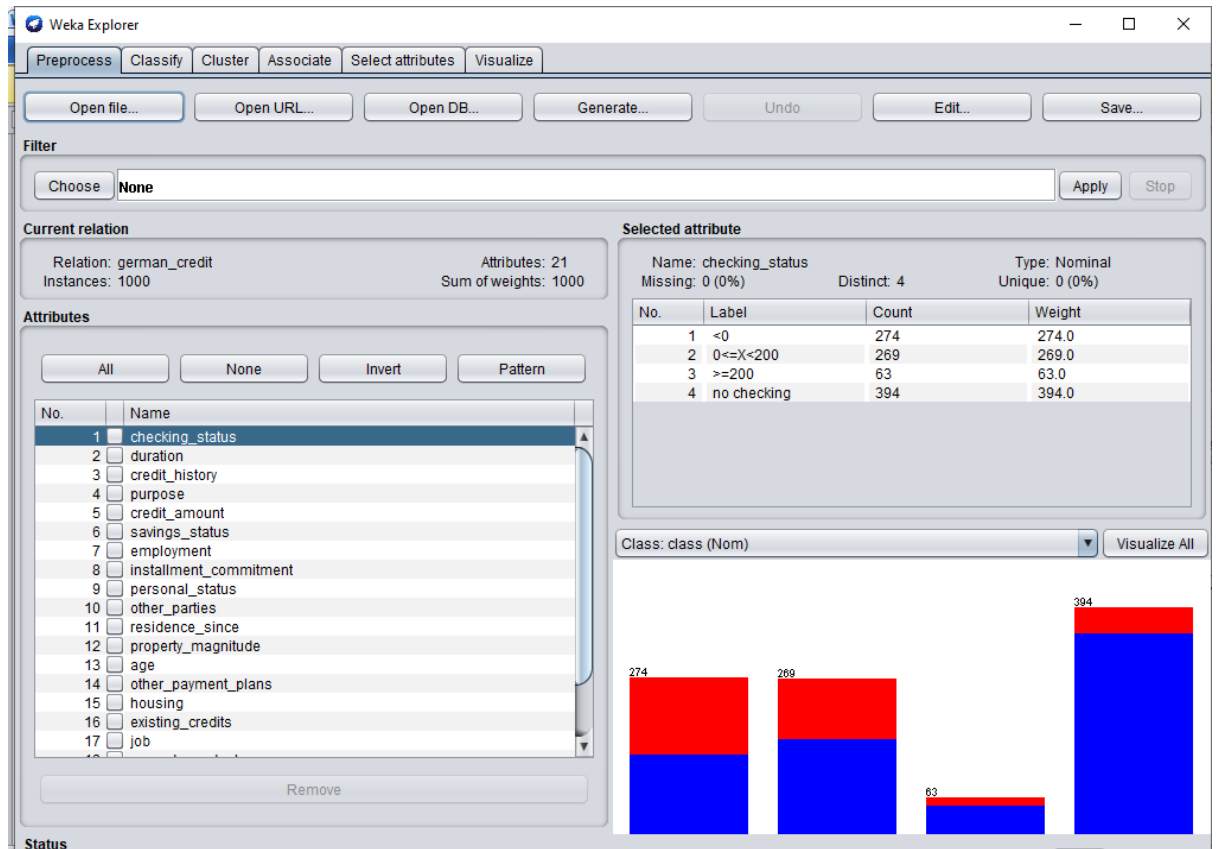


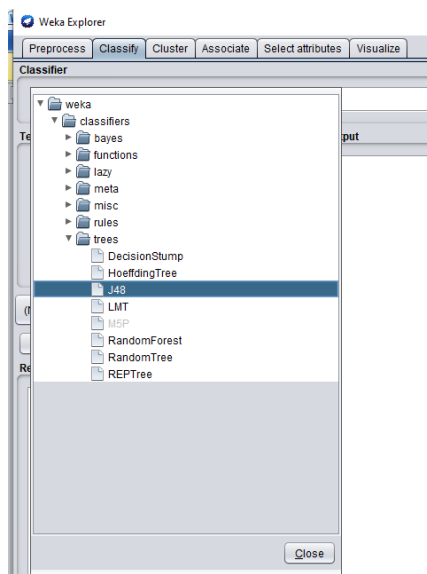
ASSIGNMENT-2

Deriving a Decision Tree from Given Dataset:

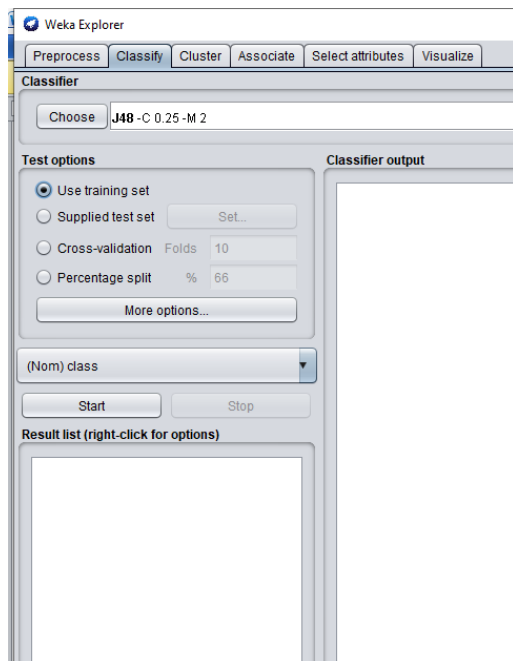
- Open the given Dataset in Weka Here we are considering German Credit Dataset



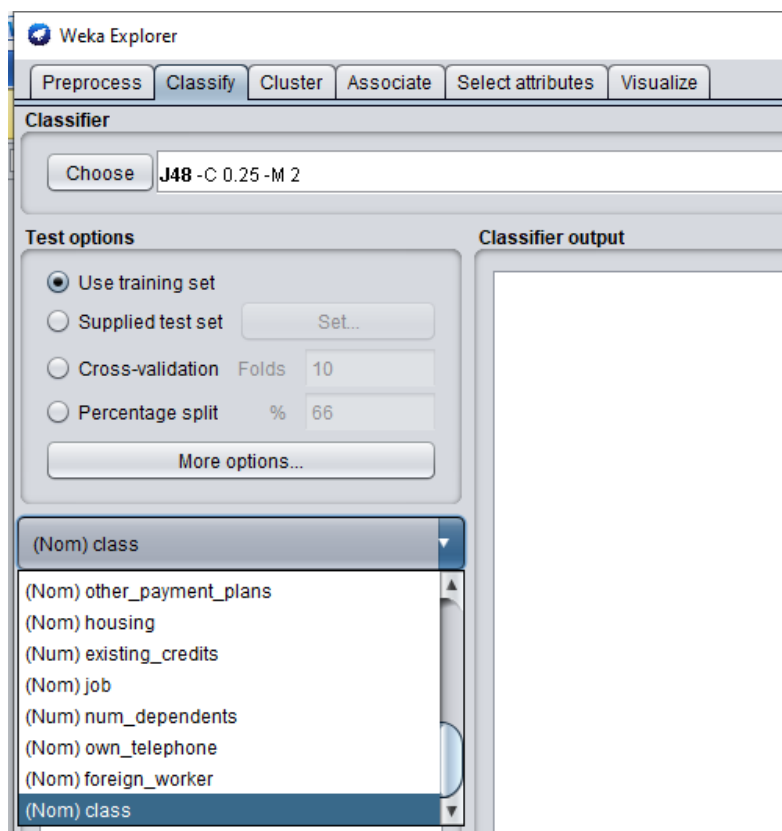
- Click on classify tab and Choose J48 from trees



