

Experiment 6

Aditya Sawant

Subject: DWM

Class: TE4

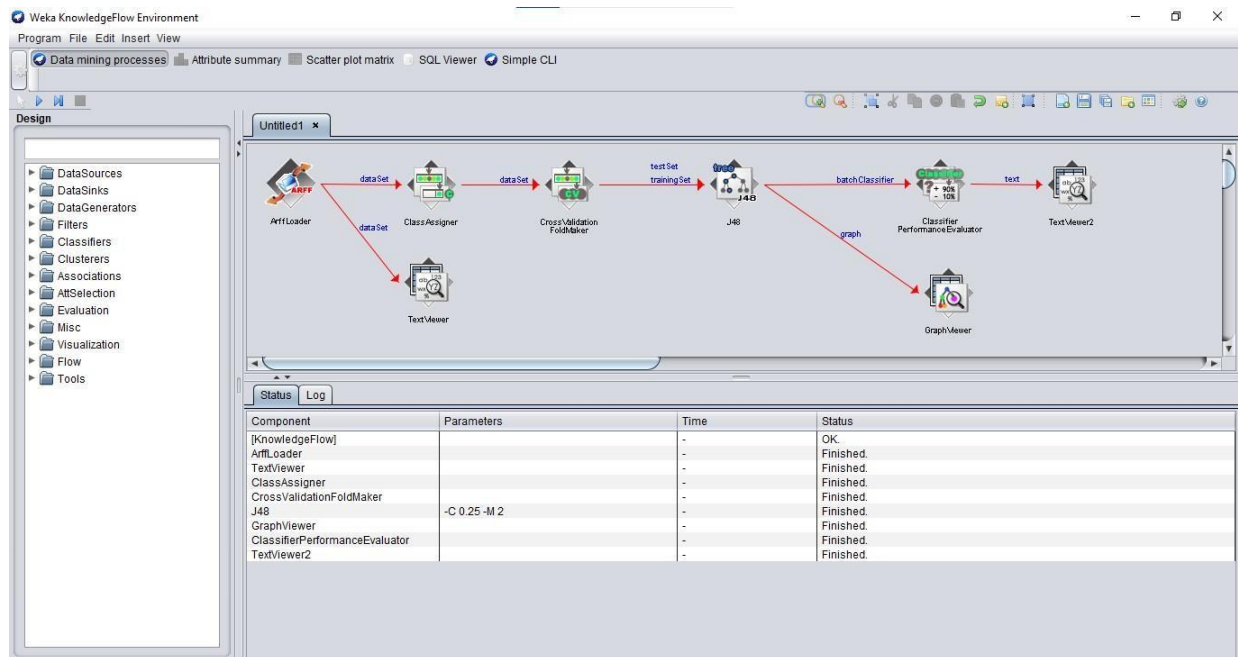
Roll no: 46

Batch: D

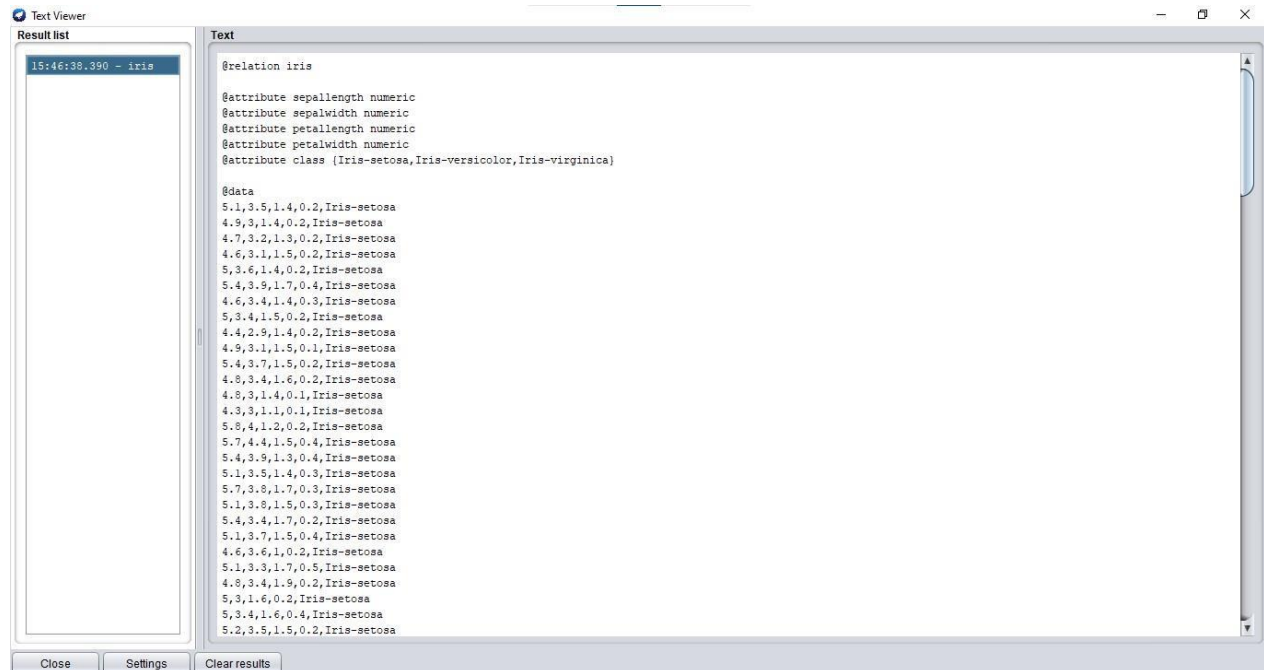
Aim: Perform data Pre-processing task and demonstrate Classification, Clustering, Association algorithm on data sets using data mining tool WEKA.

Knowledge Flow:

