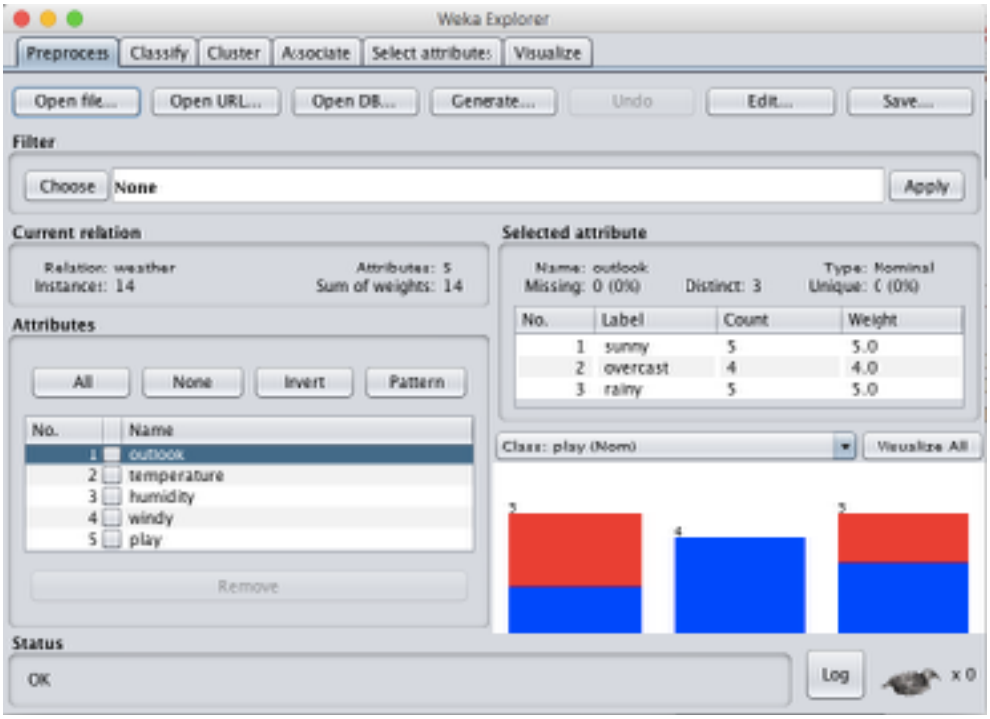
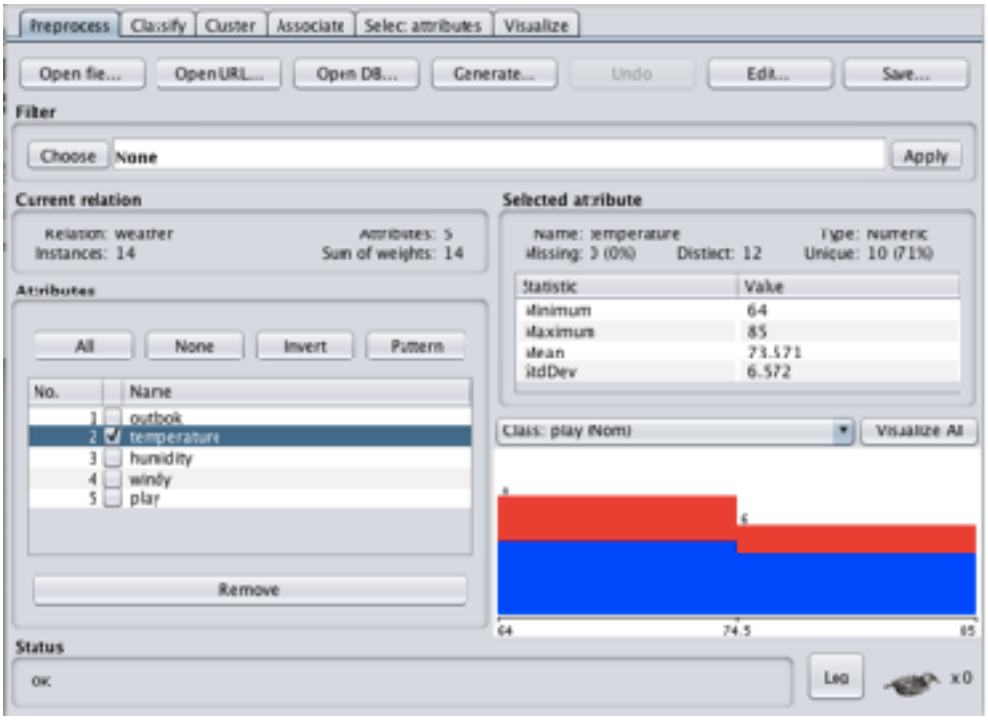
Lab 1: Introduction to the Weka Explorer