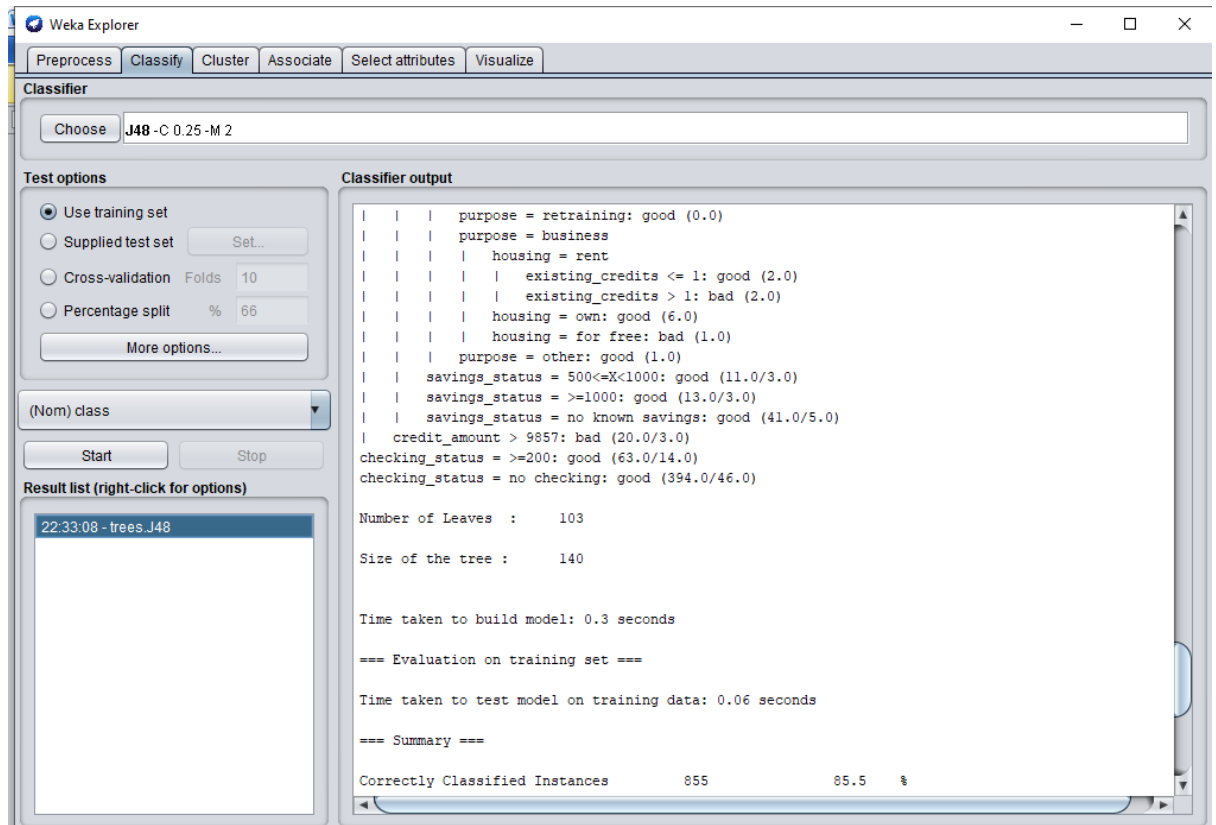
- Choose the Use training set from Test options



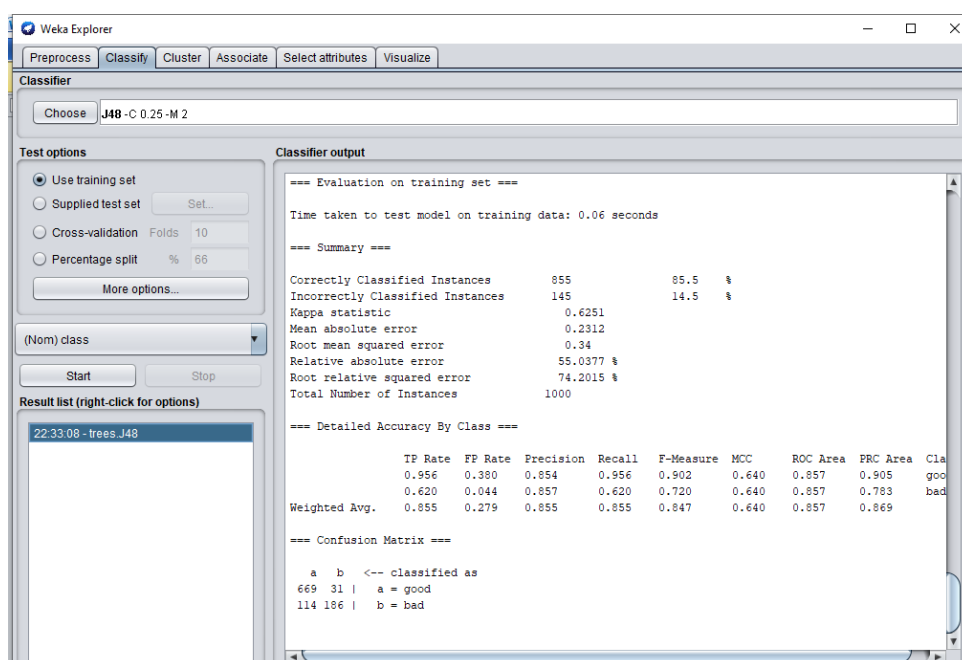
- Now, choose the attribute **class** from the dropdown



