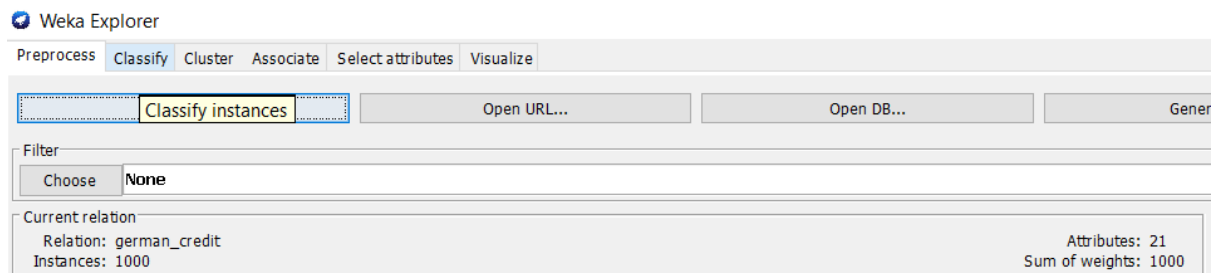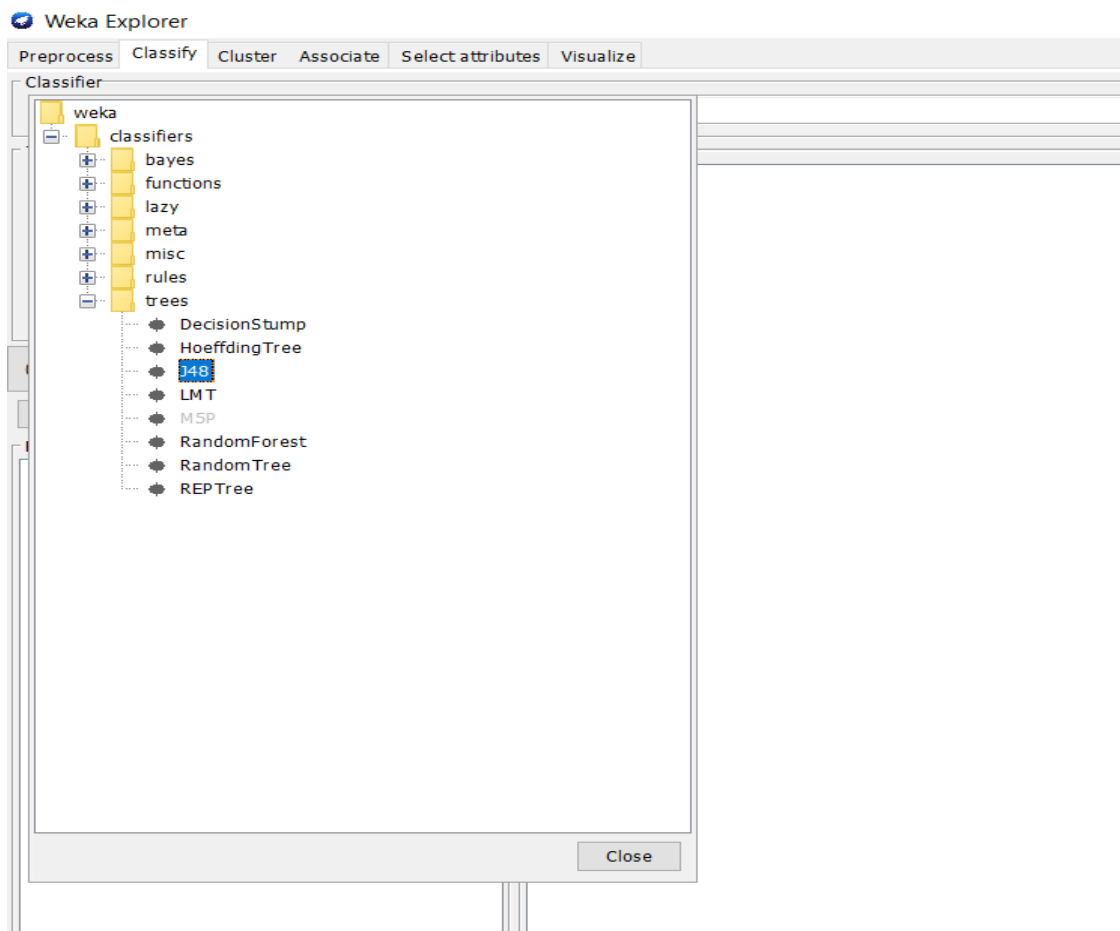# Steps to derive a Decision tree :-

