

Steps to derive decision tree:

- 1.open weka
- 2.click on classify
- 3.click on use training data set
- 4.click on class and select 14:29:45 – trees.j48 and click start to get confusion matrix and number of leaves and size of the tree and time taken to build model details.

The screenshot shows the Weka Explorer Classifier window. The 'Classifier' tab is selected, and the 'J48 -C 0.25 -M 2' model is chosen. The 'Test options' section shows 'Use training set' selected. The 'Classifier output' section displays the following information:

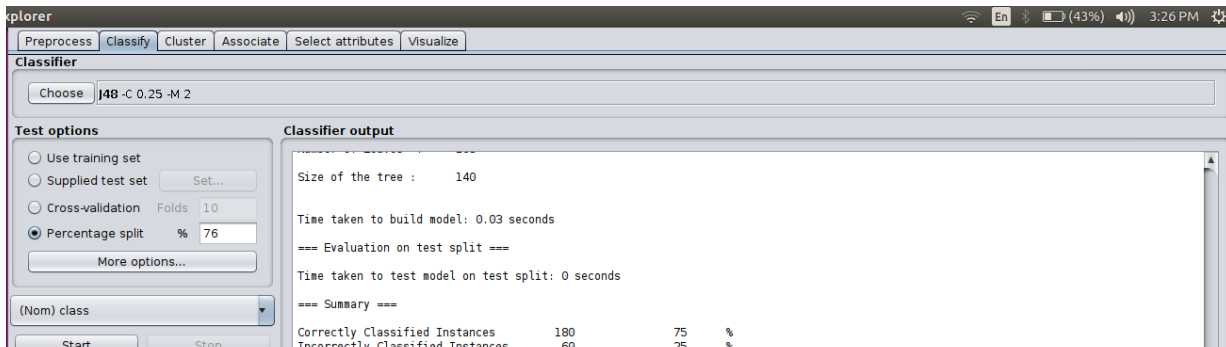
Number of Leaves : 103
Size of the tree : 140
Time taken to build model: 0.17 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0.04 seconds
=== Summary ===
Correctly Classified Instances 855 85.5 %
Incorrectly Classified Instances 145 14.5 %
Kappa statistic 0.6251
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

The 'Result list (right-click for options)' section shows the file '14:29:45 - trees.j48' selected. The 'Status' bar at the bottom shows 'OK'.

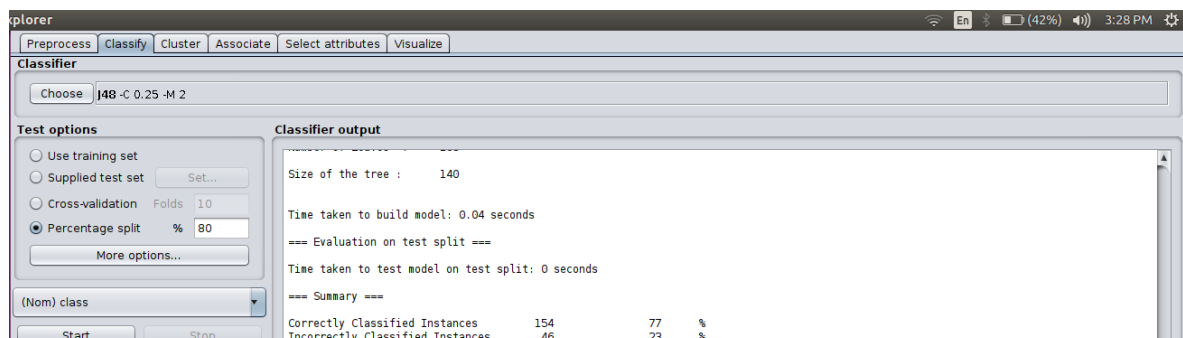
Time taken to build model is 0.17 seconds, and size of tree is 140 and number of leaves are 103.

- 5.Right click on 14:29:45-trees.j48 and select visualize tree to get the tree.