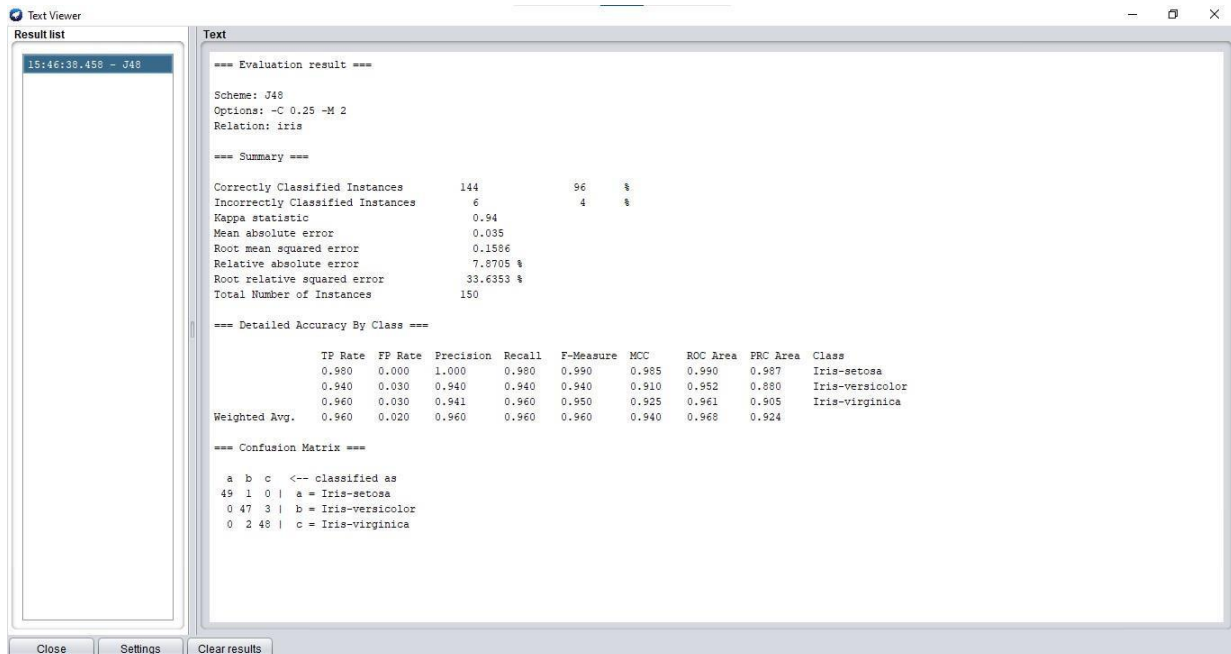
J48



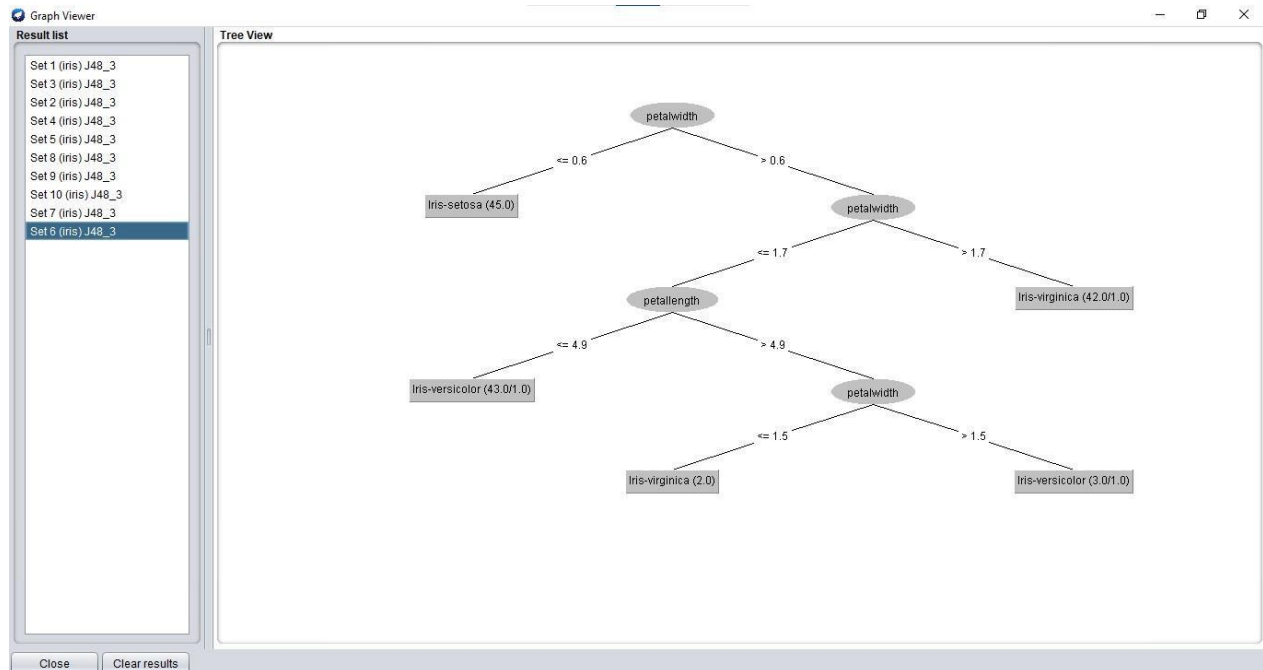
Iris Data Set Text Viewer



