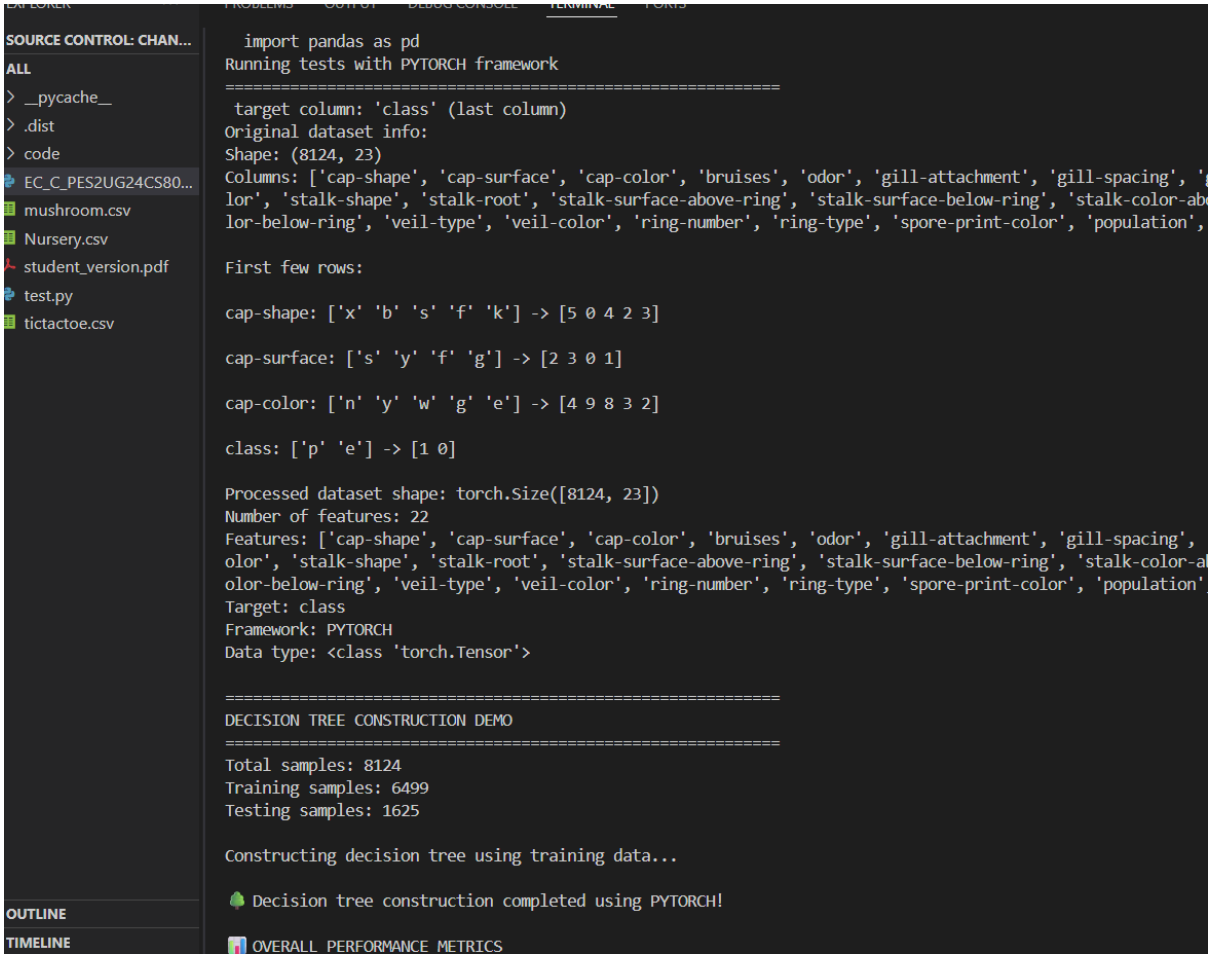
Lab Report – Lab 3

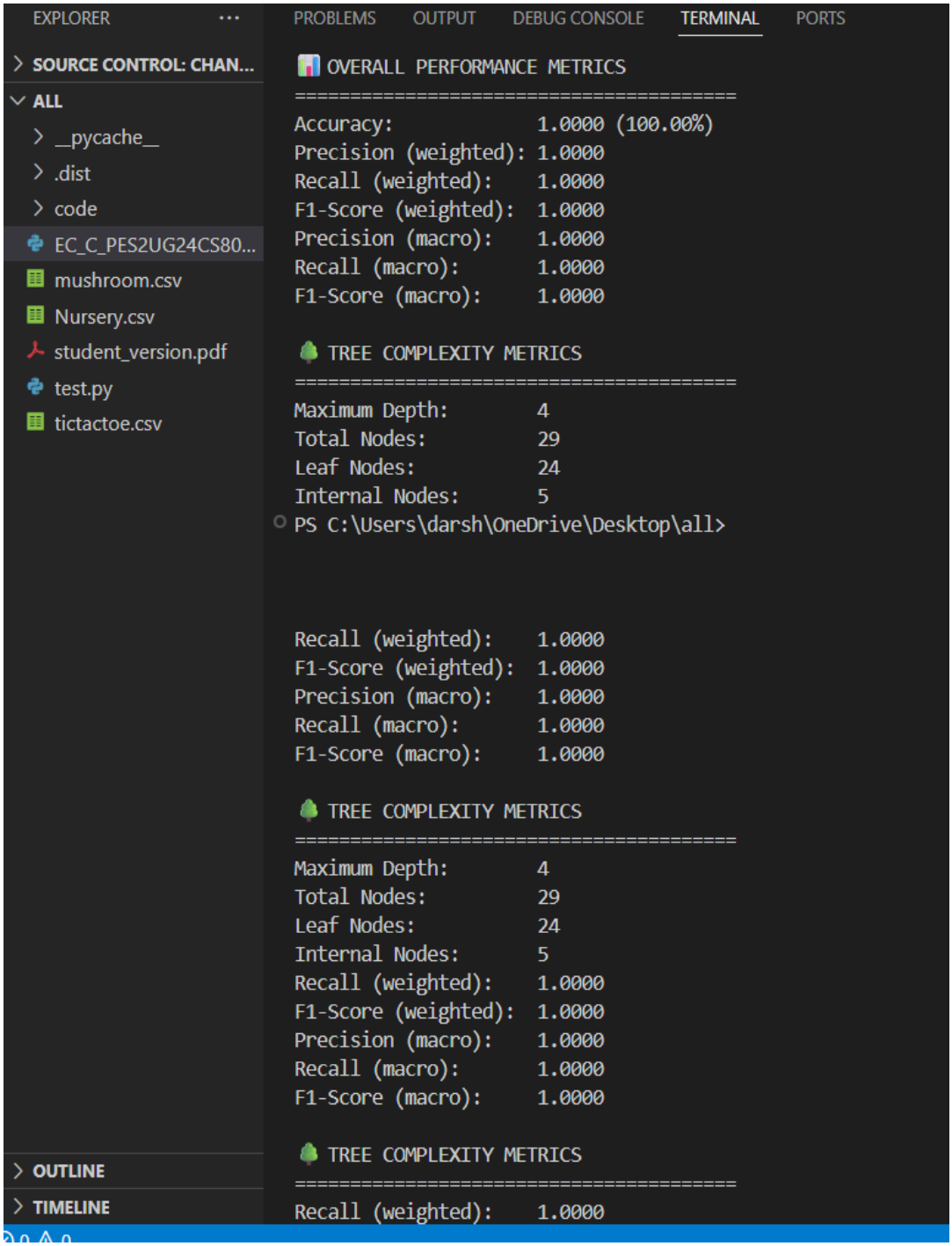
Course: Machine Learning

