**Report file HW1**

**(Xingyu Huang, Chunlei Wang, Wenbo Zhao, and Soyeon Park)**

**# Dataset 1**

* Accuracy before pruning for entropy heuristic: 74.65%
* Accuracy before pruning for variance impurity heuristic: 77.0%
* Accuracies after pruning for different values of L and K:

Accuracy after pruning for entropy heuristic

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **L** | **K** | **Accuracy (%)** | **L** | **K** | **Accuracy (%)** |
| 8 | 3 | 74.75 | 23 | 5 | 75.15 |
| 21 | 6 | 76.2 | 14 | 17 | 76.05 |
| 15 | 9 | 75.95 | 16 | 2 | 75.35 |
| 12 | 1 | 75.6 | 19 | 7 | 75.15 |
| 25 | 5 | 76.8 | 24 | 1 | 75.9 |

Accuracy after pruning for variance impurity heuristic

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **L** | **K** | **Accuracy (%)** | **L** | **K** | **Accuracy (%)** |
| 8 | 3 | 78.35 | 23 | 5 | 77.3 |
| 21 | 6 | 78.25 | 14 | 17 | 78.15 |
| 15 | 9 | 78.05 | 2 | 10 | 77.3 |
| 12 | 1 | 77.35 | 19 | 7 | 77.35 |
| 25 | 5 | 78.9 | 24 | 1 | 78.55 |

**# Dataset 2**

* Accuracy before pruning for entropy heuristic: 72.5%
* Accuracy before pruning for variance impurity heuristic: 72.67%
* Accuracies after pruning for different values of L and K:

Accuracy after pruning for entropy heuristic

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **L** | **K** | **Accuracy (%)** | **L** | **K** | **Accuracy (%)** |
| 8 | 3 | 72.67 | 23 | 5 | 76.0 |
| 21 | 6 | 73.0 | 14 | 17 | 73.5 |
| 15 | 9 | 73.33 | 2 | 10 | 72.67 |
| 12 | 1 | 73.0 | 19 | 7 | 75.5 |
| 25 | 5 | 76.17 | 24 | 1 | 74.83 |

Accuracy after pruning for variance impurity heuristic

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **L** | **K** | **Accuracy (%)** | **L** | **K** | **Accuracy (%)** |
| 8 | 3 | 72.83 | 6 | 29 | 73.0 |
| 21 | 6 | 73.5 | 14 | 17 | 76.5 |
| 15 | 9 | 76.5 | 2 | 11 | 73.5 |
| 12 | 1 | 74.17 | 19 | 7 | 75.0 |
| 25 | 5 | 77.0 | 24 | 1 | 75.83 |

According to the result of accuracy before pruning and after pruning table, we see accuracy increase in dataset1 and dataset 2 after different combination of L and K pruning. Therefore, the pruned tree is better and fit for datasets.