**1. Performance Comparison**

The Mushroom dataset gave the best results with 100% accuracy. This means the model predicted every mushroom correctly because some features like odor and gill size strongly determine if a mushroom is edible or poisonous.

The Tic-Tac-Toe dataset had the lowest accuracy at about 87%. This is because game states are complex, and the decision tree needs to check many positions before predicting the outcome.

The Nursery dataset also performed very well, with about 98.7% accuracy. The slight drop compared to Mushroom is because the dataset has five classes instead of two, making it harder for the tree to classify.

**2. Tree Characteristics**

The Mushroom dataset created the smallest tree with a depth of 4. This shows that only a few features were enough to make correct predictions.

The Tic-Tac-Toe dataset created a much bigger tree with a depth of 7 and many nodes because it needs to consider several board positions.

The Nursery dataset created the largest tree, also with a depth of 7 but with hundreds of nodes. This happened because it has many attributes with multiple values and more classes to predict.

Overall, the tree size grows when the dataset has many features or multiple class labels.

**3. Insights for Each Dataset**

For the Mushroom dataset, features like odor and gill size were the most important. The classes were balanced, and the tree was small and very accurate with no signs of overfitting.

For Tic-Tac-Toe, the center and corner squares were the most important features. The tree had to check many features to classify correctly, which made it deep and complex. This could cause overfitting because the tree tries to memorize game patterns.

For Nursery, attributes like parents, has\_nurs, and housing appeared early in the tree. The dataset is imbalanced since most cases belong to “not\_recom.” The tree was very large, which makes it harder to interpret and could lead to overfitting.

**4.a) Overall Analysis**

**Which dataset performed best and why?**  
Mushroom performed best because it has clear patterns in the data.

**Does dataset size affect performance?**  
Yes, a bigger dataset like Nursery helps accuracy, but it also increases tree complexity.

**What about the number of features?**  
More features do not always mean better results. Tic-Tac-Toe has fewer features but still created a large tree because the relationships between features are complex.

**b)Data Characteristics Impact:**

* How does class imbalance affect tree construction?

Class imbalance affects accuracy for minority classes. For example, in Nursery, the smaller classes were harder to predict.

* Which types of features (binary vs multi-valued) work better?

Binary features usually work better because they create smaller, simpler trees. Multi-valued features often lead to larger trees with more branches, as seen in the Nursery dataset.

**c)Practical Applications:**

• For which real-world scenarios is each dataset type most relevant?

* Mushroom: Useful in food safety to identify poisonous mushrooms.
* Tic-Tac-Toe: Can be used for building simple game AI.
* Nursery: Helpful in making school admission recommendations.

•What are the interpretability advantages for each domain?

Decision trees are easy to interpret, which makes them useful in fields like education and healthcare where transparency is important.

• How would you improve performance of each dataset?

* Mushroom: No improvement needed since accuracy is perfect.
* Tic-Tac-Toe: Apply pruning or limit the depth to avoid overfitting.
* Nursery: Use pruning and handle class imbalance with weighting to make the tree smaller and more balanced.