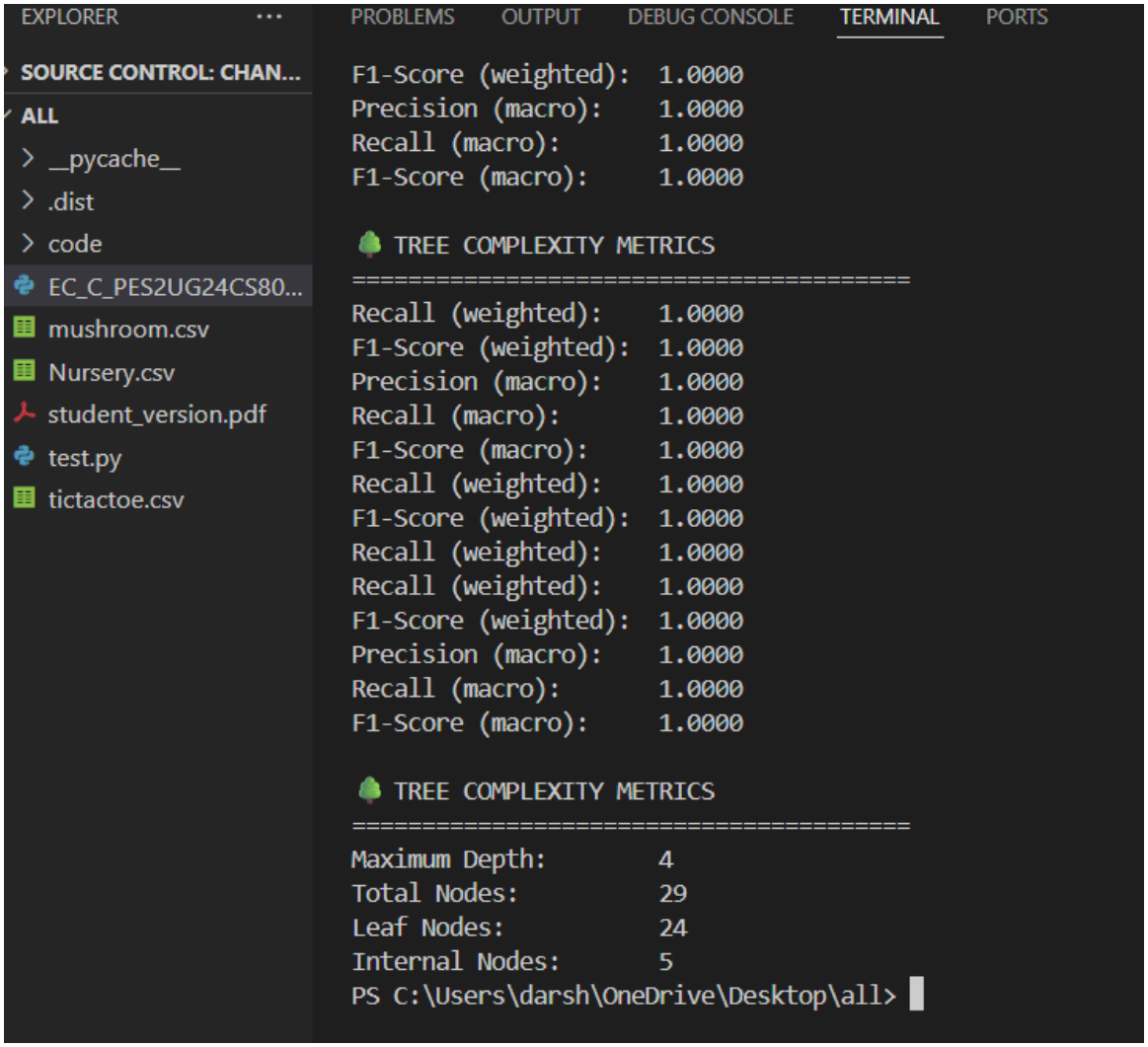
Student Name: Darshan N

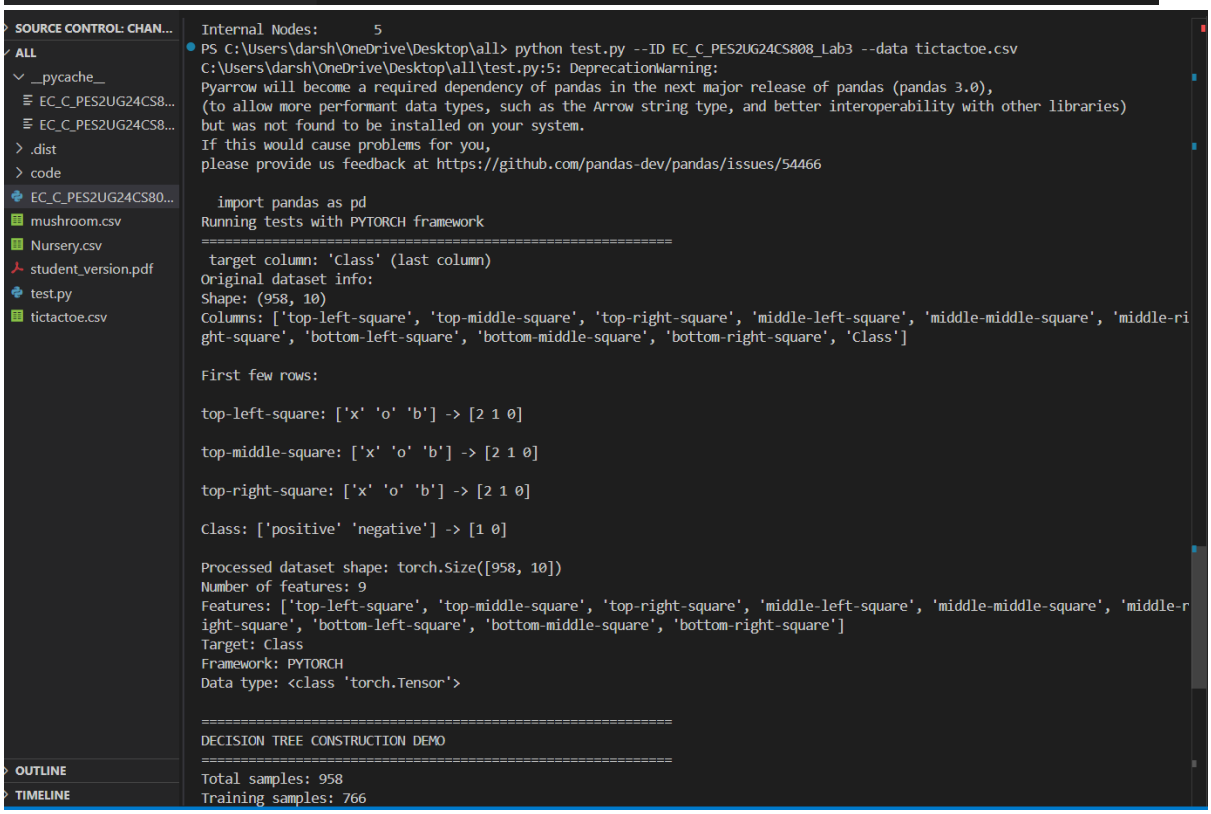
USN: PES2UG24CS808