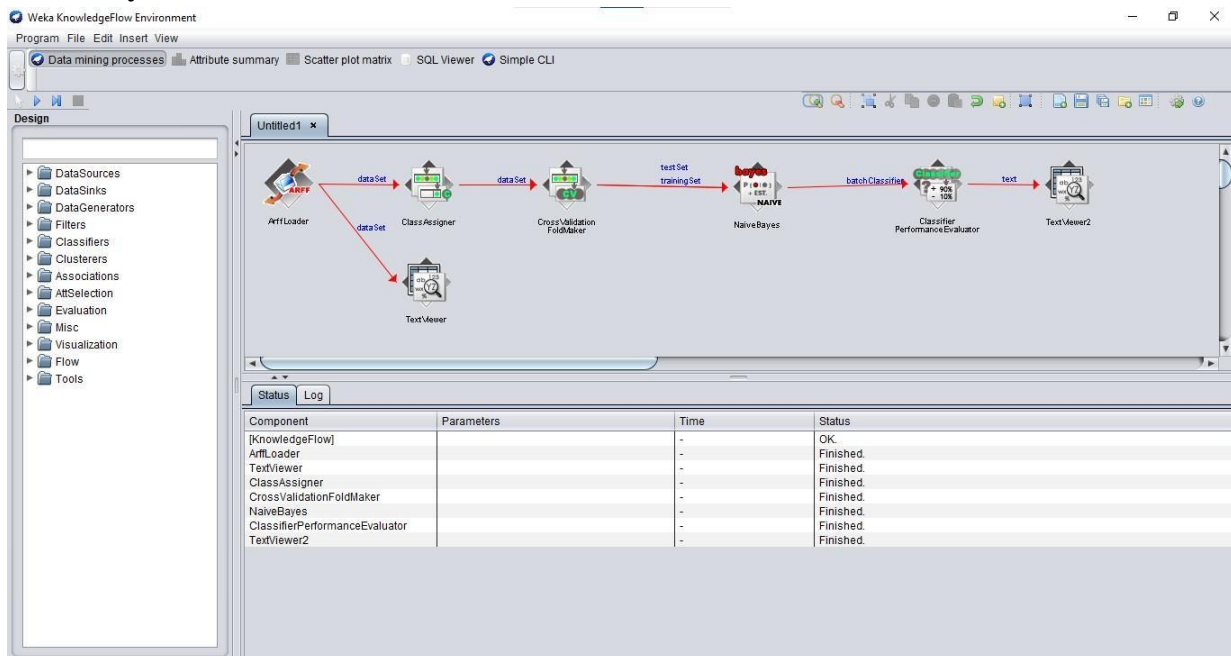
Evaluation Result Text Viewer



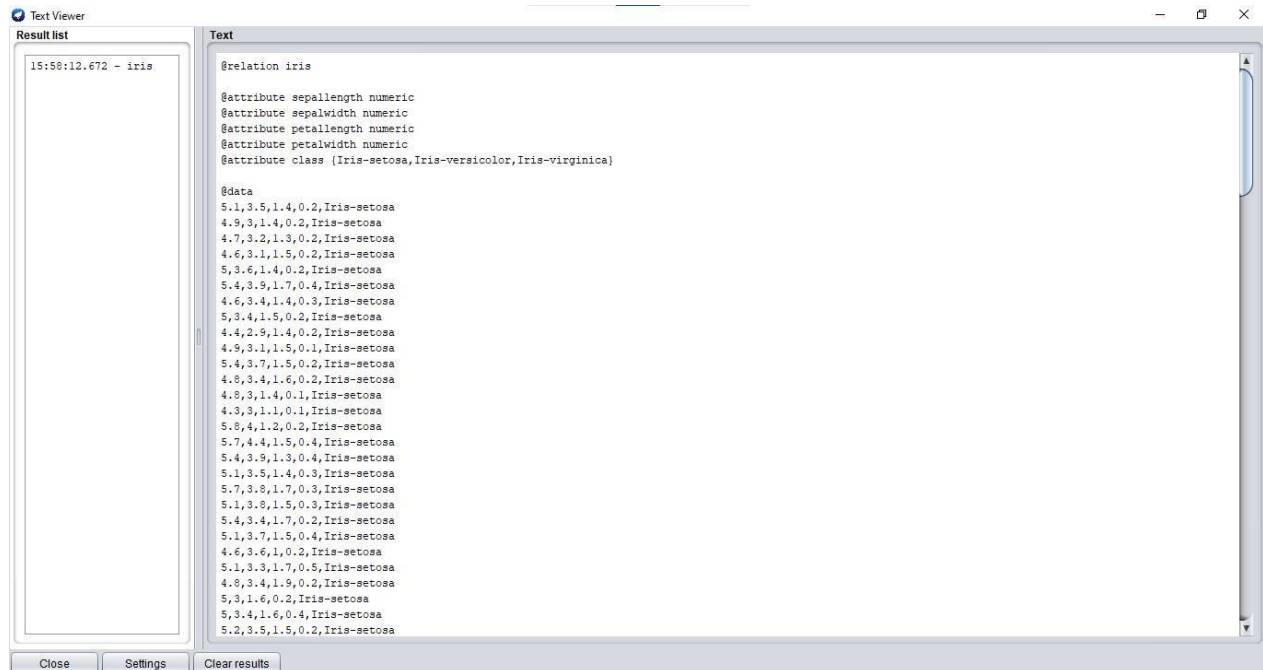
