



DATA WAREHOUSING AND DATA MINING

NAME : MOTHISHWARAN C.

REG No : 19MID0017

COURSE CODE : CSI3010

SLOT : L39+L40

FACULTY : DR. CHELLATHAMILAN T.

DECISION TREE CLASSIFIER USING WEKA

1. Draw Decision Tree using following dataset.

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Creation of arff file:

computerBuy.arff - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

customer.arff x computerBuy.arff x location.arff x

```

1  @relation computerBuy
2  @attribute RID numeric
3  @attribute age {youth,middle_aged,senior}
4  @attribute income {high,medium,low}
5  @attribute student {no,yes}
6  @attribute credits_rating {fair,excellent}
7  @attribute clas:buys_computer {no,yes}
8  @data
9  1,youth,high,no,fair,no
10 2,youth,high,no,excellent,no
11 3,middle_aged,high,no,fair,yes
12 4,senior,medium,no,fair,yes
13 5,senior,low,yes,fair,yes
14 6,senior,low,yes,excellent,no
15 7,middle_aged,low,yes,excellent,yes
16 8,youth,medium,no,fair,no
17 9,youth,low,yes,fair,yes
18 10,senior,medium,yes,fair,yes
19 11,youth,medium,yes,excellent,yes
20 12,middle_aged,medium,no,excellent,yes
21 13,middle_aged,high,yes,fair,yes
22 14,senior,medium,no,excellent,no
23

```

In WEKA:

Viewer

Relation: computerBuy-weka.filters.unsupervised.attribute.Remove-R1

No.	1: age	2: income	3: student	4: credits_rating	5: clas:buys_computer
	Nominal	Nominal	Nominal	Nominal	Nominal
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Add instance Undo OK Cancel

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose None Apply Stop

Current relation

Relation: computerBuy-weka.filters.unsupervised.attribute.Remove-R1 Attributes: 5 Instances: 14 Sum of weights: 14

Attributes

All None Invert Pattern

No.	Name
1	age
2	income
3	student
4	credits_rating
5	clas:buys_computer

Remove

Selected attribute

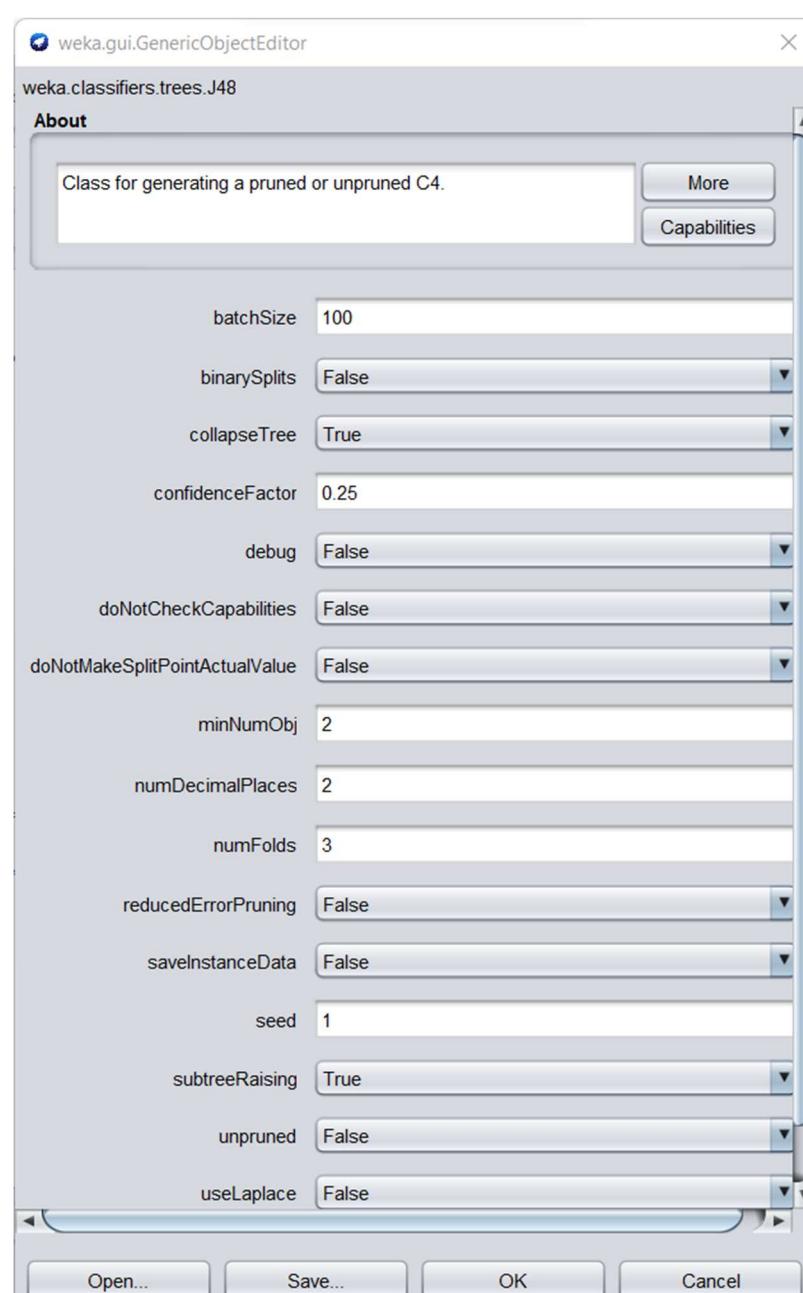
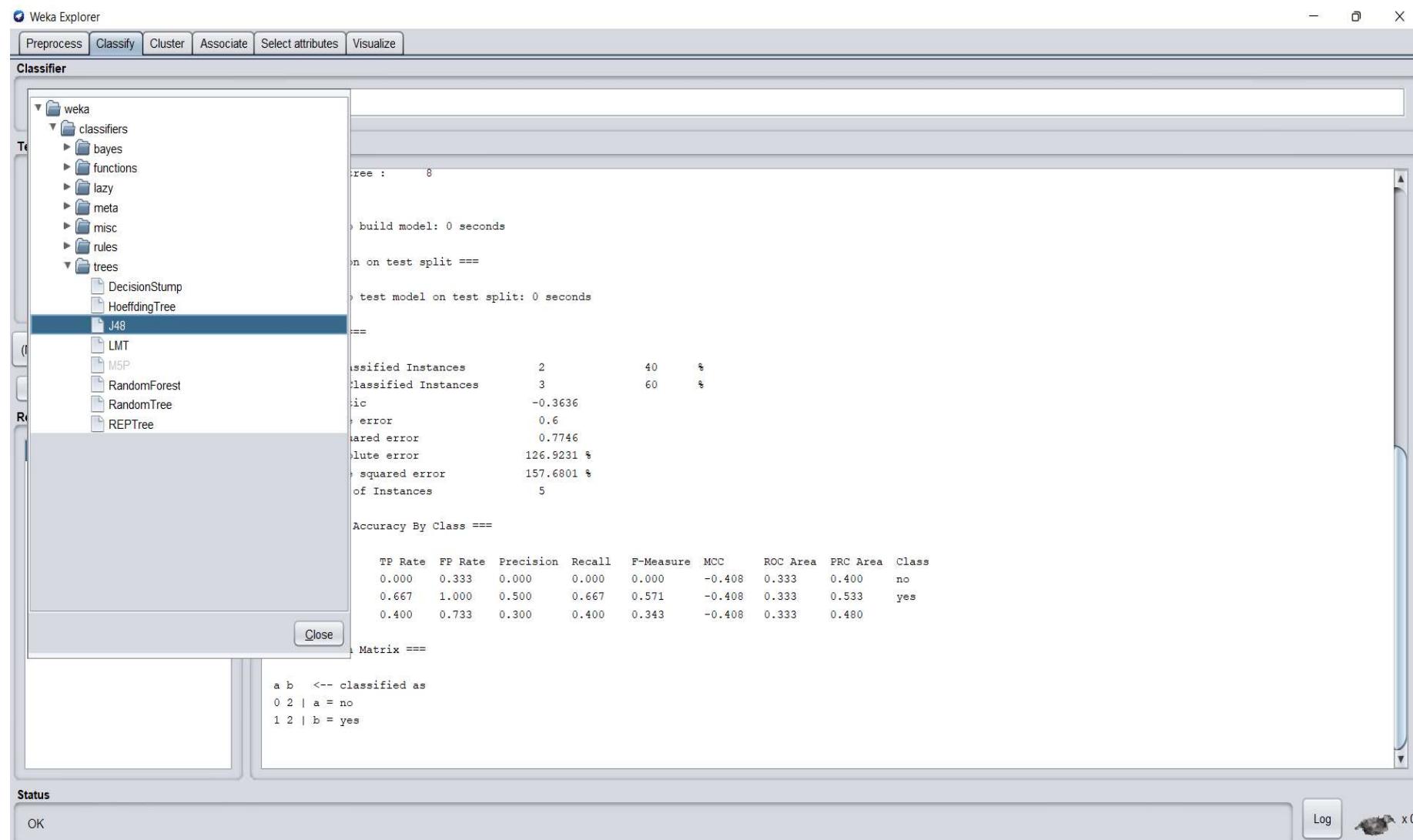
Name: age Missing: 0 (0%) Distinct: 3 Unique: 0 (0%) Type: Nominal

