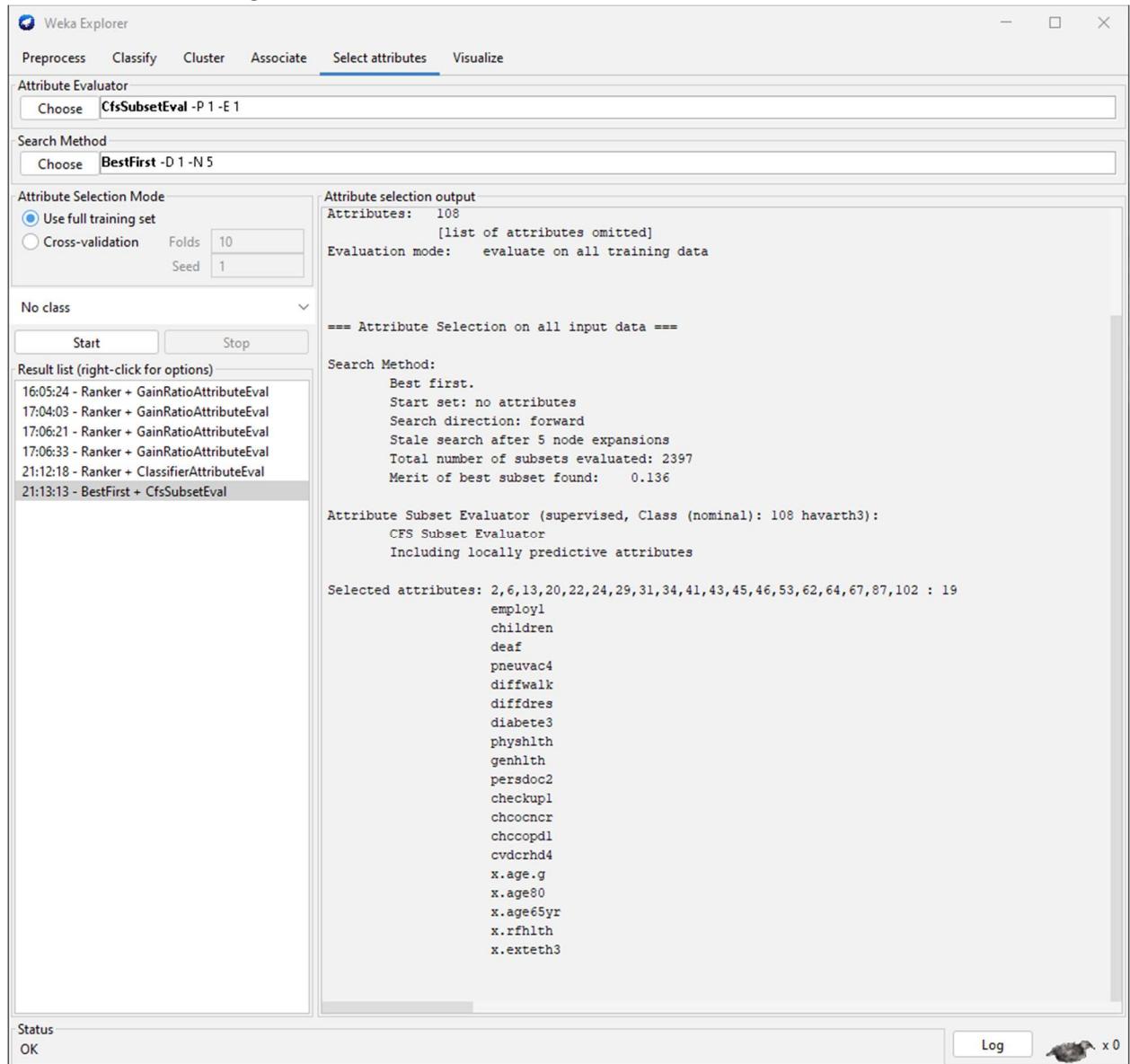


- 1) Choose attributes using CFSSubsetEval and BestFirst search method:



Selected attributes:

employ1, children, deaf, pneuvac4, diffwalk, diffdres, diabete3, physhlth, genhlth, persdoc2, checkup1, chcocnrcr, chccopd1, cvdcrhd4, x.age.g, x.age80, x.age65yr, x.rfhlth, x.exteth3

Save Training and Test sets with these attributes and run 5 different classifier models:

- Naïve Bayes:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. Under 'Classifier', 'NaiveBayes' is chosen. In the 'Test options' section, 'Supplied test set' is selected with a size of 10. The 'Classifier output' pane displays the following results:

Time taken to build model: 0.01 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.03 seconds

==== Summary ===

	Correctly Classified Instances	2937	72.3934 %
Incorrectly Classified Instances	1120	27.6066 %	
Kappa statistic	0.4143		
Mean absolute error	0.2839		
Root mean squared error	0.4771		
Relative absolute error	63.4834 %		
Root relative squared error	100.9023 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	IP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.735	0.299	0.829	0.735	0.779	0.420	0.794	0.879	0.798	2
0.701	0.265	0.575	0.701	0.632	0.420	0.794	0.638	0.798	1
Weighted Avg.	0.724	0.287	0.743	0.724	0.729	0.420	0.794	0.798	

==== Confusion Matrix ===

	a	b	<-- classified as
1977	711	1	a = 2
409	960	1	b = 1

Status: OK

Log x 0

- Logistic Regression:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Logistic -R 1.0E-8 -M -1 -num-decimal-places 4**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