Graph Viewer



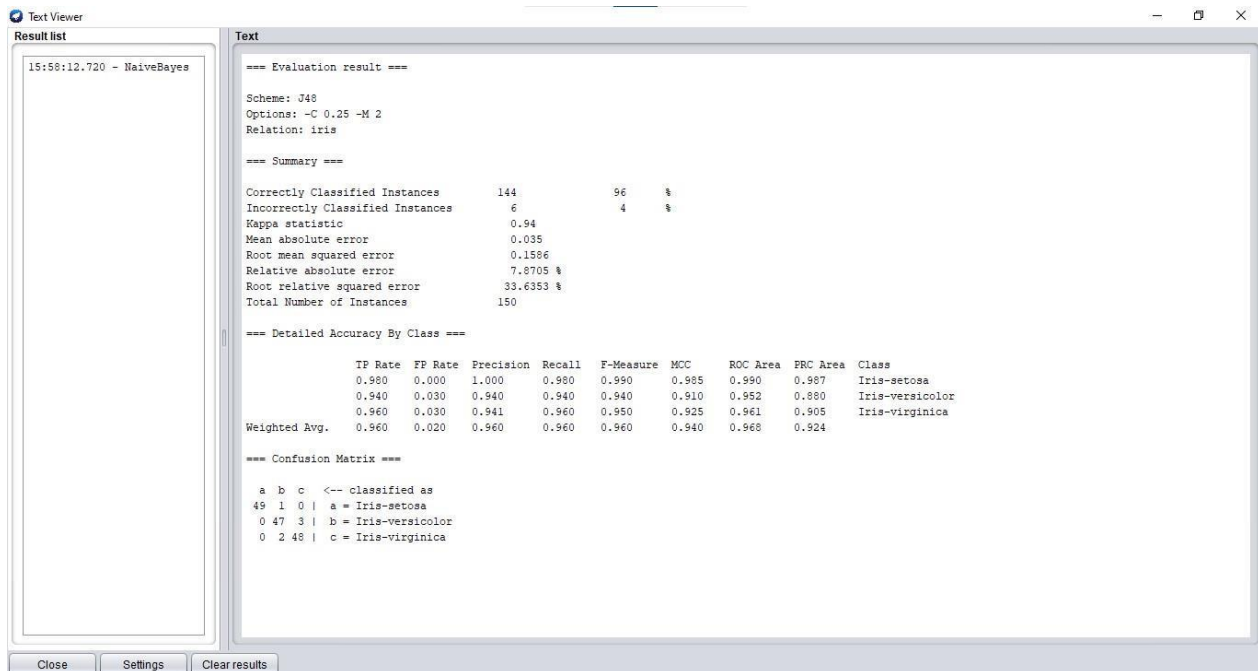
Naïve Bayes



Iris Data Set Text Viewer

