ANALYSIS REPORT

1. **Performance Comparison**

   Nursery Dataset:

   - Large dataset, highly multi-class, with strong relationships between family/finance attributes and target.
   - Accuracy: Approx 90% achievable with trees, but may drop slightly due to many class labels.
   - Precision/Recall/F1: Lower than mushrooms because of **imbalanced class distribution**.

   Mushrooms Dataset:

   - Clean, binary classification (edible vs poisonous).
   - Certain features like **odor** split the dataset almost perfectly.
   - Accuracy: Close to 100%.
   - Precision/Recall/F1: Near-perfect, since splits strongly correlate with the class.

   Tic-Tac-Toe Dataset:

   - Medium-sized, binary target.
   - Accuracy: Around 85–90%, since patterns are deterministic but representation is discrete.
   - Precision/Recall/F1: Balanced, but sometimes misclassifies unusual board states.

   Ranking by performance, Mushrooms greater than Nursery. Nursery and Tic Tac Toe are almost the same.

2. **Tree Characteristics Analysis**

   Tree Depth:

   - Nursery: Deep. Depth around 10+.
   - Mushrooms: Shallow, since features like odor separate almost immediately. Depth approximately 4–6.
   - Tic-Tac-Toe: Medium depth, depending on win conditions.

Number of Nodes:

- Nursery: High, due to many attributes and values.
- Mushrooms: Low, since 1–2 key features decide most splits.
- Tic-Tac-Toe: Medium, correlating with possible winning states.

Most Important Features:

- Nursery: finance, family, and social factors are usually top splits.
- Mushrooms: odor, gill size, spore print color.
- Tic-Tac-Toe: Central square, followed by corners.

Tree Complexity:

- Nursery: High complexity.
- Mushrooms: Low complexity.
- Tic-Tac-Toe: Medium complexity.

3. Dataset-Specific Insights

Nursery Dataset:

- **Feature Importance**: Financial stability and parental preference dominate splits.
- **Class Distribution**: Imbalanced (some decisions like "recommend" are rare).
- **Decision Patterns**: "Good financial + supportive family → priority admission."
- **Overfitting**: Risk is higher due to many features/values; pruning is necessary.

Mushrooms Dataset:

- **Feature Importance**: Odor is the single strongest indicator (almost perfect split).
- **Class Distribution**: Balanced (edible vs poisonous).
- **Decision Patterns**: "Foul odor means poisonous" emerges early.
- **Overfitting**: Minimal, since dataset is clean and separable.

Tic-Tac-Toe Dataset:

- **Feature Importance**: Center cell is most predictive.
- **Class Distribution**: Slight imbalance depending on X/O placements.
- **Decision Patterns**: "X in center and X in corner implies positive outcome."

- **Overfitting**: Possible if tree memorizes board positions rather than general patterns.

4. Comparative Analysis Report

a) Algorithm Performance:

- **Highest Accuracy**: Mushrooms dataset, because of strong attribute-class correlation.
- **Dataset Size Effect**: Nursery implies longer training time, deeper trees. Mushrooms imply efficient, clean splits. Tic-Tac-Toe implies manageable.
- **Role of Features**: More features (nursery) increase tree depth and complexity, while fewer decisive features (mushrooms) yield simpler, more accurate trees.

b) Data Characteristics Impact:

- **Class Imbalance**: Nursery suffers from imbalance, some minority classes harder to predict. Mushrooms balanced, strong results. Tic-Tac-Toe moderately imbalanced.
- **Feature Types**: Binary features (mushrooms, tic-tac-toe) lead to cleaner splits. Multi-valued categorical features (nursery), more complex splits, higher overfitting risk.

c) Practical Applications:

- **Nursery**: Admission decision support systems. Interpretable, but may need pruning for real-world use.
- **Mushrooms**: Food safety classification. High accuracy, easily interpretable rules.
- **Tic-Tac-Toe**: Game AI explanation. Good interpretability for explaining strategies.

Interpretability Advantages:

- Nursery: Explains how financial/social factors affect admission.
- Mushrooms: Clear, human-readable rules.
- Tic-Tac-Toe: Explains why certain moves are critical.

Improvements:

- Nursery: Use tree pruning and possibly convert to Random Forest to reduce overfitting.
- Mushrooms: Already near-perfect, little improvement needed. Could compress rules.

- Tic-Tac-Toe: Use feature engineering to reduce depth.

Observation:

- Mushrooms dataset has the highest accuracy, simplest tree.
- Nursery dataset has the largest, complex tree, risk of overfitting.
- Tic-Tac-Toe dataset has medium complexity, interpretable decision paths.