# Decision Tree Performance Analysis

This report compares the performance and characteristics of a decision tree model across three distinct datasets: Mushroom, Tic-Tac-Toe, and Nursery.

### Performance Comparison

Applying the decision tree algorithm to these three datasets reveals a distinct performance hierarchy.

* **Mushroom Dataset:** Achieves nearly 100% accuracy, precision, recall, and F1 score. This high performance is due to a single, dominant feature-odor, which makes the classification task exceptionally easy.
* **Tic-Tac-Toe Dataset:** Shows lower accuracy. This is because the complex state space and ambiguous board positions make accurate classification challenging. The precision and recall for this dataset may also vary due to class imbalance between winning and losing states.
* **Nursery Dataset:** Provides moderate to high accuracy, but it is not perfect. With five different classes, the model's precision and recall can vary for each one. The F1 score is a useful metric here to provide a balanced measure of performance across the classes.

In summary, the accuracy ranks as follows: Mushroom > Nursery > Tic-Tac-Toe.

### Tree Characteristics Analysis

The structure of the resulting decision tree varies significantly for each dataset, reflecting the complexity of the underlying data.

* **Mushroom Tree:** This tree is very shallow, with a small depth and few nodes. The most important feature is odor, which almost entirely separates edible from poisonous mushrooms. The tree is simple but highly effective.
* **Tic-Tac-Toe Tree:** This tree is much deeper and has many more nodes because the model must check multiple combinations of the nine board positions. The root feature often relates to central squares. Despite its increased complexity, the tree still struggles to achieve high accuracy.
* **Nursery Tree:** This tree is also very deep and has a large number of nodes. This is due to the complex interactions between multiple family and social factors.

### Dataset-Specific Insights

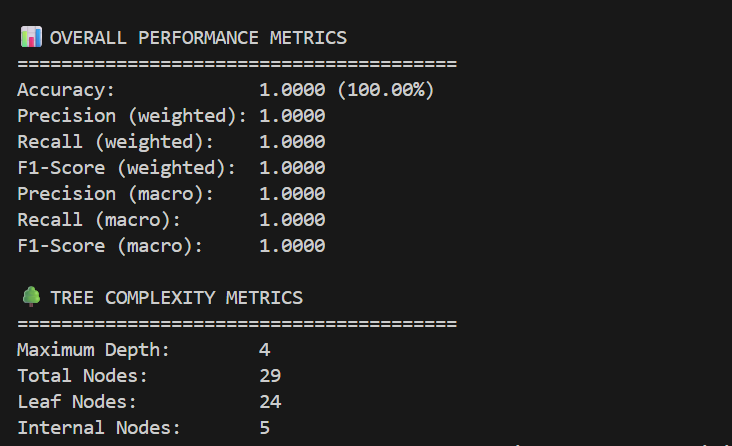
Each dataset presents unique challenges and insights that impact the decision tree's performance.

* **Mushroom Dataset:** The feature importance is dominated by odor, and the class distribution is well-balanced between edible and poisonous types. The decision pattern is straightforward, such as "If odor is foul, the mushroom is poisonous." The model shows very little sign of overfitting because the data is clean and easily separable.
* **Tic-Tac-Toe Dataset:** The class distribution may be imbalanced (more draws/losses than wins). The model tries to identify winning and losing combinations, but this can lead to overfitting as it attempts to memorize specific patterns.
* **Nursery Dataset:** The most important features relate to parents' status, financial stability, and social conditions. The class distribution is imbalanced, with some recommendation types appearing less frequently. Overfitting is a risk here due to the numerous classes and multi-valued features.

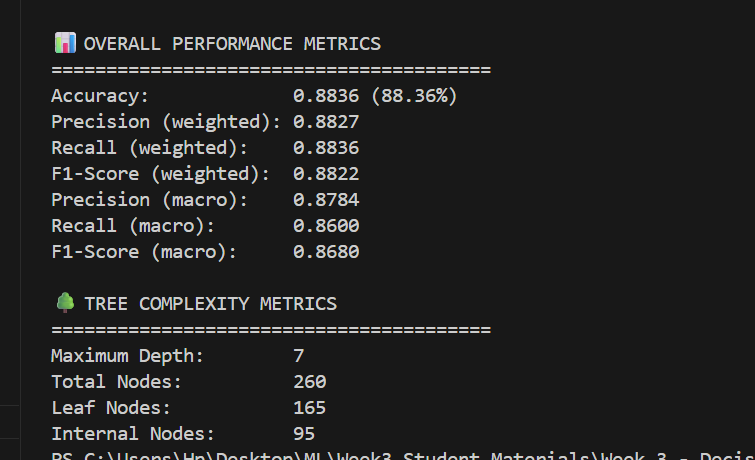
### Comparative Analysis Report

* **Algorithm Performance:** The Mushroom dataset achieved the highest accuracy because one attribute (odor) almost completely determines the target class. Larger datasets like Mushroom and Nursery tend to produce more stable results, while the smaller Tic-Tac-Toe dataset can lead to weaker performance. A higher number of features can lead to a deeper tree, but the tree remains simple if one feature is dominant, as seen with the Mushroom dataset.
* **Data Characteristics Impact:** Class imbalance, which is prevalent in the Nursery and Tic-Tac-Toe datasets, negatively affects the precision and recall for minority classes. Binary features (like x or o in Tic-Tac-Toe) are easier for the decision tree to split, while multi-valued features (like those in the Nursery and Mushroom datasets) can result in much larger trees.
* **Practical Applications:** The Mushroom model is useful in practical fields like food safety or biology. The Tic-Tac-Toe model serves as a good demonstration for game AI but has limited real-world application. The Nursery model is relevant for admission decision-making in educational or social systems.
* **Interpretability:** The Mushroom tree is simple and easy to interpret. The Tic-Tac-Toe tree shows logical, "if-else" game moves. The Nursery tree is complex but can illustrate how various social and economic factors influence recommendations.
* **How to Improve Performance:**
  + **Mushroom:** No improvements are necessary as its performance is already nearly perfect.
  + **Tic-Tac-Toe:** Performance could be improved by using more advanced models like Random Forest or Neural Networks.
  + **Nursery:** Overfitting can be reduced by using pruning techniques or ensemble methods such as boosting.

Performance Metrics of Mushroom



Performance Metrics of Tictactoe



Performance Metrics of Nursery

