

# Decision Tree

Prashanth.S 19MID0020

## Question-I

Viewer

Relation: customer

| No. | 1: age<br>Nominal | 2: income<br>Nominal | 3: student<br>Nominal | 4: credit_rating<br>Nominal | 5: buys_computer<br>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                          |

```
<> relation_customer.arff x
1  @relation customer
2
3  @attribute age {youth,middle_aged,senior}
4  @attribute income {high,medium,low}
5  @attribute student {no,yes}
6  @attribute credit_rating {fair, excellent}
7  @attribute buys_computer {no,yes}
8
9
10 @data
11 youth, high, no, fair, no
12 youth, high, no, excellent, no
13 middle_aged, high, no, fair, yes
14 senior, medium, no, fair, yes
15 senior, low, yes, fair, yes
16 senior, low, yes, excellent, no
17 middle_aged, low, yes, excellent, yes
18 youth, medium, no, fair, no
19 youth, low, yes, fair, yes
20 senior, medium, yes, fair, yes
21 youth, medium, yes, excellent, yes
22 middle_aged, medium, no, excellent, yes
23 middle_aged, high, yes, fair, yes
24 senior, medium, no, excellent, no
```

# DECISION TREE

Prashanth.S 19MID0020

The screenshot shows the Weka Explorer Classifier window. The 'Classifier' dropdown is set to 'J48 -C 0.25 -M 2'. Under 'Test options', 'Percentage split' is selected with a percentage of 66. The 'Test set' is '(Nom) buys\_computer'. The 'Classifier output' pane shows the following information:

```
=== Run information ===  
Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2  
Relation:     customer  
Instances:    14  
Attributes:   5  
              age  
              income  
              student  
              credit_rating  
              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  
| credit_rating = fair: yes (3.0)  
| credit_rating = excellent: no (2.0)
```

The screenshot shows the Weka Explorer Classifier window. The 'Classifier' dropdown is set to 'J48 -C 0.25 -M 2'. Under 'Test options', 'Percentage split' is selected with a percentage of 66. The 'Test set' is '(Nom) buys\_computer'. The 'Classifier output' pane shows the following information:

```
Number of Leaves :    5  
Size of the tree :    8  
  
Time taken to build model: 0.01 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 ===

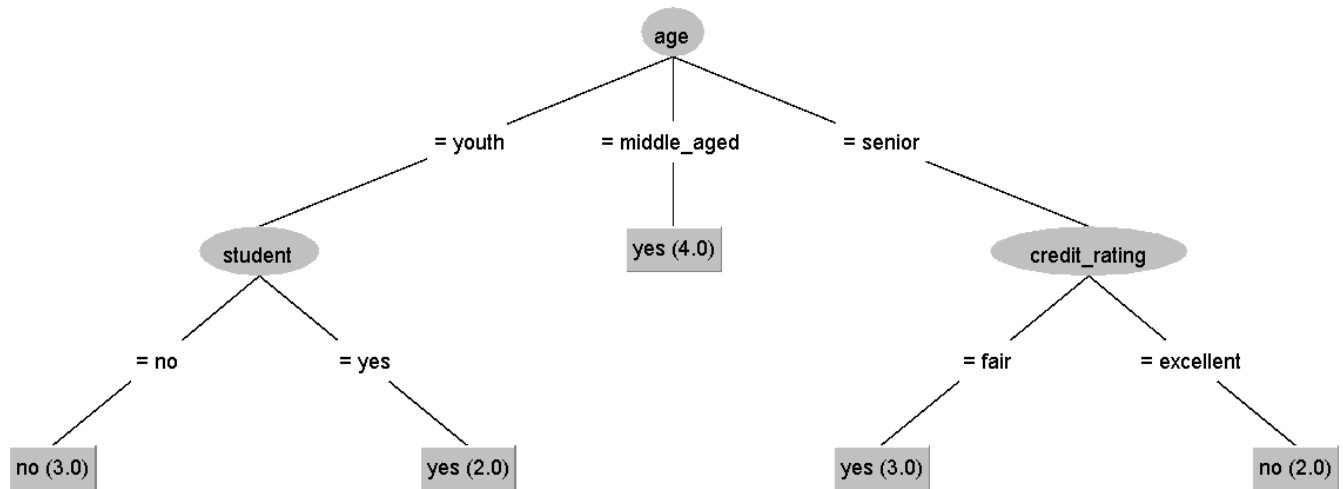
|               | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC    | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|--------|----------|----------|-------|
|               | 0.000   | 0.333   | 0.000     | 0.000  | 0.000     | -0.408 | 0.333    | 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
```

## DECISION TREE

Prashanth.S 19MID0020



Taking the 10<sup>th</sup> record With age=senior and credit\_rating=fair we are getting  
buys\_computer = yes

## Question-2

```

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
7 @data
8 x,youth,high,A
9 y,youth,low,B
10 z,middle,high,A
11 u,middle,low,B
12 v,senior,high,A
13 l,senior,low,B
14 w,youth,high,A
15 q,youth,low,B
16 r,middle,high,A
17 n,senior,high,A
18

```

### Viewer

| Relation: customer |         |         |           |          |
|--------------------|---------|---------|-----------|----------|
| 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        |

# DECISION TREE

Prashanth.S 19MID0020

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 % 53

More options...

