Lab 3 – Decision Tree Classifier: Multi-Dataset Analysis

Name: Nidhi C R
SRN: PES2UG23CS382
Section: F

Introduction

Decision trees are one of the most widely used supervised learning algorithms because of their simplicity, interpretability, and ability to handle categorical features. In this lab, we evaluated decision tree classifiers on three different datasets — Mushrooms, Nursery, and Tic-Tac-Toe — to compare performance, interpret tree structures, and analyze dataset-specific challenges.

Datasets Overview

- Mushrooms.csv → Predict edible (0) vs poisonous (1) using features like odor, color, gill characteristics.
- Nursery.csv → Predict admission recommendations across 5 classes using family/social attributes.
- TicTacToe.csv → Predict if a given board configuration is a win (1) or not (0).

Performance Metrics

The following table shows accuracy, precision, recall, and F1 scores (weighted and macro):

Dataset	Accuracy	Precision (Weighted)	Precision (Macro)	Recall (Weighted)	Recall (Macro)	F1 (Weighted)	F1 (Macro
Mushrooms	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Nursery	0.987	0.987	0.760	0.987	0.765	0.987	0.763
Tic-Tac-Toe	0.873	0.874	0.859	0.873	0.864	0.873	0.861

Tree Characteristics

Mushrooms tree is shallow and simple. Nursery tree is the largest due to multiple classes and high-cardinality features. Tic-Tac-Toe trees are deep and moderately large, reflecting game complexity.

Dataset-Specific Insights

- Mushrooms: Odor perfectly separates classes.
- Nursery: Parents, finance, and social factors dominate splits; class imbalance is an issue.
- Tic-Tac-Toe: Center and corner positions are most decisive; symmetry leads to redundant splits.

Comparative Analysis

Performance depends on dataset size, discriminative features, and class balance. Larger datasets (Mushrooms, Nursery) yield better generalization than smaller ones (Tic-Tac-Toe).

Practical Applications

- Mushrooms: Food safety classification.
- Nursery: Admission decision support (with caution about bias).
- Tic-Tac-Toe: Teaching rule-based AI and strategy learning.

Overfitting & Improvements

- Mushrooms: Minimal overfitting risk.
- Nursery: Very large tree → overfitting risk.
- Tic-Tac-Toe: Redundant symmetric branches.

Improvements: pruning, ensemble methods, handling class imbalance.

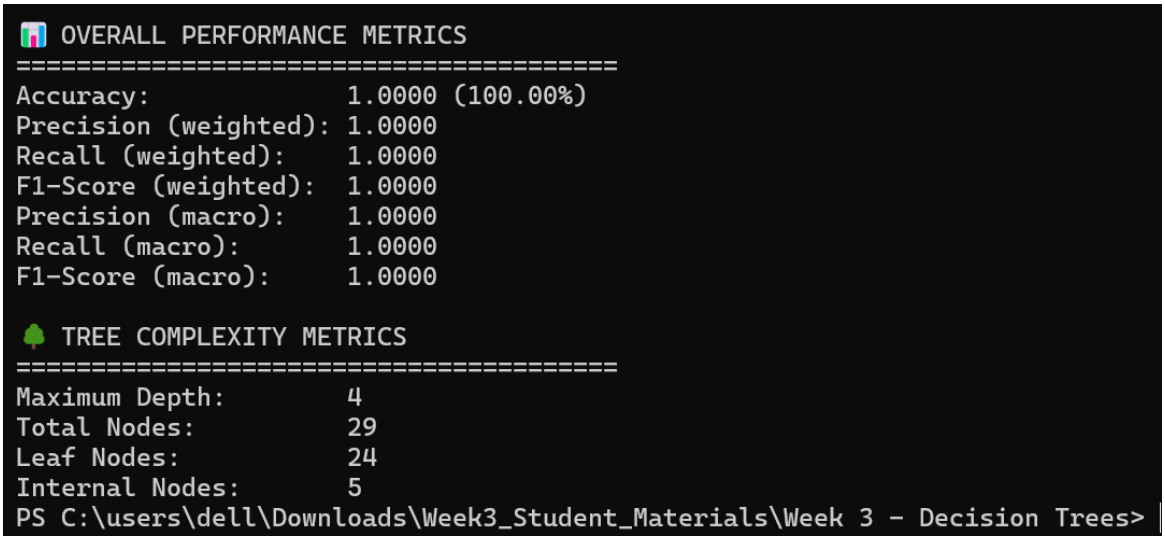
Decision Patterns

- Mushrooms: If odor=foul → poisonous.
- Nursery: If parents=usual and finance=convenient → priority.
- Tic-Tac-Toe: If middle-middle=x AND top-left=x AND bottom-right=x → win.

Conclusion

Decision trees work best when strong discriminative features exist. Dataset quality, size, and balance strongly influence performance. Pruning, ensembles, and class-balancing techniques improve results on complex or imbalanced datasets.

mushrooms.csv



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
PS C:\users\dell\Downloads\Week3_Student_Materials\Week 3 - Decision Trees>
```

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
PS C:\users\dell\Downloads\Week3_Student_Materials\Week 3 - Decision Trees>
```