PROJECT 1

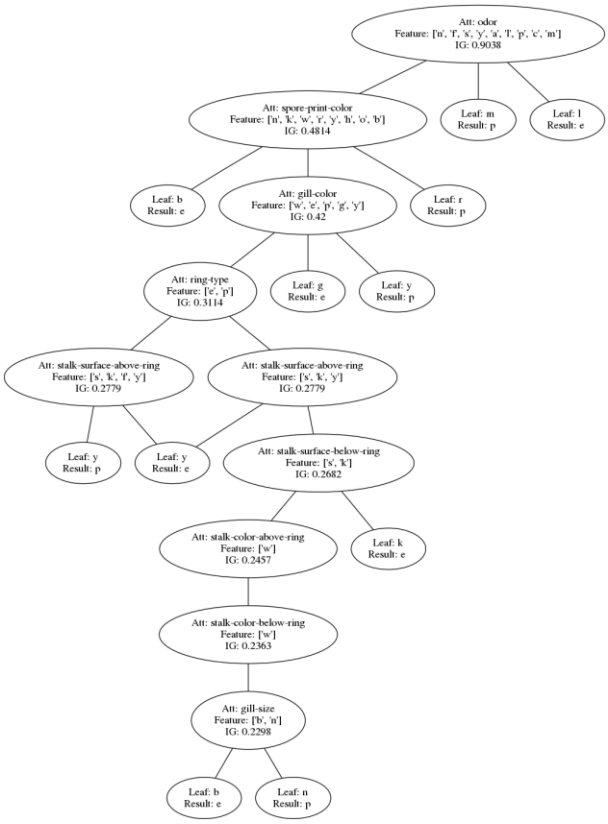
CIS 6930 – Applied Machine Learning Using Python

Program Description

This program reads the mushroom dataset and produces decision tree models using both Quinlan's ID3 and C4.5 decision tree learning algorithms. The program implements a 10-fold cross validation and computes the precision, recall and F1 score for each of the 10 folds. Additionally, the program produces a visual representation of each decision tree in two formats (Networkx and GraphVis) and writes a dot file for later use. In this view, the leaves currently stack on top of each other, removing this makes the graph extremely cluttered. I have considered removing all leaves to make it more compact.

Results

The decision trees produced by my program all have an F1 score of 1.0. This was initially concerning because I felt that it was unlikely to get this result. So, I produced the visualizations that show the decision trees taking into account all the attributes. The trees produced by the C4.5 algorithm appear deeper and have fewer leaves early in the decision process. Below is a sample of two trees on same data split.



Program Output

Below is the console output of the program.

```
====================================
ID3 Algorithm data set number: 1
====================================
--------------------------------------
Precision for decision tree 1:  1.0
Recall for decision tree    1:  1.0
F1 Score for decision tree  1:  1.0
--------------------------------------

====================================
ID3 Algorithm data set number: 2
====================================
--------------------------------------
Precision for decision tree 2:  1.0
Recall for decision tree    2:  1.0
F1 Score for decision tree  2:  1.0
--------------------------------------

====================================
ID3 Algorithm data set number: 3
====================================
--------------------------------------
```

```
Precision for decision tree 3:  1.0
Recall for decision tree    3:  1.0
F1 Score for decision tree  3:  1.0
---------------------------------------

=====================================
ID3 Algorithm data set number: 4
=====================================
---------------------------------------
Precision for decision tree 4:  1.0
Recall for decision tree    4:  1.0
F1 Score for decision tree  4:  1.0
---------------------------------------

=====================================
ID3 Algorithm data set number: 5
=====================================
---------------------------------------
Precision for decision tree 5:  1.0
Recall for decision tree    5:  1.0
F1 Score for decision tree  5:  1.0
---------------------------------------

=====================================
ID3 Algorithm data set number: 6
=====================================
---------------------------------------
Precision for decision tree 6:  1.0
Recall for decision tree    6:  1.0
F1 Score for decision tree  6:  1.0
---------------------------------------

=====================================
ID3 Algorithm data set number: 7
=====================================
---------------------------------------
Precision for decision tree 7:  1.0
Recall for decision tree    7:  1.0
F1 Score for decision tree  7:  1.0
---------------------------------------

=====================================
ID3 Algorithm data set number: 8
=====================================
---------------------------------------
Precision for decision tree 8:  1.0
Recall for decision tree    8:  1.0
F1 Score for decision tree  8:  1.0
---------------------------------------

=====================================
ID3 Algorithm data set number: 9
=====================================
---------------------------------------
Precision for decision tree 9:  1.0
```

```
Recall for decision tree    9:  1.0
F1 Score for decision tree  9:  1.0
---------------------------------------

=====================================
ID3 Algorithm data set number: 10
=====================================
---------------------------------------
Precision for decision tree 10:  1.0
Recall for decision tree    10:  1.0
F1 Score for decision tree  10:  1.0
---------------------------------------

=====================================
C4.5 Algorithm data set number: 1
=====================================
---------------------------------------
Precision for decision tree 1:  1.0
Recall for decision tree    1:  1.0
F1 Score for decision tree  1:  1.0
---------------------------------------

=====================================
C4.5 Algorithm data set number: 2
=====================================
---------------------------------------
Precision for decision tree 2:  1.0
Recall for decision tree    2:  1.0
F1 Score for decision tree  2:  1.0
---------------------------------------

=====================================
C4.5 Algorithm data set number: 3
=====================================
---------------------------------------
Precision for decision tree 3:  1.0
Recall for decision tree    3:  1.0
F1 Score for decision tree  3:  1.0
---------------------------------------

=====================================
C4.5 Algorithm data set number: 4
=====================================
---------------------------------------
Precision for decision tree 4:  1.0
Recall for decision tree    4:  1.0
F1 Score for decision tree  4:  1.0
---------------------------------------

=====================================
C4.5 Algorithm data set number: 5
=====================================
---------------------------------------
Precision for decision tree 5:  1.0
Recall for decision tree    5:  1.0
```

```
F1 Score for decision tree  5:  1.0
--------------------------------------

====================================
C4.5 Algorithm data set number: 6
====================================
--------------------------------------
Precision for decision tree 6:  1.0
Recall for decision tree    6:  1.0
F1 Score for decision tree  6:  1.0
--------------------------------------

====================================
C4.5 Algorithm data set number: 7
====================================
--------------------------------------
Precision for decision tree 7:  1.0
Recall for decision tree    7:  1.0
F1 Score for decision tree  7:  1.0
--------------------------------------

====================================
C4.5 Algorithm data set number: 8
====================================
--------------------------------------
Precision for decision tree 8:  1.0
Recall for decision tree    8:  1.0
F1 Score for decision tree  8:  1.0
--------------------------------------

====================================
C4.5 Algorithm data set number: 9
====================================
--------------------------------------
Precision for decision tree 9:  1.0
Recall for decision tree    9:  1.0
F1 Score for decision tree  9:  1.0
--------------------------------------

====================================
C4.5 Algorithm data set number: 10
====================================
--------------------------------------
Precision for decision tree 10:  1.0
Recall for decision tree    10:  1.0
F1 Score for decision tree  10:  1.0
--------------------------------------

************************************************
Precision for decision tree test ID3: 1.0
Recall for decision tree test ID3:    1.0
F1 Score for decision tree test ID3:  1.0
************************************************
************************************************
Precision for decision tree test C45: 1.0
```

```
Recall for decision tree test C45:    1.0
F1 Score for decision tree test C45:  1.0
**************************************************
```