Report

Dataset-Specific Insights:

- Mushroom Dataset
  a) Attributes that contribute most to classification- Odor, gill-size, spore-print-colour, cap-shape
  b) Class Distribution- 2 classes edible vs poisonous.
  Almost Balanced
  c) Decision pattern- odour -> foul=poisonous
  d) Overfitting Indicators – Tree depth >7; Leaf nodes with few samples.

- Nursery Dataset
  a) Attributes that contribute most to classification – Parents, housing, financial status

  b) Class Distribution – Multiple classes : Recommend, Not recommend, High recommend**. Imbalanced classes**

  c) Decision pattern– Parents-> financial status = proper-> recommend;

  Housing-> critical= not recommend

  d) Overfitting Indicators – Deep branches

- tictactoe Dataset

  a) Attributes that contribute most to classification – Board position which has immediate winning pattern.

  b) Class Distribution – Win/lose/draw; **Imbalanced**

  c) Decision pattern- Row filled X-> win

  d) Overfitting Indicators –Deep tree for small feature.

Comparative Analysis Report:

1. Highest Accuracy Dataset : Among the three datasets , the mushrooms dataset has the maximum Accuracy which is 1(100%).

2. Small dataset achieve high accuracy but if features are predictable and not Overfitting.
   Eg. Tictactoe dataset is small but Overfitting.
   Large dataset provide better performance as its easy for generalization and reliable tree construction but takes too much time.

3. More features- Increase time complexity as well as computation time.
   But if the features are not informative decrease performance.
4. Imbalanced dataset leads to biased tree.

5. Multi-valued is preferred for dataset with complex decision tree. Mainly about features it depends on whether it is informative or not.

- Mushroom Dataset – To determine whether the mushroom is poisonous or not
- Nursery Dataset- Decision making for child admission in nursery
- Tictactoe dataset- Game Outcome prediction

6. Decision trees are easy to interpret the outcome and understand the dataset.
    - Mushroom- Understand features of mushroom to predict whether edible.
    - Nursery – Make decision for the admission of a child.
    - Tictactoe – Pattern analysis of the game to predict its outcome.

7. Improvement of each dataset

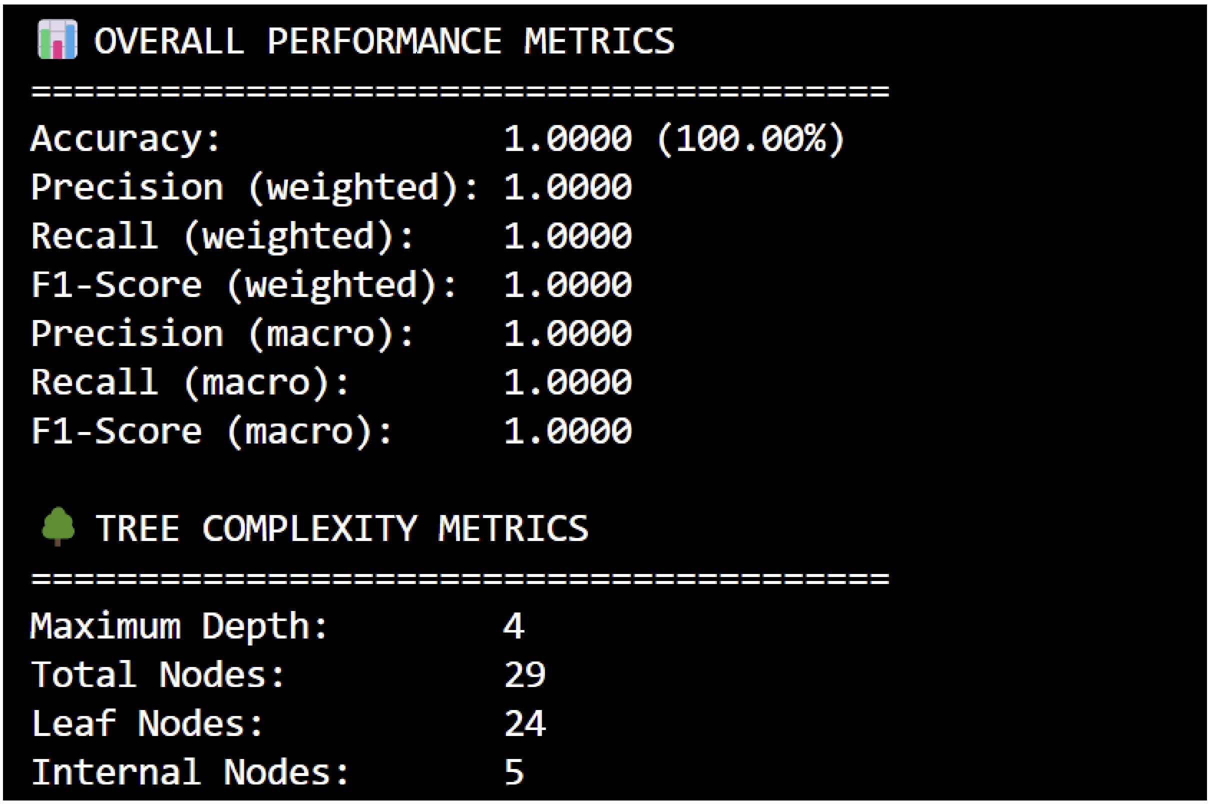
- Mushroom- Dataset perfect not much improvement as accuracy 1.
- Nursery – Rectify the imbalance and remove irrelevant features
- Tictactoe – Cross validation for generalization

Outputs of each dataset for reference for the analysis report given below .