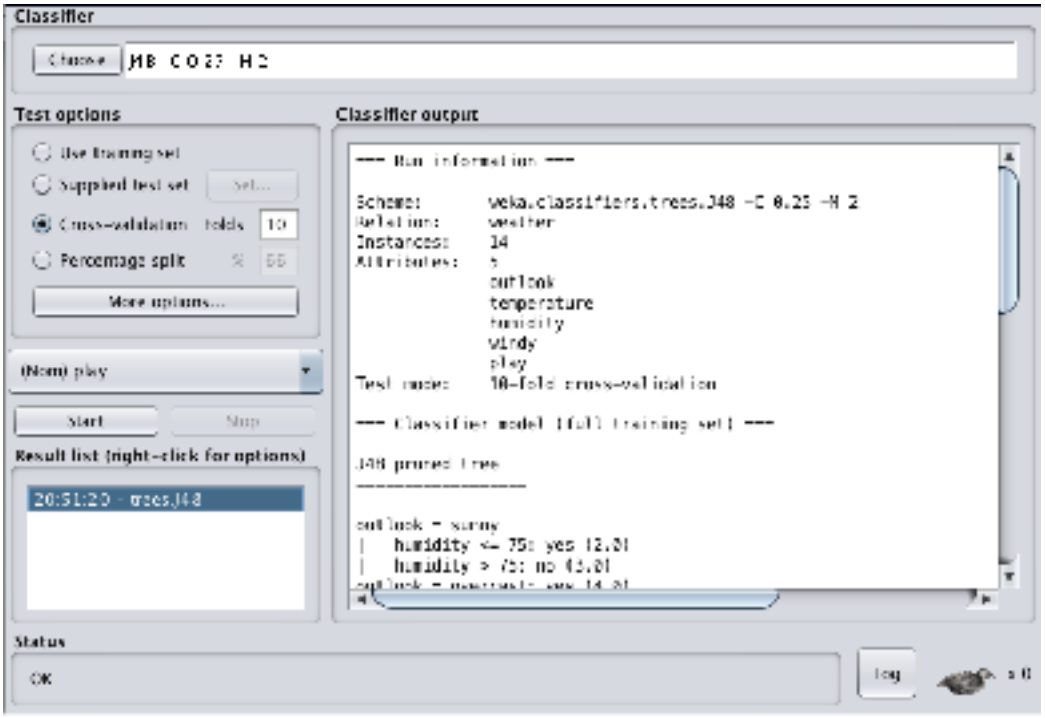
Weather data loaded



Investigating temperature attribute



Run j48 classifier with cross validation



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: 0.02 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	

