

Name: Allan Rodrigues  
Class: TE IT A  
Roll no: 59  
Pid:191104

St. Francis Institute of Technology, Mumbai-400 103  
**Department Of Information Technology**

A.Y. 2021-2022

Class: TE-ITA/B, Semester: VI

Subject: **Business Intelligence Lab**

**Experiment – 4: To Implement any one of the classifiers using WEKA (Decision Tree, Naïve Bayes, Random Forest)**

1. **Aim:** To Implement any one of the classifiers using WEKA (Decision Tree, Naïve Bayes, Random Forest)
2. **Objectives:** After study of this experiment, the students will be able to Understand and knew about all the three classifiers.
3. **Outcomes:** After study of this experiment, the students will be able to  
CO4: Design and Implement the appropriate data mining methods like classification, clustering or Frequent Pattern mining on large data sets.  
CO5: Define and apply metrics to measure the performance of various data mining algorithms
4. **Prerequisite:** Introduction to all the three classifiers through algorithms & Problem solving approach.
5. **Requirements:** Personal Computer, Windows XP operating system/Windows 7, Internet Connection, Microsoft Word, WEKA tool, Java/R/Python.
6. **Theory:**
  - a. What is Classification Data Mining?
  - b. Difference between supervised and unsupervised
  - c. Solve numerical on decision tree

Q.6

Q

1) What is classification in Data mining?  
→ Classification is the type of data analysis in which models defining relevant data classes are extracted. Classification models called classifiers, predict categorical class labels and prediction models predict continuous valued functions.

2) How classification is different from Association.  
→ Classification rules have only one attribute in their consequent (Then part), whereas association can have more than one attribute in their consequent.

c.  $C_1 = \text{Profit} = \text{Down} = 5$  samples.

$C_2 = \text{Profit} = \text{Up} = 5$  samples.

$$P(C_1) = 5/10 = P(C_2) = 5/10.$$

Entropy before split.

$$1) H(D) = \sum_{i=1}^2 P_i \log_2 \left( \frac{1}{P_i} \right).$$

$$H(D) = \frac{5}{10} \log_2 \left( \frac{10}{5} \right) + \frac{5}{10} \log_2 \left( \frac{10}{5} \right) = 0.5 + 0.5 = 1$$

2) choosing Age as splitting Attribute.

$$H(\text{Age}) = \frac{3}{10} I(3,0) = \frac{3}{10} \left( \frac{3}{3} \log_2 \frac{3}{3} + 0 \cdot 0 \right)$$

$$(\text{for mid}) H(\text{Age}) = \frac{4}{10} I(2,2) = \frac{2}{4} \log_2 \left( \frac{4}{2} \right) + \frac{2}{4} \log_2 \left( \frac{4}{2} \right) = 1$$

For new,

$$H(\text{Age}) = \frac{2}{10} I(0,3) = 0.$$

$$\therefore H(\text{Age}) = \frac{3}{10} \times 0 + \frac{4}{10} \times 1 + \frac{3}{10} \times 0 = 0.4$$

$$\text{Gain}(\text{Age}) = H(D) - H(\text{Age}) = 1 - 0.4 = 0.6$$

3) choosing competition.

for yes,

$$\therefore H(\text{comp}) = \frac{4}{10} I(3,1) = \frac{3}{4} \log_2 \left( \frac{4}{3} \right) + \frac{1}{4} \log_2 (4) = 0.8113$$

for no

$$H(\text{comp}) = \frac{6}{10} I(2,4) = \frac{2}{6} \log_2 \left( \frac{6}{2} \right)$$

$$+ \frac{4}{6} \log_2 \left( \frac{6}{4} \right) = 0.918.$$



$$H(\text{comp}) = \frac{4}{10} \times 0.8113 + \frac{6}{10} \times 0.918 = 0.875$$

$$\text{Gain}(\text{completion}) = H(0) - H(\text{comp}) = 1 - 0.875 = 0.125$$

iii) choosing Type  
for software

$$H(\text{Type}) = \frac{6}{10} (3,3) = \frac{3}{6} \log_2\left(\frac{6}{3}\right) + \frac{3}{6} \log_2\left(\frac{6}{3}\right) = 1$$

For hardware

$$H(\text{Type}) = \frac{4}{10} (2,2) = \frac{2}{4} \log_2\left(\frac{4}{2}\right) + \frac{2}{4} \log_2\left(\frac{4}{2}\right) = 1$$

$$H(\text{Type}) = \frac{6}{10} \times 1 + \frac{4}{10} \times 1 = 1$$

$$\text{Gain}(\text{Type}) = H(0) - H(\text{Type}) = 1 - 1 = 0$$

Summary : Gain Age = 0.6  
Gain (completion) = 0.125  
Gain (Type) = 0

Age attribute has highest gain hence it is used as decision attribute.



Iteration 2

considering only Age = mid values from table  
Hence now,

$$c_1 = \text{Profit} = \text{Down} = 2/4$$

$$c_2 = \text{Profit} = \text{Up} = 2/4$$

i) Entropy  $H(D) = \frac{2}{4} \log_2 \frac{4}{2} + \frac{2}{4} \log_2 \frac{4}{2} = 1$

ii) choosing competition

For Yes

$$H(\text{comp}) = \frac{2}{4} I(2,0) + \frac{2}{4} I(2,0) = 0$$

$$\text{gain}(\text{comp}) = 1 - 0 = 1$$

iii) choose Type.