When percentage split is 76%, The Accuracy is 75%

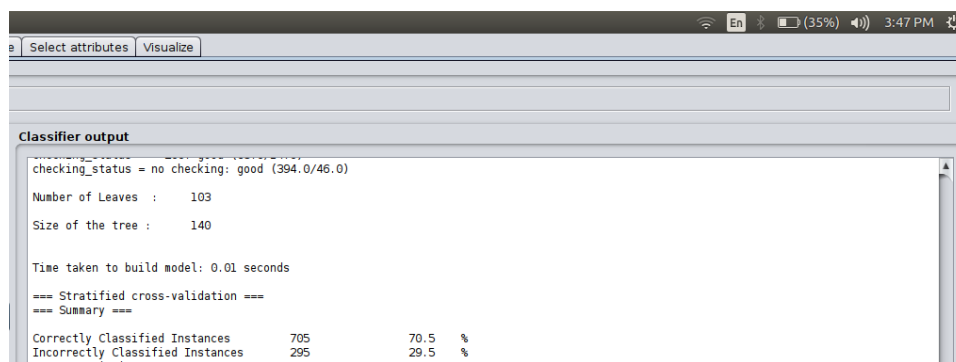


When Percentage split 80% , The Accuracy is 77%

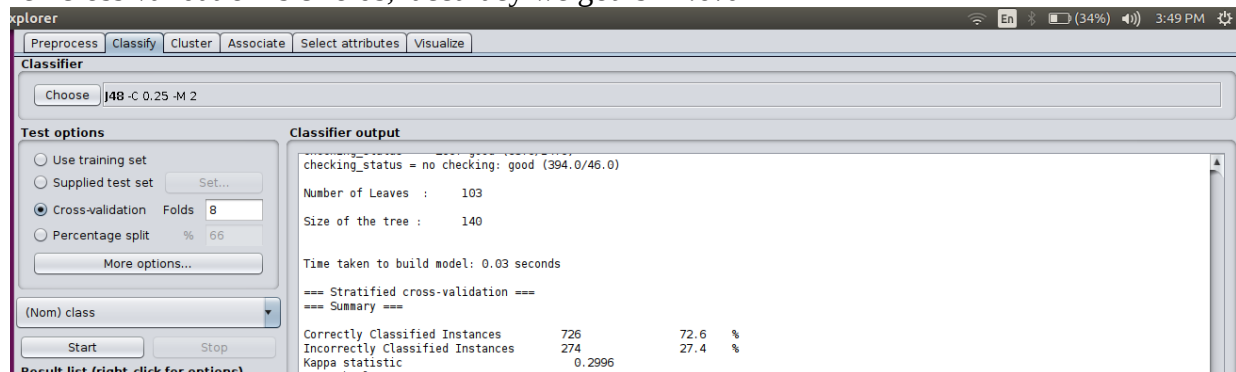


So, the maximum accuracy we got is 77% when percentage split is 80%

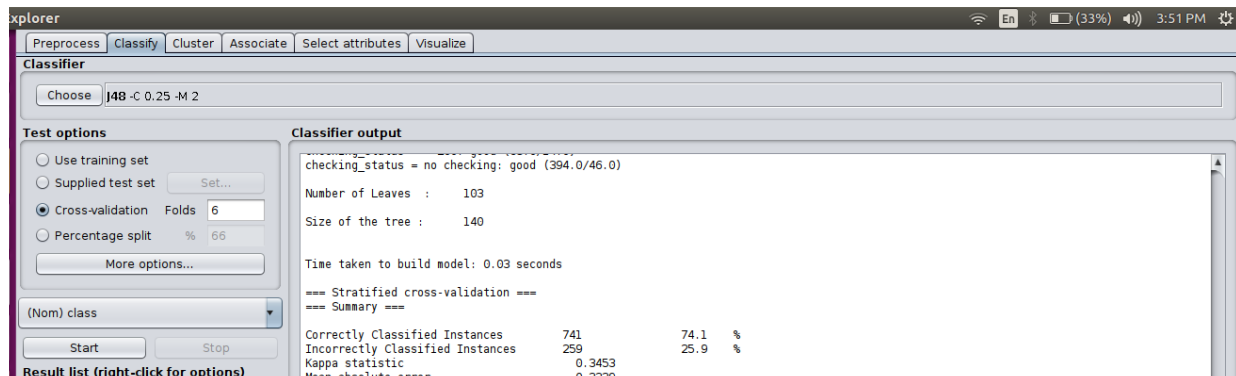
when cross validation is 10 folds , the accuracy we got is 70.5%