No.	Label	Count	Weight
1	youth	5	5.0
2	middle_aged	4	4.0
3	senior	5	5.0

Class: clas:buys_computer (Nom) Visualize All

Status: OK Log x 0

Classifier:



Classification report: (Train split: 66%)

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 % 66
-

(Nom) clas:buys_computer

Result list (right-click for options)

- 17:28:59 - trees.J48
- 17:38:00 - trees.J48

Classifier output

```
== Run information ==
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: computerBuy-weka.filters.unsupervised.attribute.Remove-R1
Instances: 14
Attributes: 5
age
income
student
credits_rating
clas:buys_computer
Test mode: split 66.0% train, remainder test

== Classifier model (full training set) ==
J48 pruned tree
-----
age = youth
| student = no: no (3.0)
| student = yes: yes (2.0)
age = middle_aged: yes (4.0)
age = senior
| credits_rating = fair: yes (3.0)
| credits_rating = excellent: no (2.0)

Number of Leaves : 5
Size of the tree : 8

Time taken to build model: 0 seconds
== Evaluation on test split ==

```

Status

OK

Log x 0

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 % 66
-

(Nom) clas:buys_computer

Result list (right-click for options)

- 17:28:59 - trees.J48
- 17:38:00 - 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 2 40 %
Incorrectly Classified Instances 3 60 %
Kappa statistic -0.3636
Mean absolute error 0.6
Root mean squared error 0.7746
Relative absolute error 126.9231 %
Root relative squared error 157.6801 %
Total Number of Instances 5

