# 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
  + 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 %
* DATASET 2 – Test Set Accuracy
  + Initial accuracy without pruning with Information Gain on Test Set: 74.5 %
  + Initial accuracy without pruning with Variance Impurity on Test Set: 74.16 %
  + Initial accuracy without pruning with Information Gain on Validation Set: 77 %
  + Initial accuracy without pruning with Variance Impurity on Validation Set: 75.33 %
* **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
  + 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 | 75.35 | 75.94 | 76.35 |
| 2 | 3 | 10 | 74.3 | 75.75 | 74.65 | 76.2 |
| 3 | 5 | 10 | 74.95 | 75.25 | 75.3 | 76.9 |
| 4 | 5 | 20 | 75.55 | 76.5 | 75.0 | 77.2 |
| 5 | 5 | 30 | 74.3 | 75.7 | 75.94 | 77.14 |
| 6 | 6 | 10 | 75.4 | 75.7 | 75.4 | 76.75 |
| 7 | 6 | 20 | 74.55 | 75.85 | 74.65 | 76.1 |
| 8 | 6 | 30 | 74.3 | 75.14 | 74.75 | 76.3 |
| 9 | 7 | 30 | 74.3 | 75.14 | 75.0 | 76.1 |
| 10 | 7 | 40 | 74.3 | 75.14 | 74.8 | 76.2 |

* DATASET 2 – Post Pruning Accuracy
  + Initial accuracy without pruning with Information Gain on Test Set: 74.5 %
  + Initial accuracy without pruning with Variance Impurity on Test Set: 74.16 %
  + Initial accuracy without pruning with Information Gain on Validation Set: 77 %
  + Initial accuracy without pruning with Variance Impurity on Validation Set: 75.33 %

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 | 77.0 | 74.66 | 75.83 | 75.33 |
| 2 | 3 | 10 | 77.0 | 74.5 | 77.0 | 74.16 |
| 3 | 5 | 10 | 77.5 | 75.5 | 76.5 | 75.16 |
| 4 | 5 | 20 | 77.0 | 74.83 | 75.33 | 75.33 |
| 5 | 5 | 30 | 77.66 | 74.5 | 75.33 | 74.16 |
| 6 | 6 | 10 | 78.0 | 74.5 | 76.5 | 74.5 |
| 7 | 6 | 20 | 77.0 | 77.5 | 76.16 | 76.66 |
| 8 | 6 | 30 | 77.0 | 74.5 | 75.66 | 74.5 |
| 9 | 7 | 30 | 79.33 | 74.5 | 75.33 | 74.16 |
| 10 | 7 | 40 | 77.33 | 74.5 | 77.5 | 76.33 |