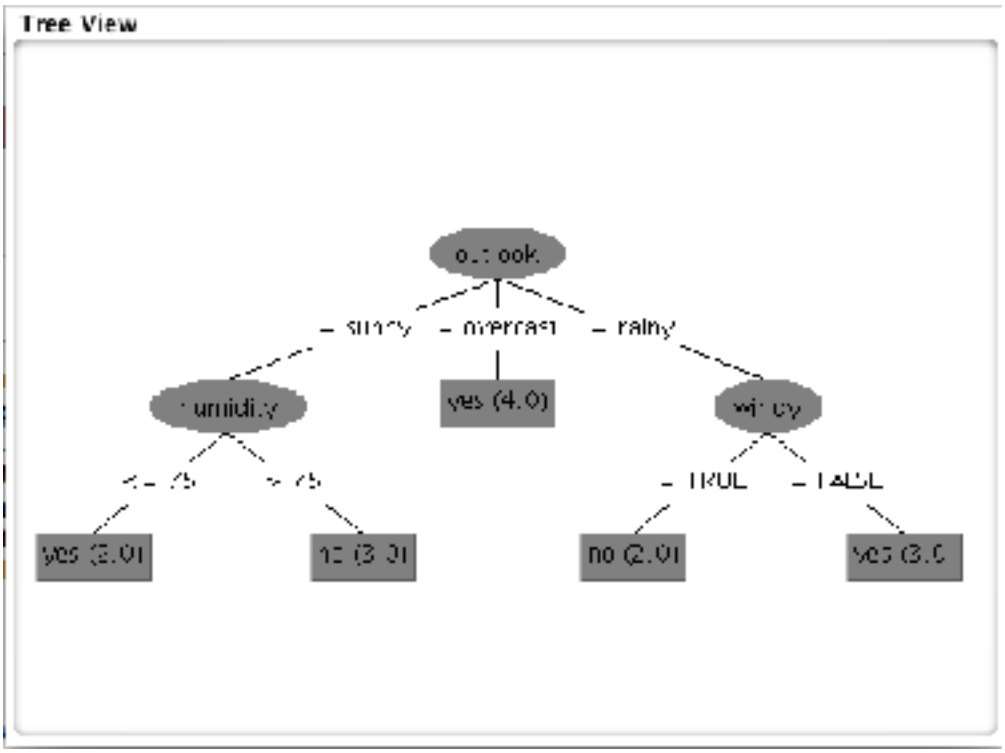
=== Detailed Accuracy By Class ===

	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	l a = yes
3	2	l b = no

Visualized Tree

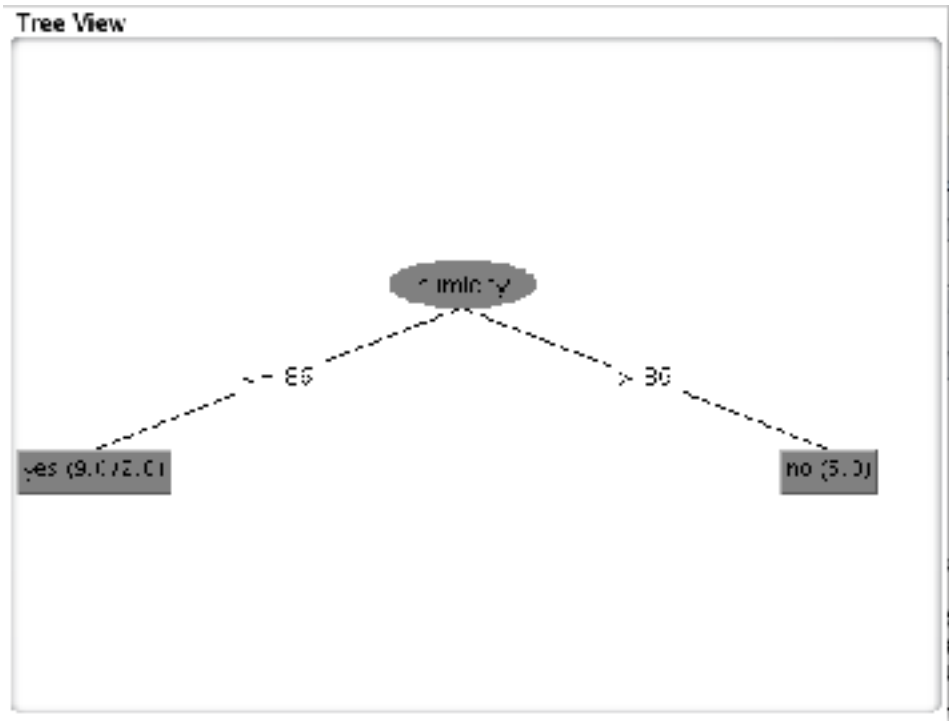


Edit weather data

Relation: weather

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	sunny	72.0	95.0	FALSE	no
4	sunny	69.0	70.0	FALSE	yes
5	sunny	75.0	70.0	TRUE	yes
6	overcast	83.0	86.0	FALSE	yes
7	overcast	64.0	65.0	TRUE	yes
8	overcast	72.0	90.0	TRUE	no
9	overcast	81.0	75.0	FALSE	yes
10	rainy	70.0	96.0	FALSE	no
11	rainy	68.0	80.0	FALSE	yes
12	rainy	65.0	70.0	TRUE	no
13	rainy	75.0	80.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

Run J48 again



New output for “Use training set” :

=== 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: evaluate on training data

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

J48 pruned tree
---

humidity <= 86: yes (9.0/2.0)
humidity > 86: no (5.0)