== Detailed Accuracy By Class ==

```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.333	0.000	0.000	0.000	-0.408	0.333	0.400	no
0.667	1.000	0.500	0.667	0.571	-0.408	0.333	0.533	yes
Weighted Avg.	0.400	0.733	0.300	0.400	0.343	-0.408	0.333	0.480

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

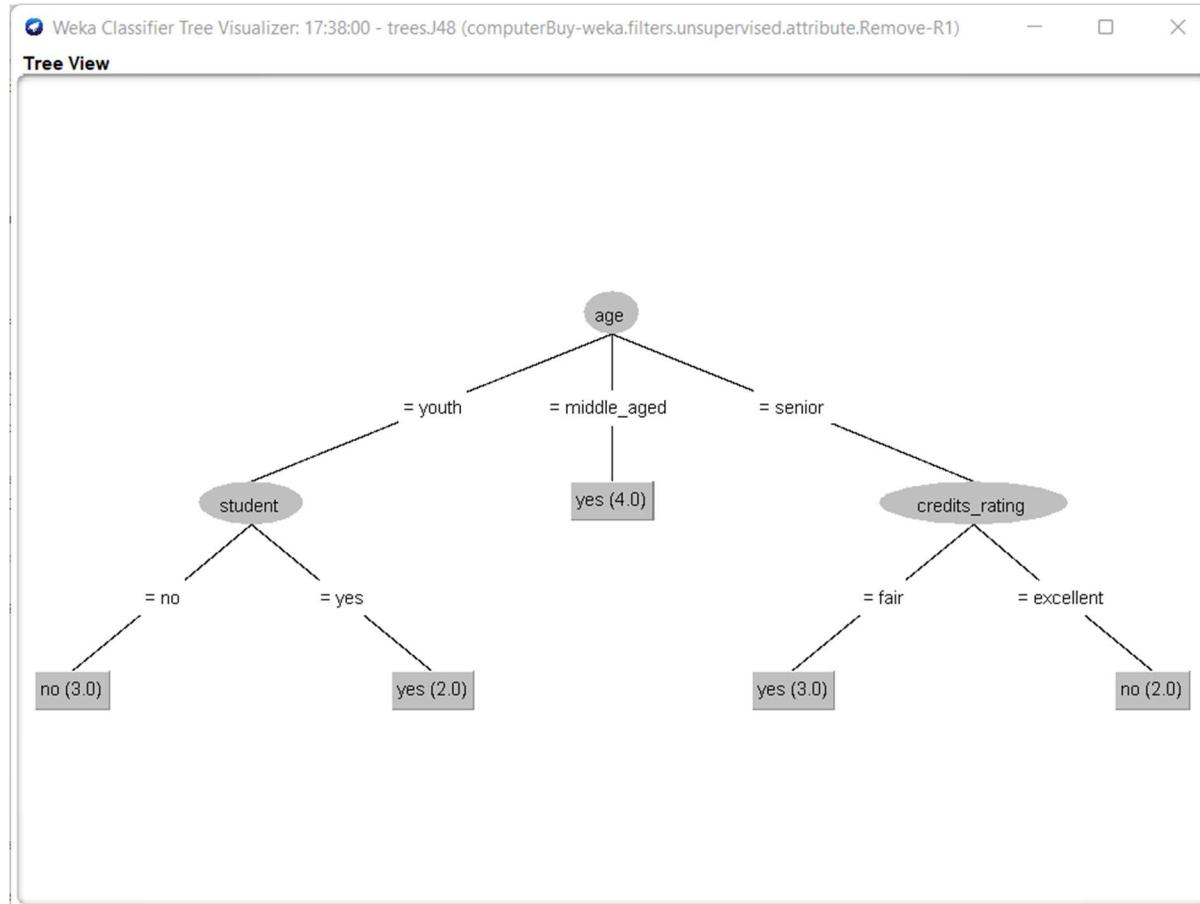
```

Status

OK

Log x 0

Tree:



Instance: age: youth, Income: medium, student: yes, credits rating: excellent.

Prediction: Yes

Actual: Yes

2. Construct the Decision Tree using the following customer dataset

```
customer.arff - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help

customer.arff      ×      computerBuy.arff      ×      location.arff      ×

1 @relation customer
2 @attribute name {x,y,z,u,v,l,w,q,r,n}
3 @attribute age {youth,middle,senior}
4 @attribute income {high,medium,low}
5 @attribute class {A,B}
6 @data
7 x,youth,high,A
8 y,youth,low,B
9 z,middle,high,A
10 u,middle,low,B
11 v,senior,high,A
12 l,senior,low,B
13 w,youth,high,A
14 q,youth,low,B
15 r,middle,high,A
16 n,senior,high,A
17
```

In WEKA:

Viewer

Relation: customer-weka.filters.unsupervised.attribute.Remove-R1

No.	1: age	2: income	3: class
	Nominal	Nominal	Nominal
1	youth	high	A
2	youth	low	B
3	middle	high	A
4	middle	low	B
5	senior	high	A
6	senior	low	B
7	youth	high	A
8	youth	low	B
9	middle	high	A
10	senior	high	A

Add instance Undo OK

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose None Apply Stop

Current relation

Relation: customer-weka.filters.unsupervised.attribute.Remove-R1 Attributes: 3 Instances: 10 Sum of weights: 10

Attributes

No.	Name
1	age
2	income
3	class

Selected attribute

Name	age	Missing	0 (0%)	Distinct	3	Type	Nominal	Unique	0 (0%)
No.	Label			No.	Count				
1	youth			1	4				4.0
2	middle			2	3				3.0
3	senior			3	3				3.0

Class: class (Nom) Visualize All

Status OK Log x 0

Classifier:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

weka classifiers bayes functions lazy meta misc rules trees DecisionStump HoeffdingTree J48 LMT M5P RandomForest RandomTree REPTree

Class for building and using a decision stump

Usually used in conjunction with a boosting algorithm. Does regression (based on mean-squared error) or classification (based on entropy). Missing is treated as a separate value.

CAPABILITIES

Class – Binary class, Date class, Missing class values, Nominal class, Numeric class