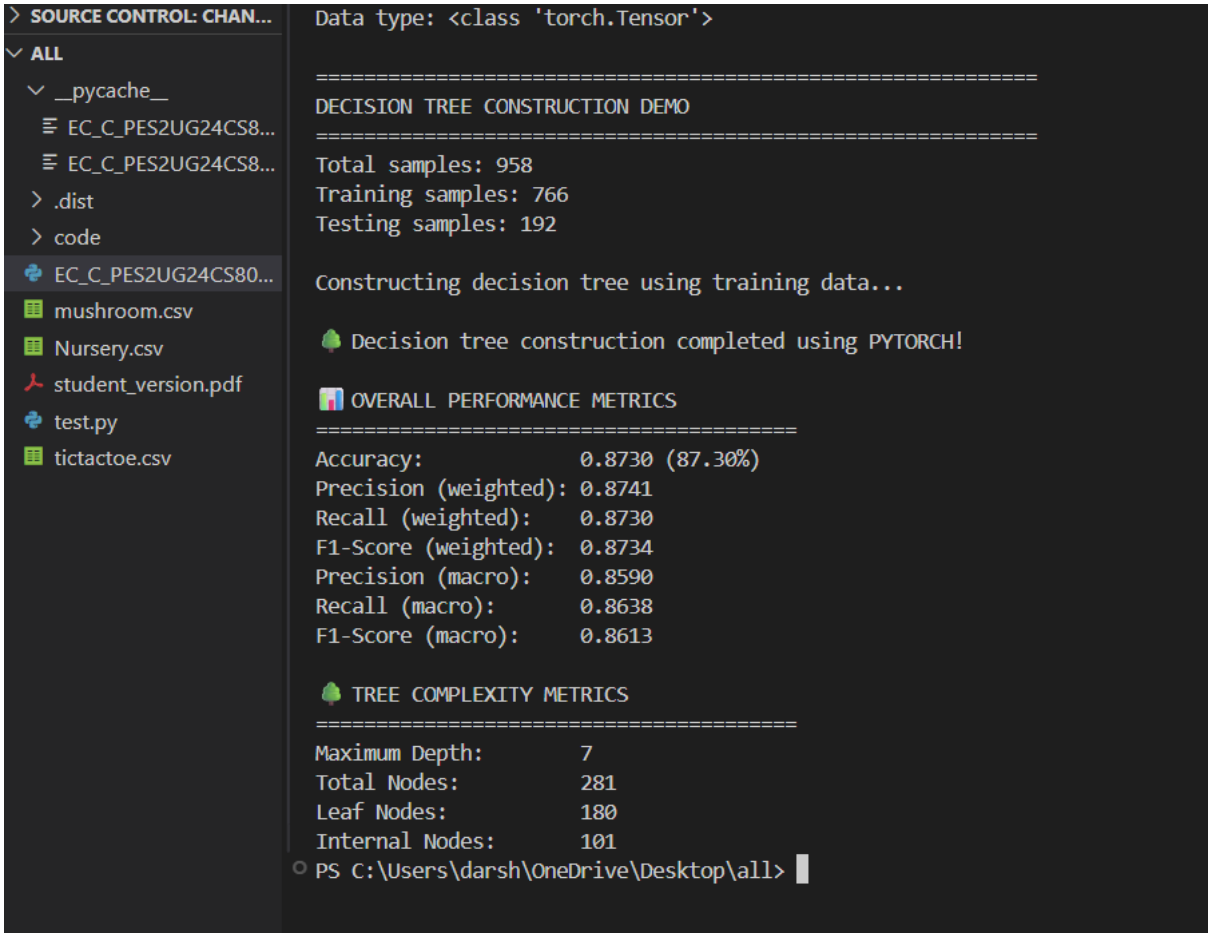
1. **Entropy and information gain calculated for Mushroom , tic tac toe and nursery dataset.**

