Report

The accuracies of the tree using Information gain heuristic are:

Dataset 1 = 75.15%
Dataset 2 = 76.83%

The accuracies of the tree using Variance impurity heuristic are:

Dataset 1 = 77%
Dataset 2 = 77.16%

The accuracies of the pruned trees for different values of L,K are shown below.

Heuristic 1: Information gain
Dataset 1:

Accuracy on the test dataset 1:

L	K	Accuracy (in %)
3	3	75.15
3	4	75.15
2	4	75.15
4	6	75.15
2	2	75.15
4	5	75.15
3	5	75.15
5	5	75.15
5	4	75.15
5	3	75.15

Accuracy on the test dataset 2:

Dataset 2:

L	K	Accuracy (in %)
3	3	76.83
3	4	76.83
2	4	76.83
4	6	76.83
2	2	76.83
4	5	76.83

3	5	76.83
5	5	76.83
5	4	76.83
5	3	76.83

Heuristic 2: Variance Impurity
Dataset 1:

Accuracy on the test dataset 1:

L	K	Accuracy (in %)
3	3	77
3	4	77
2	4	77
4	6	77
2	2	77
4	5	77
3	5	77
5	5	77
5	4	77
5	3	77

Accuracy on the test dataset 2:

Dataset 2:

L	K	Accuracy (in %)
3	3	77.16
3	4	77.16
2	4	77.16
4	6	77.16
2	2	77.16
4	5	77.16
3	5	77.16
5	5	77.16
5	4	77.16
5	3	77.16

All the cases we tried for different combinations of L and K, we do not get better accuracy for the post-pruned trees. This is due to the fact that the in the pruning process selecting a random integer P might not always give a tree with the better accuracy than the original one. A better heuristic for the pruning could result in pruned-trees with higher accuracy.