VIT®

UNIVERSITY

(Estd. u/s 3 of UGC Act 1956)

Vellore-632 014, Tamil Nadu, India.

www.vit.ac.in

S.THARUN

19MID0031

CSI3010 – DATA WAREHOUSING AND DATA MINING

FACULTY : CHELLATAMILAN T

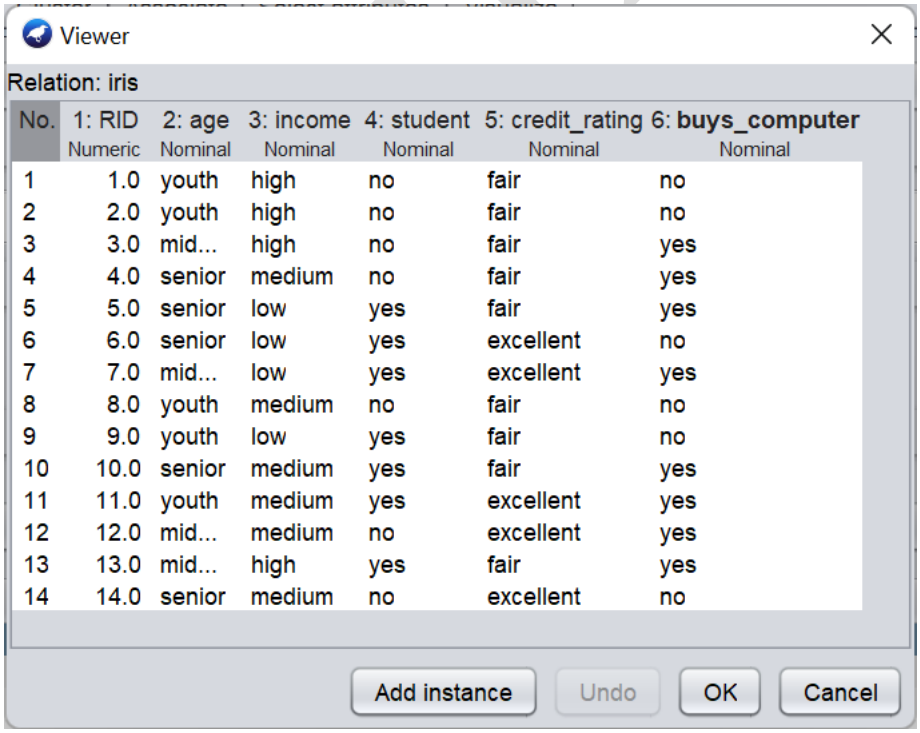
01-02-2022

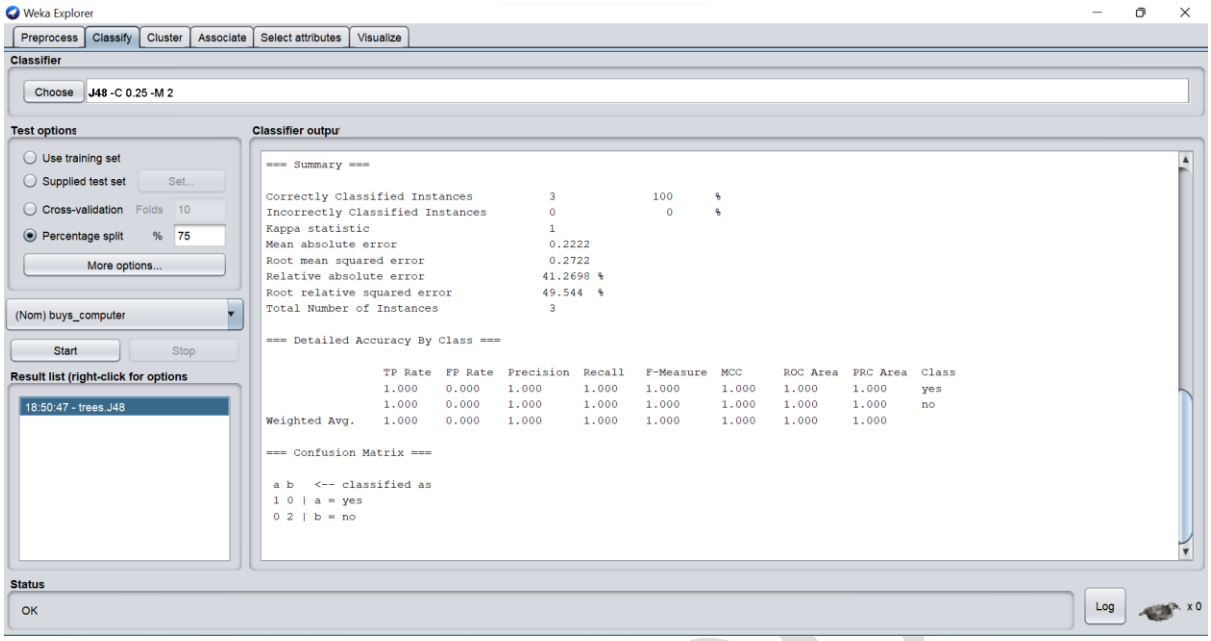