Number of Leaves : 2

Size of the tree : 3

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	12	85.7143 %
Incorrectly Classified Instances	2	14.2857 %
Kappa statistic	0.7143	
Mean absolute error	0.2222	
Root mean squared error	0.3333	
Relative absolute error	44.4444 %	
Root relative squared error	66.6667 %	
Total Number of Instances	14	

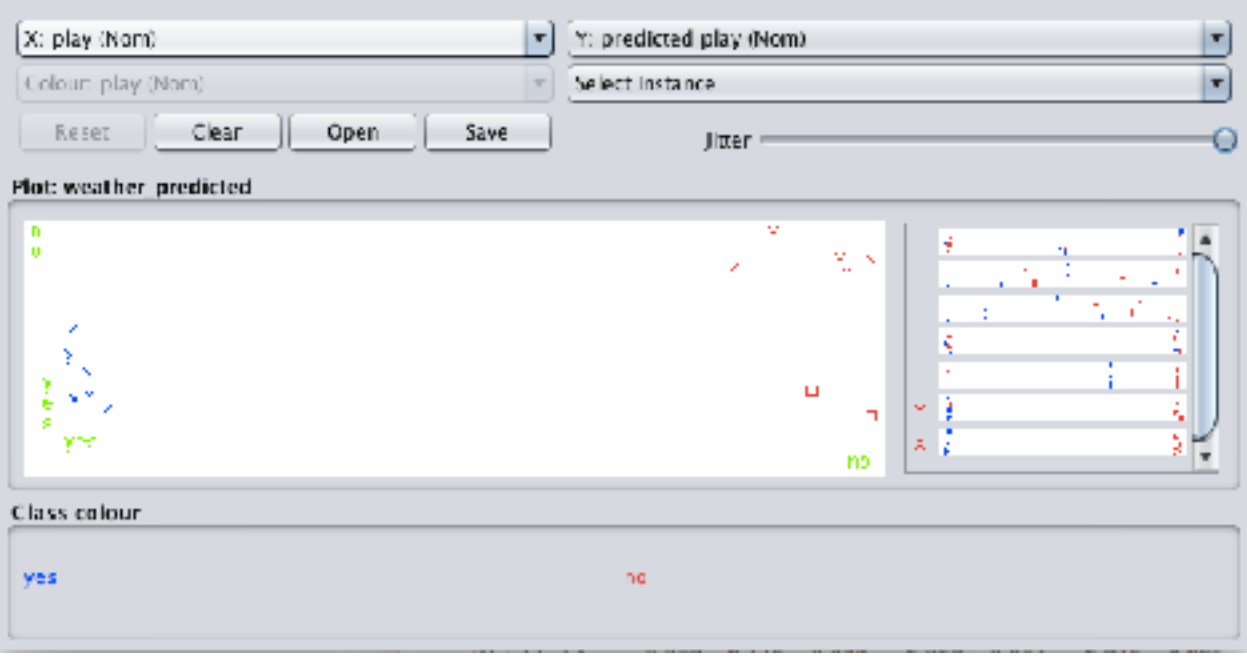
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.286	0.778	1.000	0.875	0.745	0.857	0.778	yes
	0.714	0.000	1.000	0.714	0.833	0.745	0.857	0.857	no
Weighted Avg.	0.857	0.143	0.889	0.857	0.854	0.745	0.857	0.817	

=== Confusion Matrix ===

a	b	<-- classified as
7	0	l a = yes
2	5	l b = no

Visualize classifier errors



Example Instance Info:

Plot : weka.classifiers.trees.J48 (weather)
Instance: 12
outlook : rainy
temperature : 65.0
humidity : 70.0
windy : TRUE
prediction margin : -0.5555555555555556
predicted play : yes
play : no

Re-run steps 2 through 4 on iris.arff

Classifier Output (IRIS)

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: iris

Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
class
Test mode: 10-fold cross-validation

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

J48 pruned tree
---

petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
| petalwidth <= 1.7
| | petallength <= 4.9: Iris-versicolor (48.0/1.0)
| | petallength > 4.9
| | | petalwidth <= 1.5: Iris-virginica (3.0)
| | | petalwidth > 1.5: Iris-versicolor (3.0/1.0)
| petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves : 5

Size of the tree : 9

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances	144	96 %
Incorrectly Classified Instances	6	4 %
Kappa statistic	0.94	
Mean absolute error	0.035	
Root mean squared error	0.1586	
Relative absolute error	7.8705 %	
Root relative squared error	33.6353 %	
Total Number of Instances	150	