Attributes – Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes

Interfaces – Sourcable, WeightedInstancesHandler

Additional

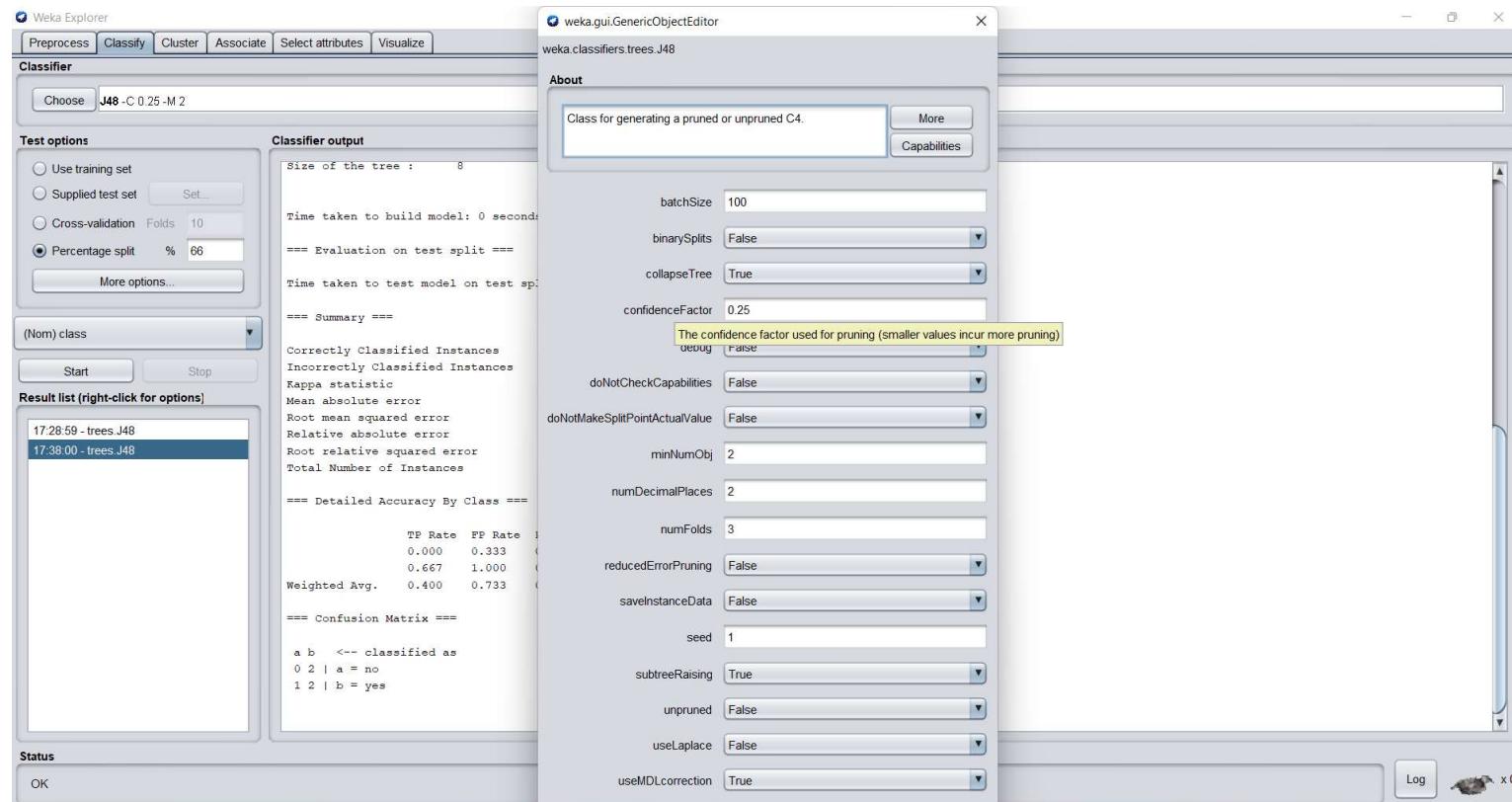
Minimum number of instances: 1

TP	Rate	FP	Rate	Precision	Recall	F-Measure	MCC	ROC Area	FPR Area	Class