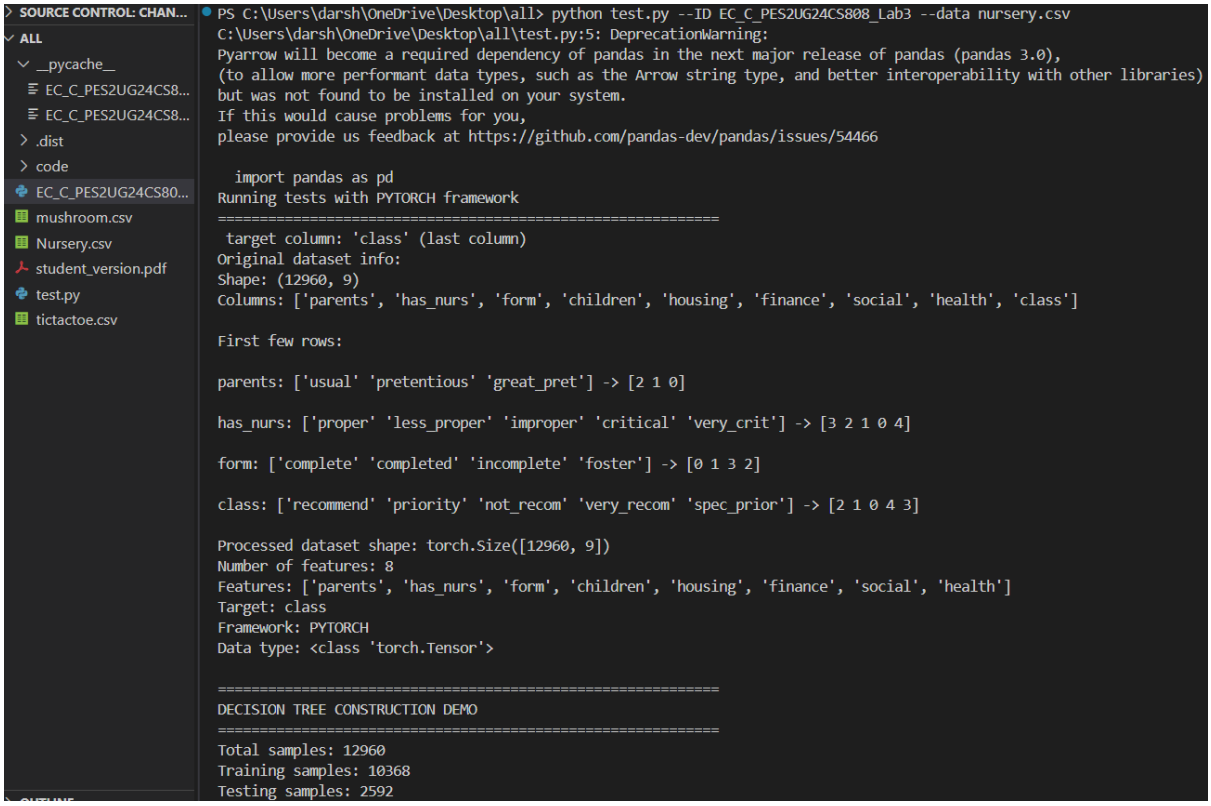


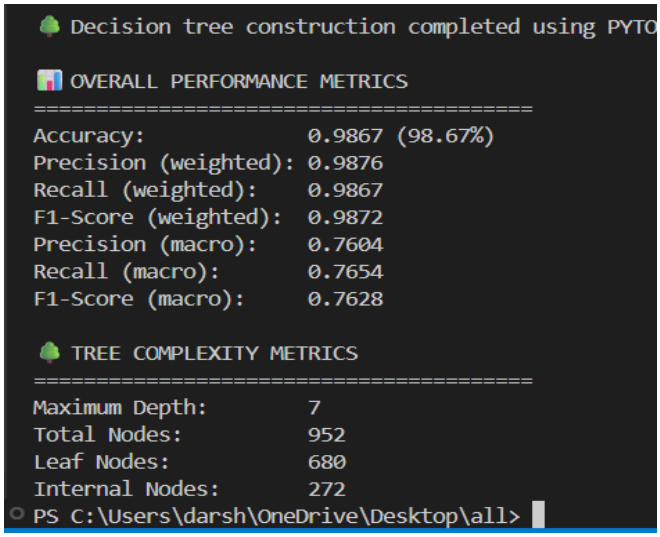

EXPLORER
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
SOURCE CONTROL: CHAN...
ALL
__pycache__
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
OUTLINE
TIMELINE
OVERALL PERFORMANCE METRICS
Accuracy: 1.0000 (100.00%)
Precision (weighted): 1.0000
Recall (weighted): 1.0000
F1-Score (weighted): 1.0000
Precision (macro): 1.0000
Recall (macro): 1.0000
F1-Score (macro): 1.0000
TREE COMPLEXITY METRICS
Maximum Depth: 4
Total Nodes: 29
Leaf Nodes: 24
Internal Nodes: 5
PS C:\Users\darsh\OneDrive\Desktop\all>
Recall (weighted): 1.0000
F1-Score (weighted): 1.0000
Precision (macro): 1.0000
Recall (macro): 1.0000
F1-Score (macro): 1.0000
TREE COMPLEXITY METRICS
Maximum Depth: 4
Total Nodes: 29
Leaf Nodes: 24
Internal Nodes: 5
Recall (weighted): 1.0000
F1-Score (weighted): 1.0000
Precision (macro): 1.0000
Recall (macro): 1.0000
F1-Score (macro): 1.0000
TREE COMPLEXITY METRICS
Recall (weighted): 1.0000


