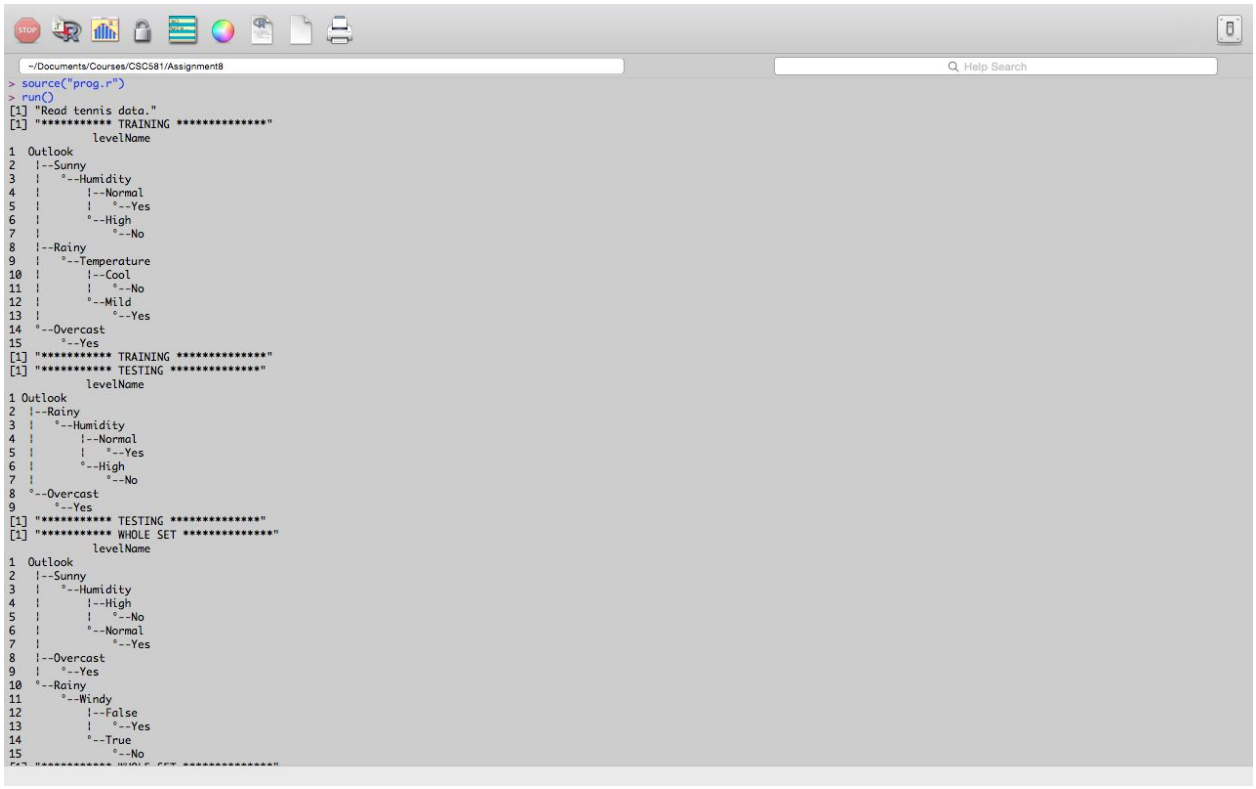

Figure 1.

Looking at the confusion matrix (Figure 2), the training set shows a 100% accuracy rate. The decision tree seems to be a good model for classifying whether or not someone plays tennis given weather-related factors.

	Predicted	
Observed	+1	-1
+1	6	0
-1	0	4

Figure 2.