0.000		0.333		0.000	0.000	0.000	-0.408	0.333	0.400	no
0.667		1.000		0.500	0.667	0.571	-0.408	0.333	0.533	yes
0.400		0.733		0.300	0.400	0.343	-0.408	0.333	0.480	

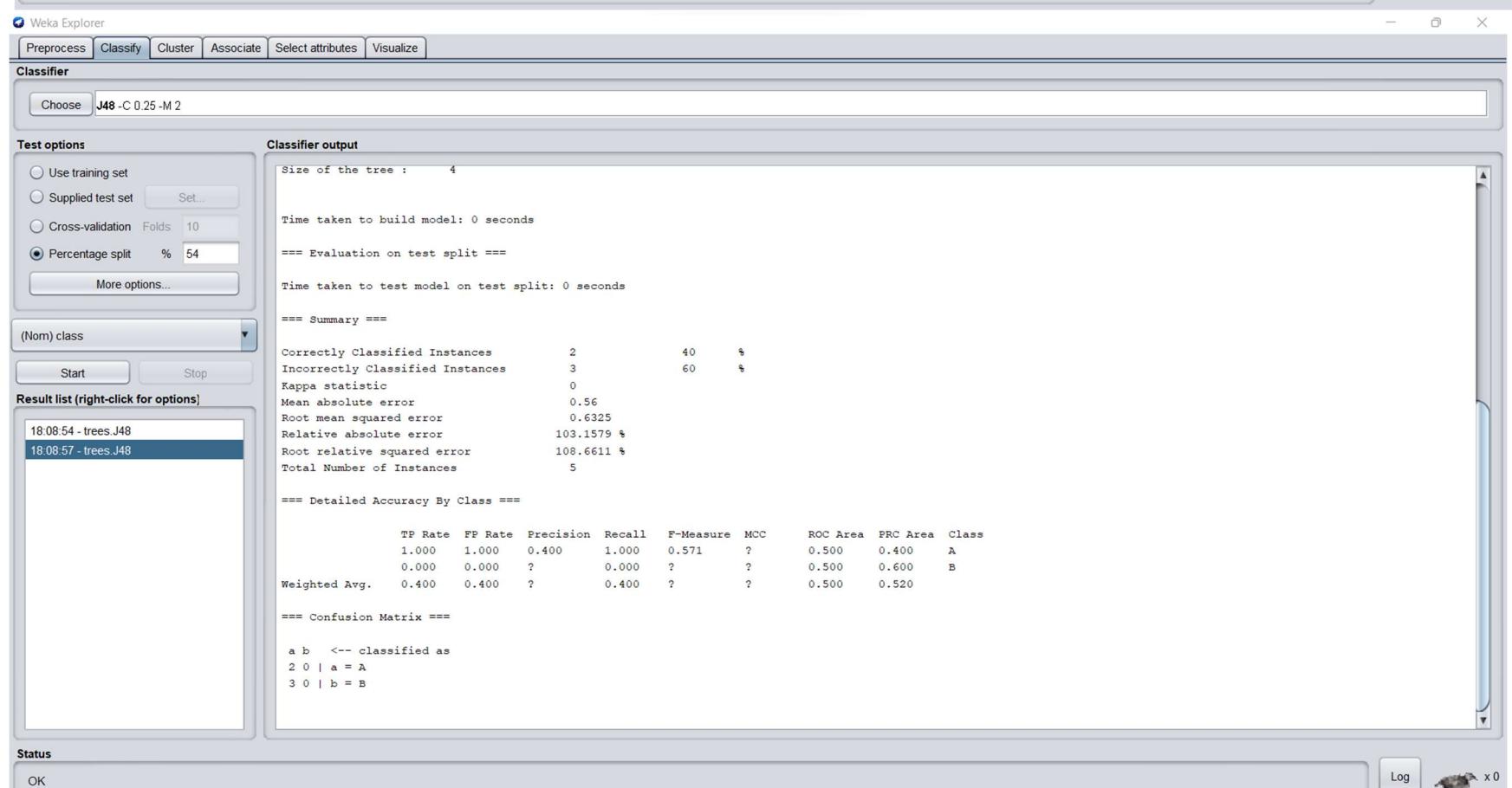
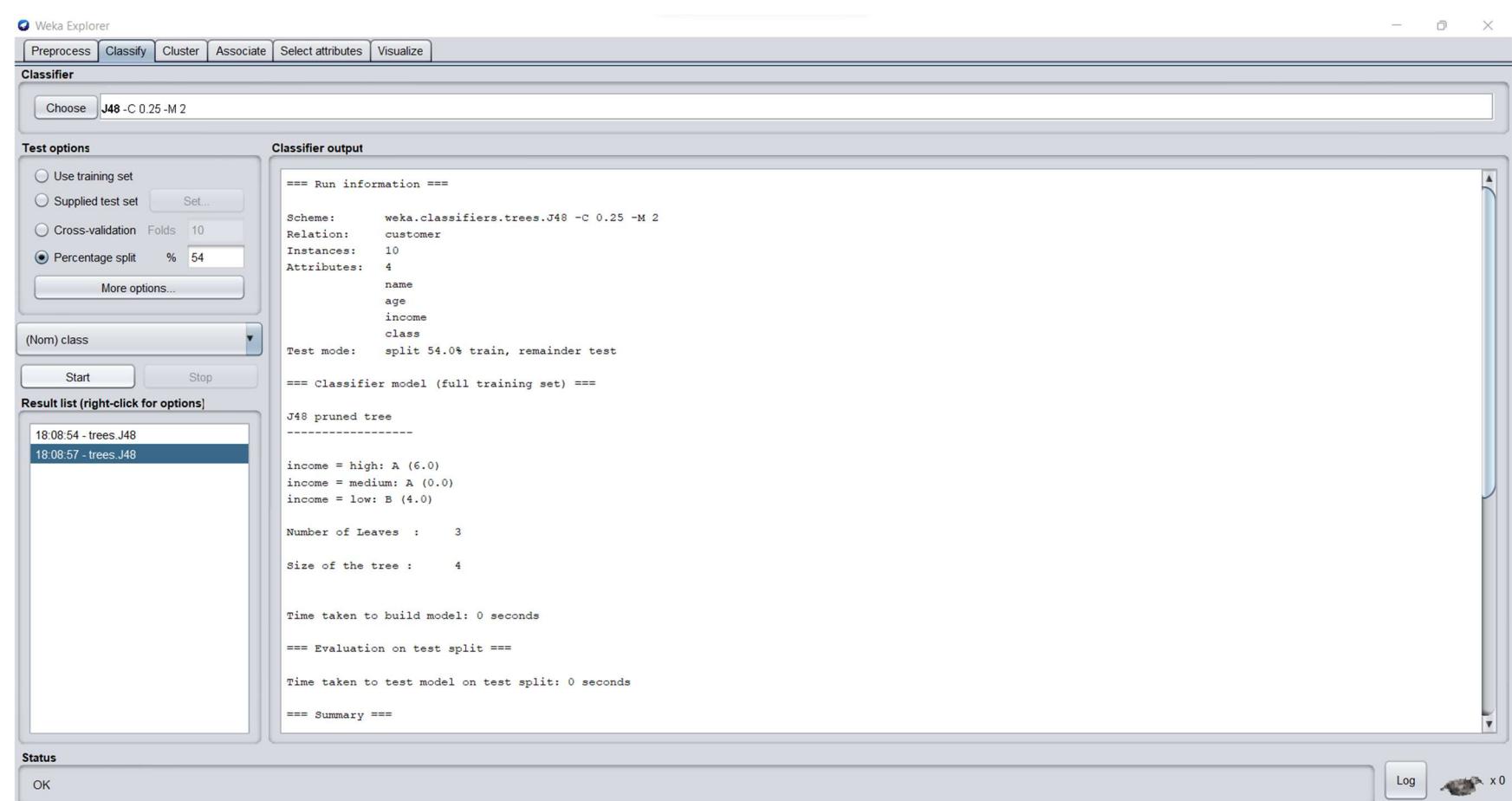
Matrix ==>