CONSTRUCTING AND EVALUATING DECISION TREES

1. Draw Decision Tree using following dataset Dataset.

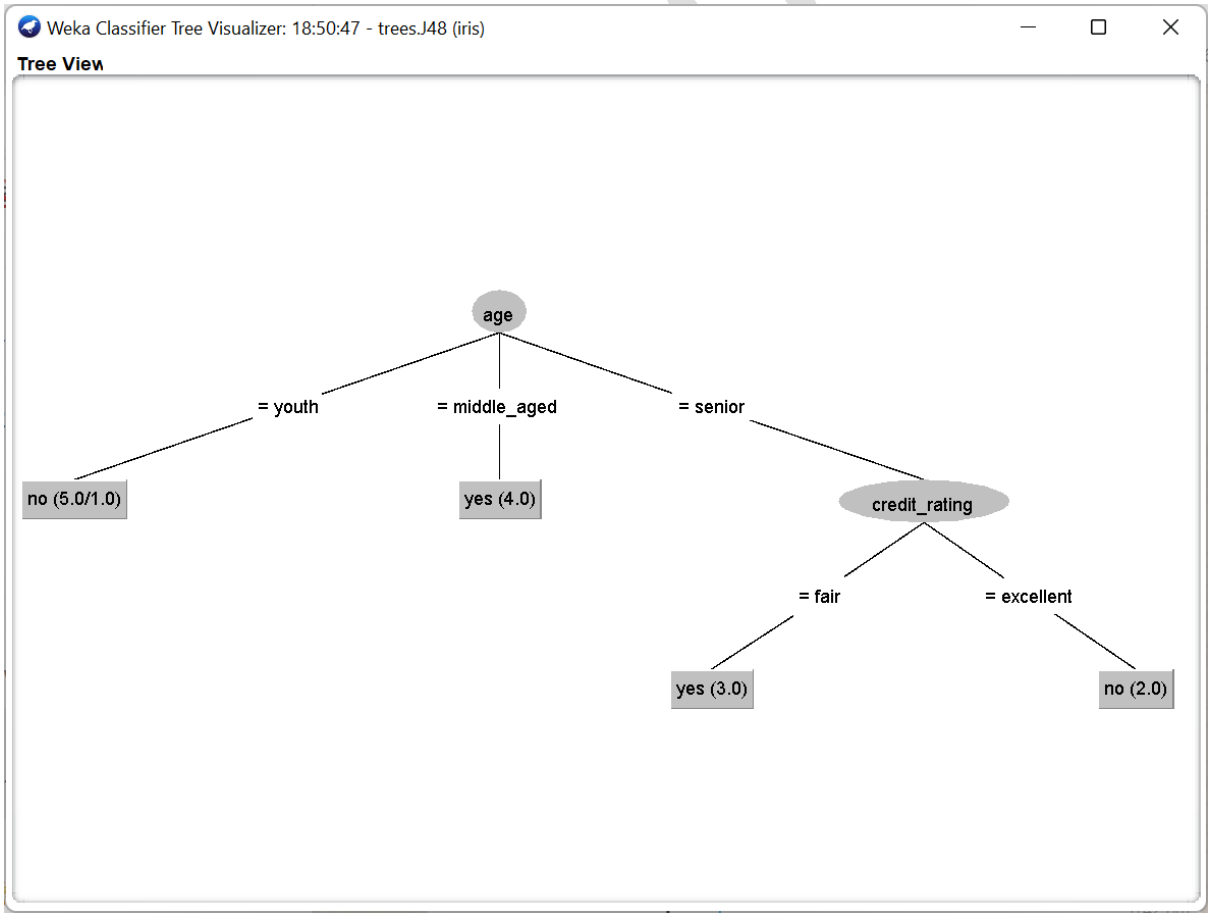
RID	*age*	*income*	*student*	*credit_rating*	*Class: buys_computer*
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth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

