**Experiment 1**

**Effect of Early Stopping:**

With the early stopping of Decision Tree at particular depth has following stats:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Depth** | **Total Nodes** | **Number of Terminal Nodes** | **Training Accuracy** | **Test Accuracy** |
| 4 | 169 | 85 | 0.963 | 0.962 |
| 5 | 295 | 148 | 0.96 | 0.958 |
| 6 | 287 | 144 | 0.973 | 0.958 |
| 7 | 265 | 133 | 0.971 | 0.96 |
| 8 | 253 | 127 | 0.968 | 0.957 |

So with stopping the execution early, the no. of terminal nodes first increases and then decreases. This trend is due to creation of entirely new decision tree which is based on random numbers. The same trend can be seen in accuracy results. This is because intermediate trees give much accuracy then fully developed Tree.

**Number of times an attribute is used**

|  |  |  |
| --- | --- | --- |
| Attribute 0 Counts: 14.0  Attribute 1 Counts: 0.0  Attribute 2 Counts: 5.0  Attribute 3 Counts: 1.0  Attribute 4 Counts: 0.0  Attribute 5 Counts: 1.0  Attribute 6 Counts: 5.0  Attribute 7 Counts: 2.0  Attribute 8 Counts: 4.0  Attribute 9 Counts: 6.0  Attribute 10 Counts: 2.0  Attribute 11 Counts: 0.0  Attribute 12 Counts: 1.0  Attribute 13 Counts: 2.0  Attribute 14 Counts: 1.0  Attribute 15 Counts: 1.0  Attribute 16 Counts: 0.0  Attribute 17 Counts: 1.0  Attribute 18 Counts: 1.0  Attribute 19 Counts: 0.0  Attribute 20 Counts: 0.0  Attribute 21 Counts: 1.0  Attribute 22 Counts: 1.0  Attribute 23 Counts: 0.0  Attribute 24 Counts: 0.0  Attribute 25 Counts: 2.0  Attribute 26 Counts: 0.0  Attribute 27 Counts: 0.0  Attribute 28 Counts: 0.0 | Attribute 29 Counts: 0.0  Attribute 30 Counts: 1.0  Attribute 31 Counts: 0.0  Attribute 32 Counts: 0.0  Attribute 33 Counts: 0.0  Attribute 34 Counts: 0.0  Attribute 35 Counts: 0.0  Attribute 36 Counts: 0.0  Attribute 37 Counts: 0.0  Attribute 38 Counts: 0.0  Attribute 39 Counts: 0.0  Attribute 40 Counts: 0.0  Attribute 41 Counts: 0.0  Attribute 42 Counts: 0.0  Attribute 43 Counts: 0.0  Attribute 44 Counts: 0.0  Attribute 45 Counts: 0.0  Attribute 46 Counts: 0.0  Attribute 47 Counts: 0.0  Attribute 48 Counts: 0.0  Attribute 49 Counts: 0.0  Attribute 50 Counts: 0.0  Attribute 51 Counts: 0.0  Attribute 52 Counts: 0.0  Attribute 53 Counts: 0.0  Attribute 54 Counts: 0.0  Attribute 55 Counts: 0.0  Attribute 56 Counts: 0.0 | Attribute 57 Counts: 0.0  Attribute 58 Counts: 0.0  Attribute 59 Counts: 0.0  Attribute 60 Counts: 0.0  Attribute 61 Counts: 0.0  Attribute 62 Counts: 0.0  Attribute 63 Counts: 0.0  Attribute 64 Counts: 0.0  Attribute 65 Counts: 0.0  Attribute 66 Counts: 0.0  Attribute 67 Counts: 0.0  Attribute 68 Counts: 0.0  Attribute 69 Counts: 0.0  Attribute 70 Counts: 0.0  Attribute 71 Counts: 0.0  Attribute 72 Counts: 0.0  Attribute 73 Counts: 0.0  Attribute 74 Counts: 0.0  Attribute 75 Counts: 0.0  Attribute 76 Counts: 0.0  Attribute 77 Counts: 0.0  Attribute 78 Counts: 0.0  Attribute 79 Counts: 1.0  Attribute 80 Counts: 0.0  Attribute 81 Counts: 0.0  Attribute 82 Counts: 0.0  Attribute 83 Counts: 0.0  Attribute 84 Counts: 0.0 |

**Experiment 2**

**Effect of adding Noise to data on Decision Tree**

The effect of adding noise by changing class of training variable is can be seen in this table. This noise has been added by random number generator at random indices of training data

|  |  |
| --- | --- |
| **Noise (in %)** | **Number of Nodes** |
| 0 | 173 |
| 0.5 | 305 |
| 1 | 279 |
| 5 | 343 |
| 10 | 405 |

* The results are not linear in starting but later it follows linear trend
* This is entirely due to noise added through random numbers so every time it will come up with different results
* Sometimes the change in class label happens at the right indices and sometimes at wrong indices of training data and it behaves accordingly
* But as we can see from stats there is upward lift with increasing noise in data. So further increase of noise in data probably will increase complexity of decision tree and hence number of nodes will increase
* The more number of nodes used also indicate the over fitting of data.

\*The results in next execution can be different as noise is added by random number generator

**Experiment No. 3**

**Post Pruning of tree**

1. **Given data**

With the given data the number of nodes reduced to only root node as data was highly unbalanced with the same accuracy as with fully developed tree or higher

|  |  |  |
| --- | --- | --- |
| **Tree** | **Number of Nodes** | **Accuracy** |
| No Pruning | 211 | 0.964 |
| Post Pruning | 1 | 0.964 |

1. **Down Sampled Data**

With the down sampled data number of nodes decreases as shown in table with enhance in accuracy

|  |  |  |
| --- | --- | --- |
| **Tree** | **Number of Nodes** | **Accuracy** |
| No Pruning | 311 | .558 |
| Post Pruning | 157 | .608 |