- Next click on Start. We will get Output as below:



- Number of leaves = 103
- Number of tree = 140
- Time Taken to build Model = 0.3 sec
- We also get the Confusion Matrix as below:



VARYING THE PERCENTAGE SPLIT FOR ACCURACY:

1) FOR J48 CLASSIFIER:

- Accuracy when Percentage split is 66%

The screenshot shows the Weka Explorer interface with the J48 classifier selected. The 'Test options' section has 'Percentage split' set to 66%. The 'Classifier output' pane displays the following results:

```
| 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: 0.03 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      247           72.6471 %
Incorrectly Classified Instances     93           27.3529 %
Kappa statistic                     0.2687
Mean absolute error                  0.3351
Root mean squared error              0.4836
Relative absolute error              80.8066 %
Root relative squared error          108.7935 %
Total Number of Instances           340

=== Detailed Accuracy By Class ===
```

The 'Result list' on the left shows three entries for 'trees.J48' with timestamps 22:33:08, 22:36:58, and 22:40:37. The bottom of the output pane shows a table header for 'TP Rate', 'FP Rate', 'Precision', 'Recall', 'F-Measure', 'MCC', 'ROC Area', 'PRC Area', and 'Cla'.

- Accuracy when Percentage split is 80%

The screenshot shows the Weka Explorer interface with the J48 classifier selected. The 'Test options' section has 'Percentage split' set to 80%. The 'Classifier output' pane displays the following results:

```
| 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

=== 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

=== Detailed Accuracy By Class ===
```

The 'Result list' on the left shows three entries for 'trees.J48' with timestamps 22:33:08, 22:36:58, and 22:41:26. The bottom of the output pane shows a table header for 'TP Rate', 'FP Rate', 'Precision', 'Recall', 'F-Measure', 'MCC', 'ROC Area', 'PRC Area', and 'Cla'.

2) FOR NAIVEBAYES CLASSIFIER:

- Accuracy when Percentage split is 66%

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section has 'Percentage split' set to 66%. The 'Classifier output' pane displays the following results:

```
=== Evaluation on test split ===
Time taken to test model on test split: 0.01 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  Cla
               0.852    0.478    0.832     0.852    0.842     0.383    0.804    0.921    goo
               0.522    0.148    0.560     0.522    0.540     0.383    0.804    0.592    bad
Weighted Avg.   0.765    0.390    0.760     0.765    0.762     0.383    0.804    0.834

=== Confusion Matrix ===
  a  b  <-- classified as
213 37 | a = good
 43 47 | b = bad
```

- Accuracy when Percentage split is 80%

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

```
=== Evaluation on test split ===
Time taken to test model on test split: 0.01 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

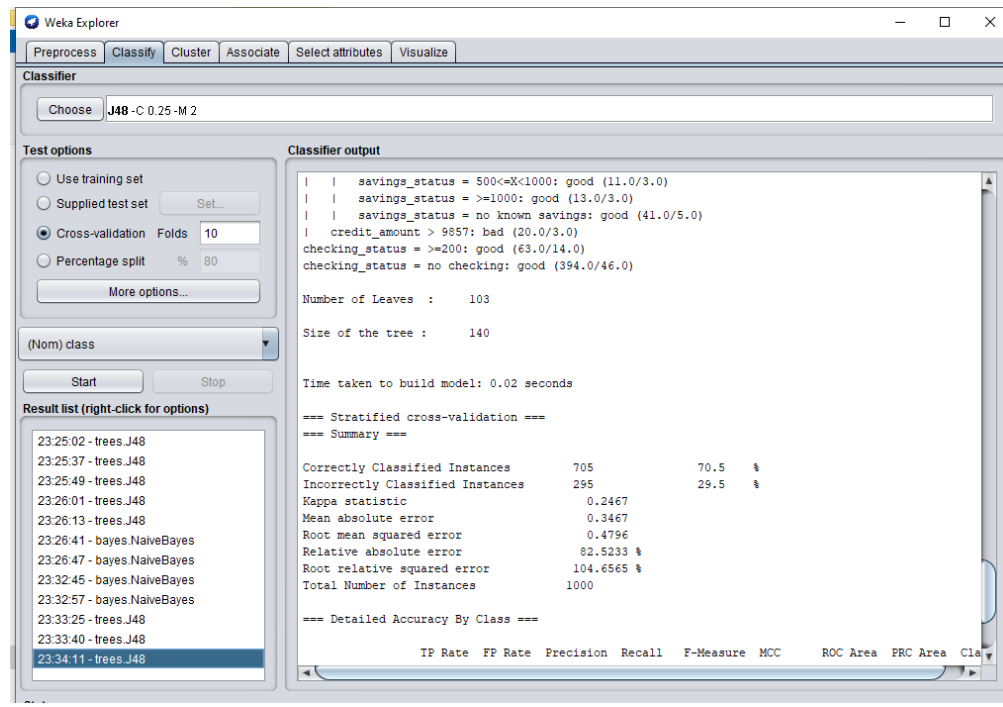
=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
               0.799    0.412    0.850     0.799    0.824     0.368    0.796    0.923    goo
               0.588    0.201    0.500     0.588    0.541     0.368    0.796    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
```