Step 1 : Open Weka and open the required dataset, here, german credit dataset.
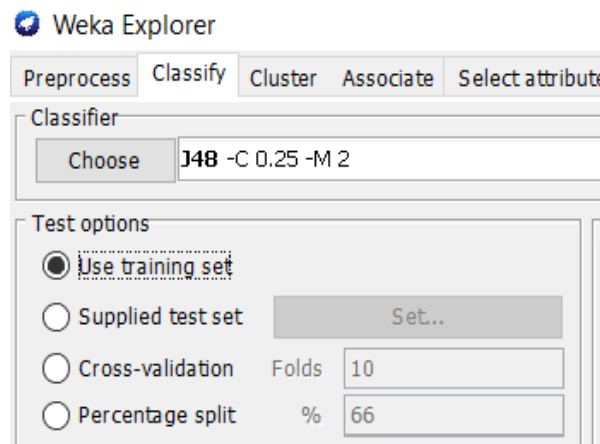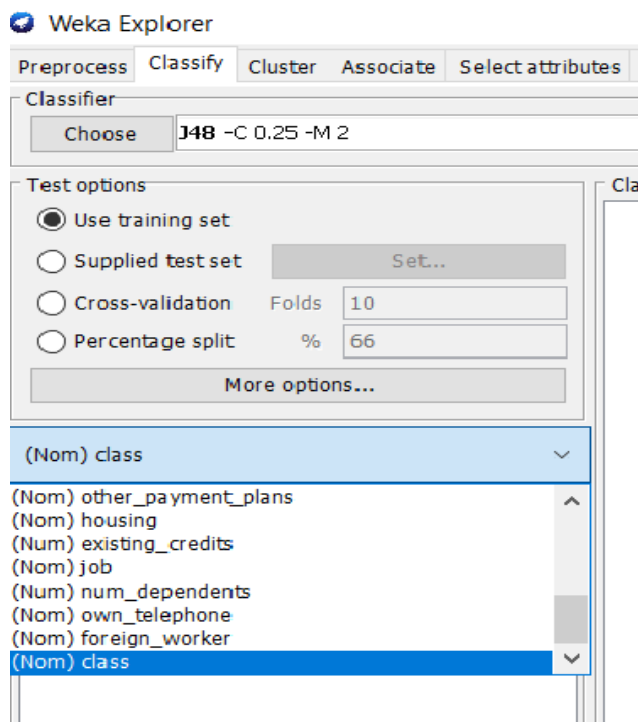
Step 2 : Click on the Classify tab.

Step 3 : After clicking on the Classify tab, click on the choose option on the left to select J48 classifier from trees.

Step 4 : Choose the **Use training set** from Test options.

Step 5 : Now, choose the attribute **class** from the dropdown.

Step 6 : Next click on Start. We will get a confusion matrix as the output.

Number of leaves : 103

Size of the tree : 140

Time taken to build model : 0.2 seconds

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) class

Start | Stop

Result list (right-click for options)
12:17:24 - trees.J48

Classifier output

```
checking_status = >=200: good (63.0/14.0)

checking_status = no checking: good (394.0/46.0)


Number of Leaves  :      103

Size of the tree :      140


Time taken to build model: 0.2 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.05 seconds

=== Summary ===

Correctly Classified Instances         855               85.5    %
Incorrectly Classified Instances       145               14.5    %
Kappa statistic                          0.6251
Mean absolute error                      0.2312
Root mean squared error                  0.34
Relative absolute error                 55.0377 %
Root relative squared error             74.2015 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC Area  Class
                 0.956    0.380    0.854      0.956   0.902      0.640   0.857     0.905     good
                 0.620    0.044    0.857      0.620   0.720      0.640   0.857     0.783     bad
Weighted Avg.    0.855    0.279    0.855      0.855   0.847      0.640   0.857     0.869

=== Confusion Matrix ===

   a    b   <-- classified as
 669   31 |   a = good
 114  186 |   b = bad
```

Status
OK

Step 7 : Right click on the option in the Result list and select the **visualize tree** option.

