MACHINE LEARNING LAB-3

Name: Navyata Venkatesh

SRN: PES2UG23CS375

Section: F

REPORT FOR LAB-3

- Mushroom dataset: 100% accuracy, precision, recall, and F1-score. Perfectly separable due to strong features like odor and spore-print-color.
- Tic-Tac-Toe dataset: 89.36% accuracy, with precision, recall, and F1 around 0.89. Performance lower because outcomes depend on interdependent board states.
- Nursery dataset: 98.67% accuracy, weighted metrics near 0.98, but macro metrics around 0.76. Minority classes such as *very_recom* and *spec_prior* were harder to classify.
- Mushroom tree: Depth 4, 59 total nodes. Small and efficient structure achieving perfect results.
- Tic-Tac-Toe tree: Depth 7, 306 nodes. Larger and deeper due to many possible board patterns.
- Nursery tree: Depth 7, 992 nodes. Largest tree, influenced by multi-class output and multi-valued attributes.
- Most important features: odor and spore-print-color (Mushroom), middle-square and corners (Tic-Tac-Toe), finance, parents, and social (Nursery).
- Mushroom dataset: Balanced classes, strong separability, decision patterns like “odor = foul → poisonous.” No overfitting observed.
- Tic-Tac-Toe dataset: Balanced, middle-square highly predictive, but deeper structure suggested possible overfitting to rare cases.
- Nursery dataset: High predictive performance, but imbalance lowered macro scores. Large tree indicated high complexity and minor overfitting on rare classes.
- Mushroom dataset achieved the highest accuracy due to highly predictive features.
- Tic-Tac-Toe required deeper trees and showed lower accuracy because of complex interdependencies.

- Nursery achieved strong overall accuracy but showed class imbalance issues.
- Dataset size and feature richness shaped the results: Mushroom was simple and separable, Tic-Tac-Toe interdependent and less generalizable, Nursery large and multi-class.
- Applications: Mushroom → food safety, Tic-Tac-Toe → strategy modelling, Nursery → recommendation systems.
- Improvements: pruning for Mushroom and Nursery, balancing classes for Nursery, and ensemble methods (Random Forests, Boosting) for Tic-Tac-Toe.