For software  $\frac{2}{4} I(1,1) = \frac{1}{2} \log_2(2) + \frac{1}{2} \log_2(2) = 1$

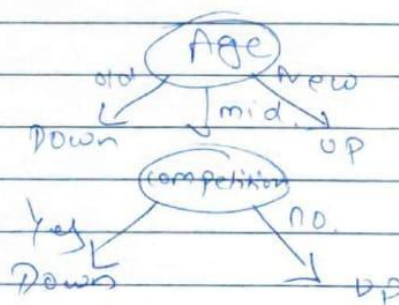
For hardware  $= \frac{2}{4} I(1,1) = 1$

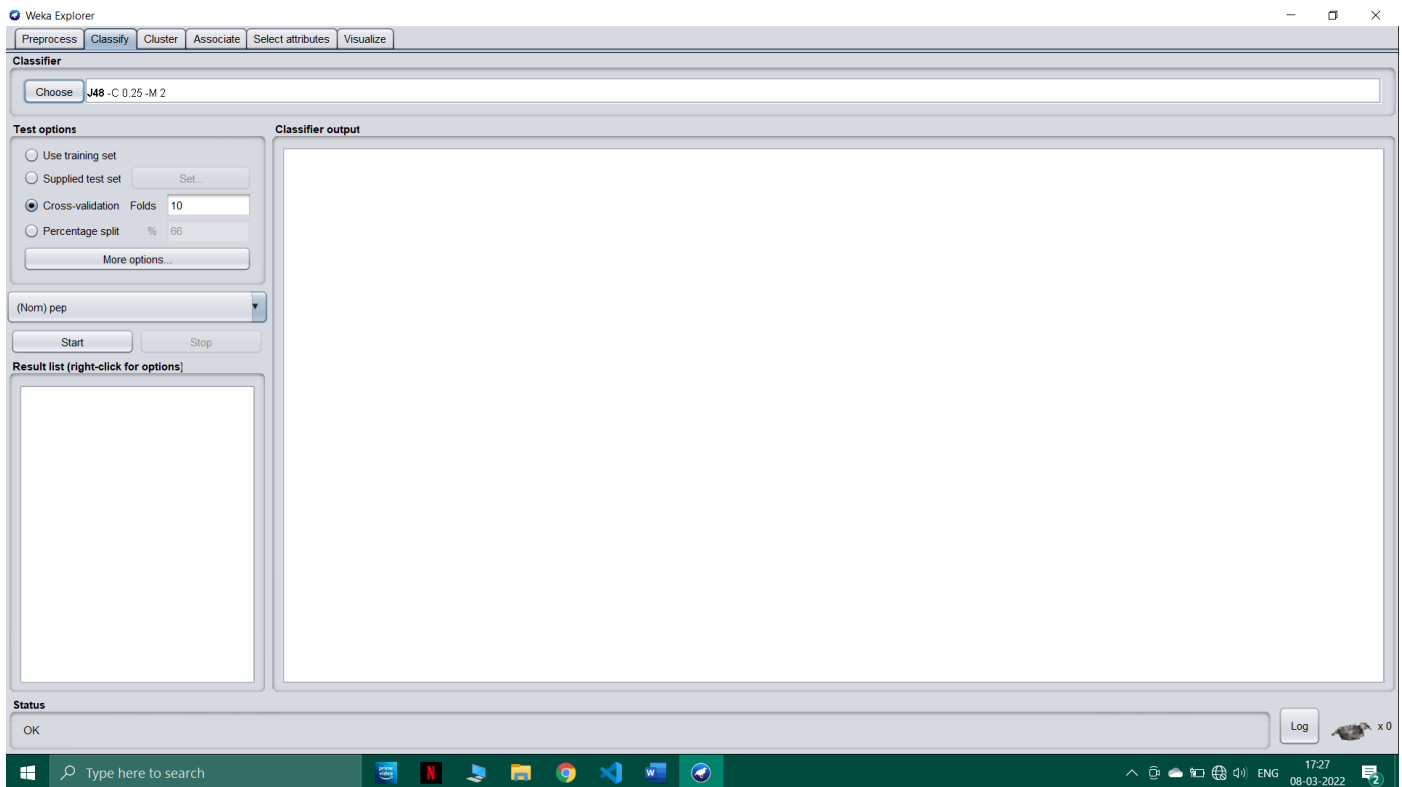
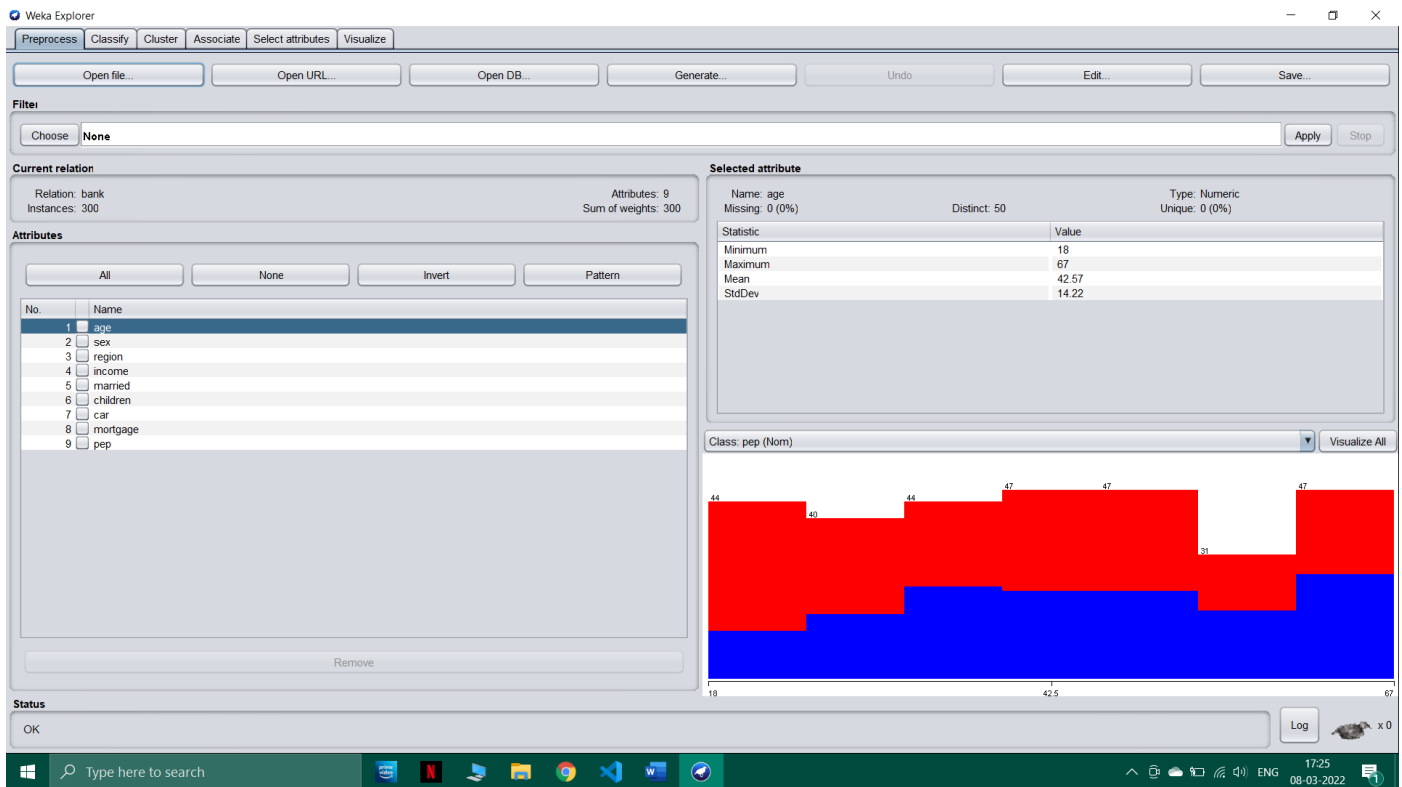
$$\therefore H(\text{Type}) = \frac{2}{4} + \frac{2}{4} = 1$$