18:28:55 - bayes.NaiveBayes
18:33:49 - functions.Logistic

Classifier output

```
Time taken to build model: 0.53 seconds
==== Evaluation on test set ====
Time taken to test model on supplied test set: 0.03 seconds
==== Summary ====
Correctly Classified Instances      3053      75.2526 %
Incorrectly Classified Instances   1004      24.7474 %
Kappa statistic                   0.4054
Mean absolute error               0.3303
Root mean squared error          0.4086
Relative absolute error           73.8614 %
Root relative squared error     86.4235 %
Total Number of Instances        4057

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


|               | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| 0.885         | 0.508   | 0.774   | 0.885     | 0.826  | 0.416     | 0.804 | 0.882    | 0.805    | 2     |
| 0.492         | 0.115   | 0.686   | 0.492     | 0.573  | 0.416     | 0.804 | 0.654    | 0.805    | 1     |
| Weighted Avg. | 0.753   | 0.375   | 0.744     | 0.753  | 0.741     | 0.416 | 0.804    | 0.805    |       |


==== Confusion Matrix ====

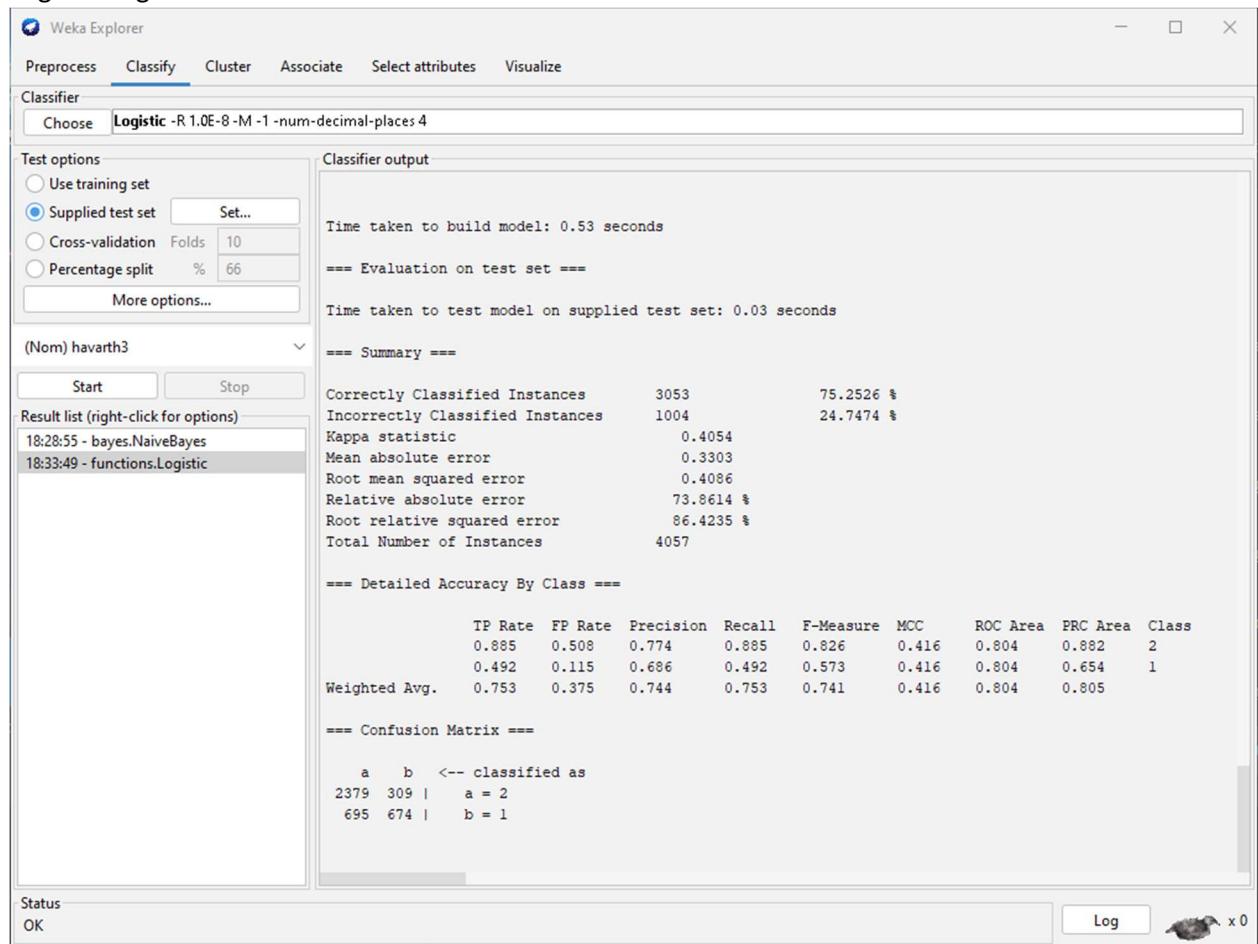

```
a b <-- classified as
2379 309 | a = 2
695 674 | b = 1
```



Status OK



Log


```

- Multilayer Perceptron (Neural Network):

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a'. The 'Test options' section has 'Supplied test set' selected. The 'Classifier output' pane displays the following results:

```
Time taken to build model: 177.5 seconds
===
Evaluation on test set ===
Time taken to test model on supplied test set: 0.09 seconds
===
Summary ===
Correctly Classified Instances      2943      72.5413 %
Incorrectly Classified Instances   1114      27.4587 %
Kappa statistic                      0.3198
Mean absolute error                  0.3141
Root mean squared error              0.4576
Relative absolute error              70.2448 %
Root relative squared error        96.7757 %
Total Number of Instances           4057

===
Detailed Accuracy By Class ===
IP Rate    FP Rate    Precision    Recall    F-Measure    MCC    ROC Area    PRC Area    Class
0.894     0.605     0.744     0.894     0.812     0.339     0.743     0.837     2
0.395     0.106     0.654     0.395     0.493     0.339     0.743     0.579     1
Weighted Avg.    0.725     0.437     0.713     0.725     0.704     0.339     0.743     0.750

===
Confusion Matrix ===
a      b      <-- classified as
2402  286 |  a = 2
828  541 |  b = 1
```

The status bar at the bottom left shows 'OK'.