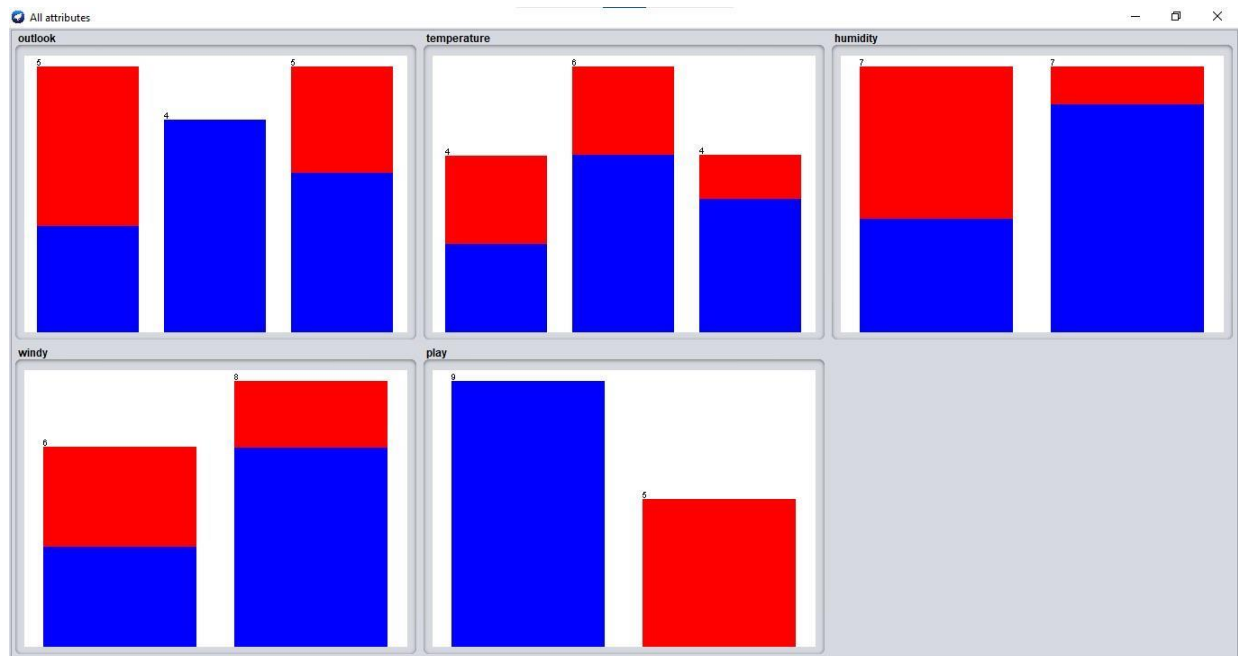
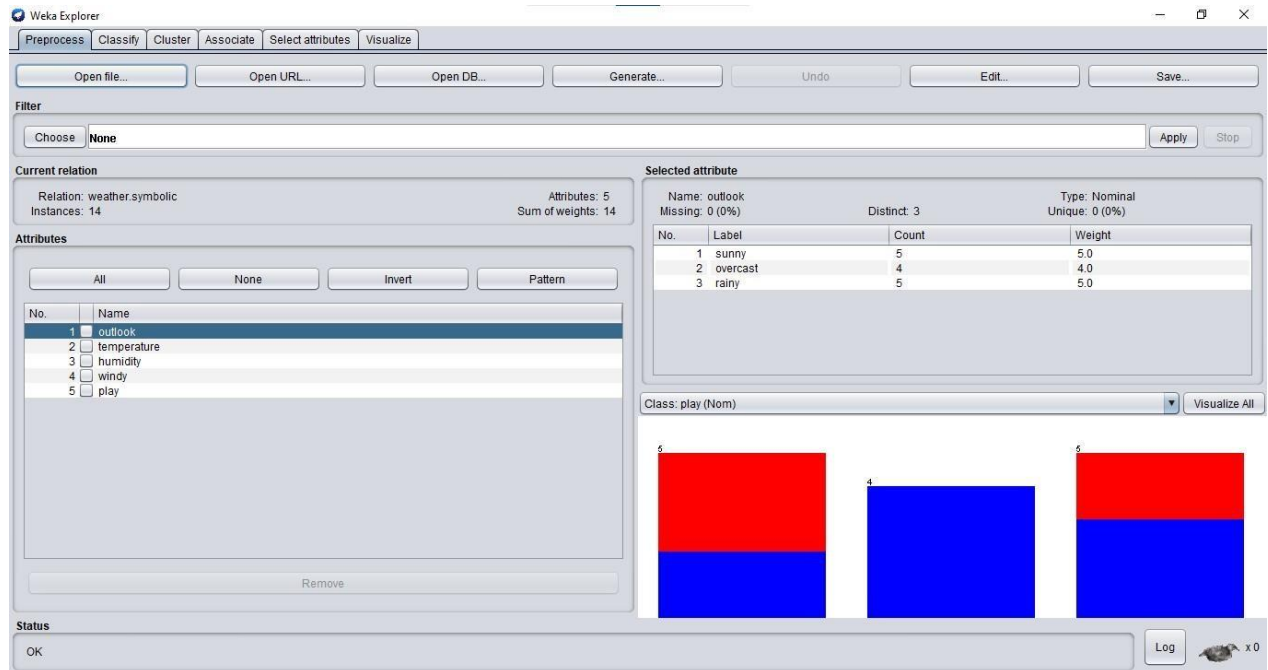


Evaluation Result



Explorer:

Weather Data Set



J48 Classifier

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 Set...