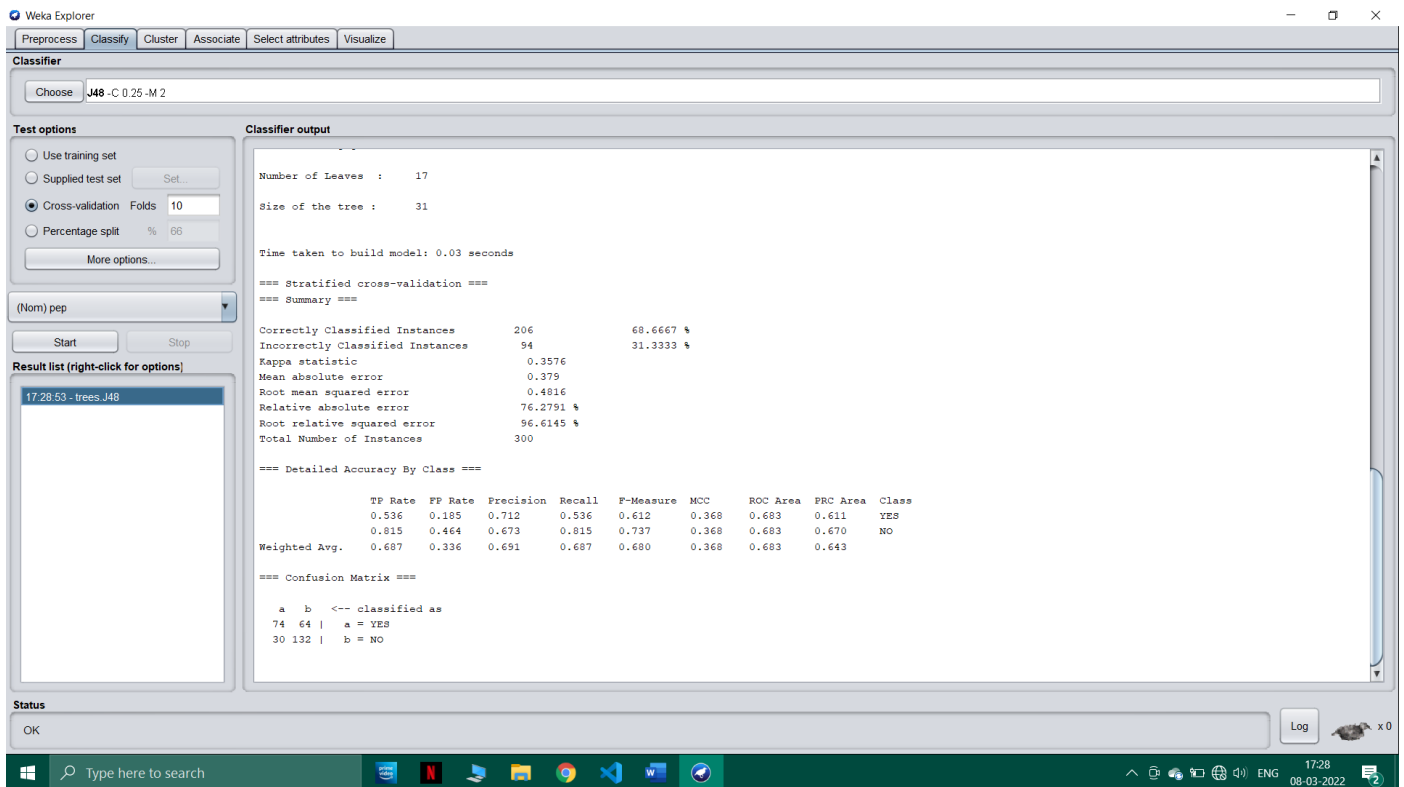
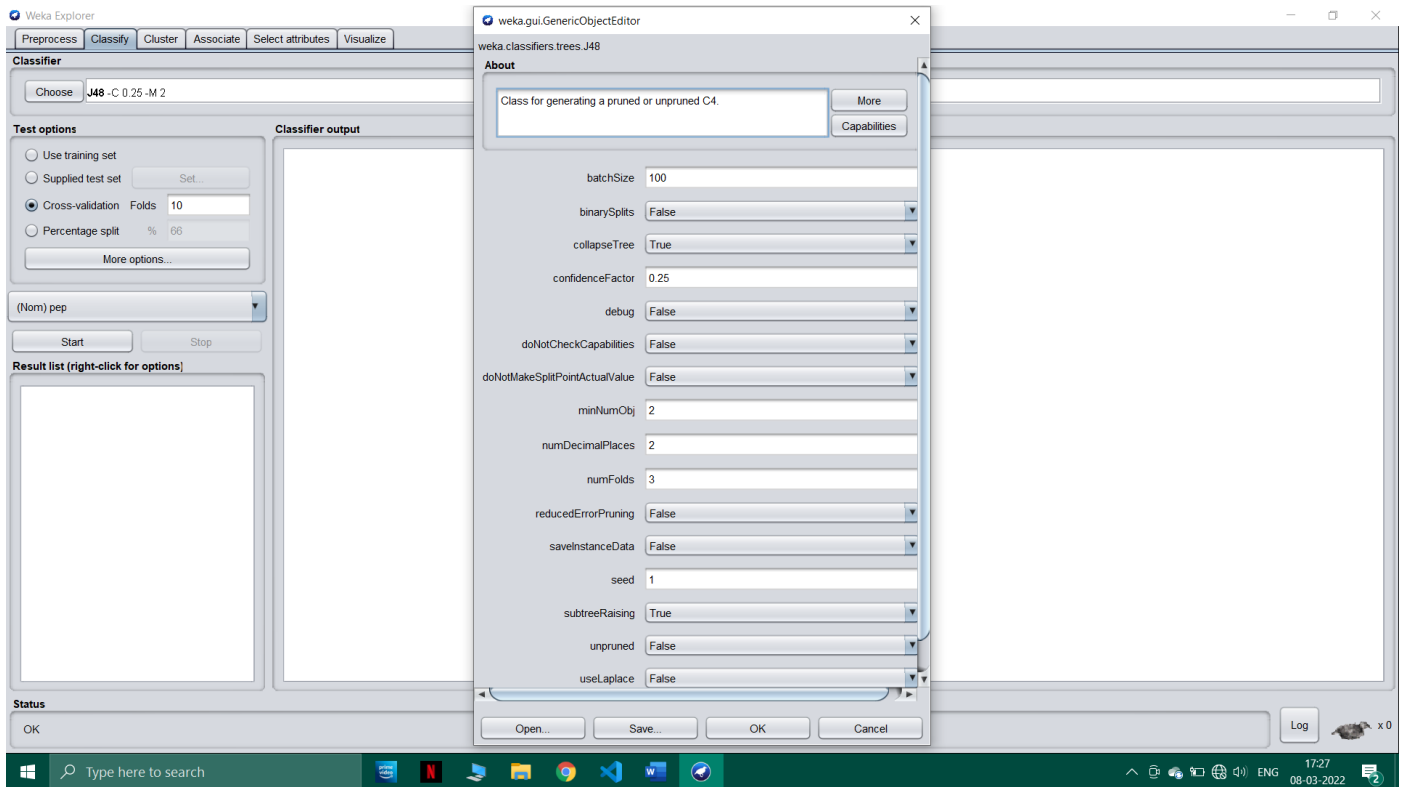
$$\text{gain}(\text{competition}) = H(D) - H(\text{Type}) = 1 - 1 = 0$$