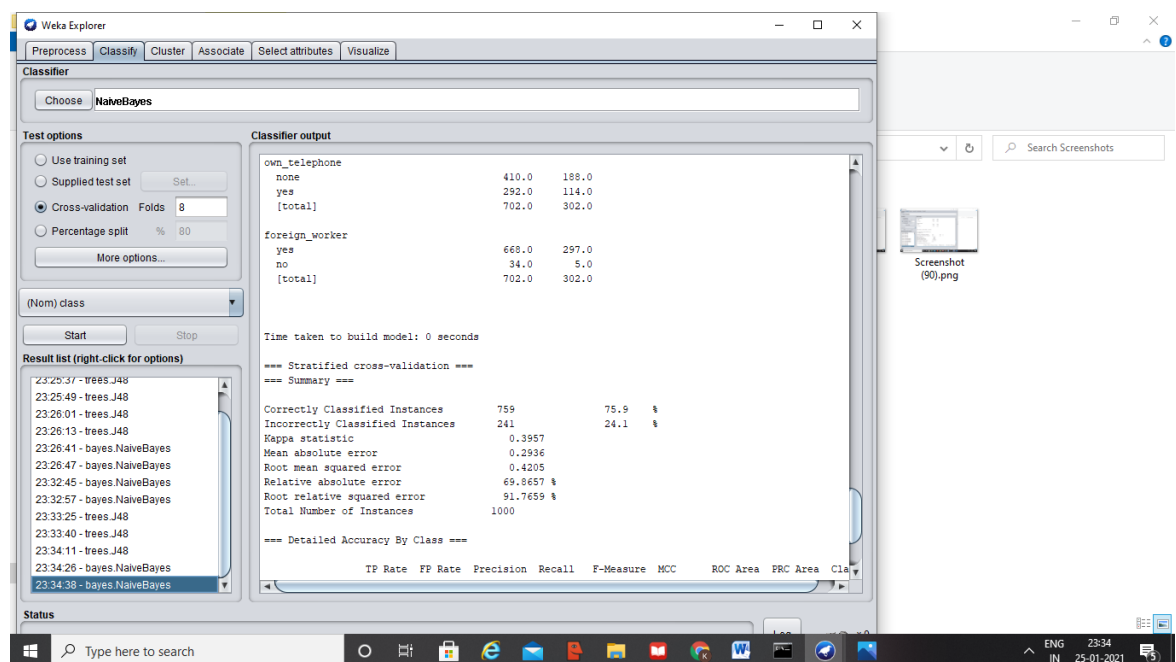
VARYING THE CROSS VALIDATIONS FOR ACCURACY:

1) FOR J48 CLASSIFIER:

- Accuracy when Cross – validation is 10



- Accuracy when Cross – validation is 8



2) FOR NAIVEBAYES CLASSIFIER:

- Accuracy when Cross – validation is 10

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section has 'Cross-validation' set to 10 folds. The 'Classifier output' pane displays the following data:

Class	Actual \ Predicted	none	yes	[total]
own_telephone	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 seconds

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

Metric	Value	%
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
--	---------	---------	-----------	--------	-----------	-----	----------	----------	-------

- Accuracy when Cross – validation is 8

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section has 'Cross-validation' set to 8 folds. The 'Classifier output' pane displays the following data:

Class	Actual \ Predicted	none	yes	[total]
own_telephone	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 seconds

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

Metric	Value	%
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
--	---------	---------	-----------	--------	-----------	-----	----------	----------	-------

