

Lab3: Decision Tree Report

Name: Muskan Goenka | SRN: PES2UG23CS355 | Section: F

Performance Comparison

Dataset	Accuracy	Precision (weighted)	Precision (macro)	Recall (weighted)	Recall (macro)	F1-Score (weighted)	F1-Score (macro)
Mushrooms	1.0000 (100%)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Nursery	0.9887 (98.87%)	0.9888	0.9577	0.9887	0.9576	0.9887	0.9576
Tic-Tac-Toe	0.8836 (88.36%)	0.8827	0.8784	0.8836	0.8600	0.8822	0.8680

Observations:

- The **Mushrooms dataset** achieved a perfect score across all metrics, indicating that its features (e.g., odor) provide a very strong separation between edible and poisonous mushrooms.
- **Nursery** also performed very well, but the macro precision/recall values are slightly lower, which shows some imbalance in the dataset.
- **Tic-Tac-Toe** had the weakest results, with ~88% accuracy, which makes sense because board game states are more complicated to classify.

Tree Characteristics Analysis

Dataset	Max Depth	Total Nodes	Leaf Nodes	Internal Nodes
Mushrooms	4	29	24	5
Nursery	7	983	703	280
Tic-Tac-Toe	7	260	165	95

Insights:

- The **Mushrooms tree** is very shallow and compact. A few strong features allow for fast and accurate classification.
- The **Nursery tree** is the largest, reflecting the dataset's many possible outcomes and higher feature diversity.
- Tic-Tac-Toe sits in between, but still shows longer paths since game boards need more rules to be covered.

Dataset-Specific Insights

Mushrooms:

- *Feature Importance:* Odor is the most decisive feature.
- *Class Distribution:* Balanced dataset.
- *Decision Patterns:* Simple rules (e.g., foul odor → poisonous).
- *Overfitting:* No signs, since accuracy is perfect with a small tree.

Nursery:

- *Feature Importance:* Finance, social, and health strongly influence decisions.
- *Class Distribution:* Imbalanced, leading to weaker macro scores.
- *Decision Patterns:* Rules like "poor finance → not recommended" dominate.

- *Overfitting*: Tree is large but generalizes well (high accuracy).

Tic-Tac-Toe:

- *Feature Importance*: Center and corner positions on the board matter most.
 - *Class Distribution*: More negative outcomes than positive ones, leading to imbalance.
 - *Decision Patterns*: Long paths, reflecting complex game scenarios.
 - *Overfitting*: Tree size suggests risk of overfitting, pruning could help.
-

Comparative Analysis Report

Algorithm Performance:

- Best performance on **Mushrooms (100%)**, worst on Tic-Tac-Toe (88%).
- Larger datasets (Nursery) increase complexity but maintain high performance.
- Feature quality is more critical than quantity—Mushrooms worked well because of strong discriminatory features.

Impact of Data Characteristics:

- Class imbalance reduces macro precision/recall, especially in Nursery.
- Multi-valued categorical features improve classification when they align well with the target.

Real-World Relevance:

- **Mushrooms**: Suitable for binary classification problems (e.g., safe vs unsafe, fraud detection).
- **Nursery**: Useful for admission/recommendation systems involving multiple factors.
- **Tic-Tac-Toe**: Demonstrates game decision-making and rule-based AI.

Improvements:

- Mushrooms: Already optimal, no change needed.
- Nursery: Address class imbalance with resampling or weighting.
- Tic-Tac-Toe: Could use pruning or more advanced algorithms (Random Forest, SVM) to improve performance.