Summary  $\text{gain}(\text{competition}) = 1$   
 $\text{gain}(\text{Type}) = 0$

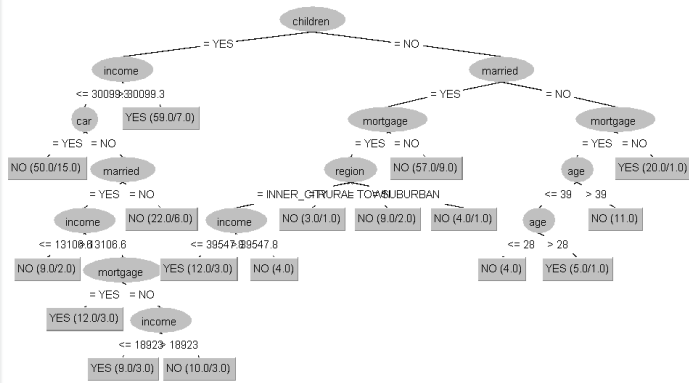
Competition attribute has highest gain hence placed before Age = 'Mid'.







## Tree View



bank-data1.csv.arff x bank-data2.csv.arff x index.html x bank.arff x bank-new.arff

```
1 relation bank
2
3 @attribute age numeric
4 @attribute sex {MALE,FEMALE}
5 @attribute region {INNER_CITY,RURAL,TOWN,SUBURBAN}
6 @attribute income numeric
7 @attribute married {YES,NO}
8 @attribute children {YES,NO}
9 @attribute car {YES,NO}
10 @attribute mortgage {YES,NO}
11 @attribute pep {YES,NO}
12
13 @data
14 48,FEMALE,INNER_CITY,17546,NO,YES,NO,NO,YES
15 40,MALE,TOWN,30085.1,YES,YES,YES,YES,NO
16 51,FEMALE,INNER_CITY,16575.4,YES,NO,YES,NO,NO
17 23,FEMALE,TOWN,20375.4,YES,YES,NO,NO,NO
18 57,FEMALE,RURAL,50576.3,YES,NO,NO,NO,NO
19 57,FEMALE,TOWN,37869.6,YES,YES,NO,NO,YES
20 22,MALE,RURAL,8877.07,NO,NO,NO,NO,YES
21 58,MALE,TOWN,24946.6,YES,NO,YES,NO,NO
22 37,FEMALE,SUBURBAN,25304.3,YES,YES,YES,NO,NO
23 54,MALE,TOWN,24212.1,YES,YES,YES,NO,NO
24 66,FEMALE,TOWN,59803.9,YES,NO,NO,NO,NO
25 52,FEMALE,INNER_CITY,26658.8,NO,NO,YES,YES,NO
26 44,FEMALE,TOWN,15735.8,YES,YES,NO,YES,YES
27 66,FEMALE,TOWN,55204.7,YES,YES,YES,YES,YES
28 36,MALE,RURAL,19474.6,YES,NO,NO,YES,NO
29 38,FEMALE,INNER_CITY,22342.1,YES,NO,YES,YES,NO
30 37,FEMALE,TOWN,17729.8,YES,YES,NO,YES,NO
31 46,FEMALE,SUBURBAN,41016,YES,NO,NO,YES,NO
32 62,FEMALE,INNER_CITY,26909.2,YES,NO,NO,NO,YES
33 31,MALE,TOWN,22522.8,YES,NO,YES,NO,NO
34 61,MALE,INNER_CITY,57880.7,YES,YES,NO,NO,YES
35 50,MALE,TOWN,16497.3,YES,YES,NO,NO,NO
36 54,MALE,INNER_CITY,38446.6,YES,NO,NO,NO,NO
37 27,FEMALE,TOWN,15538.8,NO,NO,YES,YES,NO
38 22,MALE,INNER_CITY,12640.3,NO,YES,YES,NO,NO
39 56,MALE,INNER_CITY,41034,YES,NO,YES,YES,NO
40 45,MALE,INNER_CITY,20809.7,YES,NO,NO,YES,NO
41 39,FEMALE,TOWN,20114,YES,YES,NO,NO,YES
```

Line 1, Column 1

Tab Size: 4 Plain Text



```
C:\Users\HP\Desktop\Allan\sem-V\BIL\Exp4\bank-new.arff - Sublime Text (UNREGISTERED)
bank-data1.csv.arff x bank-data2.csv.arff x index.html x bank.arff x bank-new.arff x
1 @relation bank-new
2
3 @attribute age numeric
4 @attribute sex {MALE,FEMALE}
5 @attribute region {INNER_CITY,RURAL,TOWN,SUBURBAN}
6 @attribute income numeric
7 @attribute married {YES,NO}
8 @attribute children {YES,NO}
9 @attribute car {YES,NO}
10 @attribute mortgage {YES,NO}
11 @attribute pep {YES,NO}
12
13 @data
14 23,MALE,INNER_CITY,18766.9,YES,NO,YES,YES,?
15 30,MALE,RURAL,9915.67,NO,YES,NO,YES,?
16 45,FEMALE,RURAL,21881.6,NO,NO,YES,NO,?
17 50,MALE,TOWN,46794.4,YES,YES,NO,YES,?
18 41,FEMALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
19 20,MALE,INNER_CITY,16688.5,NO,YES,NO,YES,?
20 46,FEMALE,RURAL,39068,YES,NO,YES,YES,?
21 50,FEMALE,INNER_CITY,27740.8,YES,YES,YES,YES,?
22 42,MALE,INNER_CITY,33584.9,NO,YES,YES,NO,?
23 57,FEMALE,TOWN,19621.3,YES,YES,YES,NO,?
24 63,FEMALE,INNER_CITY,47630.9,YES,NO,NO,YES,?
25 26,FEMALE,INNER_CITY,22378.5,NO,NO,YES,YES,?
26 62,FEMALE,RURAL,20837.1,YES,NO,YES,NO,?
27 26,FEMALE,SUBURBAN,23912.7,YES,NO,YES,NO,?
28 19,MALE,RURAL,8005.13,YES,YES,NO,NO,?
29 44,MALE,TOWN,34961.7,YES,YES,NO,YES,?
30 32,FEMALE,INNER_CITY,24627.6,YES,NO,YES,YES,?
31 56,FEMALE,RURAL,47315.3,YES,YES,YES,NO,?
32 26,MALE,TOWN,13196.2,YES,YES,NO,NO,?
33 43,FEMALE,TOWN,20528.9,NO,YES,YES,NO,?
34 40,MALE,TOWN,37227.8,NO,YES,YES,NO,?
35 29,FEMALE,INNER_CITY,21022.6,YES,YES,NO,NO,?
36 54,MALE,RURAL,49986.7,YES,YES,YES,NO,?
37 51,MALE,RURAL,44199.9,YES,NO,YES,YES,?
38 24,FEMALE,INNER_CITY,20050.2,YES,YES,YES,NO,?
39 20,FEMALE,SUBURBAN,13013.9,NO,YES,NO,YES,?
40 48,FEMALE,TOWN,39465.8,YES,NO,YES,NO,?
41 52,FEMALE,RURAL,39387.7,YES,NO,NO,YES,?
42 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
43 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
44 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
45 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
46 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
47 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
48 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
49 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
50 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
51 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
52 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
53 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
54 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
55 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
56 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
57 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
58 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
59 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
60 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
61 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
62 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
63 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
64 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
65 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
66 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
67 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
68 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
69 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
70 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
71 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
72 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
73 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
74 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
75 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
76 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
77 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
78 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
79 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
80 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
81 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
82 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
83 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
84 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
85 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
86 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
87 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
88 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
89 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
90 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
91 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
92 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
93 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
94 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
95 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
96 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
97 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
98 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
99 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
100 50,MALE,INNER_CITY,20721.1,YES,NO,YES,NO,?
```