(Nom) class

Start Stop

Result list (right-click for options)

18:11:17 - trees.J48

Classifier output

```
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    customer
Instances:   10
Attributes:  4
             name
             age
             income
             class
Test mode:   split 53.0% train, remainder test

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

J48 pruned tree
-----

income = high: A (6.0)
income = medium: A (0.0)
income = low: B (4.0)

Number of Leaves :    3
Size of the tree :    4
```

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 % 53

More options...

(Nom) class

Start Stop

Result list (right-click for options)

18:11:17 - 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      2           40 %
Incorrectly Classified Instances    3           60 %
Kappa statistic                     0
Mean absolute error                  0.56
Root mean squared error              0.6325
Relative absolute error              103.1579 %
Root relative squared error          108.6611 %
Total Number of Instances           5

=== Detailed Accuracy By Class ===

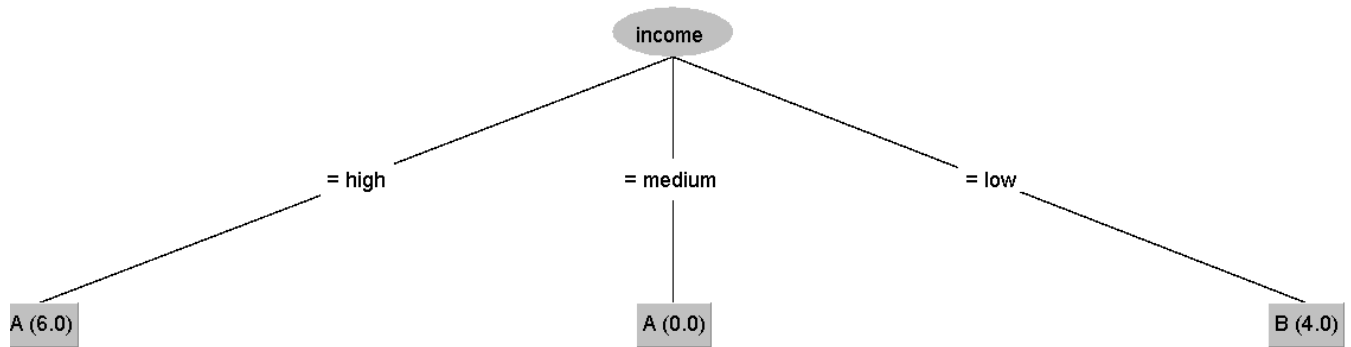
              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              1.000    1.000    0.400     1.000    0.571      ?      0.500    0.400    A
              0.000    0.000    ?         0.000    ?         ?      0.500    0.600    B
Weighted Avg.   0.400    0.400    ?         0.400    ?         ?      0.500    0.520
```

=== Confusion Matrix ===

```
a b  <-- classified as
2 0 | a = A
3 0 | b = B
```

## DECISION TREE

Prashanth.S 19MID0020



With income=high, the person belongs to the class A

## Question-3

```
location.arff
1  @relation location
2
3  @attribute age {21,24,25}
4  @attribute location {hyd,blr,kdp}
5
6  @data
7  21,hyd
8  21,hyd
9  24,blr
10 24,blr
11 24,blr
12 24,blr
13 21,hyd
14 25,kdp
15 25,kdp
16 25,kdp
```

 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         |

# DECISION TREE

Prashanth.S 19MID0020

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'J48 -C 0.25 -M 2'. The 'Test options' section has 'Percentage split' set to 60%. The 'Classifier output' pane displays the following information:

```
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    location
Instances:   10
Attributes:  2
              age
              location
Test mode:   split 60.0% 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
```

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'J48 -C 0.25 -M 2'. The 'Test options' section has 'Percentage split' set to 60%. The 'Classifier output' pane displays the following information:

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      4          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          4

=== 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    |       |

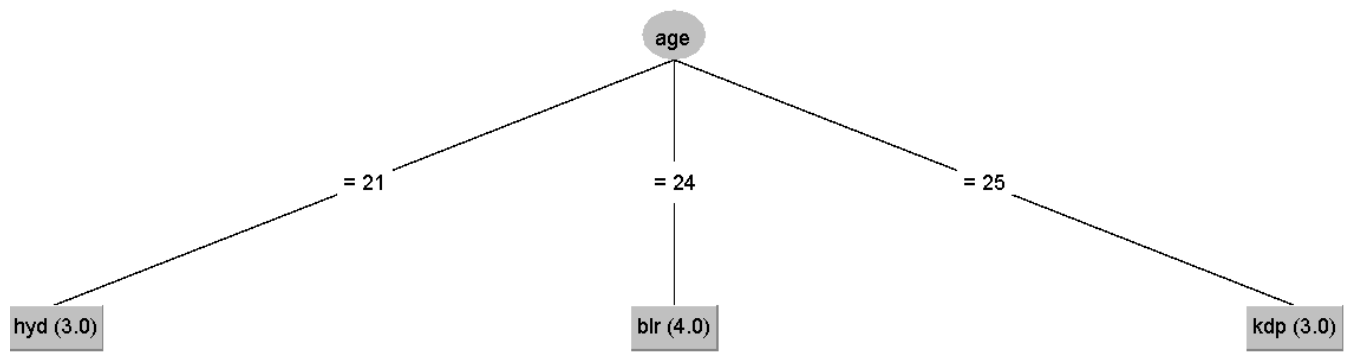
```
=== Confusion Matrix ===

a b c  <-- classified as
2 0 0 | a = hyd
0 2 0 | b = blr
0 0 0 | c = kdp
```



## DECISION TREE

Prashanth.S 19MID0020



With the 2<sup>nd</sup> record Age=21, will get Location=hyd