when cross validation is 8 folds, accuracy we got is 72.6%



when cross validation is 6 folds , the accuracy we got is: 74.1%

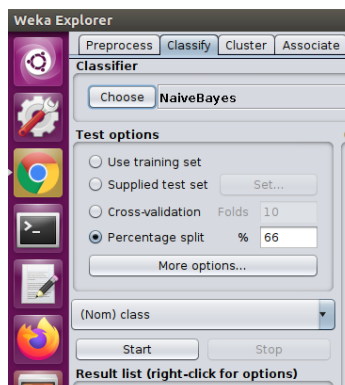


so the maximum accuracy we got is 74.1% when cross validation is 6 folds.

For Bayes:

steps:

- 1.open weka
- 2.click on explore
- 3.select g-credit file
- 4.click on classify
- 5.click on choose and select bayes and choose Naive Bayes



time taken to build model is:0.05 seconds and confusion matrix is:

The screenshot shows the Orange3 Classifier window with the NaiveBayes classifier selected. The 'Test options' section on the left has 'Percentage split' set to 66%. The 'Classifier output' section on the right displays the following information:

[total] 702.0 302.0

Time taken to build model: 0.05 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.04 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	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.852	0.478	0.832	0.852	0.842	0.383	0.804	0.921	good
	0.522	0.148	0.560	0.522	0.540	0.383	0.804	0.592	bad

=== Confusion Matrix ===

a \ b	classified as	
213	a = good	37
43	b = bad	47

The 'Result list' on the left shows the model '12:02:36 - bayes.NaiveBayes' is selected. The 'Status' bar at the bottom shows 'OK'.

when percentage split is 66%,the accuracy we got is:76.4706%

This is a duplicate of the screenshot above, showing the same Orange3 Classifier window with the NaiveBayes classifier selected and the same results displayed.

when percentage split is 71% the accuracy we got is: 76.2069%

The screenshot shows the Orange3 Classifier window with the NaiveBayes classifier selected. The 'Test options' tab is active, showing 'Percentage split' at 71%. The 'Classifier output' tab displays the following results:

Summary

Metric	Value	Percentage
Correctly Classified Instances	221	76.2069 %
Incorrectly Classified Instances	69	23.7931 %
Kappa statistic	0.3773	
Mean absolute error	0.2847	
Root mean squared error	0.4101	
Relative absolute error	68.8121 %	
Root relative squared error	92.0178 %	
Total Number of Instances	290	

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.762	0.395	0.756	0.762	0.758	0.378	0.792	0.820	good
									bad

Confusion Matrix

a \ b	good	bad
good	181	31
bad	38	40

when percentage split is 76% the accuracy we got is: 76.25%

The screenshot shows the Orange3 Classifier window with the NaiveBayes classifier selected. The 'Test options' tab is active, showing 'Percentage split' at 76%. The 'Classifier output' tab displays the following results:

Summary

Metric	Value	Percentage
Correctly Classified Instances	183	76.25 %
Incorrectly Classified Instances	57	23.75 %
Kappa statistic	0.3958	
Mean absolute error	0.2817	
Root mean squared error	0.4075	
Relative absolute error	68.4051 %	
Root relative squared error	91.6924 %	
Total Number of Instances	240	

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.763	0.365	0.764	0.763	0.763	0.396	0.800	0.826	good
									bad

