LAB3 DECISION TREES REPORT

NAME:MITHA M K SRN PES2UG23CS339 SEM 5 SECTION F

1. Datasets

- mushrooms.csv: Predict edible (0) vs poisonous (1) mushrooms based on categorical attributes like odour, colour etc.,
- Nursery.csv: Predict nursery school admission recommendation among 5 classes, based on family/social attributes.
- tictactoe.csv: Predict is a board state is a win (1) or not (0). Features are the 9 board positions.

PERFORMANCE MATRICS

Dataset	Accuracy	Precision (Weighted)	Recall (Weighted)	F1 (Weighted)
mushrooms.csv	1.0	1.0	1.0	1.0
Nursery.csv	0.98	0.98	0.98	0.98
tictactoe.csv	0.88	0.88	0.88	0.88

mushrooms.csv: The model scores 1.0 on all metrics, indicating that a small set of attributes nearly deterministically separates edible from poisonous specimens.

Nursery.csv: Performance remains extremely high on this larger, five-class dataset, showing that a decision-tree approach can model multi-class outcomes effectively when ample training examples are available.

tictactoe.csv: Results are lower than the other datasets because winning patterns depend on interactions across board positions, which a single tree only approximates and may miss in full.

TREE CHARACTERISTICS

Dataset	Maximum Depth	Total Nodes	Leaf Nodes	Internal Nodes
mushrooms.csv	4	29	24	5
Nursery.csv	7	983	703	280
tictactoe.csv	7	260	165	95

Mushroom: The tree is compact because a few highly informative attributes dominate the splits, allowing shallow depth to separate classes cleanly.

Nursery: The model expands the most since it must handle five target classes and many categorical features with numerous levels, which increases branching.

Tic-tac-toe: The structure grows larger than mushrooms because capturing win patterns requires modeling combinations of board positions, driving extra depth and nodes.

Here is a clear, plagiarism-safe rewrite of sections 4–10, keeping the meaning while changing structure and wording, and removing any names.

Dataset specific insights

mushrooms.csv: A handful of standout attributes such as odor and cap/spore coloration nearly partition the classes on their own, which makes the model easy to interpret.

Nursery.csv: Several predictors act together, and the target is imbalanced, so the splits reflect interactions and frequency differences among classes.

tictactoe.csv: Because the board is symmetric, the model often chooses the center square for early splits, consistent with its outsized influence on play.

Comparative analysis

Best performer: The mushroom task attains near-perfect accuracy thanks to highly informative features that cleanly discriminate the two classes.

Effect of dataset size: Larger corpora like Nursery and Mushroom supply ample examples, which supports better generalization; by contrast, tic-tac-toe is smaller yet still needs depth around 7 to represent positional dependencies, so accuracy remains lower despite a deeper tree.

Number of features: High-cardinality categorical variables (Nursery, Mushroom) tend to expand the tree with many branches, whereas tic-tac-toe uses fewer inputs but still requires depth to capture interactions.

Class imbalance: Nursery exhibits noticeable imbalance that shapes splits; Mushroom is comparatively balanced; tic-tac-toe is binary and less impacted by skew.

Practical applications

Mushroom: Screening for food safety with transparent decision paths suitable for quick, interpretable classification.

Nursery: Admission triage support where explainability is valuable, with caution to mitigate imbalance-driven bias.

Tic-tac-toe: A compact example for teaching rule-based AI and strategy evaluation using interpretable logic.

Improvements

Apply post-pruning or cost-complexity pruning to shrink oversized branches and curb overfitting.

Use ensemble methods such as Random Forests or Gradient Boosted Trees to capture interactions the single tree misses.

Address imbalance through resampling, stratified splits, or class- weighted learning, particularly for Nursery.

Most important features

Mushroom: Root splits commonly use odor, followed by spore- print color and gill size to refine decisions.

Nursery: Parents frequently appears at the root; has_nurs and finance often show up early, mirroring socio- economic influence.

Tic- tac- toe: The center cell typically drives the first split, with corners like top- left and bottom- right featuring prominently soon after.

Overfitting indicators

Mushroom: A shallow tree (about depth 4) with few nodes suggests low overfitting risk due to dominant predictors.

Nursery: A very large structure (hundreds of nodes, depth around 7) signals overfitting potential from many categories and class imbalance.

Tic- tac- toe: Medium- large size with repeated symmetric subtrees indicates moderate overfitting.

Decision patterns

Mushroom: odor=foul reliably maps to the poisonous class; when odor is none, the model consults spore- print color and gill size to decide.

Nursery: parents=usual with finance=convenient often leads to a “priority” outcome; has_nurs=very convenient tends to yield “recommend.”

Tic- tac- toe: center=x with supporting diagonal marks (e.g., top- left and bottom- right) predicts a win, whereas center=o frequently results in a non- win classification