Assignment 1

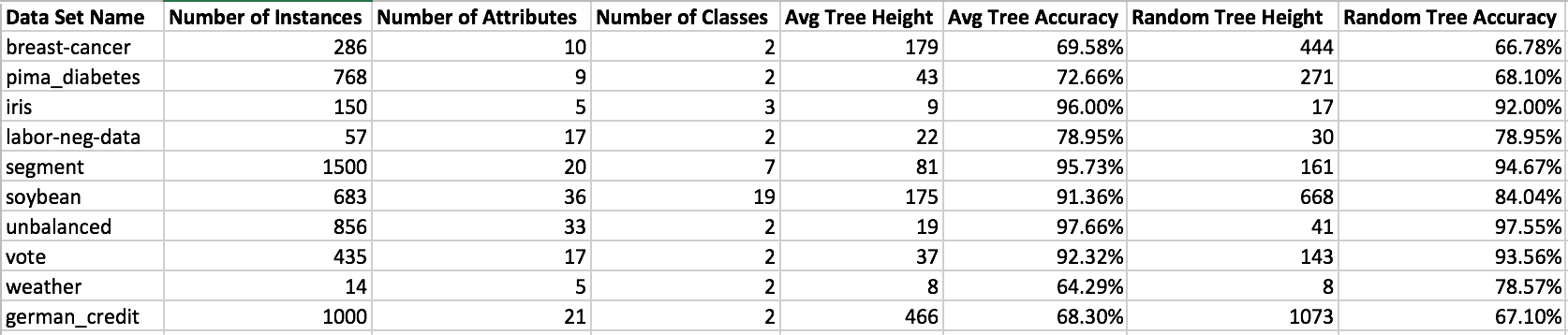
Joshua Galindo

Feb 13th 2017

Describe the details (e.g., implementation if any, software/packages used, parameter setting, etc.) of your work/experiments. Include all raw experiment results, and several screenshots as supporting evidence.

1. I used Weka as the software to collect my data from. All my test were ran using J48 which includes C4.5. All my raw data has been included in the Zip file. Also note that the Gain Ratio results are included at the bottom of the files without \_random in the title.

Use a table to present your experiment results. The table should include the following information: name of the dataset, number of instances, number of attributes, number of classes, C4.5 average height, C4.5 average accuracy, Random average height, and Random average accuracy.



Observe and report the tree heights and accuracy performance of C4.5 and Random. Would C4.5 generate shorter trees than Random on average? Which is more accurate on average? Provide a summative analysis on your experiment results to support your answers. There are no correct answers to these questions.

1. During my tests I found that the random tree height was larger than the average tree height in all 10 cases. The tree accuracy varied in all ten cases. But in most cases the C4.5 tree accuracy was higher. In fact, the C4.5 accuracy was higher in 7/10 cases and was the same in percent in 1 case.

I believe the Random Tree results had a higher height because they the Weka program doesn’t allow you to control the pruning results like the J48 results do.