

Prajakta Deokule

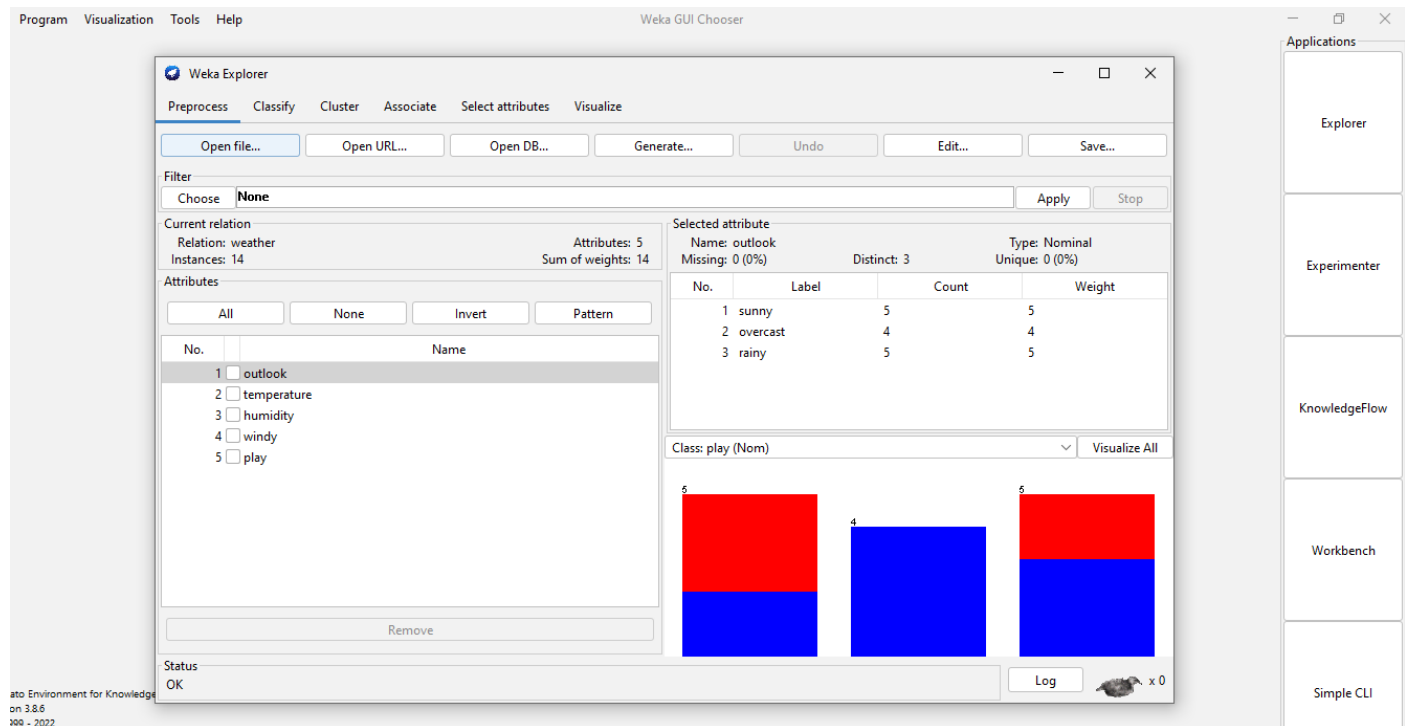
3330

A1

ASSIGNMENT 4

Problem Statement: Implement Decision Tree on a Dataset using WEKA.

Data set used: weather.arff



The process of creating a decision tree works by greedily selecting the best split point in order to make predictions and repeating the process until the tree is a fixed depth.

For implementing decision tree algorithm in WEKA

Algorithm used is REPTree

Another more advanced decision tree algorithm that you can use is the J48 in Weka.

Using REPTree

The screenshot shows the Weka Explorer interface with the REPTree classifier selected. The 'Classify' tab is active, and the 'Test options' section shows 'Cross-validation' with 10 folds. The 'Classifier output' pane displays the following results:

Size of the tree : 1

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	9	64.2857 %
Incorrectly Classified Instances	5	35.7143 %
Kappa statistic	0	
Mean absolute error	0.4725	
Root mean squared error	0.4958	
Relative absolute error	99.2308 %	
Root relative squared error	100.4896 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	1.000	1.000	0.643	1.000	0.783	?	0.178	0.555
	0.000	0.000	?	0.000	?	?	0.178	0.318
Weighted Avg.	0.643	0.643	?	0.643	?	?	0.178	0.470

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

5 0 | b = no

The screenshot shows the Weka Classifier Tree Visualizer window. The tree structure is empty, and the 'Tree View' pane displays the following information:

1 : yes (9/3) [5/2]

Using J48

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds
- ☐ Percentage split %

(Nom) play

Result list (right-click for options)

23:23:32 - trees.J48

Classifier output

```
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    weather
Instances:   14
Attributes:  5
    outlook
    temperature
    humidity
    windy
    play
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

outlook = sunny
|  humidity <= 75: yes (2.0)
|  humidity > 75: no (3.0)
outlook = overcast: yes (4.0)
outlook = rainy
|  windy = TRUE: no (2.0)
|  windy = FALSE: yes (3.0)

Number of Leaves :    5

Size of the tree :    8

Time taken to build model: 0.04 seconds
```

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds
- ☐ Percentage split %

(Nom) play

Result list (right-click for options)

23:23:32 - trees.J48

Classifier output

```
Size of the tree :    8

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      9           64.2857 %
Incorrectly Classified Instances    5           35.7143 %
Kappa statistic                    0.186
Mean absolute error                 0.2857
Root mean squared error             0.4818
Relative absolute error             60 %
Root relative squared error         97.6586 %
Total Number of Instances          14

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.778    0.600    0.700     0.778    0.737      0.189    0.789    0.847    yes
                0.400    0.222    0.500     0.400    0.444      0.189    0.789    0.738    no
Weighted Avg.   0.643    0.465    0.629     0.643    0.632      0.189    0.789    0.808

=== Confusion Matrix ===

a b  <-- classified as
7 2 | a = yes
3 2 | b = no
```

Tree View