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

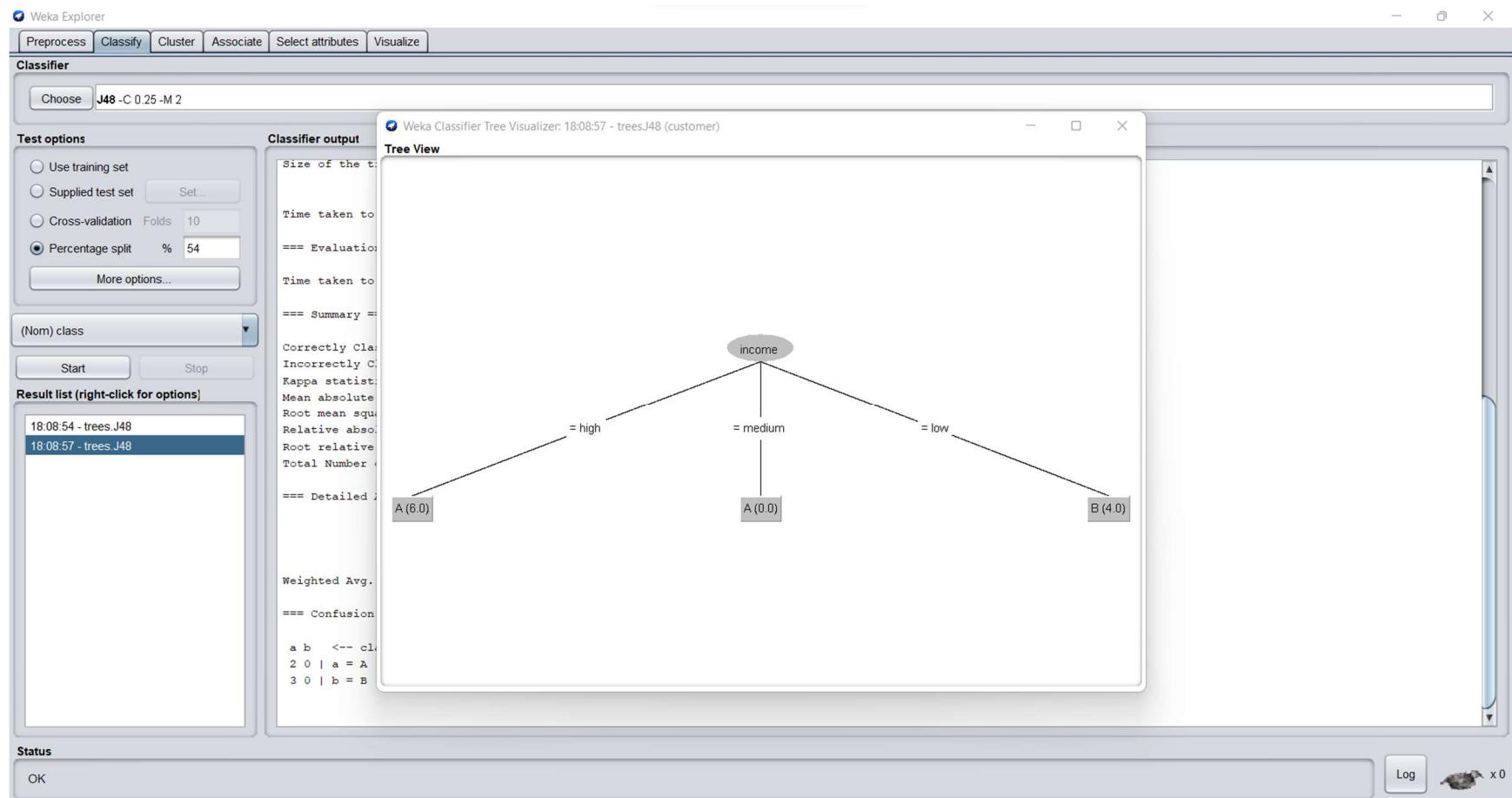
Status OK Log x 0



Classification report: (Train split: 54%)



Tree:



Instance: age: senior, Income: low.

Predicted Class: B

Actual Class: B

3. Construct the Decision Tree using the following location dataset

The screenshot shows a Sublime Text editor window with three tabs: 'customer.arff', 'computerBuy.arff', and 'location.arff'. The 'customer.arff' tab is active and contains the following ARFF code:

```
1 @relation customer
2 @attribute name {x,y,z,u,v,l,w,q,r,n}
3 @attribute age {youth,middle,senior}
4 @attribute income {high,medium,low}
5 @attribute class {A,B}
6 @data
7 x,youth,high,A
8 y,youth,low,B
9 z,middle,high,A
10 u,middle,low,B
11 v,senior,high,A
12 l,senior,low,B
13 w,youth,high,A
14 q,youth,low,B
15 r,middle,high,A
16 n,senior,high,A
17
```

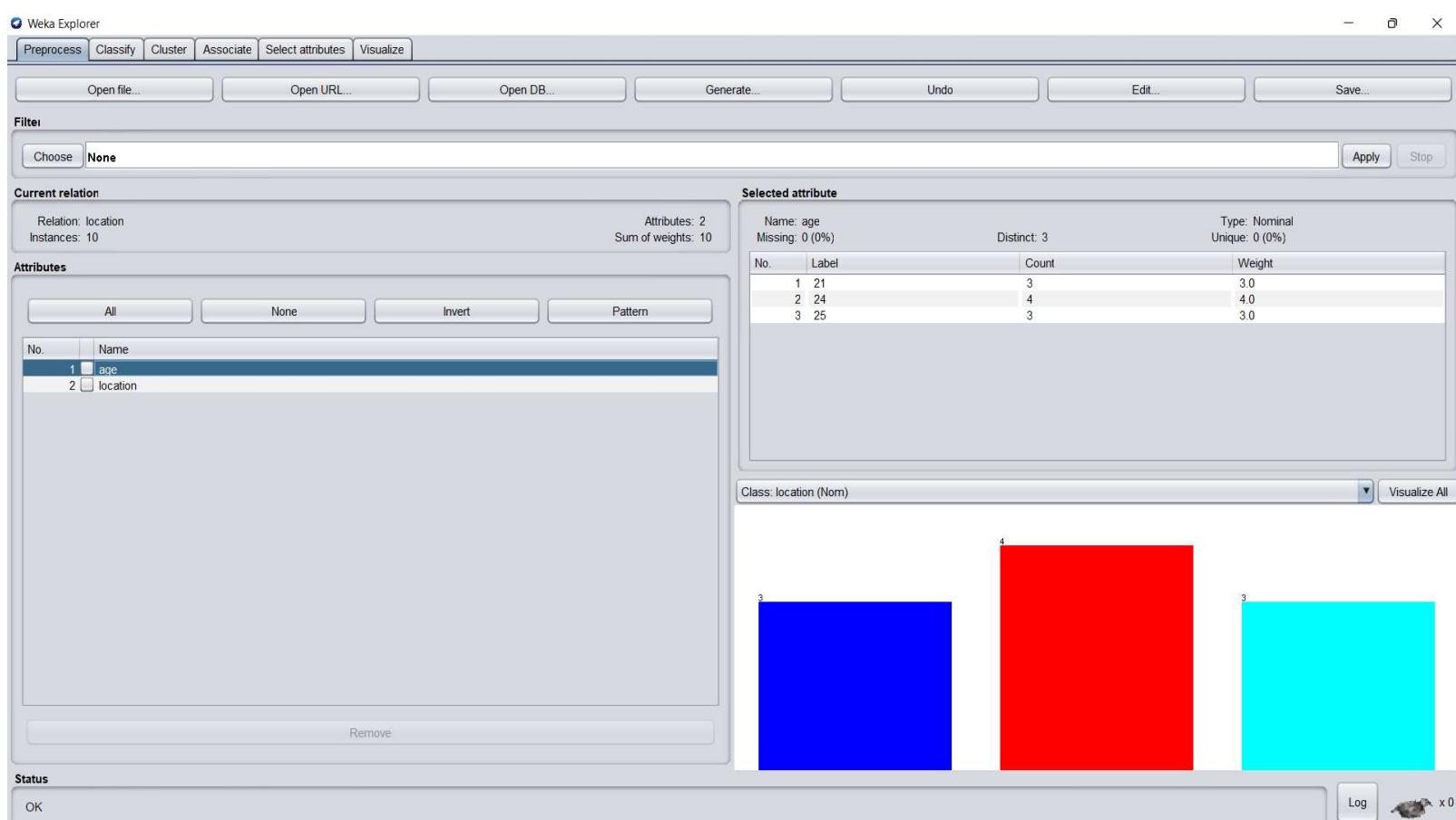
In WEKA:

