Decision Tree Performance Analysis

1. Performance Comparison

- Mushroom → Accuracy - 100%, Precision/Recall/F1 = 1.0
- TicTacToe → Accuracy – 88.36%, Precision ≈ 0.883, Recall ≈ 0.884, F1 ≈ 0.882
- Nursery → Accuracy – 98.87%, Precision ≈ 0.989, Recall ≈ 0.989, F1 ≈ 0.989

The accuracy ranks as follows: Mushroom > Nursery > Tic-Tac-Toe.

2. Tree Characteristics Analysis

- **Tree Depth**:
  Mushroom = 4
  Tic-Tac-Toe = 7
  Nursery = 7
- **Number of Nodes**:
  Mushroom = 29 (very small tree)
  Tic-Tac-Toe = 260 (medium tree)
  Nursery = 983 nodes (very large tree)
- **Most Important Features**:
  Mushroom = odor (clear root split)
  Tic-Tac-Toe = center square, top-right, top-left
  Nursery = children, form, housing
- **Tree Complexity**:
  Mushroom → Simple, easy to interpret.
  Tic-Tac-Toe → Moderate complexity.
  Nursery → Very complex due to many features and class categories.

3. Dataset-Specific Insights

Mushroom

- **Feature Importance**: Odor dominates.
- **Class Distribution**: Balanced edible vs poisonous.
- **Decision Patterns**: Simple one-step rules.
- **Overfitting**: None (tree generalizes perfectly).

Tic-Tac-Toe

- **Feature Importance**: Center and corners of board.
- **Class Distribution**: Balanced (win vs non-win).
- **Decision Patterns**: Tree learns "three-in-a-row" rules.
- **Overfitting**: Slight risk due to board states, but performance still high.

Nursery

- **Feature Importance**: Parents, finance, housing features matter most.
- **Class Distribution**: Imbalanced (some categories appear more often).
- **Decision Patterns**: Multi-step rules, many layers of decisions.
- **Overfitting**: High risk (big tree: 983 nodes, depth 7).

4. Comparative Analysis Report

a) Algorithm Performance:
   a. The **mushroom** dataset has highest accuracy because the target variable is strongly linked to a single feature (odor). Nursery dataset also has high accuracy but has complex and imbalanced classes and tictactoe dataset has patterns which were complex.
   b. Decision tree can learn more from large datasets which improves accuracy, but the tree also becomes large and complex. But if the smaller datasets have clear pattern, their accuracy will also be good.
   c. Fewer features results in simple, accurate tree. Many features can cause the tree to be large, risking of overfitting.

b) Data Characteristics Impact:
   - **Class Imbalance**: In Nursery, some classes dominate more than others. This causes the tree to focus on majority classes and may reduce accuracy for minority classes. Mushroom and Tic-Tac-Toe are more balanced, so trees perform better without bias.
   - **Binary vs Multi-Valued Features**: Binary features are easier for trees (like Tic-Tac-Toe moves or Mushroom edible/poisonous).
   - **Multi-valued features** (housing, finance, parents in Nursery) make the tree grow larger and harder to interpret.

c) Practical Applications

   - **Real-World Relevance**: Mushroom dataset → Food safety (detecting poisonous vs edible mushrooms). Tic-Tac-Toe dataset → Game strategy analysis, AI game bots. Nursery dataset → Decision support for childcare/nursery admission systems.
   - **Interpretability Advantages**: Mushroom → Very interpretable, a single feature (odor) decides. Tic-Tac-Toe → Shows game-winning strategies (center and corner importance). Nursery → Explains how financial, housing, and family conditions affect decisions, though the tree is very complex.
   - **Improving Performance**: Mushroom: Already perfect, no changes needed. Tic-Tac-Toe: Could improve by using pruning or ensemble methods (Random Forest, boosting). Nursery**:** Reduce overfitting by pruning, feature selection, or combining with ensemble models.