**IMPLEMENTATION OF DECISION TREE USING INFORMATION GAIN & VARIANCE IMPURITY HEURISTICS**

**DATASET1:** **INFORMATION GAIN**

|  |  |  |
| --- | --- | --- |
| [L,K] | INFORMATION GAIN | |
| PRE-PRUNING | POST-PRUNING |
| [8,5] | 0.751624187906047 | 0.7531234382808596 |
| [12,15] | 0.751624187906047 | 0.7566216891554223 |
| [20,10] | 0.751624187906047 | 0.7566216891554223 |
| [10,15] | 0.751624187906047 | 0.7561219390304847 |
| [8,15] | 0.751624187906047 | 0.7551224387806097 |
| [15,8] | 0.751624187906047 | 0.7646176911544228 |
| [25,15] | 0.751624187906047 | 0.7551224387806097 |
| [10,25] | 0.751624187906047 | 0.7591204397801099 |
| [30,15] | 0.751624187906047 | 0.7501249375312344 |
| [15,5] | 0.751624187906047 | 0.7591204397801099 |

**DATASET1:** **VARIANCE IMPURITY**

|  |  |  |
| --- | --- | --- |
| [L,K] | VARIANCE IMPURITY | |
|  | PRE-PRUNING | POST-PRUNING |
| [8,5] | 0.7611194402798601 | 0.7631184407796102 |
| [12,15] | 0.7611194402798601 | 0.7666166916541729 |
| [20,10] | 0.7611194402798601 | 0.7631194402798601 |
| [10,15] | 0.7611194402798601 | 0.7636176911544228 |
| [8,15] | 0.7611194402798601 | 0.7631179410294853 |
| [15,8] | 0.7611194402798601 | 0.7641169415292354 |
| [25,15] | 0.7611194402798601 | 0.7641164417791104 |
| [10,25] | 0.7611194402798601 | 0.7731134432783608 |
| [30,15] | 0.7611194402798601 | 0.7641159420289855 |
| [15,5] | 0.7611194402798601 | 0.7651194402798601 |

**DATASET2:** **INFORMATION GAIN**

|  |  |  |
| --- | --- | --- |
| [L,K] | INFORMATION GAIN | |
| PRE-PRUNING | POST-PRUNING |
| [8,5] | 0.7454242928452579 | 0.7587687188019967 |
| [12,15] | 0.7454242928452579 | 0.7566216891554223 |
| [20,10] | 0.7454242928452579 | 0.7566216891554223 |
| [10,15] | 0.7454242928452579 | 0.7561219390304847 |
| [8,15] | 0.7454242928452579 | 0.7551224387806097 |
| [15,8] | 0.7454242928452579 | 0.7646176911544228 |
| [25,15] | 0.7454242928452579 | 0.7551224387806097 |
| [10,25] | 0.7454242928452579 | 0.7591204397801099 |
| [30,15] | 0.7454242928452579 | 0.7501249375312344 |
| [15,5] | 0.7454242928452579 | 0.7591204397801099 |

**DATASET2:** **VARIANCE IMPURITY**

|  |  |  |
| --- | --- | --- |
| [L,K] | VARIANCE IMPURITY | |
|  | PRE-PRUNING | POST-PRUNING |
| [8,5] | 0.7420965058236273 | 0.7303993344425957 |
| [12,15] | 0.7420965058236273 | 0.7303826955074875 |
| [20,10] | 0.7420965058236273 | 0.7311194402798601 |
| [10,15] | 0.7420965058236273 | 0.7346176911544228 |
| [8,15] | 0.7420965058236273 | 0.7341179410294853 |
| [15,8] | 0.7420965058236273 | 0.7361169415292354 |
| [25,15] | 0.7420965058236273 | 0.7371164417791104 |
| [10,25] | 0.7420965058236273 | 0.7331134432783608 |
| [30,15] | 0.7420965058236273 | 0.7381159420289855 |
| [15,5] | 0.7420965058236273 | 0.7311194402798601 |

To run the program first compile the DTree and node java files. Then,

java Dtree 15 8 training\_set.csv validation\_set.csv test\_set.csv yes

By using the above line in the command prompt you can observe the decision tree implementation along with the pruning for information gain and variance impurity values in each case.