☐ Cross-validation Folds 10

☒ Percentage split % 80

More options...

(Nom) play

Start Stop

Result list (right-click for options)

16:22:54 - trees.J48

Classifier output

```
=== Run information ===
Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    weather.symbolic
Instances:    14
Attributes:   5
    outlook
    temperature
    humidity
    windy
    play
Test mode:    split 80.0% train, remainder test
=== Classifier model (full training set) ===
J48 pruned tree
-----
outlook = sunny
| humidity = high: no (3.0)
| humidity = normal: yes (2.0)
outlook = overcast: yes (4.0)
outlook = rainy
| windy = TRUE: no (2.0)
| windy = FALSE: yes (3.0)
Number of Leaves :    5

```

Status

OK

Log x0

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 Set...

☐ Cross-validation Folds 10

☒ Percentage split % 80

More options...

(Nom) play

Start Stop

Result list (right-click for options)

16:22:54 - trees.J48

Classifier output

```
Size of the tree :    8

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===
Correctly Classified Instances      0          0 %
Incorrectly Classified Instances    3         100 %
Kappa statistic                    -0.8
Mean absolute error                  1
Root mean squared error              1
Relative absolute error             177.2727 %
Root relative squared error         168.7695 %
Total Number of Instances           3

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
0.000      1.000      0.000      0.000    0.000    -1.000    0.000    0.333    yes
0.000      1.000      0.000      0.000    0.000    -1.000    0.000    0.667    no
Weighted Avg.  0.000      1.000      0.000      0.000    0.000    -1.000    0.000    0.556

=== Confusion Matrix ===

```

Status

OK

Log x0

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 Set...
☐ Cross-validation Folds 10
☒ Percentage split % 80
 More options...

(Nom) play

Start Stop

Result list (right-click for options)

16:22:54 - trees.J48

Classifier output

```

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      0      0 %
Incorrectly Classified Instances    3      100 %
Kappa statistic                    -0.8
Mean absolute error                 1
Root mean squared error             1
Relative absolute error             177.2727 %
Root relative squared error         168.7695 %
Total Number of Instances          3

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.000    1.000    0.000     0.000    0.000     -1.000    0.000    0.333    yes
      0.000    1.000    0.000     0.000    0.000     -1.000    0.000    0.667    no
Weighted Avg.  0.000    1.000    0.000     0.000    0.000     -1.000    0.000    0.556

=== Confusion Matrix ===

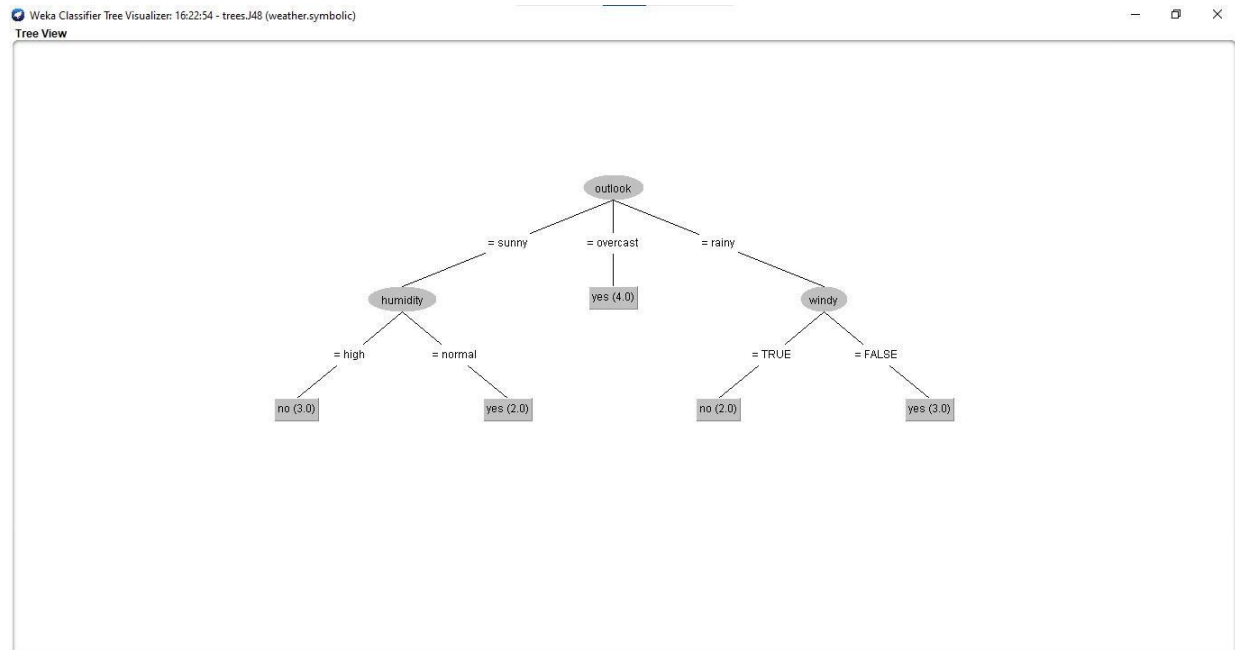
a b  <-- classified as
0 1 | a = yes
2 0 | b = no
  
