Analysis Report

1.

Accuracy:

- Mushroom dataset achieved perfect accuracy (100%), indicating highly separable features.
- Nursery followed with 98.67%, showing strong performance despite multi-class complexity.
- TicTacToe had 87.30%, which is solid for a binary classification task with spatial dependencies.

Precision:

- Mushroom again leads with perfect precision, meaning no false positives.
- Nursery's weighted precision is high (0.9876), but macro precision (0.7604) reveals class imbalance.
- TicTacToe maintains balanced precision across classes.

Recall:

- Mushroom's recall is perfect, indicating no false negatives.
- Nursery's macro recall (0.7654) is lower than weighted, again due to minority class performance.
- TicTacToe shows consistent recall across both metrics.

F1-Score:

- Mushroom's F1-score is 1.0000, confirming ideal balance between precision and recall.
- Nursery's weighted F1 is excellent (0.9872), but macro F1 (0.7628) suggests some classes are harder to predict.
- TicTacToe's F1-score (0.8734 weighted) reflects balanced performance for a moderately complex binary task.

2.

- Mushroom tree is shallow and compact, yet perfectly accurate—indicating strong feature separability.
- Nursery tree is large and deep, reflecting the complexity of multi-class classification.
- TicTacToe tree is moderately complex, likely due to the spatial nature of board positions.

3.

TicTacToe

- **Feature Importance:** Board positions (e.g., center square) likely more influential.
- **Class Distribution:** Binary (positive/negative), likely balanced.
- **Decision Patterns:** Winning combinations (rows, columns, diagonals) drive splits.
- **Overfitting Indicators:** Moderate depth and node count suggest generalization is reasonable.

Nursery

- **Feature Importance:** Attributes like "has_nurs," "form," and "finance" likely dominate.
- **Class Distribution:** Five classes, with imbalance evident from macro vs weighted scores.
- **Decision Patterns:** Complex paths due to multi-valued categorical features.
- **Overfitting Indicators:** High node count and depth may hint at overfitting, especially with rare classes.

Mushroom

- **Feature Importance:** "odor" is the dominant root feature with high information gain.
- **Class Distribution:** Binary (edible/poisonous), likely well-separated.
- **Decision Patterns: Simple, direct paths from odor to class.**
- **Overfitting Indicators: None—perfect accuracy with minimal tree depth**

4.

a) Algorithm Performance

- **Highest Accuracy**: Mushroom dataset, due to strong feature separability.
- **Dataset Size Impact**: Larger datasets (Nursery) allow better generalization but increase tree complexity.
- **Feature Count Role**: Mushroom has many features, but only a few are truly informative—highlighting the importance of feature selection.

b) Data Characteristics Impact

- **Class Imbalance**: Nursery's lower macro scores show how imbalance affects recall and precision for minority classes.
- **Feature Types**: Binary features (Mushroom, TicTacToe) yield simpler trees; multi-valued features (Nursery) increase complexity and depth.

c) Practical Applications

TicTacToe Dataset

- **Real-World Relevance**:
  The TicTacToe dataset is ideal for modeling strategic decision-making in turn-based games. It can be used to train AI agents for board games, simulate player behavior, or analyze winning patterns in competitive environments.
- **Interpretability Advantages**:
  Decision trees offer clear visual paths that mirror human reasoning in gameplay. Each split corresponds to a move or board position, making it easy to trace how a game outcome is predicted. This transparency is valuable for educational tools and AI explainability in gaming.

Nursery Dataset

- **Real-World Relevance**:
  This dataset simulates a recommendation system for child care admissions based on multiple socio-economic and familial factors. It can be adapted for decision support in education, social services, or resource allocation systems.
- **Interpretability Advantages**:
  Despite its complexity, decision trees allow stakeholders to understand how different attributes (e.g., parental status, housing, health) influence recommendations. This is crucial in sensitive domains where fairness, accountability, and transparency are essential.

Mushroom Dataset

- **Real-World Relevance**:
  The Mushroom dataset is directly applicable to food safety, toxicology, and environmental biology. It helps classify mushrooms as edible or poisonous based on observable traits, which is critical for public health and ecological research.
- **Interpretability Advantages**:
  The tree structure is exceptionally simple and accurate, with dominant features like "odor" leading to immediate classification. This makes it highly interpretable for non-technical users such as field biologists or for use in mobile identification apps.
  Would you like help drafting the final conclusion or formatting this into a full report layout?

d) Performance Improvement Suggestions

- **TicTacToe**: Consider ensemble methods (e.g., Random Forest) to capture complex spatial patterns.
- **Nursery**: Apply feature selection or pruning to reduce overfitting and improve macro scores.
- **Mushroom**: Already optimal—future work could test robustness under noisy or missing data conditions.