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) class

Start | Stop

Result list (right-click for options)
12:17:24 - trees.J48

View in main window
View in separate window
Save result buffer
Delete result buffer(s)

Load model
Save model
Re-evaluate model on current test set
Re-apply this model's configuration

Visualize classifier errors
Visualize tree
Visualize margin curve
Visualize threshold curve   >
Cost/Benefit analysis   >
Visualize cost curve   >

```
checking_status = >=200: good (63.0/1
checking_status = no checking: good (

Number of Leaves  :      103

Size of the tree :      140

Time taken to build model: 0.2 second

=== Evaluation on training set ===
```

```
   a    b   <-- classified as
 669   31 |   a = good
 114  186 |   b = bad
```

Step 8 : The output will be as follows :

Weka Classifier Tree Visualizer: 12:17:24 - trees.J48 (german_credit)
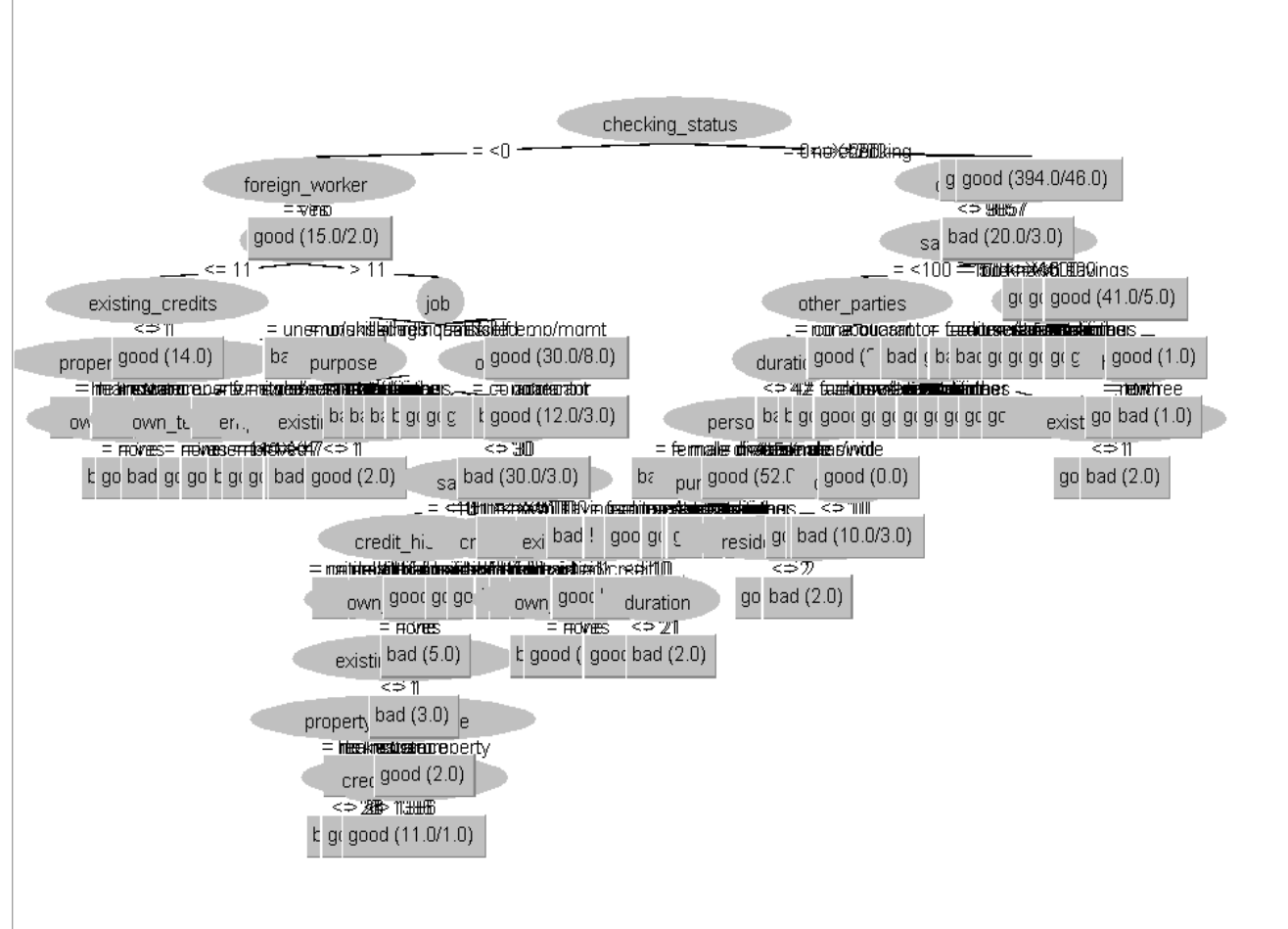
Accuracy measures vary with Percentage split :