CONCLUSION:

Decision trees has high performance on datasets with informative features and balanced classes. Dataset size, feature type, and class distribution all significantly affect performance and interpretability. Proper preprocessing, pruning, and feature selection can further enhance accuracy and generalization across domains.

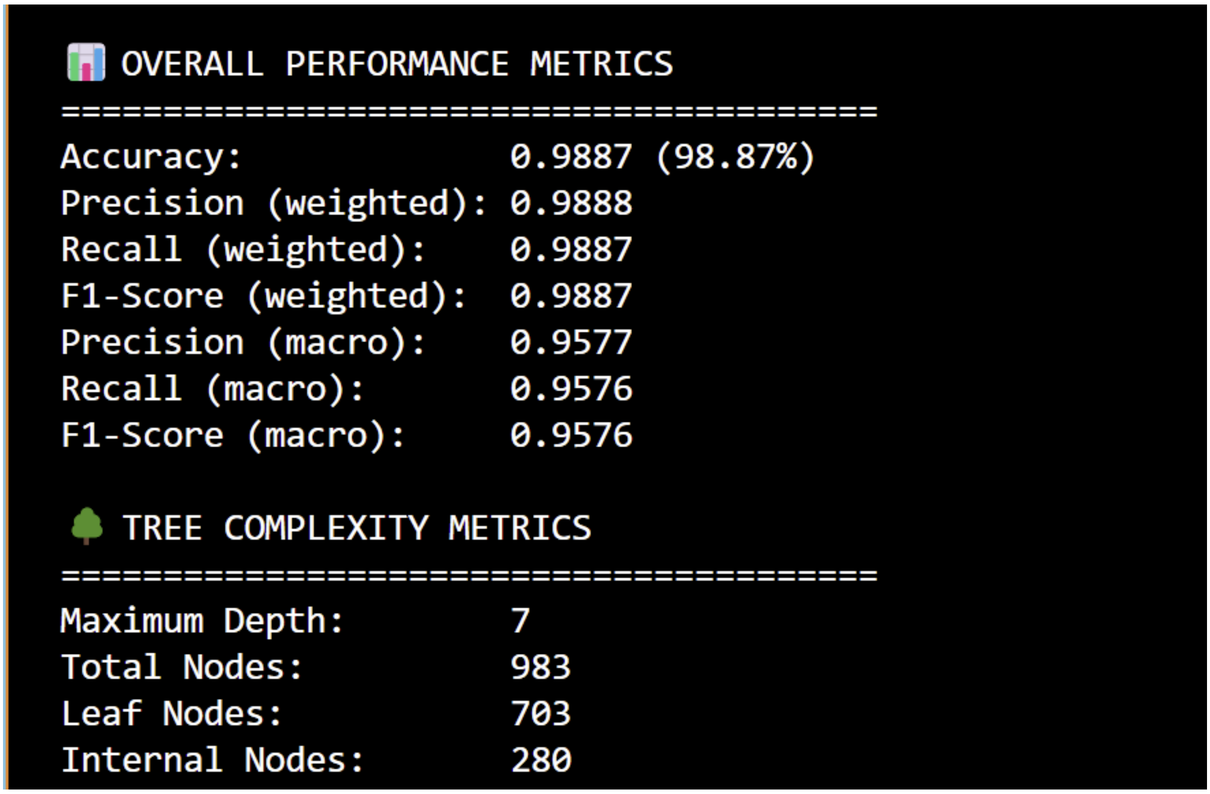
OUTPUT for musrooms.csv



```
📊 OVERALL PERFORMANCE METRICS
========================================
Accuracy:              1.0000 (100.00%)
Precision (weighted): 1.0000
Recall (weighted):     1.0000
F1-Score (weighted):   1.0000
Precision (macro):     1.0000
Recall (macro):        1.0000
F1-Score (macro):      1.0000

🌳 TREE COMPLEXITY METRICS
========================================
Maximum Depth:         4
Total Nodes:           29
Leaf Nodes:            24
Internal Nodes:        5
```