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) pep

Start

Stop

Result list (right-click for options)

17:28:53 - trees.J48

Classifier output

Number of Leaves : 17

Size of the tree : 31

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 20

Incorrectly Classified Instances 9

Kappa statistic

Mean absolute error

Root mean squared error

Relative absolute error

Root relative squared error

Total Number of Instances 30

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.536	0.185	0.712	0.536	0.612	0.368	0.683	0.611	YES
	0.815	0.464	0.673	0.815	0.737	0.368	0.683	0.670	NO
Weighted Avg.	0.687	0.336	0.691	0.687	0.680	0.368	0.683	0.643	

=== Confusion Matrix ===

a	b	<-- classified as	
74	64	a = YES	
30	132	b = NO	

Test Instances

Relation: None

Instances: None

Attributes: None

Sum of weights: None

Open file...

Open URL...

Class No class

Close

Status

OK

Log

x 0

Type here to search 17:41 08-03-2022

**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) pep

Start Stop

**Result list (right-click for options)**

- 17:28:53 - trees.J48
- 17:42:43 - trees.J48

**Classifier output**

```

| | | | age > 28: YES (5.0/1.0)
| | | | age > 39: NO (11.0)
| | | mortgage = NO: YES (20.0/1.0)

Number of Leaves : 17
Size of the tree : 31

Time taken to build model: 0.01 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Total Number of Instances 0
Ignored Class Unknown Instances 50

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
      ?      ?      ?      ?      ?      ?      ?      ?      YES
Weighted Avg.  ?      ?      ?      ?      ?      ?      ?      ?      NO

=== Confusion Matrix ===

a b  <-- classified as
0 0 | a = YES
0 0 | b = NO
  
```

**Status**

OK Log x 0

**Weka Classifier Visualize: 17:42:43 - trees.J48 (bank)**

X: pep (Nom) Y: predicted pep (Nom)