COMPUTER DATA



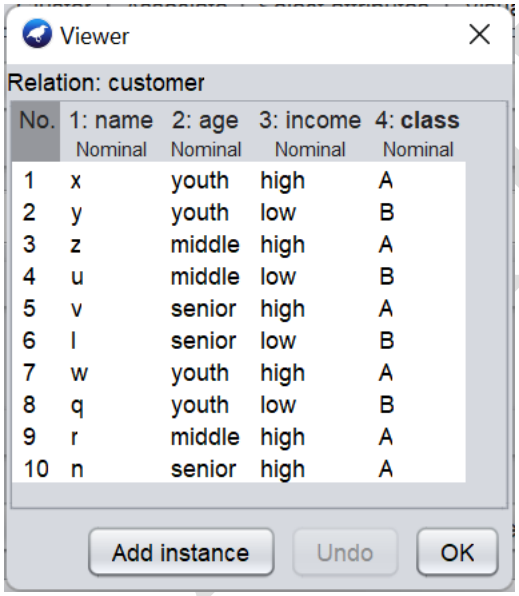



```
=== Summary ===

Correctly Classified Instances           3               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.2222
Root mean squared error                  0.2722
Relative absolute error                 41.2698 %
Root relative squared error             49.544  %
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     yes
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     no
Weighted Avg.  1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000

=== Confusion Matrix ===

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

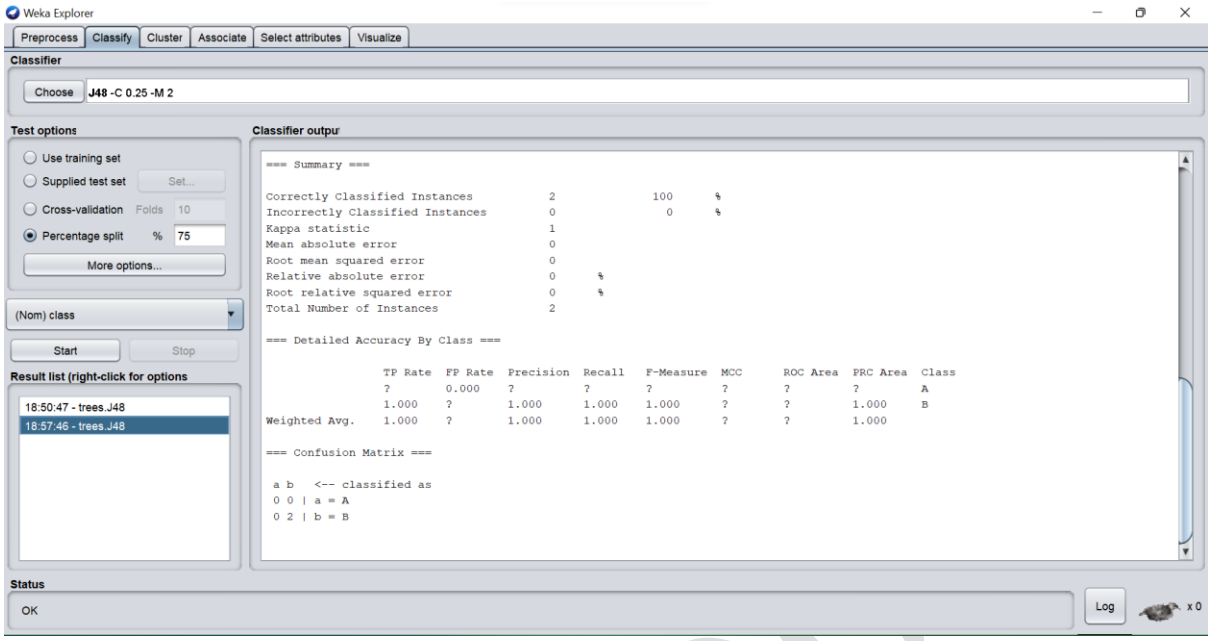


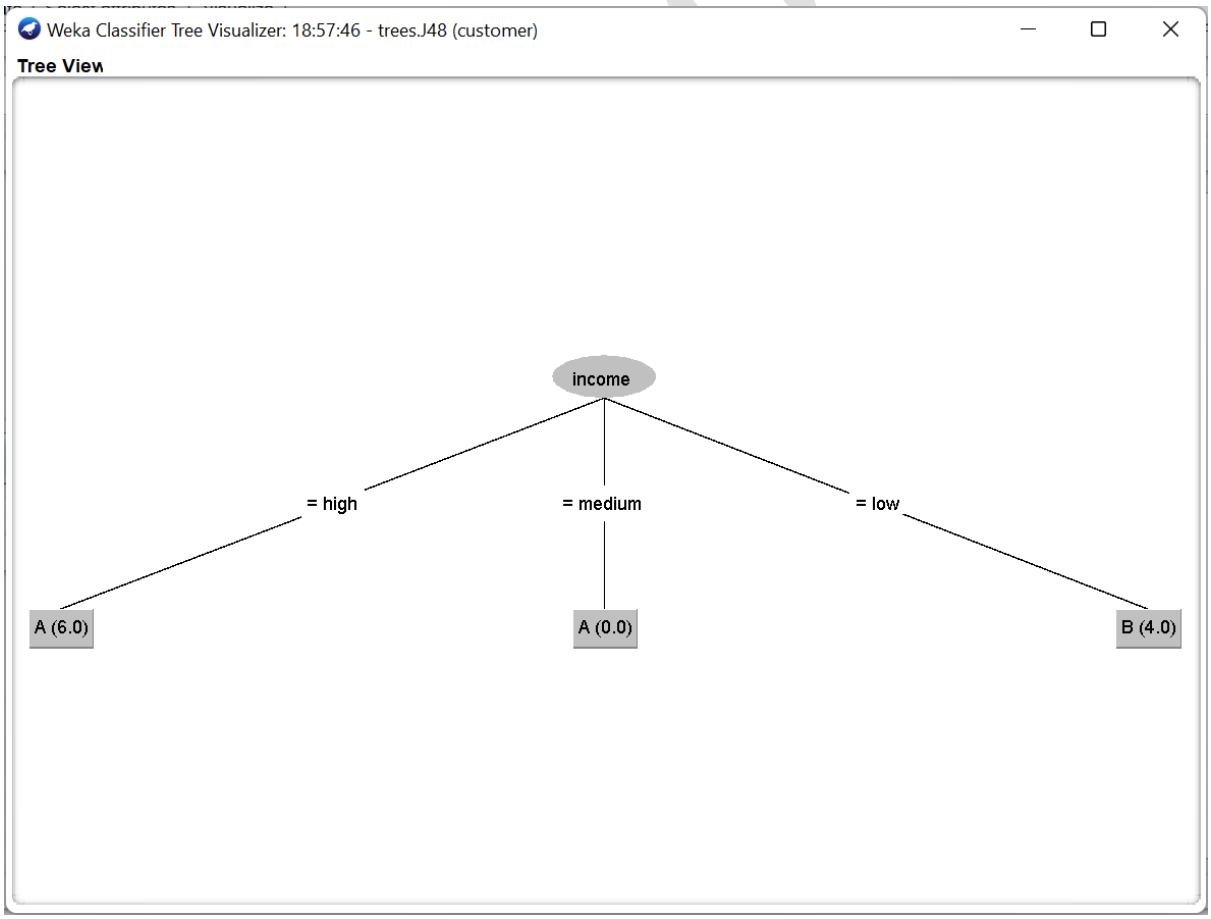