- J48:

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) havarth3

Start Stop

Result list (right-click for options)
18:28:55 - bayes.NaiveBayes
18:33:49 - functions.Logistic
18:34:27 - functions.MultilayerPerceptron
18:39:10 - trees;J48

Classifier output

```
Time taken to build model: 0.23 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.02 seconds

===
Summary ===

Correctly Classified Instances      2958      72.911 %
Incorrectly Classified Instances   1099      27.089 %
Kappa statistic                   0.3673
Mean absolute error               0.3437
Root mean squared error          0.4417
Relative absolute error           76.8654 %
Root relative squared error     93.4236 %
Total Number of Instances        4057

===
Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.840    0.488    0.772    0.840    0.804    0.371    0.742    0.823    2
          0.512    0.160    0.619    0.512    0.561    0.371    0.742    0.548    1
Weighted Avg.       0.729    0.377    0.720    0.729    0.722    0.371    0.742    0.730

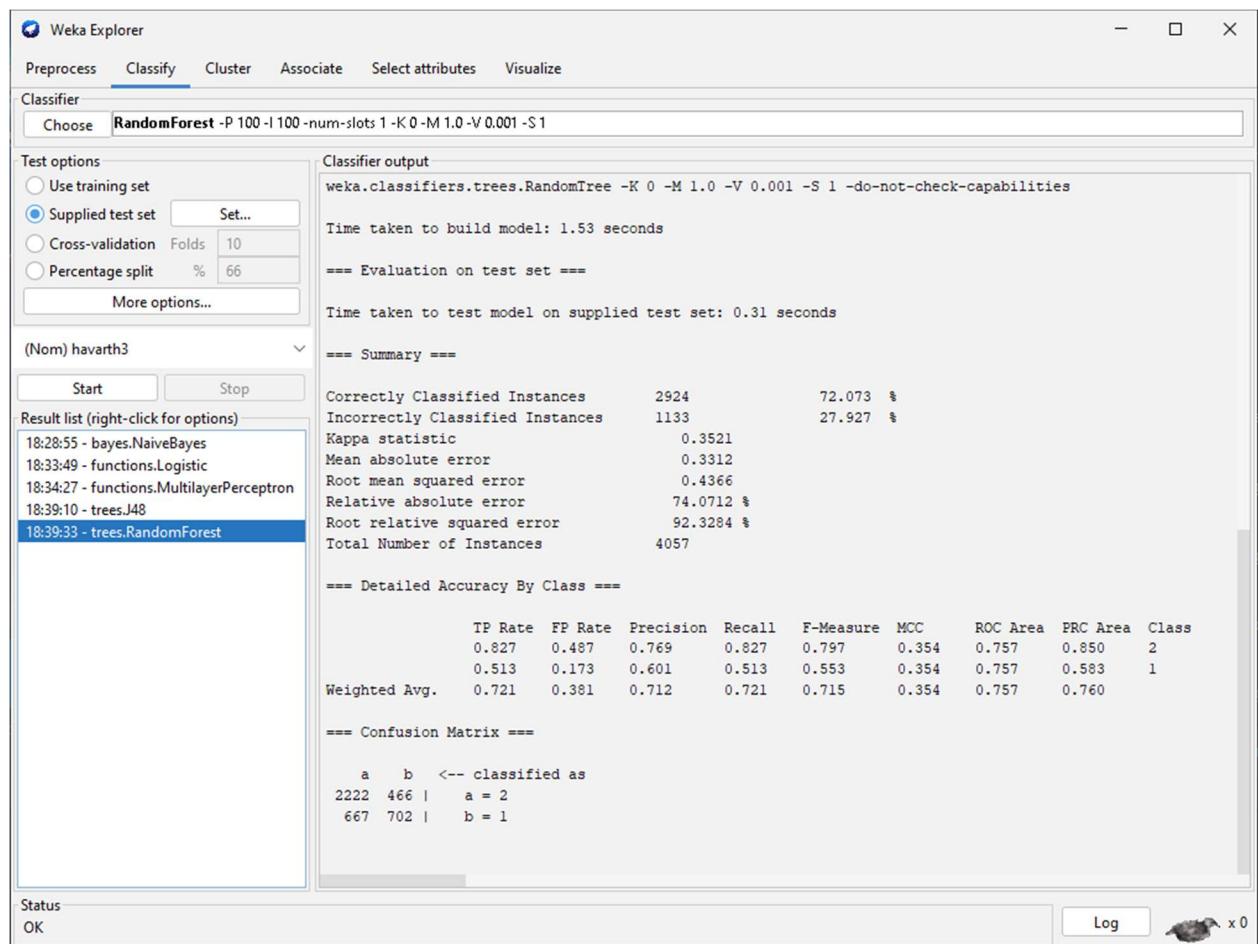
===
Confusion Matrix ===

      a     b  <- classified as
2257  431 |  a = 2
 668  701 |  b = 1
```

