

Week 3 Report - ID3 Decision Tree Analysis

1. Performance Comparison

Dataset	Accuracy	Precision (weighted)	Recall (weighted)	F1-score (weighted)
Mushroom	100%	1.000	1.000	1.000
TicTacToe	87.3%	0.874	0.873	0.873
Nursery	98.67%	0.988	0.987	0.987

2. Tree Characteristics Analysis

Dataset	Max Depth	Total Nodes	Leaf / Internal Nodes
Mushroom	4	29	24 / 5
TicTacToe	7	281	180 / 101
Nursery	7	952	680 / 272

Most important features observed:

- Mushroom: Odor strongly separates edible vs poisonous.
- TicTacToe: Middle and corner squares are key early splits.
- Nursery: Parents, children, and social attributes are most influential.

Tree complexity grows with dataset size and number of classes.

3. Dataset-Specific Insights

Mushroom: Very balanced dataset, clear separation by odor. No overfitting seen.

TicTacToe: Some ambiguous states cause misclassification. Slight overfitting due to large tree relative to data size.

Nursery: High accuracy overall, but class imbalance lowers macro precision. Large tree reflects dataset size and multiple classes.

4. Comparative Analysis Report

a) Algorithm Performance:

- Mushroom achieved highest accuracy (100%) due to highly predictive features like odor.
 - Dataset size affects performance: Nursery (large dataset) gave very high accuracy, TicTacToe (small dataset) had lower performance.
 - More features increase tree complexity but also help achieve higher accuracy when predictive (e.g., Nursery).
-

b) Data Characteristics Impact:

- Class imbalance reduces macro metrics (seen in Nursery).
 - Binary features (Mushroom, TicTacToe) are simpler but sometimes less expressive. Multi-valued features (Nursery) add richness but increase tree size.
-

c) Practical Applications:

- Mushroom dataset: Useful for food safety applications.
- TicTacToe: Illustrates game-state prediction and AI decision-making.
- Nursery: Relevant to education systems and admission recommendations.

Interpretability is an advantage of trees: we can clearly see which features drive decisions.

d) Improvements:

- Apply pruning to reduce overfitting (TicTacToe, Nursery).
- Use class balancing or cost-sensitive learning for Nursery dataset.
- Compare with advanced algorithms (C4.5, CART) for robustness.