Assignment 1

# Daniel Ortiz

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Number of Instances | Number of Attributes | Number of Classes | C4.5  Height | C4.5  Accuracy | Random  Height | Random  Accuracy |
| BalanceScale | 625 | 5 | 3 | 11 | 79.20% | 13 | 77.28% |
| Balloons | 20 | 5 | 2 | 3 | 100.0% | 3 | 100.0% |
| Car | 1728 | 7 | 4 | 6 | 94.15% | 7 | 83.16% |
| Contraceptive | 1473 | 10 | 3 | 17 | 48.88% | 21 | 46.63% |
| Ecoli | 336 | 8 | 8 | 10 | 83.63% | 12 | 78.27% |
| Glass | 214 | 10 | 7 | 11 | 67.29% | 14 | 70.10% |
| Iris | 150 | 5 | 3 | 5 | 96.00% | 6 | 92.00% |
| Nursery | 12960 | 9 | 5 | 8 | 98.79% | 9 | 94.62% |
| TicTacToe | 958 | 10 | 2 | 7 | 86.43% | 9 | 81.21% |
| Trains | 10 | 33 | 2 | 3 | 90.00% | 2 | 50.00% |

|  |  |  |  |
| --- | --- | --- | --- |
| C4.5 Average Height | C4.5 Average Accuracy | Random Average Height | Random Average Accuracy |
| 8.1 | 84.44% | 9.6 | 67.87% |

In this experiment, I compared the height and accuracy averages of two different decision tree algorithms called C4.5 and Random Tree. To achieve the results listed above I used the software program called WEKA and computed decision trees based off some databases from the ICU website. Using C4.5 algorithm without pruning and the Random Tree algorithm I was able to conduct this comparison between the two algorithms.

On average C4.5 algorithm produced shorter trees at an average height of 8.1 while the Random Tree algorithm produced an average height of 9.6. The C4.5 algorithm was not always the best choice to produce shorter trees as with the Balloons database it produced the same exact tree as the Random Tree did and in the Train database it produced a taller tree than the Random Tree algorithm. C4.5 had a higher average accuracy (84.44%) than the Random Tree (67.87%) but on occasion C4.5 performed just a little better, within 10%, than the Random Tree. C4.5 did have a worse accuracy percentage than the Random Tree during the Glass database.

Because of Random Tree’s simpler algorithm, it would produce trees at a quicker rate when dealing with large instances of data with large number of attributes. Though typically speed is not always the main factor when choosing the appropriate decision tree algorithm. C4.5 with its 84.44% accuracy and average height of 8.1 makes it the better choice for creating efficient, accurate, and short decision trees.