Confusion Matrix

a \ b	good	bad
good	147	29
bad	28	36

when percentage split is 80%, the accuracy we got is: 74.5%

The screenshot shows the Orange3 Classifier window with the NaiveBayes classifier selected. The 'Test options' section has 'Percentage split' set to 80%. The 'Classifier output' pane displays the following results:

```
[total] 702.0 302.0

Time taken to build model: 0.07 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.05 seconds

=== Summary ===
Correctly Classified Instances 149 74.5 %
Incorrectly Classified Instances 51 25.5 %
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

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0.799  0.412  0.850  0.799  0.824  0.368  0.796  0.923  good
0.588  0.201  0.500  0.588  0.541  0.368  0.796  0.539  bad
Weighted Avg.  0.745  0.358  0.761  0.745  0.751  0.368  0.796  0.825

=== Confusion Matrix ===
  a  b  <-- classified as
119 30 | a = good
 21 30 | b = bad
```

The maximum accuracy we got is 76.25% when percentage split is 76%

when cross validation is 10 folds the accuracy we got is: 75.4%

The screenshot shows the Orange3 Classifier window with the NaiveBayes classifier selected. The 'Test options' section has 'Cross-validation' set to 10 folds. The 'Classifier output' pane displays the following results:

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

Time taken to build model: 0.04 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

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0.854  0.503  0.800  0.854  0.831  0.385  0.787  0.891  good
0.497  0.136  0.611  0.497  0.548  0.385  0.787  0.577  bad
Weighted Avg.  0.754  0.393  0.743  0.754  0.746  0.385  0.787  0.797

=== Confusion Matrix ===
  a  b  <-- classified as
605 95 | a = good
151 149 | b = bad
```

when cross validation is 8 folds the accuracy we got is : 75.9%

lorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 8

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:56:17 - rules.ZeroR

12:02:36 - bayes.NaiveBayes

12:07:45 - bayes.NaiveBayes

12:08:56 - bayes.NaiveBayes

12:13:59 - bayes.NaiveBayes

12:16:11 - bayes.NaiveBayes

12:17:21 - 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: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

```
Correctly Classified Instances    759      75.9 %
Incorrectly Classified Instances  241      24.1 %
Kappa statistic                   0.3957
Mean absolute error               0.2936
Root mean squared error          0.4205
Relative absolute error          69.8657 %
Root relative squared error      91.7659 %
Total Number of Instances       1000
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.759	0.383	0.749	0.759	0.752	0.399	0.789	0.798	good
	0.866	0.490	0.805	0.866	0.834	0.399	0.789	0.893	good
	0.510	0.134	0.619	0.510	0.559	0.399	0.789	0.575	bad

=== Confusion Matrix ===

```

a  b  <-- classified as
606 94 | a = good
147 153 | b = bad
```

when cross validation is 6 folds the accuracy we got is: 75.4%

explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 6

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:56:17 - rules.ZeroR

12:02:36 - bayes.NaiveBayes

12:07:45 - bayes.NaiveBayes

12:08:56 - bayes.NaiveBayes

12:13:59 - bayes.NaiveBayes

12:16:11 - bayes.NaiveBayes

12:17:21 - bayes.NaiveBayes

12:18:30 - 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: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

```
Correctly Classified Instances    754      75.4 %
Incorrectly Classified Instances  246      24.6 %
Kappa statistic                   0.3813
Mean absolute error               0.2955
Root mean squared error          0.4222
Relative absolute error          70.3237 %
Root relative squared error      92.1377 %
Total Number of Instances       1000
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.754	0.393	0.743	0.754	0.746	0.385	0.785	0.795	good
	0.864	0.503	0.800	0.864	0.831	0.385	0.785	0.890	good
	0.497	0.136	0.611	0.497	0.548	0.385	0.785	0.573	bad

=== Confusion Matrix ===

```

a  b  <-- classified as
605 95 | a = good
151 149 | b = bad
```

So the maximum accuracy we got is 75.9% when cross validation is 8 folds.