EXPLORER
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
SOURCE CONTROL: CHAN...
ALL
__pycache__
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
F1-Score (weighted): 1.0000
Precision (macro): 1.0000
Recall (macro): 1.0000
F1-Score (macro): 1.0000
TREE COMPLEXITY METRICS
Recall (weighted): 1.0000
F1-Score (weighted): 1.0000
Precision (macro): 1.0000
Recall (macro): 1.0000
F1-Score (macro): 1.0000
Recall (weighted): 1.0000
F1-Score (weighted): 1.0000
Recall (weighted): 1.0000
Recall (weighted): 1.0000
F1-Score (weighted): 1.0000
Precision (macro): 1.0000
Recall (macro): 1.0000
F1-Score (macro): 1.0000
TREE COMPLEXITY METRICS
Maximum Depth: 4
Total Nodes: 29
Leaf Nodes: 24
Internal Nodes: 5
PS C:\Users\darsh\OneDrive\Desktop\all>

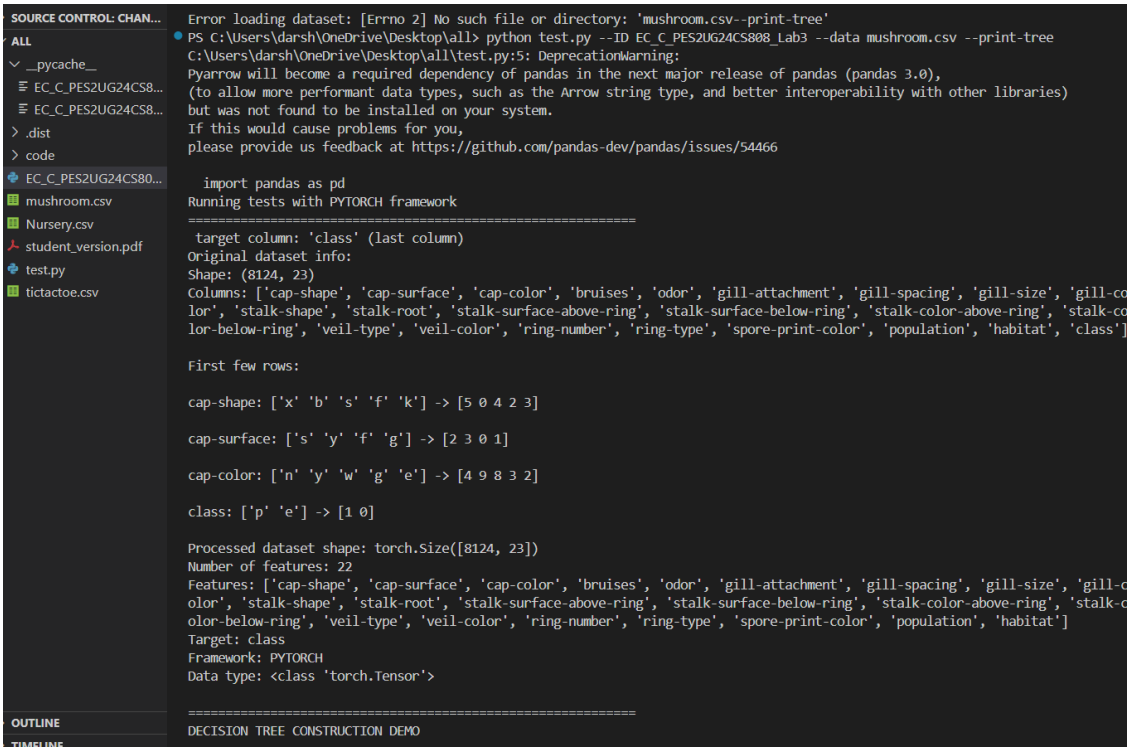

SOURCE CONTROL: CHAN...
ALL
__pycache__
EC_C_PES2UG24CS8...
EC_C_PES2UG24CS8...
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
OUTLINE
TIMELINE
Internal Nodes: 5
PS C:\Users\darsh\OneDrive\Desktop\all> python test.py --ID EC_C_PES2UG24CS808_Lab3 --data tictactoe.csv
C:\Users\darsh\OneDrive\Desktop\all\test.py:5: DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major release of pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and better interoperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54466
import pandas as pd
Running tests with PYTORCH framework
target column: 'Class' (last column)
Original dataset info:
Shape: (958, 10)
Columns: ['top-left-square', 'top-middle-square', 'top-right-square', 'middle-left-square', 'middle-middle-square', 'middle-right-square', 'bottom-left-square', 'bottom-middle-square', 'bottom-right-square', 'Class']
First few rows:
top-left-square: ['x' 'o' 'b'] -> [2 1 0]
top-middle-square: ['x' 'o' 'b'] -> [2 1 0]
top-right-square: ['x' 'o' 'b'] -> [2 1 0]
Class: ['positive' 'negative'] -> [1 0]
Processed dataset shape: torch.Size([958, 10])
Number of features: 9
Features: ['top-left-square', 'top-middle-square', 'top-right-square', 'middle-left-square', 'middle-middle-square', 'middle-right-square', 'bottom-left-square', 'bottom-middle-square', 'bottom-right-square']
Target: Class
Framework: PYTORCH
Data type: <class 'torch.Tensor'>
DECISION TREE CONSTRUCTION DEMO
Total samples: 958
Training samples: 766