3. Construct the Decision Tree using the following customer dataset

@relation customer
@attribute name {x,y,z,u,v,l,w,q,r,n}
@attribute age {youth,middle,senior}
@attribute income {high,medium,low}
@attribute class {A,B}
@data
x,youth,high,A
y,youth,low,B
z,middle,high,A
u,middle,low,B
v,senior,high,A
l,senior,low,B
w,youth,high,A
q,youth,low,B
r,middle,high,A
n,senior,high,A



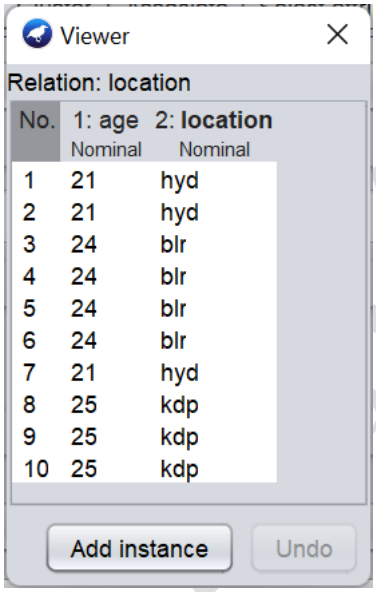
No.	1: name	2: age	3: income	4: class
	Nominal	Nominal	Nominal	Nominal
1	x	youth	high	A
2	y	youth	low	B
3	z	middle	high	A
4	u	middle	low	B
5	v	senior	high	A
6	l	senior	low	B
7	w	youth	high	A
8	q	youth	low	B
9	r	middle	high	A
10	n	senior	high	A

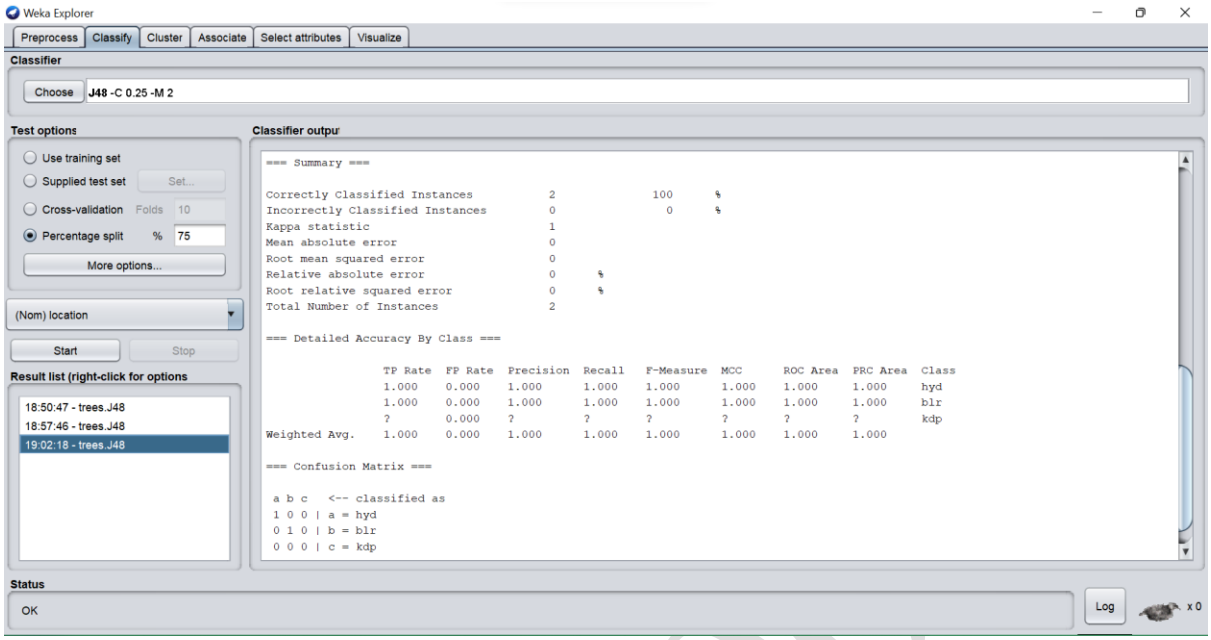

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