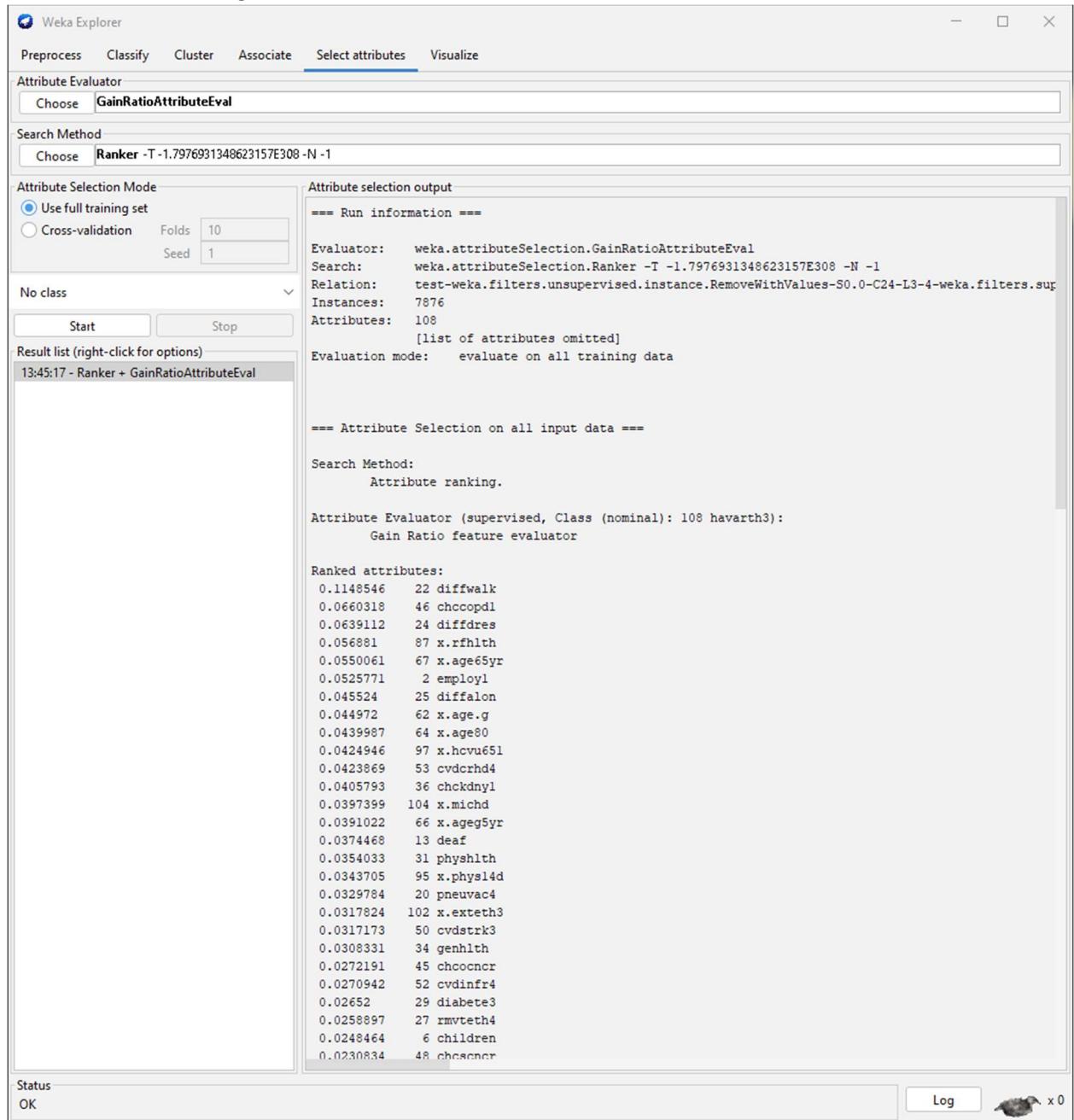
Status
OK

Log x 0

- Random Forest:



- 2) Choose attributes using GainRatioAttributeEval and the Ranker Search Method:



We select the top 20 attributes selected by this selection method. This is roughly 20% of the data, which is enough to make accurate classifications.

Selected attributes:

employ1, deaf, pneuvac4, diffwalk, diffdres, diffalon, physhlth, chckdny1, chccopd1, cvdstrk3, cvdcrhd4, x.age.g, x.age80, x.ageg5yr, x.age65yr, x.rfhlth, x.phys14d, x.hcvu651, x.exteth3, x.michd

Run 5 different classifiers:

- Naïve Bayes:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)
13:55:45 - bayes.NaiveBayes

Classifier output

```
x.exteth3
1      3068.0  941.0
2      2063.0  1641.0
9      90.0    79.0
[total] 5221.0  2661.0

x.michd
2      4884.0  2213.0
1      301.0   412.0
[total] 5185.0  2625.0
```

Time taken to build model: 0.02 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

	Correctly Classified Instances	2869	70.7173 %
Incorrectly Classified Instances	1188	29.2827 %	
Kappa statistic	0.385		
Mean absolute error	0.291		
Root mean squared error	0.4957		
Relative absolute error	65.0804 %		
Root relative squared error	104.8397 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.712	0.302	0.822	0.712	0.763	0.392	0.787	0.879	2	
0.698	0.288	0.552	0.698	0.617	0.392	0.787	0.624	1	
Weighted Avg.	0.707	0.297	0.731	0.707	0.714	0.392	0.787	0.793	

==== Confusion Matrix ===

	a	b	<-- classified as
1913	775		a = 2
413	956		b = 1

Status OK Log x 0

- Logistic

