Machine Learning
Report of Assignment - 2
George Koshy - gxk140830

Question1 ID3 algorithm:

Data Set1 :
Accuracy observed

Total Number of nodes pruned	Accuracy Observed
0	75.85%
5	76.8%
10	77.45%
15	76.35%
20	76.3%
25	75.85%

Data Set2 :
Accuracy observed

Total Number of nodes pruned	Accuracy Observed
0	72.33%
5	74%
10	75.6%
15	74.66%
20	73.5%
25	73.5%

Observations: It can be observed that the accuracy increases on pruning of nodes and after reaching a specific number the accuracy of the decision tree get poorer. This notion could be wrong for different datasets.

Bonus assignment question:

Specify the tree parameters built

Data Set 1:

	Average Depth	Number of Nodes
Decision Tree with ID3	7.438	177
Tree With Random Attributes	9.257	427

Data Set 2:

	Average Depth	Number of Nodes
Decision Tree with ID3	7.958	191
Tree With Random Attributes	9.238	485

Accuracy Comparison bonus question:

Data Set1 :
Accuracy observed

Total Number of nodes pruned	Accuracy Observed in ID3	Accuracy in Random Att Selec
0	75.85%	67.45%
5	76.8%	80.4%
10	77.45%	64.1%
15	76.35%	70.7%
20	76.3%	71.05%
25	75.85%	83.8%

Data Set2 :
Accuracy observed

Total Number of nodes pruned	Accuracy Observed	Accuracy in Random Att Selec
0	72.33%	71.5%
5	74%	80.6%
10	75.6%	81%
15	74.66%	77.6%
20	73.5%	65.5%
25	73.5%	71.8%

Observation: Each of trees generated by random attribute selection are completely different in each run and the nodes pruned are also completely random. Once can observe in multiple runs the accuracy going to as low was 53% and high as 85% as well.

Random selection of nodes created a tree of almost twice the number of nodes than ID3. Hence ID3 gives the smallest tree possible.