a) Algorithm Performance:

• Which dataset achieved the highest accuracy and why?

The Mushroom dataset achieved the highest accuracy because of features that it has such as cap shape, odor, and gill colour, which are very effective at predicting whether a mushroom is edible.

• How does dataset size affect performance?

Larger datasets take more time to train but are better at capturing patterns. Smaller datasets might lead to overfitting.

• What role does the number of features play?

Adding more features can improve accuracy, but irrelevant features can create noise, messy data and make analysis difficult to perform.

b)Data Characteristics Impact:

• How does class imbalance affect tree construction?

If one class is larger than the other, the decision tree might become biased, making it harder to analyse the data.

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

Multi-valued features worked better than binary features.

c) Practical Applications:

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

The mushroom dataset can be used for agriculture field and food safety and stuff related to food or products. The Tic-Tac-Toe dataset can be used for strategy prediction in maybe some mind games. The Nursery dataset can be used to make decisions in a school environment.

•What are the interpretability advantages for each domain?

The mushroom dataset is simple and easy to interpret. The Tic-Tac-Toe dataset uses binary values (win/lose) so it is also easy to interpret. The Nursery dataset is more complex, but demonstrates how a decision tree can be created from multiple attributes and values.

• How would you improve performance for each dataset?

In the mushroom dataset, pruning could reduce class imbalance. In the Tic-Tac-Toe dataset, overfitting can be reduced by pruning. In the Nursery dataset, choosing only relevant and required attributes would improve performance.