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **Logistic -R 1.0E-8 -M -1 -num-decimal-places 4**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

13:55:45 - bayes.NaiveBayes
13:57:00 - functions.Logistic

Classifier output

x.phys14d=1	1.0951
x.phys14d=2	1.0511
x.phys14d=9	0.7605
x.phys14d=3	0.8114
x.hcvu65l=1	0.952
x.hcvu65l=9	1.0085
x.hcvu65l=2	1.1574
x.exteth3=1	1.0919
x.exteth3=2	0.92
x.exteth3=9	0.9447
x.michd=1	1.3323

Time taken to build model: 0.32 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.02 seconds

==== Summary ===

Correctly Classified Instances	3027	74.6118 %
Incorrectly Classified Instances	1030	25.3882 %
Kappa statistic	0.3834	
Mean absolute error	0.3337	
Root mean squared error	0.4104	
Relative absolute error	74.6156 %	
Root relative squared error	86.7989 %	
Total Number of Instances	4057	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.891	0.538	0.765	0.891	0.823	0.398	0.801	0.880	0.2	
0.462	0.109	0.683	0.462	0.551	0.398	0.801	0.656	0.1	
Weighted Avg.	0.746	0.393	0.737	0.746	0.731	0.398	0.801	0.804	

==== Confusion Matrix ===

a	b	<-- classified as
2394	294	a = 2
736	633	b = 1

Status OK Log x 0

- Multilayer Perceptron (Neural Network):

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. Under 'Classifier', 'MultilayerPerceptron' is chosen with parameters: -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a. The 'Test options' section shows 'Supplied test set' selected. The 'Classifier output' pane displays the model's structure and evaluation results.