```

Status

OK

Log x 0

Tree View



Naïve Bayes Classifier

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 80

More options...

(Nom) play

Start Stop

Result list (right-click for options)

16:22:54 - trees.J48

16:27:27 - bayes.NaiveBayes

Classifier output

```
=== Run information ===

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    weather.symbolic
Instances:   14
Attributes:  5
    outlook
    temperature
    humidity
    windy
    play
Test mode:   10-fold cross-validation

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

Naive Bayes Classifier

      Class
Attribute  yes  no
      (0.63) (0.38)

=====
outlook
sunny      3.0  4.0
overcast   5.0  1.0
rainy      4.0  3.0
[total]    12.0  8.0

temperature
hot        3.0  3.0
mild       5.0  3.0
cool       4.0  2.0
[total]    12.0  8.0

humidity
high       4.0  5.0
normal     7.0  2.0
[total]    11.0  7.0

windy
TRUE       4.0  4.0
FALSE      7.0  3.0
[total]    11.0  7.0

Time taken to build model: 0 seconds

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

Correctly Classified Instances      8          57.1429 %
Incorrectly Classified Instances    6          42.8571 %
Kappa statistic                    -0.0244
Mean absolute error                 0.4374
Root mean squared error            0.4916
Relative absolute error            91.8631 %
```

Status

OK

Log x0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 80

More options...

(Nom) play

Start Stop

Result list (right-click for options)

16:22:54 - trees.J48

16:27:27 - bayes.NaiveBayes

Classifier output

```
temperature
hot        3.0  3.0
mild       5.0  3.0
cool       4.0  2.0
[total]    12.0  8.0

humidity
high       4.0  5.0
normal     7.0  2.0
[total]    11.0  7.0

windy
TRUE       4.0  4.0
FALSE      7.0  3.0
[total]    11.0  7.0

Time taken to build model: 0 seconds

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

Correctly Classified Instances      8          57.1429 %
Incorrectly Classified Instances    6          42.8571 %
Kappa statistic                    -0.0244
Mean absolute error                 0.4374
Root mean squared error            0.4916
Relative absolute error            91.8631 %
```

Status

OK

Log x0

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

NaiveBayes

Test options

Use training set

Supplied test set

Cross-validation

Folds

10

Percentage split

%

80

More options...

(Nom) play

Start

Stop

Result list (right-click for options)

16:22:54 - trees.J48

16:27:27 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances

8

57.1429 %

Incorrectly Classified Instances

6

42.8571 %

Kappa statistic

-0.0244

Mean absolute error

0.4374

Root mean squared error

0.4916

Relative absolute error

91.8631 %

Root relative squared error

99.6492 %

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	0.800	0.636	0.778	0.700	-0.026	0.578	0.697	yes
	0.200	0.222	0.333	0.200	0.250	-0.026	0.578	0.557	no
Weighted Avg.	0.571	0.594	0.528	0.571	0.539	-0.026	0.578	0.647	

=== Confusion Matrix ===

a b

<-- classified as

7 2 | a = yes

4 1 | b = no

Status

OK

Log

x 0