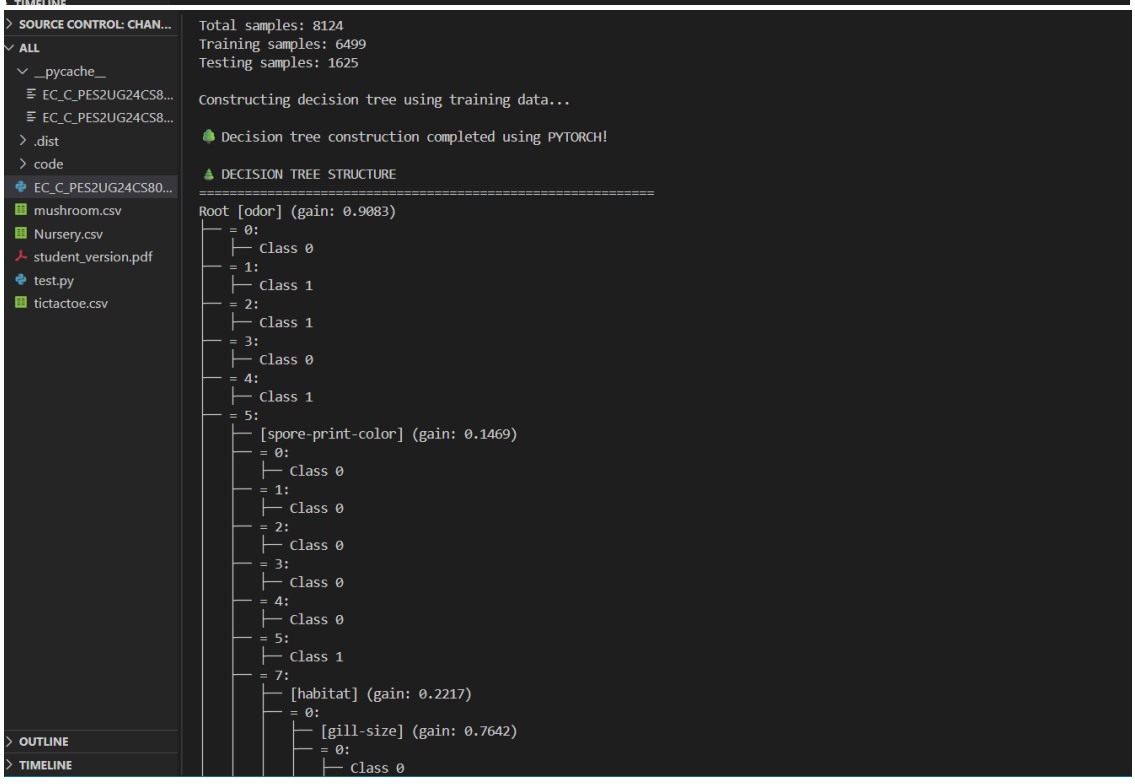

SOURCE CONTROL: CHAN...
ALL
__pycache__
EC_C_PES2UG24CS8...
EC_C_PES2UG24CS8...
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
Data type: <class 'torch.Tensor'>
==
DECISION TREE CONSTRUCTION DEMO
==
Total samples: 958
Training samples: 766
Testing samples: 192
Constructing decision tree using training data...
Decision tree construction completed using PYTORCH!
OVERALL PERFORMANCE METRICS
====================================
Accuracy: 0.8730 (87.30%)
Precision (weighted): 0.8741
Recall (weighted): 0.8730
F1-Score (weighted): 0.8734
Precision (macro): 0.8590
Recall (macro): 0.8638
F1-Score (macro): 0.8613
TREE COMPLEXITY METRICS
====================================
Maximum Depth: 7
Total Nodes: 281
Leaf Nodes: 180
Internal Nodes: 101
PS C:\Users\darsh\OneDrive\Desktop\all>

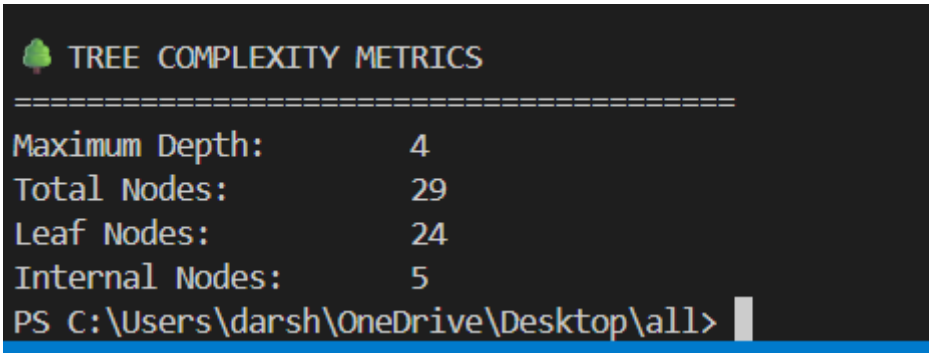

SOURCE CONTROL: CHAN...
ALL
__pycache__
EC_C_PES2UG24CS8...
EC_C_PES2UG24CS8...
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
PS C:\Users\darsh\OneDrive\Desktop\all> python test.py --ID EC_C_PES2UG24CS808_Lab3 --data nursery.csv
C:\Users\darsh\OneDrive\Desktop\all\test.py:5: DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major release of pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and better interoperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54466
import pandas as pd
Running tests with PYTORCH framework
==
target column: 'class' (last column)
Original dataset info:
Shape: (12960, 9)
Columns: ['parents', 'has_nurs', 'form', 'children', 'housing', 'finance', 'social', 'health', 'class']
First few rows:
parents: ['usual' 'pretentious' 'great_pret'] -> [2 1 0]
has_nurs: ['proper' 'less_proper' 'improper' 'critical' 'very_crit'] -> [3 2 1 0 4]
form: ['complete' 'completed' 'incomplete' 'foster'] -> [0 1 3 2]
class: ['recommend' 'priority' 'not_recom' 'very_recom' 'spec_prior'] -> [2 1 0 4 3]
Processed dataset shape: torch.Size([12960, 9])
Number of features: 8
Features: ['parents', 'has_nurs', 'form', 'children', 'housing', 'finance', 'social', 'health']
Target: class
Framework: PYTORCH
Data type: <class 'torch.Tensor'>
==
DECISION TREE CONSTRUCTION DEMO
==
Total samples: 12960
Training samples: 10368
Testing samples: 2592