Colour: pep (Nom) Select Instance

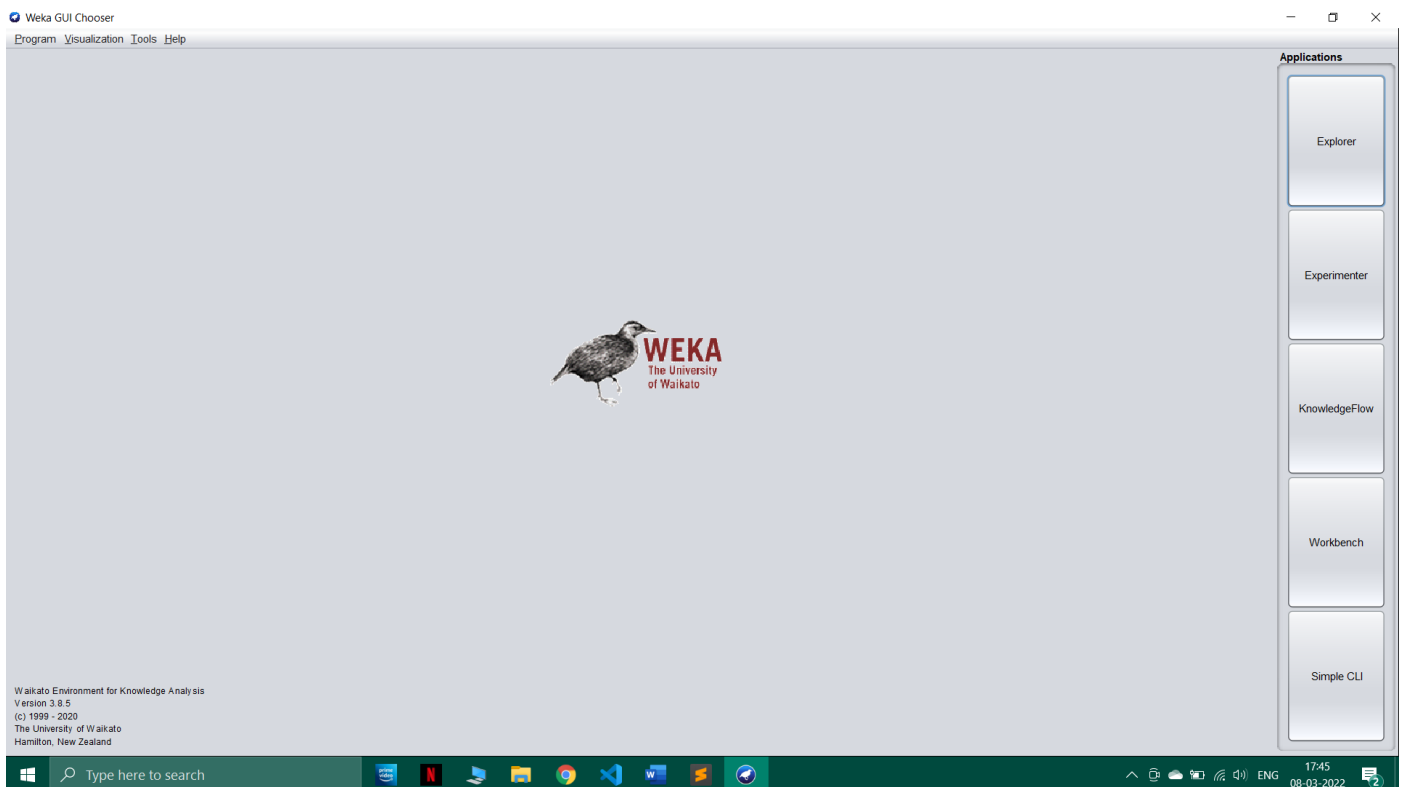
Reset Clear Open Save Jitter

**Plot: bank-new\_predicted**

**Class colour**

YES NO

```
C:\Users\HP\Desktop\Allan\sem-VI\BIL\Exp-4\bank_predicted.arff - Sublime Text (UNREGISTERED)
bank-data1.csv.arff x bank-data2.csv.arff x index.html x bank.arff x bank_predicted.arff x bank-new.arff x
1 @relation bank-new_predicted
2
3 @attribute age numeric
4 @attribute sex {MALE,FEMALE}
5 @attribute region {INNER_CITY,RURAL,TOWN,SUBURBAN}
6 @attribute income numeric
7 @attribute married {YES,NO}
8 @attribute children {YES,NO}
9 @attribute car {YES,NO}
10 @attribute mortgage {YES,NO}
11 @attribute 'prediction margin' numeric
12 @attribute 'predicted pep' {YES,NO}
13 @attribute pep {YES,NO}
14
15 @data
16 23,MALE,INNER_CITY,18766.9,YES,NO,YES,YES,0.5,YES,?
17 30,MALE,RURAL,9915.67,NO,YES,NO,YES,-0.454545,NO,?
18 45,FEMALE,RURAL,21881.6,NO,NO,YES,NO,0.9,YES,?
19 50,MALE,TOWN,46794.4,YES,YES,NO,YES,0.762712,YES,?
20 41,FEMALE,INNER_CITY,20721.1,YES,NO,YES,NO,-0.684211,NO,?
21 20,MALE,INNER_CITY,16688.5,NO,YES,NO,YES,-0.454545,NO,?
22 46,FEMALE,RURAL,39068,YES,NO,YES,YES,-0.333333,NO,?
23 50,FEMALE,INNER_CITY,27740.8,YES,YES,YES,YES,-0.4,NO,?
24 42,MALE,INNER_CITY,33584.9,NO,YES,YES,NO,0.762712,YES,?
25 57,FEMALE,TOWN,19621.3,YES,YES,YES,NO,-0.4,NO,?
26 63,FEMALE,INNER_CITY,47630.9,YES,NO,NO,YES,-1,NO,?
27 26,FEMALE,INNER_CITY,22378.5,NO,NO,YES,YES,-1,NO,?
28 62,FEMALE,RURAL,20837.1,YES,NO,YES,NO,-0.684211,NO,?
29 26,FEMALE,SUBURBAN,23912.7,YES,NO,YES,NO,-0.684211,NO,?
30 19,MALE,RURAL,8005.13,YES,YES,NO,NO,-0.555556,NO,?
31 44,MALE,TOWN,34961.7,YES,YES,NO,YES,0.762712,YES,?
32 32,FEMALE,INNER_CITY,24627.6,YES,NO,YES,YES,0.5,YES,?
33 56,FEMALE,RURAL,47315.3,YES,YES,YES,NO,0.762712,YES,?
34 26,MALE,TOWN,13196.2,YES,YES,NO,NO,0.333333,YES,?
35 43,FEMALE,TOWN,20528.9,NO,YES,YES,NO,-0.4,NO,?
36 40,MALE,TOWN,37227.8,NO,YES,YES,NO,0.762712,YES,?
37 29,FEMALE,INNER_CITY,21022.6,YES,YES,NO,NO,-0.4,NO,?
38 54,MALE,RURAL,49986.7,YES,YES,YES,NO,0.762712,YES,?
39 51,MALE,RURAL,44199.9,YES,NO,YES,YES,-0.333333,NO,?
40 24,FEMALE,INNER_CITY,20050.2,YES,YES,YES,NO,-0.4,NO,?
41 20,FEMALE,SUBURBAN,13013.9,NO,YES,NO,YES,-0.454545,NO,?
42 46,FEMALE,RURAL,39068,YES,NO,YES,YES,-0.333333,NO,?
Line 1, Column 1 Tab Size: 4 Plain Text
```



```
SimpleCLI

Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.
Command completion for classnames and files is
initiated with <Tab>. In order to distinguish
between files and classnames, file names must
be either absolute or start with './' or '~/ '
(the latter is a shortcut for the home directory).
<Alt+BackSpace> is used for deleting the text
in the commandline in chunks.

Type 'help' followed by <Enter> to see an overview
of all commands.

java weka.classifiers.trees.J48 -C 0.25 -M 2 -t C:\Users\HP\Desktop\Allan\sem-VI\BIL\Exp4\bank.arff -d C:\Users\HP\Desktop\Allan\sem-VI\BIL\Exp4\bank.model
```

```
SimpleCLI

> java weka.classifiers.trees.J48 -C 0.25 -M 2 -t directory-path\bank.arff -d directory-path \bank.model

> java weka.classifiers.trees.J48 -C 0.25 -M 2 -t C:\Users\HP\Desktop\Allan\sem-VI\BIL\Exp4\bank.arff -d C:\Users\HP\Desktop\Allan\sem-VI\BIL\Exp4\bank.model

Options: -C 0.25 -M 2

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

J48 pruned tree
-----
children = YES
| income <= 30099.3
| | car = YES: NO (50.0/15.0)
| | car = NO
| | | married = YES
| | | | income <= 13106.6: NO (9.0/2.0)
| | | | income > 13106.6
| | | | | mortgage = YES: YES (12.0/3.0)
| | | | | mortgage = NO
| | | | | | income <= 18923: YES (9.0/3.0)
| | | | | | income > 18923: NO (10.0/3.0)
| | | married = NO: NO (22.0/6.0)
| income > 30099.3: YES (59.0/7.0)
children = NO
| married = YES
| | mortgage = YES
| | | region = INNER_CITY
| | | | income <= 39547.8: YES (12.0/3.0)
| | | | income > 39547.8: NO (4.0)
| | | region = RURAL: NO (3.0/1.0)
| | | region = TOWN: NO (9.0/2.0)
| | | region = SUBURBAN: NO (4.0/1.0)
| | mortgage = NO: NO (57.0/9.0)
| married = NO
| | mortgage = YES
| | | age <= 39