Case 1 : J48 Classifier

Accuracy when Percentage split is 66%

Accuracy when Percentage split is 73%

Accuracy when Percentage split is 80%

```
checking_status = >=200: good (63.0/14.0)

checking_status = no checking: good (394.0/46.0)


Number of Leaves  :        103


Size of the tree :         140



Time taken to build model: 0.02 seconds


=== Evaluation on test split ===


Time taken to test model on test split: 0 seconds


=== Summary ===


Correctly Classified Instances         154               77      %
Incorrectly Classified Instances        46               23      %
Kappa statistic                          0.3867
Mean absolute error                      0.2947
Root mean squared error                  0.4433
Relative absolute error                 72.2746 %
Root relative squared error            100.8586 %
Total Number of Instances              200
```

Test options
- Use training set
- Supplied test set   Set...
- Cross-validation Folds 10
- Percentage split % 80
- More options...

(Nom) class

Start   Stop

Result list (right-click for options)
```
12:17:24 - trees.J48
12:41:21 - trees.J48
12:42:28 - trees.J48
12:44:14 - trees.J48
```

Case 2 : NaiveBayes Classifier

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

```
weka
  classifiers
    bayes
        BayesNet
        NaiveBayes
        NaiveBayesMultinomial
        NaiveBayesMultinomialText
        NaiveBayesMultinomialUpdateable
        NaiveBayesUpdateable
    functions
    lazy
    meta
    misc
    rules
    trees
```

Close

Accuracy when Percentage split is 66%

Classifier

Choose | NaiveBayes

Test options
- ○ Use training set
- ○ Supplied test set   Set...
- ○ Cross-validation   Folds 10
- ● Percentage split   % 66

More options...

(Nom) class

Start | Stop

Result list (right-click for options)
```
12:17:24 - trees.J48
12:41:21 - trees.J48
12:42:28 - trees.J48
12:44:14 - trees.J48
12:47:47 - bayes.NaiveBayes
```

Classifier output

```
foreign_worker
  yes                                  668.0      297.0
  no                                    34.0        5.0
  [total]                              702.0      302.0


Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.05 seconds

=== Summary ===

Correctly Classified Instances         260              76.4706 %
Incorrectly Classified Instances        80              23.5294 %
Kappa statistic                          0.3824
Mean absolute error                      0.2819
Root mean squared error                  0.4005
Relative absolute error                 67.9798 %
Root relative squared error             90.114  %
Total Number of Instances              340
```

Accuracy when Percentage split is 73%

Classifier

Choose | NaiveBayes

Test options
- ○ Use training set
- ○ Supplied test set   Set...
- ○ Cross-validation   Folds 10
- ● Percentage split   % 73

More options...

(Nom) class

Start | Stop

Result list (right-click for options)
```
12:17:24 - trees.J48
12:41:21 - trees.J48
12:42:28 - trees.J48
12:44:14 - trees.J48
12:47:47 - bayes.NaiveBayes
12:49:46 - bayes.NaiveBayes
```

Classifier output

```
foreign_worker
  yes                                  668.0      297.0
  no                                    34.0        5.0
  [total]                              702.0      302.0


Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.02 seconds

=== Summary ===

Correctly Classified Instances         204              75.5556 %
Incorrectly Classified Instances        66              24.4444 %
Kappa statistic                          0.3752
Mean absolute error                      0.2864
Root mean squared error                  0.4142
Relative absolute error                 69.153  %
Root relative squared error             92.5664 %
Total Number of Instances              270
```

Accuracy when Percentage split is 80%

```
Classifier
  Choose    NaiveBayes

Test options                          Classifier output
  ○ Use training set
                                      foreign_worker
  ○ Supplied test set    Set...          yes                          668.0      297.0
  ○ Cross-validation  Folds  10          no                            34.0        5.0
  ◉ Percentage split    %   80           [total]                      702.0      302.0
        More options...

(Nom) class                    ⌄        Time taken to build model: 0 seconds

     Start              Stop           === Evaluation on test split ===
Result list (right-click for options)
  12:17:24 - trees.J48                  Time taken to test model on test split: 0.02 seconds
  12:41:21 - trees.J48
  12:42:28 - trees.J48                  === Summary ===
  12:44:14 - trees.J48
  12:47:47 - bayes.NaiveBayes           Correctly Classified Instances         149               74.5   %
  12:49:46 - bayes.NaiveBayes           Incorrectly Classified Instances        51               25.5   %
  12:50:58 - bayes.NaiveBayes           Kappa statistic                          0.3657
                                        Mean absolute error                      0.2879
                                        Root mean squared error                  0.4129
                                        Relative absolute error                 70.6169 %
                                        Root relative squared error             93.9316 %
                                        Total Number of Instances              200
```