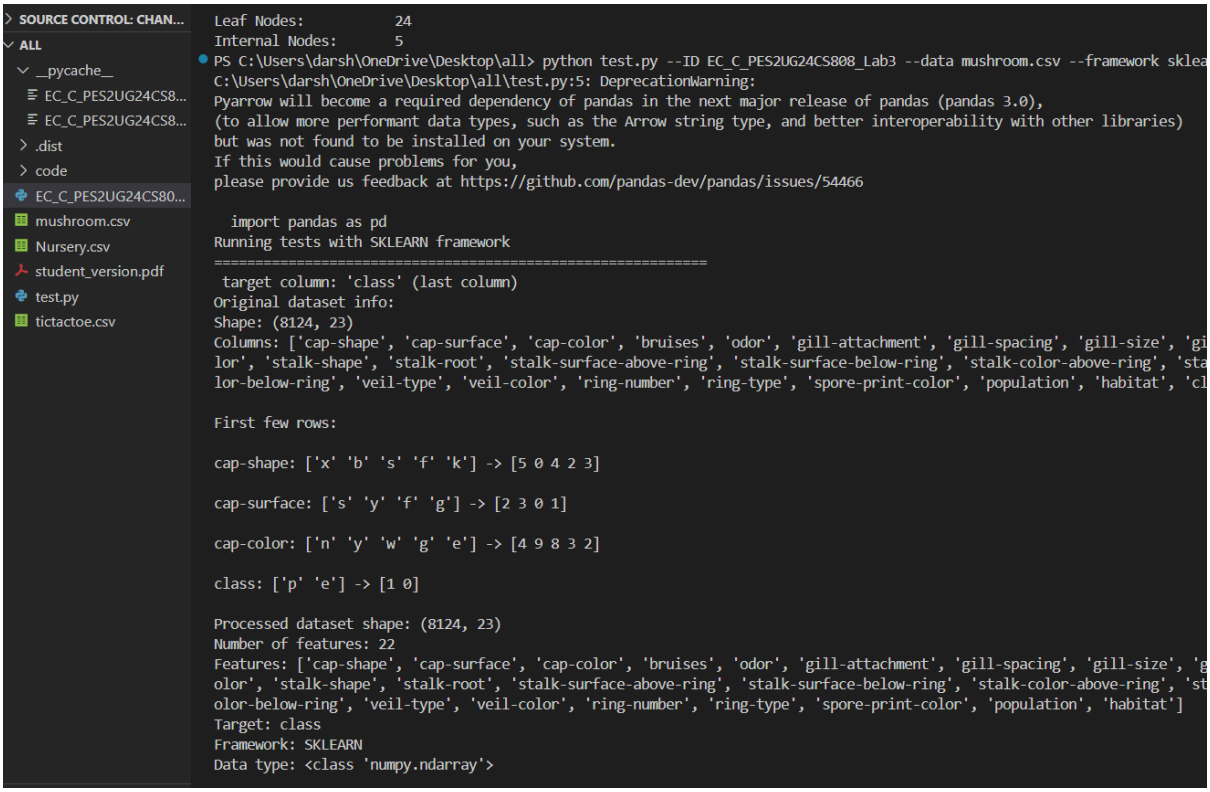
2. **Decision tree generated using ID3 algorithm for Mushroom dataset**