75
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options
18:50:47 - trees.J48
18:57:46 - trees.J48
=== Summary ===
Correctly Classified Instances 2 100 %
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 2
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
? 0.000 ? ? ? ? ? ? A
1.000 ? 1.000 1.000 1.000 ? ? 1.000 B
Weighted Avg. 1.000 ? 1.000 1.000 1.000 ? ? 1.000
=== Confusion Matrix ===
a b <-- classified as
0 0 | a = A
0 2 | b = B
Status
OK
Log
x 0

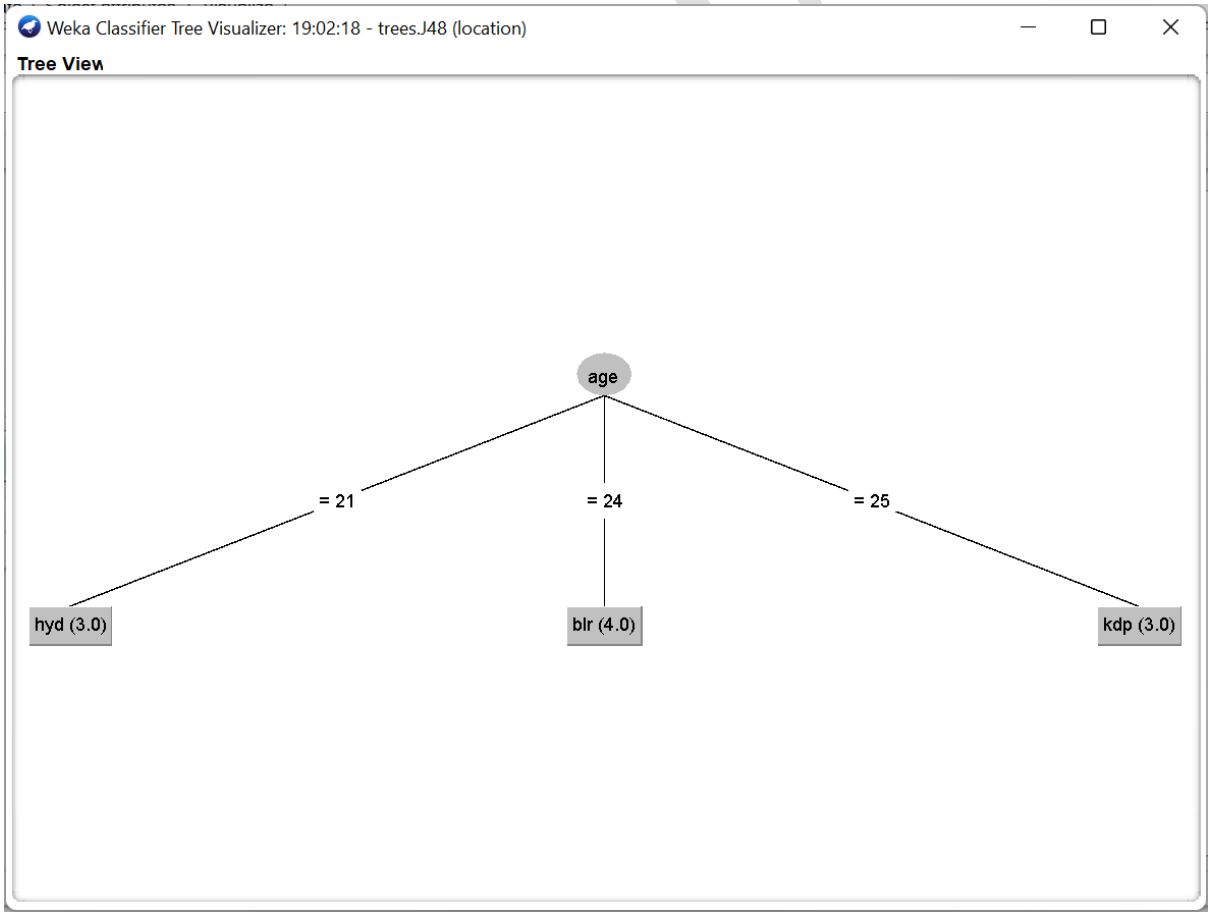

Weka Classifier Tree Visualizer: 18:57:46 - trees.J48 (customer)
Tree View
income
= high
= medium
= low
A (6.0)
A (0.0)
B (4.0)

Other Dataset : Location

```
@relation location
@attribute age {21,24,25}
@attribute location {hyd,blr,kdp}
@data
21,hyd
21,hyd
24,blr
24,blr
24,blr
24,blr
21,hyd
25,kdp
25,kdp
25,kdp
```




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
%
75
More options...
(Nom) location
Start
Stop
Result list (right-click for options
18:50:47 - trees.J48
18:57:46 - trees.J48
19:02:18 - trees.J48
Classifier output
=== Summary ===
Correctly Classified Instances 2 100 %
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 2
=== 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 hyd
1.000 0.000 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 1 0 | b = blr
0 0 0 | c = kdp
Status
OK
Log
x 0


Weka Classifier Tree Visualizer: 19:02:18 - trees.J48 (location)
Tree View
age
= 21
= 24
= 25
hyd (3.0)
blr (4.0)
kdp (3.0)