| | | | age <= 28: NO (4.0)
| | | | age > 28: YES (5.0/1.0)
| | | age > 39: NO (11.0)
| | mortgage = NO: YES (20.0/1.0)

Number of Leaves : 17
```



```
SimpleCLI
Time taken to build model: 0.02 seconds

Time taken to test model on training data: 0.01 seconds

=== Error on training data ===

Correctly Classified Instances      243          81 %
Incorrectly Classified Instances    57           19 %
Kappa statistic                     0.6132
Mean absolute error                 0.2899
Root mean squared error             0.3807
Relative absolute error             58.3575 %
Root relative squared error         76.3937 %
Total Number of Instances          300

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MDC   ROC Area  PRC Area  Class
      0.717    0.111    0.846    0.717    0.776     0.620   0.854    0.816    YES
      0.889    0.283    0.787    0.889    0.835     0.620   0.854    0.838    NO
Weighted Avg.   0.810    0.204    0.814    0.810    0.808     0.620   0.854    0.828

=== Confusion Matrix ===

  a  b  <-- classified as
  99  39 |  a = YES
 18 144 |  b = NO

Time taken to perform cross-validation: 0.05 seconds

=== Stratified cross-validation ===

Correctly Classified Instances      206          68.6667 %
Incorrectly Classified Instances    94          31.3333 %
Kappa statistic                     0.3576
Mean absolute error                 0.379
Root mean squared error             0.4816
Relative absolute error             76.2791 %
Root relative squared error         96.6145 %
Total Number of Instances          300

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MDC   ROC Area  PRC Area  Class
      0.717    0.111    0.846    0.717    0.776     0.620   0.854    0.816    YES
      0.889    0.283    0.787    0.889    0.835     0.620   0.854    0.838    NO
Weighted Avg.   0.810    0.204    0.814    0.810    0.808     0.620   0.854    0.828

=== Confusion Matrix ===

  a  b  <-- classified as
  99  39 |  a = YES
 18 144 |  b = NO

Time taken to perform cross-validation: 0.05 seconds

=== Stratified cross-validation ===

Correctly Classified Instances      206          68.6667 %
Incorrectly Classified Instances    94          31.3333 %
Kappa statistic                     0.3576
Mean absolute error                 0.379
Root mean squared error             0.4816
Relative absolute error             76.2791 %
Root relative squared error         96.6145 %
Total Number of Instances          300

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MDC   ROC Area  PRC Area  Class
      0.536    0.185    0.712    0.536    0.612     0.368   0.683    0.611    YES
      0.815    0.464    0.673    0.815    0.737     0.368   0.683    0.670    NO
Weighted Avg.   0.687    0.336    0.691    0.687    0.680     0.368   0.683    0.643

=== Confusion Matrix ===

  a  b  <-- classified as
  74  64 |  a = YES
 30 132 |  b = NO
```

## 8. Post-Experiments Exercise

### a. Extended Theory:

- Write about Decision Tree and SVM

### b. Questions:

- Compare and Contrast between Decision Tree and SVM

### c. Conclusion:

- Summary of Experiment
- Importance of Experiment
- Application of Experiment

Q8

- a) Decision tree is a structure that includes a root node, branches & leaf nodes. Each internal node denotes a test and each leaf node holds a class label. The topmost node in the tree is the root node.  
It is a popular supervised learning algorithm, which is used for classification as well as regression.
- b)
  - SVM uses kernel to solve non-linear problems whereas decision tree defines types rectangles in input space to solve the problem.
  - Decision trees are better for categorical data & it deals collinearity better than SVM.

### Conclusion

In this experiment we used WEKA tools to classify a dataset using J48 classifier. We also learnt to do the same process 'simple CLI'.

Decision trees help to evaluate your options. Helps you to choose between several courses of action. It is used in business, customer relationship management to evaluate best actions.