=== Detailed Accuracy By Class ===

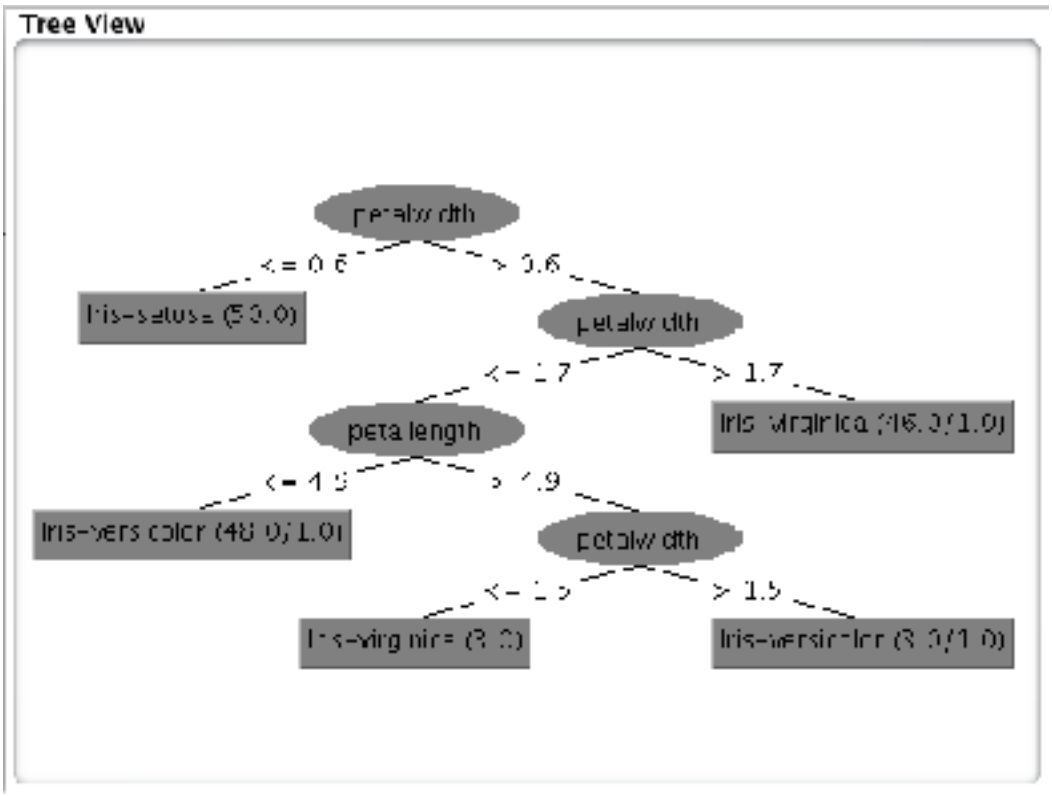
TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.980	0.000	1.000	0.980	0.990	0.985	0.990	0.987	Iris-setosa
0.940	0.030	0.940	0.940	0.940	0.910	0.952	0.880	Iris-versicolor
0.960	0.030	0.941	0.960	0.950	0.925	0.961	0.905	Iris-virginica