Accuracy measures vary with Cross - validation :

Case 1 : J48 Classifier

Accuracy when Cross – validation is 10

Accuracy when Cross – validation is 8

Accuracy when Cross – validation is 6


```
|   |     savings_status = >=1000: good (13.0/3.0)
|   |     savings_status = no known savings: good (41.0/5.0)
|   credit_amount > 9857: bad (20.0/3.0)
checking_status = >=200: good (63.0/14.0)
checking_status = no checking: good (394.0/46.0)


Number of Leaves  :      103

Size of the tree :      140


Time taken to build model: 0.02 seconds

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

Correctly Classified Instances        741             74.1   %
Incorrectly Classified Instances      259             25.9   %
Kappa statistic                        0.3453
Mean absolute error                    0.3239
Root mean squared error                0.4479
Relative absolute error               77.0782 %
Root relative squared error           97.737  %
Total Number of Instances           1000
```

Case 2 : NaiveBayes Classifier

Accuracy when Cross – validation is 10


```
   none                            410.0        188.0
   yes                             292.0        114.0
   [total]                         702.0        302.0

foreign_worker
   yes                             668.0        297.0
   no                               34.0          5.0
   [total]                         702.0        302.0


Time taken to build model: 0.01 seconds

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

Correctly Classified Instances        754             75.4   %
Incorrectly Classified Instances      246             24.6   %
Kappa statistic                        0.3813
Mean absolute error                    0.2936
Root mean squared error                0.4201
Relative absolute error               69.8801 %
Root relative squared error           91.6718 %
Total Number of Instances           1000
```

Accuracy when Cross – validation is 8

```
Classifier
  Choose   NaiveBayes

Test options                          Classifier output
  ○ Use training set                    none                          410.0      188.0
  ○ Supplied test set      Set...       yes                           292.0      114.0
  ● Cross-validation  Folds  8          [total]                       702.0      302.0
  ○ Percentage split    %   80
         More options...                foreign_worker
                                          yes                         668.0      297.0
(Nom) class                     ˅         no                           34.0        5.0
                                          [total]                     702.0      302.0
    Start              Stop

Result list (right-click for options)   Time taken to build model: 0 seconds
12:17:24 - trees.J48
12:41:21 - trees.J48                    === Stratified cross-validation ===
12:42:28 - trees.J48                    === Summary ===
12:44:14 - trees.J48
12:47:47 - bayes.NaiveBayes             Correctly Classified Instances      759             75.9   %
12:49:46 - bayes.NaiveBayes             Incorrectly Classified Instances    241             24.1   %
12:50:58 - bayes.NaiveBayes             Kappa statistic                       0.3957
12:53:52 - trees.J48                    Mean absolute error                   0.2936
12:55:08 - trees.J48                    Root mean squared error               0.4205
12:56:12 - trees.J48                    Relative absolute error              69.8657 %
12:57:33 - bayes.NaiveBayes             Root relative squared error          91.7659 %
12:58:29 - bayes.NaiveBayes             Total Number of Instances          1000
```

Accuracy when Cross – validation is 6

```
Classifier
  Choose   NaiveBayes

Test options                          Classifier output
  ○ Use training set                    none                          410.0      188.0
  ○ Supplied test set      Set...       yes                           292.0      114.0
  ● Cross-validation  Folds  6          [total]                       702.0      302.0
  ○ Percentage split    %   80          foreign_worker
         More options...                  yes                         668.0      297.0
                                          no                           34.0        5.0
(Nom) class                     ˅         [total]                     702.0      302.0

    Start              Stop

Result list (right-click for options)   Time taken to build model: 0 seconds
12:17:24 - trees.J48
12:41:21 - trees.J48                    === Stratified cross-validation ===
12:42:28 - trees.J48                    === Summary ===
12:44:14 - trees.J48
12:47:47 - bayes.NaiveBayes             Correctly Classified Instances      754             75.4   %
12:49:46 - bayes.NaiveBayes             Incorrectly Classified Instances    246             24.6   %
12:50:58 - bayes.NaiveBayes             Kappa statistic                       0.3813
12:53:52 - trees.J48                    Mean absolute error                   0.2955
12:55:08 - trees.J48                    Root mean squared error               0.4222
12:56:12 - trees.J48                    Relative absolute error              70.3237 %
12:57:33 - bayes.NaiveBayes             Root relative squared error          92.1377 %
12:58:29 - bayes.NaiveBayes             Total Number of Instances          1000
12:59:20 - bayes.NaiveBayes
```