# MACHINE LEARNING

MEHAK BERI | Net Id: MXB166430 | HomeWork-1

**On the two datasets available on the class web page:**

* **Report the accuracy on the test set for decision trees constructed using the two heuristics mentioned above.**
* DATASET 1 – Test Set Accuracy
  + Information Gain: 75.149 %
  + Variance Impurity: 76.1 %
* DATASET 2 – Test Set Accuracy
  + Information Gain
  + Variance Impurity
* **Choose 10 suitable values for L and K (not 10 values for each, just 10 combinations). For each of them, report the accuracies for the post pruned decision trees constructed using the two heuristics.**
* DATASET 1 – Post Pruning Accuracy on Validation Set
  + Initial accuracy without pruning with Information Gain on Test Set: 75.149 %
  + Initial accuracy without pruning with Variance Impurity on Test Set: 76.1 %
  + Initial accuracy without pruning with Information Gain on Validation Set: 74.3 %
  + Initial accuracy without pruning with Variance Impurity on Validation Set: 74.65 %

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| INDEX | L | K | Validation Set Accuracy- Information Gain(%) | Test Set Accuracy- Information Gain(%) | Validation Set Accuracy- Variance Impurity(%) | Test Set Accuracy- Variance Impurity(%) |
| 1 | 3 | 5 | 74.3 | 75.75 | 75.5 | 76.1 |
| 2 | 5 | 10 | 74.55 | 76 | 75.6 | 76.14 |
| 3 | 5 | 20 |  |  |  |  |
| 4 | 5 | 30 |  |  |  |  |
| 5 | 6 | 10 |  |  |  |  |
| 6 | 6 | 20 |  |  |  |  |
| 7 | 6 | 30 |  |  |  |  |
| 8 | 7 | 20 |  |  |  |  |
| 9 | 7 | 40 |  |  |  |  |
| 10 | 7 | 50 |  |  |  |  |

* DATASET 2 – Post Pruning Accuracy on Validation Set