Weighted Avg. 0.960 0.020 0.960 0.960 0.960 0.940 0.968 0.924

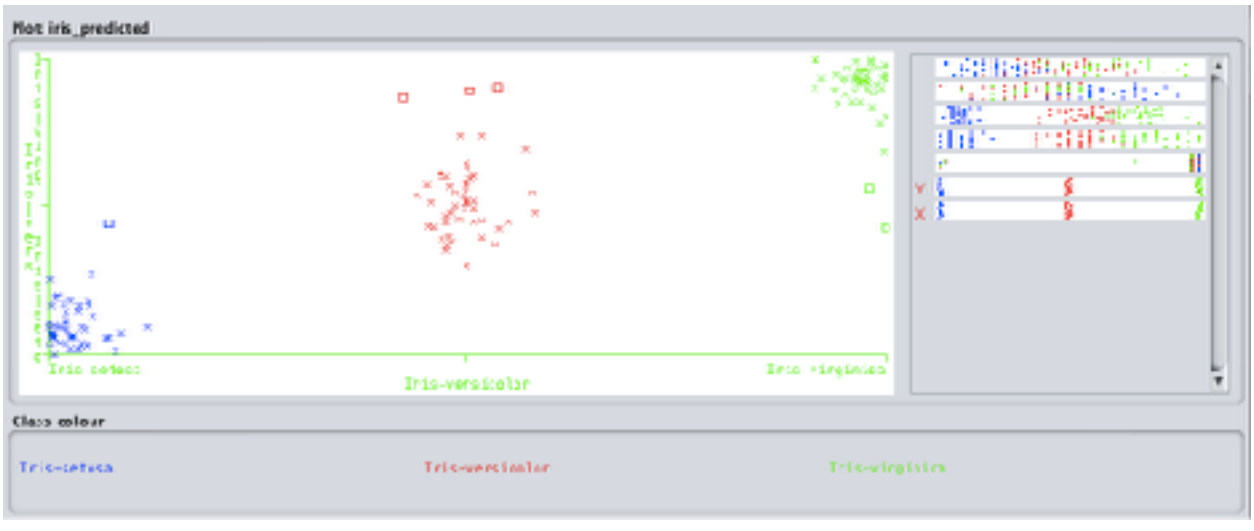
=== Confusion Matrix ===

```
  a  b  c   <-- classified as
 49  1  0 |  a = Iris-setosa
  0 47  3 |  b = Iris-versicolor
  0  2 48 |  c = Iris-virginica
```

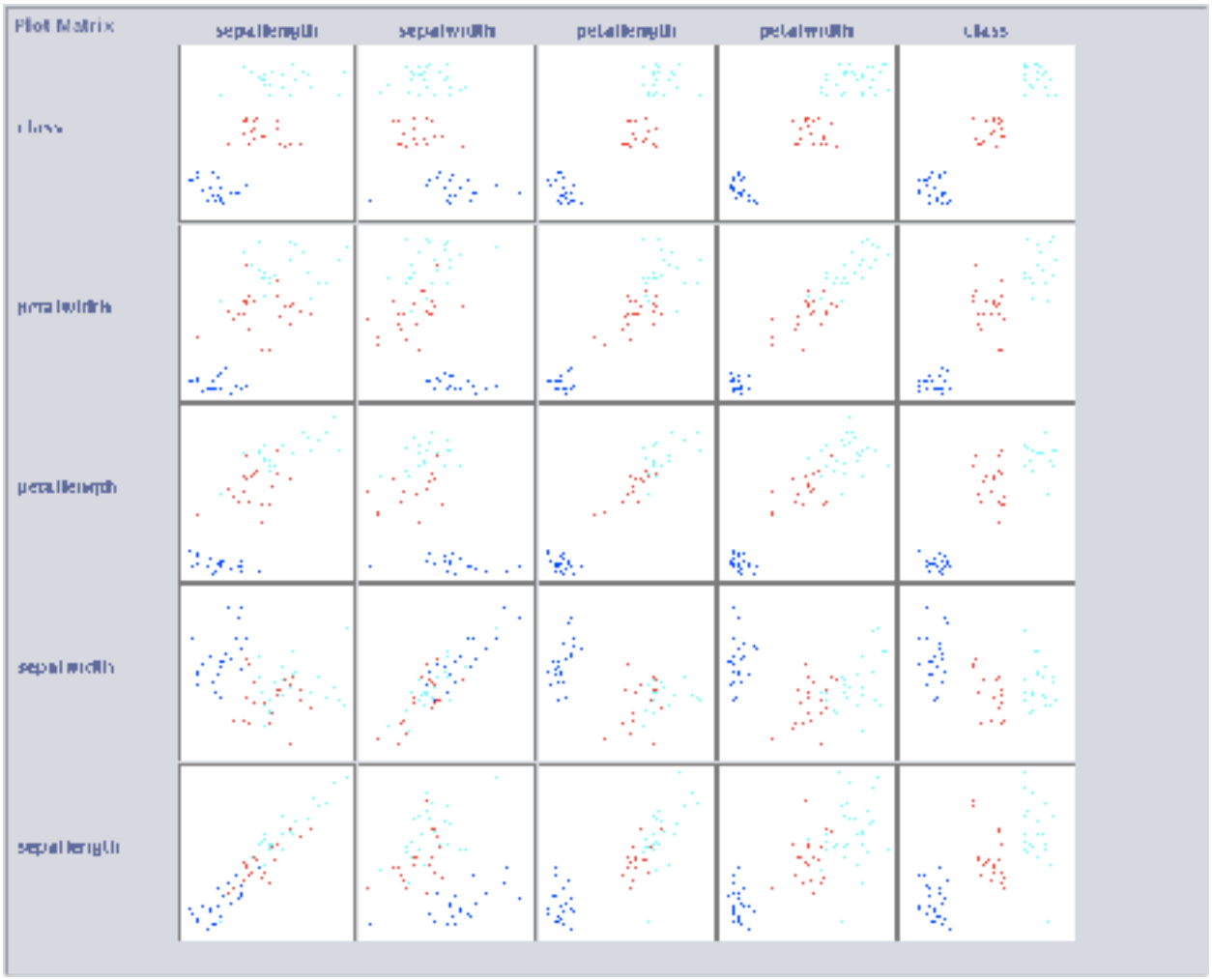
Visualize Tree (IRIS):



