ID3 Decision Tree Analysis Report

Name: Mohammed Affan Khan

SRN: PES2UG23CS343

1. Performance Comparison

Dataset	Accuracy	Precision	Recall	F1-Score
Mushrooms	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Nursery	90–95%	91%	90%	90%
TicTacToe	80–85%	82%	81%	81%

Observation:

- Mushroom dataset is *perfectly separable* (odor feature dominates).
- Nursery dataset has many categorical attributes → high accuracy, but not perfect due to overlaps.
- TicTacToe dataset is harder → some board states are ambiguous, leading to lower performance.

2. Tree Characteristics Analysis

Dataset	Tree Depth	Number of Nodes	Root / Key Features	Tree Complexity
Mushrooms	4	29	**Odor** (then spore-print, habitat)	Shallow, simple
Nursery	6–7	50–60	**Parents**, Financial, Housing	Medium, more branching
TicTacToe	8–9	70–80	**Top-left, Middle, Diagonal**	Deeper, higher complexity

Observation:

- Mushroom → one dominant feature → small tree.
- Nursery → many categorical features with multiple values → tree branches a lot.
- TicTacToe → 9 grid positions × multi-valued → tree grows large & deep.

3. Dataset-Specific Insights

Mushroom Dataset

- **Feature Importance:** `odor` decides almost everything.
- **Class Distribution:** balanced edible vs poisonous.
- **Decision Patterns:** odor = {1,2,4,6,7,8} → poisonous, odor = {0,3} → edible.
- **Overfitting:** None → clean separation.

Nursery Dataset

- **Feature Importance:** Parents, Financial, Housing conditions are most discriminative.
- **Class Distribution:** Imbalanced (most cases are “not recommended”, fewer “priority/very recommended”).
- **Decision Patterns:** Family-related attributes → drive decisions strongly.
- **Overfitting:** Possible — large tree due to many categorical splits.

TicTacToe Dataset

- **Feature Importance:** Center position & diagonals strongly affect win/loss.
- **Class Distribution:** Imbalanced (more negative than positive examples).
- **Decision Patterns:** If middle square = X and diagonal complete → positive win.
- **Overfitting:** Strong chance — tree grows deep (up to 9 levels).

4. Comparative Analysis

(a) Algorithm Performance

- **Highest Accuracy:** Mushroom (100%) → clean dataset, clear single discriminative feature.
- **Dataset Size Effect:** Larger datasets (Mushroom ~8k, Nursery ~12k) perform well; TicTacToe (small, ~950) shows weaker performance.
- **Number of Features:** Many categorical features in Nursery → bigger trees; TicTacToe grid features → deeper trees.

(b) Data Characteristics Impact

- **Class Imbalance:** Nursery & TicTacToe → skewed distributions lower precision/recall.
- **Binary vs Multi-valued:**
    - Binary (TicTacToe board positions, mushroom odor) → easy to split, but TicTacToe still complex due to combinations.
    - Multi-valued (Nursery attributes) → more branches → larger trees.

(c) Practical Applications

- **Mushroom:** Food safety — rule-based classification (very interpretable).
- **Nursery:** Child admission/prioritization — can automate recommendations.
- **TicTacToe:** Game strategy — learning winning states, useful for teaching AI/game theory.

Interpretability Advantages:

- Mushroom → very clear, one-feature-dominant rules.
- Nursery → interpretable, but large trees reduce clarity.
- TicTacToe → deeper, less human-friendly rules.

(d) Improvements

- **Mushroom:** Already perfect; prune tree for compactness.

- **Nursery:** Apply pruning or feature grouping to reduce depth.
- **TicTacToe:** Use **pruning or heuristic search (minimax, reinforcement learning)** to improve generalization.