Viewer

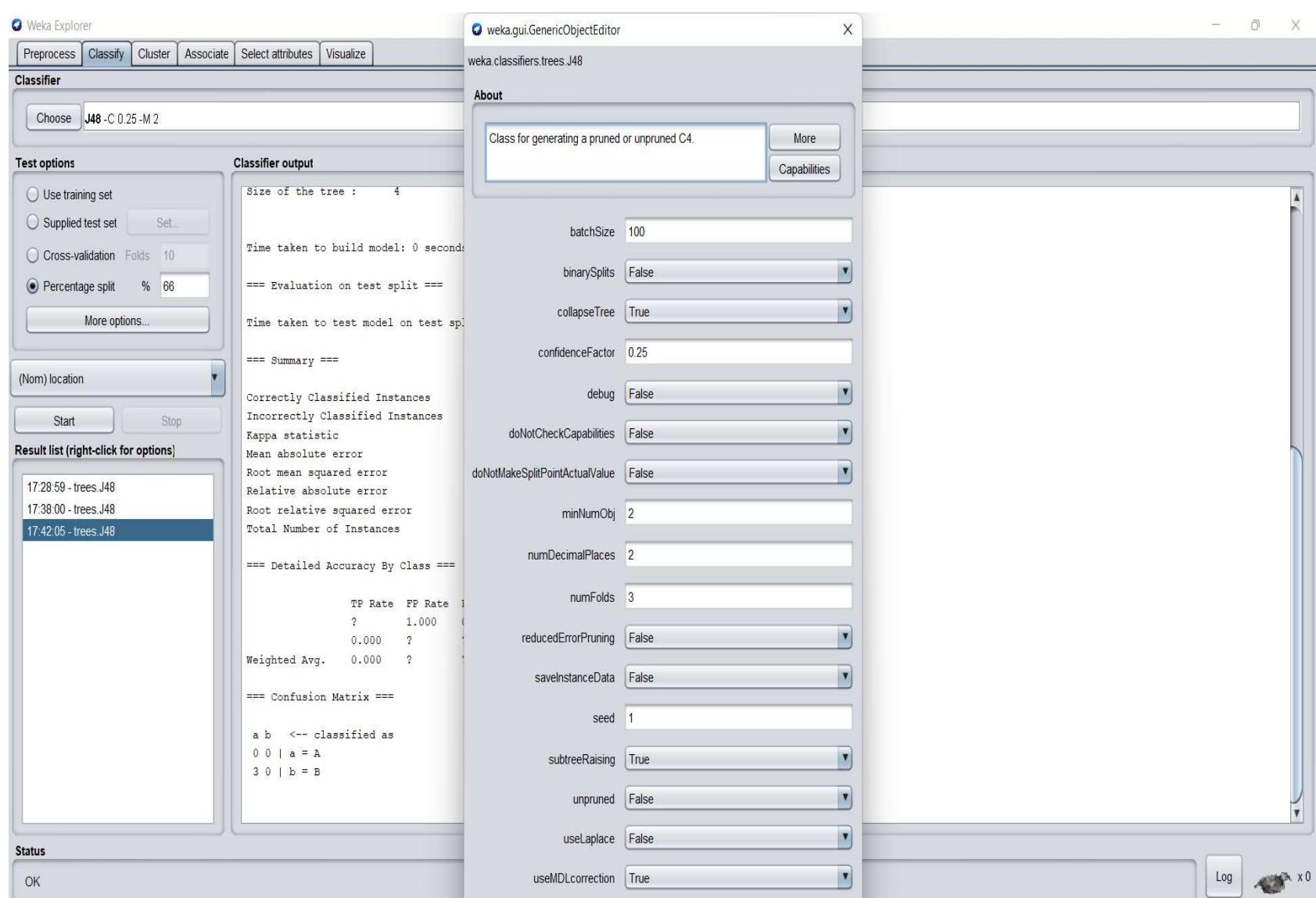
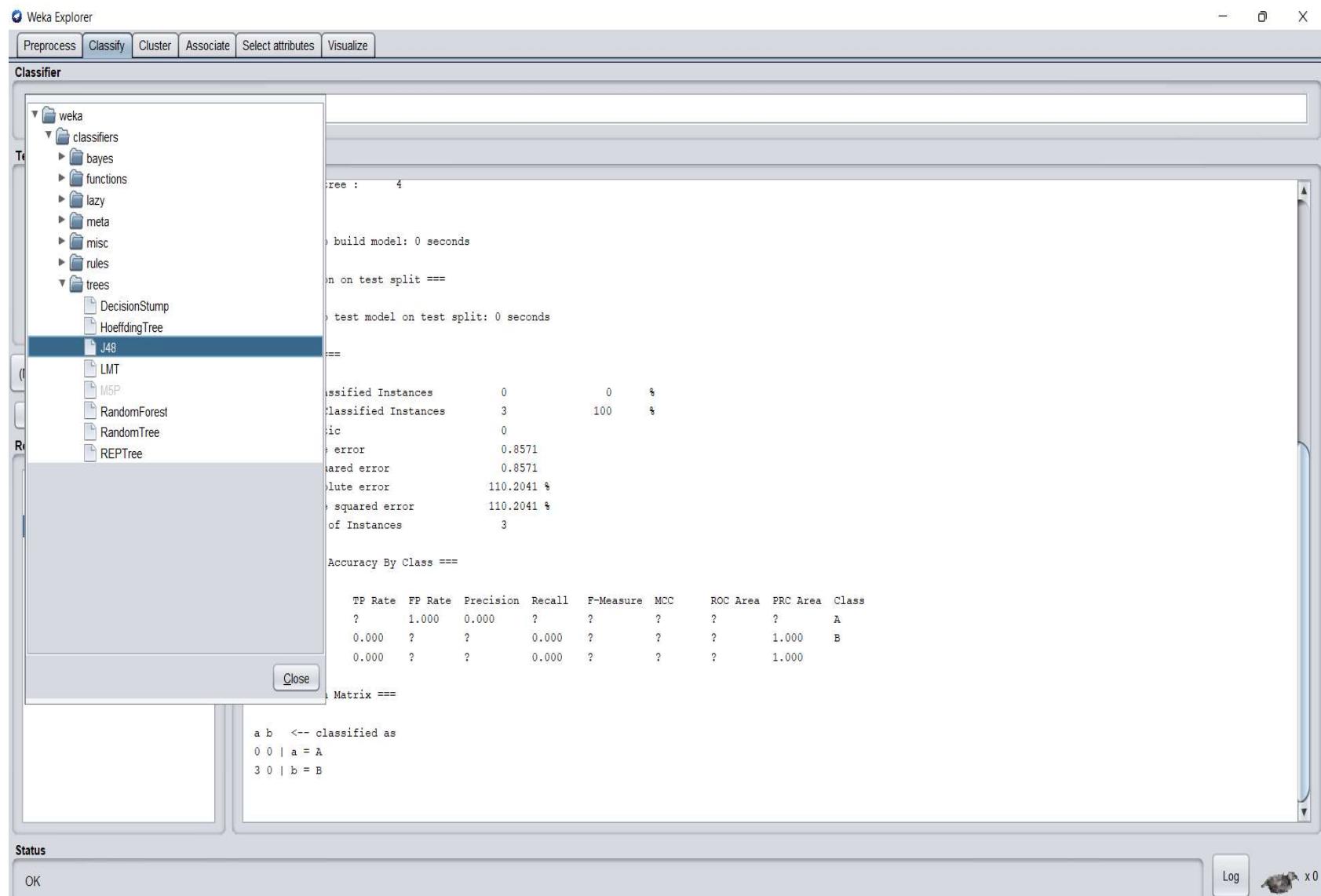
Relation: location