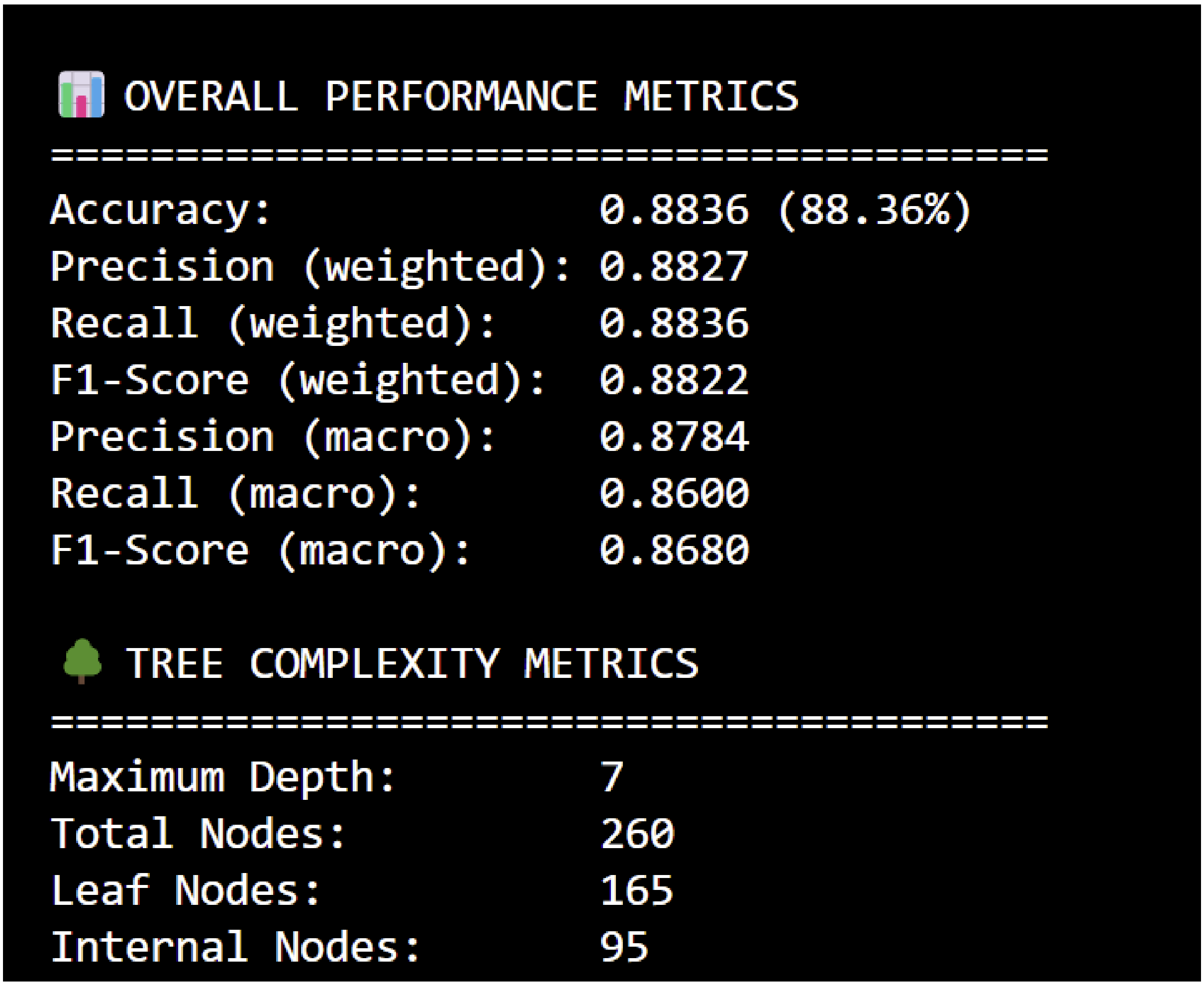
OUTPUT for Nursery.csv



```
📊 OVERALL PERFORMANCE METRICS
========================================
Accuracy:              0.9887 (98.87%)
Precision (weighted): 0.9888
Recall (weighted):     0.9887
F1-Score (weighted):   0.9887
Precision (macro):     0.9577
Recall (macro):        0.9576
F1-Score (macro):      0.9576

🌳 TREE COMPLEXITY METRICS
========================================
Maximum Depth:         7
Total Nodes:           983
Leaf Nodes:            703
Internal Nodes:        280
```

OUTPUT for tictactoe.csv



```
📊 OVERALL PERFORMANCE METRICS
========================================
Accuracy:             0.8836 (88.36%)
Precision (weighted): 0.8827
Recall (weighted):    0.8836
F1-Score (weighted):  0.8822
Precision (macro):    0.8784
Recall (macro):       0.8600
F1-Score (macro):     0.8680

🌳 TREE COMPLEXITY METRICS
========================================
Maximum Depth:        7
Total Nodes:          260
Leaf Nodes:           165
Internal Nodes:       95
```

Name: Nandana Mathew

Srn: PES2UG23CS913

Section: F

Signature : nandana