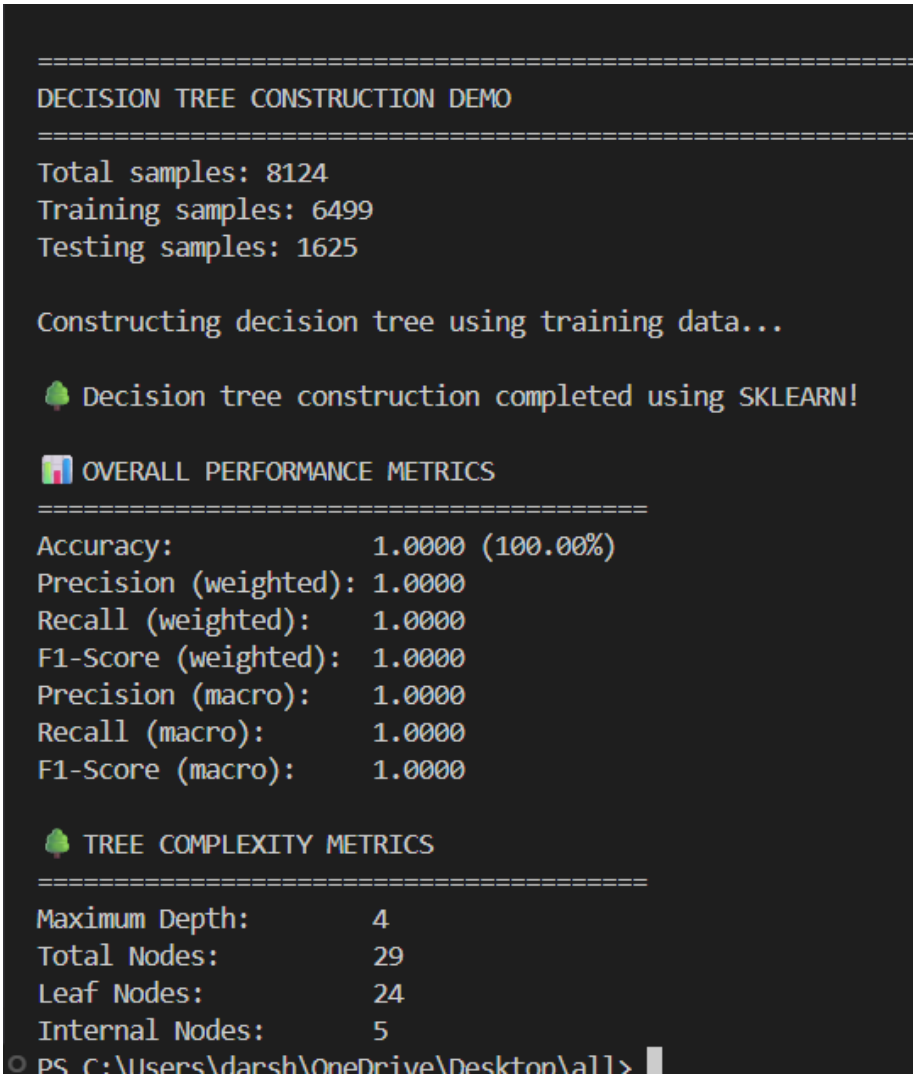

SOURCE CONTROL: CHAN...
ALL
__pycache__
EC_C_PES2UG24CS8...
EC_C_PES2UG24CS8...
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
OUTLINE
TIMELINE
Error loading dataset: [Errno 2] No such file or directory: 'mushroom.csv--print-tree'
PS C:\Users\darsh\OneDrive\Desktop\all> python test.py --ID EC_C_PES2UG24CS808_Lab3 --data mushroom.csv --print-tree
C:\Users\darsh\OneDrive\Desktop\all\test.py:5: DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major release of pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and better interoperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54466
import pandas as pd
Running tests with PYTORCH framework
target column: 'class' (last column)
Original dataset info:
Shape: (8124, 23)
Columns: ['cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor', 'gill-attachment', 'gill-spacing', 'gill-size', 'gill-color', 'stalk-shape', 'stalk-root', 'stalk-surface-above-ring', 'stalk-surface-below-ring', 'stalk-color-above-ring', 'stalk-color-below-ring', 'veil-type', 'veil-color', 'ring-number', 'ring-type', 'spore-print-color', 'population', 'habitat', 'class']
First few rows:
cap-shape: ['x' 'b' 's' 'f' 'k'] -> [5 0 4 2 3]
cap-surface: ['s' 'y' 'f' 'g'] -> [2 3 0 1]
cap-color: ['n' 'y' 'w' 'g' 'e'] -> [4 9 8 3 2]
class: ['p' 'e'] -> [1 0]
Processed dataset shape: torch.Size([8124, 23])
Number of features: 22
Features: ['cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor', 'gill-attachment', 'gill-spacing', 'gill-size', 'gill-color', 'stalk-shape', 'stalk-root', 'stalk-surface-above-ring', 'stalk-surface-below-ring', 'stalk-color-above-ring', 'stalk-color-below-ring', 'veil-type', 'veil-color', 'ring-number', 'ring-type', 'spore-print-color', 'population', 'habitat']
Target: class
Framework: PYTORCH
Data type: <class 'torch.Tensor'>
DECISION TREE CONSTRUCTION DEMO


SOURCE CONTROL: CHAN...
ALL
__pycache__
EC_C_PES2UG24CS8...
EC_C_PES2UG24CS8...
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
OUTLINE
TIMELINE
Total samples: 8124
Training samples: 6499
Testing samples: 1625
Constructing decision tree using training data...
Decision tree construction completed using PYTORCH!
DECISION TREE STRUCTURE
Root [odor] (gain: 0.9083)
├── = 0:
│ ├── Class 0
├── = 1:
│ ├── Class 1
├── = 2:
│ ├── Class 1
├── = 3:
│ ├── Class 0
├── = 4:
│ ├── Class 1
├── = 5:
│ ├── [spore-print-color] (gain: 0.1469)
│ ├── = 0:
│ │ ├── Class 0
│ ├── = 1:
│ │ ├── Class 0
│ ├── = 2:
│ │ ├── Class 0
│ ├── = 3:
│ │ ├── Class 0
│ ├── = 4:
│ │ ├── Class 0
│ ├── = 5:
│ │ ├── Class 1
│ ├── = 7:
│ │ ├── [habitat] (gain: 0.2217)
│ │ ├── = 0:
│ │ │ ├── [gill-size] (gain: 0.7642)
│ │ │ ├── = 0:
│ │ │ │ ├── Class 0



3. **Comparison with sklearn DecisionTreeClassifier implementation.**


SOURCE CONTROL: CHAN...
ALL
__pycache__
EC_C_PES2UG24CS8...
EC_C_PES2UG24CS8...
.dist
code
EC_C_PES2UG24CS80...
mushroom.csv
Nursery.csv
student_version.pdf
test.py
tictactoe.csv
Leaf Nodes: 24
Internal Nodes: 5
PS C:\Users\darsh\OneDrive\Desktop\all> python test.py --ID EC_C_PES2UG24CS808_Lab3 --data mushroom.csv --framework sklea
C:\Users\darsh\OneDrive\Desktop\all\test.py:5: DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major release of pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and better interoperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54466
import pandas as pd
Running tests with SKLEARN framework
target column: 'class' (last column)
Original dataset info:
Shape: (8124, 23)
Columns: ['cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor', 'gill-attachment', 'gill-spacing', 'gill-size', 'gi
lor', 'stalk-shape', 'stalk-root', 'stalk-surface-above-ring', 'stalk-surface-below-ring', 'stalk-color-above-ring', 'sta
lor-below-ring', 'veil-type', 'veil-color', 'ring-number', 'ring-type', 'spore-print-color', 'population', 'habitat', 'cl
First few rows:
cap-shape: ['x' 'b' 's' 'f' 'k'] -> [5 0 4 2 3]
cap-surface: ['s' 'y' 'f' 'g'] -> [2 3 0 1]
cap-color: ['n' 'y' 'w' 'g' 'e'] -> [4 9 8 3 2]
class: ['p' 'e'] -> [1 0]
Processed dataset shape: (8124, 23)
Number of features: 22
Features: ['cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor', 'gill-attachment', 'gill-spacing', 'gill-size', 'g
olor', 'stalk-shape', 'stalk-root', 'stalk-surface-above-ring', 'stalk-surface-below-ring', 'stalk-color-above-ring', 'st
olor-below-ring', 'veil-type', 'veil-color', 'ring-number', 'ring-type', 'spore-print-color', 'population', 'habitat']
Target: class
Framework: SKLEARN
Data type: <class 'numpy.ndarray'>


DECISION TREE CONSTRUCTION DEMO
Total samples: 8124
Training samples: 6499
Testing samples: 1625
Constructing decision tree using training data...
Decision tree construction completed using SKLEARN!
OVERALL PERFORMANCE METRICS
Accuracy: 1.0000 (100.00%)
Precision (weighted): 1.0000
Recall (weighted): 1.0000
F1-Score (weighted): 1.0000
Precision (macro): 1.0000
Recall (macro): 1.0000
F1-Score (macro): 1.0000
TREE COMPLEXITY METRICS
Maximum Depth: 4
Total Nodes: 29
Leaf Nodes: 24
Internal Nodes: 5
PS C:\Users\darsh\OneDrive\Desktop\all>