No.	1: age	2: location
	Nominal	Nominal
1	21	hyd
2	21	hyd
3	24	blr
4	24	blr
5	24	blr
6	24	blr
7	21	hyd
8	25	kdp
9	25	kdp
10	25	kdp

Add instance Undo



Classifier:



Classification report: (Train split: 66%)

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 % 66

More options...

(Nom) location

Start Stop

Result list (right-click for options)

- 17:28:59 - trees.J48
- 17:38:00 - trees.J48
- 17:42:05 - trees.J48
- 17:46:25 - trees.J48**

Classifier output

```
==== Run information ====
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: location
Instances: 10
Attributes: 2
age
location
Test mode: split 66.0% train, remainder test
==== Classifier model (full training set) ====
J48 pruned tree
-----
age = 21: hyd (3.0)
age = 24: blr (4.0)
age = 25: kdp (3.0)

Number of Leaves : 3
Size of the tree : 4

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 3 100 %
Incorrectly Classified Instances 0 0 %

```

Status

OK Log x 0

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 % 66

More options...

(Nom) location

Start Stop

Result list (right-click for options)

- 17:28:59 - trees.J48
- 17:38:00 - trees.J48
- 17:42:05 - trees.J48
- 17:46:25 - trees.J48**

Classifier output

```
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 3 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0 %
Total Number of Instances 3

==== Detailed Accuracy By Class ====


|               | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| 1.000         | 0.000   | 1.000   | 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    | hyd   |
| 1.000         | 0.000   | 1.000   | 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    | blr   |
| ?             | 0.000   | ?       | ?         | ?      | ?         | ?     | ?        | ?        | kdp   |
| Weighted Avg. | 1.000   | 0.000   | 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    |       |

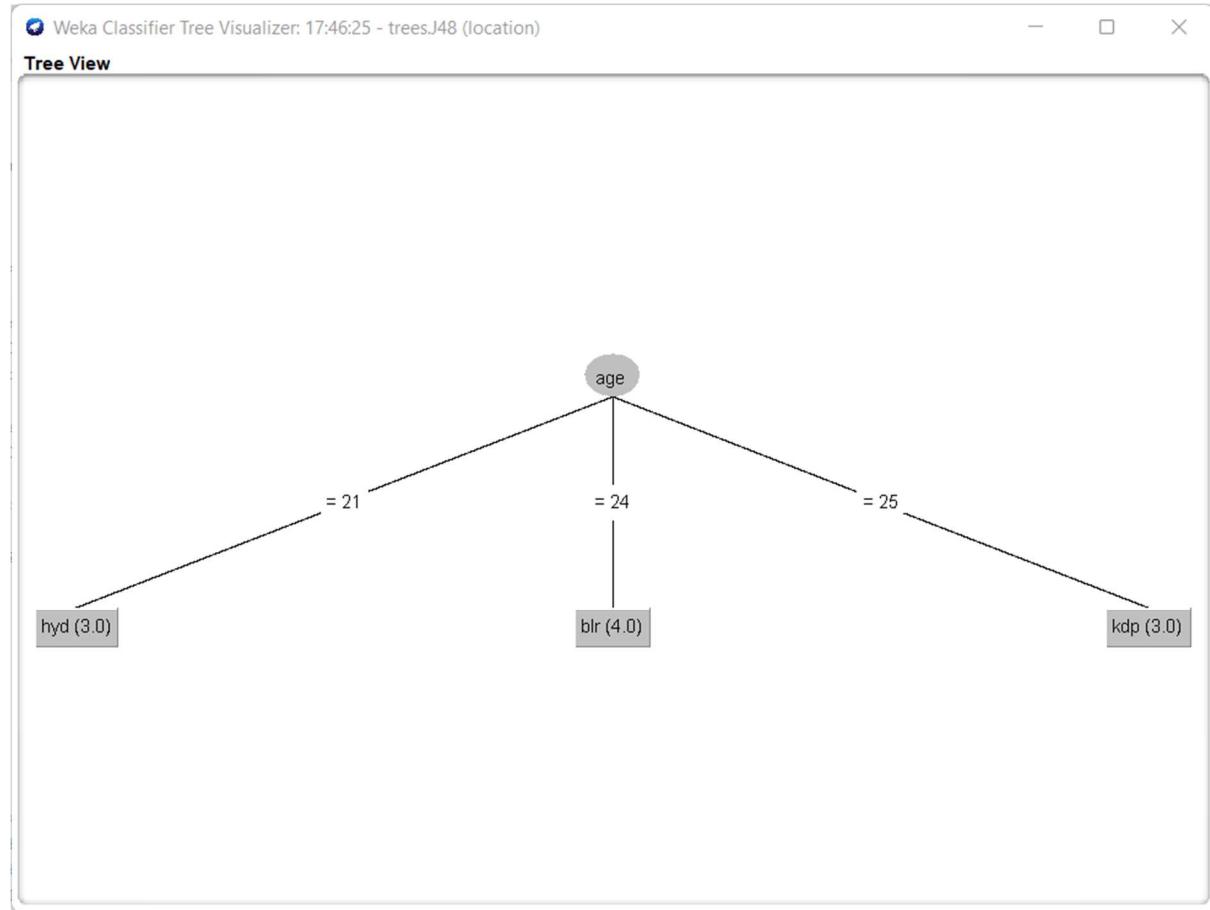

==== Confusion Matrix ====
a b c <-- classified as
1 0 0 | a = hyd
0 2 0 | b = blr
0 0 1 | c = kdp

```

Status

OK Log x 0

Tree:



Instance: age: 24.

Predicted Class: blr

Actual Class: blr