ID3 Decision Tree Analysis Report

Name: MOHIT KUMAR

SRN: PES2UG23CS350

1. Performance Comparison

Dataset	Accuracy	Precision	Recall	F1-Score
Mushroom	1.0000	1.0000	1.0000	1.0000
Tic-Tac-Toe	0.8730	0.8730	0.8730	0.8730
Nursery	0.9867	0.9867	0.9867	0.9867

2. Tree Characteristics Analysis

Dataset	Tree Depth	Number of Nodes	Most Important Features	Tree Complexity
Mushroom	Shallow (2-3 levels)	Low	Odor, spore-print-color	Simple, minimal splits
Tic-Tac-Toe	Medium (4-6 levels)	Medium	Center position, corner positions	Moderate complexity
Nursery	Deep (7+ levels)	High	Finance, social, health	High complexity, many branches

3. Dataset-Specific Insights

Mushroom Dataset

- **Feature Importance:** Odor attribute dominates classification decisions

- **Class Distribution:** Balanced between edible and poisonous classes

- **Decision Patterns:** Single attribute often determines classification

- **Overfitting Indicators:** Minimal overfitting due to strong feature predictability

Tic-Tac-Toe Dataset

- **Feature Importance:** Board positions, particularly center and corners

- **Class Distribution:** Balanced between positive and negative outcomes

- **Decision Patterns:** Logical game-winning combinations

- **Overfitting Indicators:** Potential memorization of specific board configurations

Nursery Dataset

- **Feature Importance:** Finance, social status, and health factors

- **Class Distribution:** Imbalanced, with "not_recom" class dominant

- **Decision Patterns:** Complex multi-attribute decision paths

- **Overfitting Indicators:** Deep tree structure suggests possible overfitting

4. Comparative Analysis Report

a) Algorithm Performance

Which dataset achieved the highest accuracy and why?

Mushroom dataset achieved the highest accuracy (100%) due to highly discriminative features, particularly the odor attribute which strongly correlates with mushroom toxicity.

How does dataset size affect performance?

Larger datasets like Nursery create deeper trees with more nodes, potentially leading to overfitting. Smaller, well-structured datasets like Mushroom produce simpler, more generalizable models.

What role does the number of features play?

More features increase tree complexity. Mushroom (22 features) achieved perfect accuracy with simple splits, while Nursery (8 features) required deeper branching due to multi-valued categorical attributes.

b) Data Characteristics Impact

How does class imbalance affect tree construction?

Class imbalance in Nursery dataset causes the tree to favor majority classes, resulting in biased predictions toward "not_recom" outcomes and reduced performance on minority classes.

Which types of features work better?

Binary features create simpler decision boundaries and smaller trees. Multi-valued categorical features in Nursery dataset increase tree complexity and branching factor, making interpretation difficult.

c) Practical Applications

Real-world scenarios for each dataset type:

- **Mushroom:** Botanical classification, food safety applications

- **Tic-Tac-Toe:** Game AI development, strategic decision modeling

- **Nursery:** Educational admission systems, resource allocation decisions

Interpretability advantages for each domain:

- **Mushroom:** Clear safety rules easily understood by field workers

- **Tic-Tac-Toe:** Transparent game strategy logic

- **Nursery:** Explainable admission criteria for fairness assessment

How would you improve performance for each dataset?

- **Mushroom:** Apply post-pruning to reduce tree size while maintaining accuracy

- **Tic-Tac-Toe:** Implement early stopping to prevent memorization of rare board states

- **Nursery:** Use class weighting to address imbalance and limit maximum tree depth to reduce overfitting