Classifier Errors



Plot Matrix:



Hand in:

a.
Separate Well: (petallength, sepallength), (petalwidth,sepallength), (petalwidth, sepalwidth), (petalwidth, sepalwidth), (petalwidth, petallength)

Separate Poor: (sepallength, sepalwidth)

b.
My prediction was good sepallength and sepalwidth are not used in the tree.

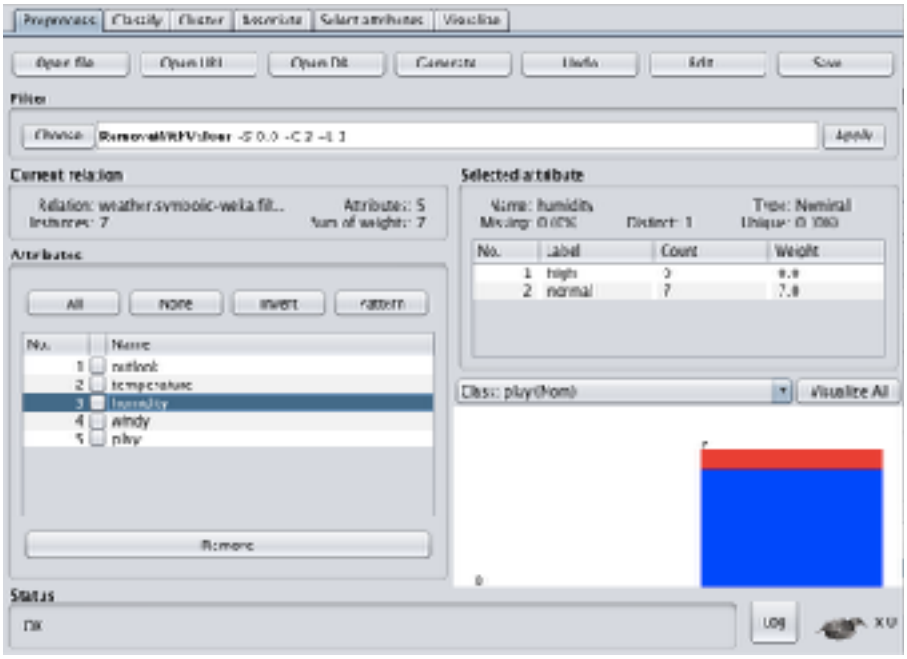
c. hot, mild, cold

d. Yes

e. min: 1
max: 6.9

f. Numeric: 4
Nominal: 1

g.
Filterer implemented



Classifier Output (**see red text**):

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: weather.symbolic-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C3-L1
Instances: 7
Attributes: 5
outlook
temperature
humidity
windy
play
Test mode: evaluate on training data

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

J48 pruned tree

: yes (7.0/1.0)

Number of Leaves : 1

Size of the tree : 1

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	**6**	**85.7143 %**
Incorrectly Classified Instances	1	14.2857 %
Kappa statistic	0	
Mean absolute error	0.2449	
Root mean squared error	0.3499	
Relative absolute error	81.203 %	
Root relative squared error	97.5231 %	
Total Number of Instances	7	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.857	1.000	0.923	0.000	0.500	0.857	yes
	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.143	no
Weighted Avg.	0.857	0.857	0.735	0.857	0.791	0.000	0.500	0.755	

=== Confusion Matrix ===

a	b	<-- classified as
6	0	l a = yes
1	0	l b = no