Classifier output:

```

Attrib x.hcvu651=2 4.641267723672673
Attrib x.exteth3=1 -0.06599242437572077
Attrib x.exteth3=2 -2.8327089553232843
Attrib x.exteth3=9 2.884267361439933
Attrib x.michd=1 -3.887841737713513

Class 2
Input
Node 0

Class 1
Input
Node 1

```

Time taken to build model: 135.37 seconds

==== Evaluation on test set ====

Time taken to test model on supplied test set: 0.05 seconds

==== Summary ====

	Correctly Classified Instances	72.2948 %
Incorrectly Classified Instances	1124	27.7052 %
Kappa statistic	0.3218	
Mean absolute error	0.3173	
Root mean squared error	0.4516	
Relative absolute error	70.9496 %	
Root relative squared error	95.5149 %	
Total Number of Instances	4057	

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.881	0.587	0.747	0.881	0.808	0.336	0.744	0.829	0.829	2
0.413	0.119	0.638	0.413	0.502	0.336	0.744	0.583	0.583	1
Weighted Avg.	0.723	0.429	0.710	0.723	0.705	0.336	0.744	0.746	

==== Confusion Matrix ====

		a	b	<-- classified as
		2367	321	a = 2
		803	566	b = 1

Status: OK

- J48

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) havarth3

Start Stop

Result list (right-click for options)

13:55:45 - bayes.NaiveBayes
13:57:00 - functions.Logistic
13:57:51 - functions.MultilayerPerceptron
14:01:30 - trees.J48

Classifier output

```
| | | x.rfhlth = 1: 1 (10.35/2.26)
| | | x.rfhlth = 2: 2 (5.11/1.06)
| | | x.rfhlth = 9: 2 (1.0)
| | chccopdl = 1: 1 (3.03/0.01)
| | chccopdl = 7: 1 (0.0)
| | chccopdl = 9: 2 (0.0)
| diffwalk = 9: 2 (6.16/2.05)
```

Number of Leaves : 355

Size of the tree : 527

Time taken to build model: 0.11 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances	3009	74.1681 %
Incorrectly Classified Instances	1048	25.8319 %
Kappa statistic	0.3869	
Mean absolute error	0.3366	
Root mean squared error	0.4333	
Relative absolute error	75.2672 %	
Root relative squared error	91.6306 %	
Total Number of Instances	4057	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.865	0.500	0.772	0.865	0.816	0.394	0.756	0.831	0.831	2
0.500	0.135	0.653	0.500	0.566	0.394	0.756	0.569	0.569	1
Weighted Avg.	0.742	0.377	0.732	0.742	0.732	0.394	0.756	0.742	

==== Confusion Matrix ===

a	b	<-- classified as
2325	363	a = 2
685	684	b = 1

Status OK Log x 0

- Random Forest:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section has 'Supplied test set' selected. The 'Classifier output' pane displays the following text:

```

x.exteth3
x.michd
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

== Classifier model (full training set) ==

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 1.48 seconds

== Evaluation on test set ==

Time taken to test model on supplied test set: 0.28 seconds

== Summary ==

Correctly Classified Instances      2923          72.0483 %
Incorrectly Classified Instances   1134          27.9517 %
Kappa statistic                   0.3504
Mean absolute error               0.3308
Root mean squared error           0.4355
Relative absolute error            73.9831 %
Root relative squared error       92.0941 %
Total Number of Instances         4057

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.828    0.491    0.768     0.828    0.797     0.353   0.758    0.851    2
      0.509    0.172    0.601     0.509    0.551     0.353   0.758    0.589    1
Weighted Avg.      0.720    0.383    0.712     0.720    0.714     0.353   0.758    0.763

== Confusion Matrix ==

  a   b  <- classified as
2226 462 |   a = 2
 672 697 |   b = 1

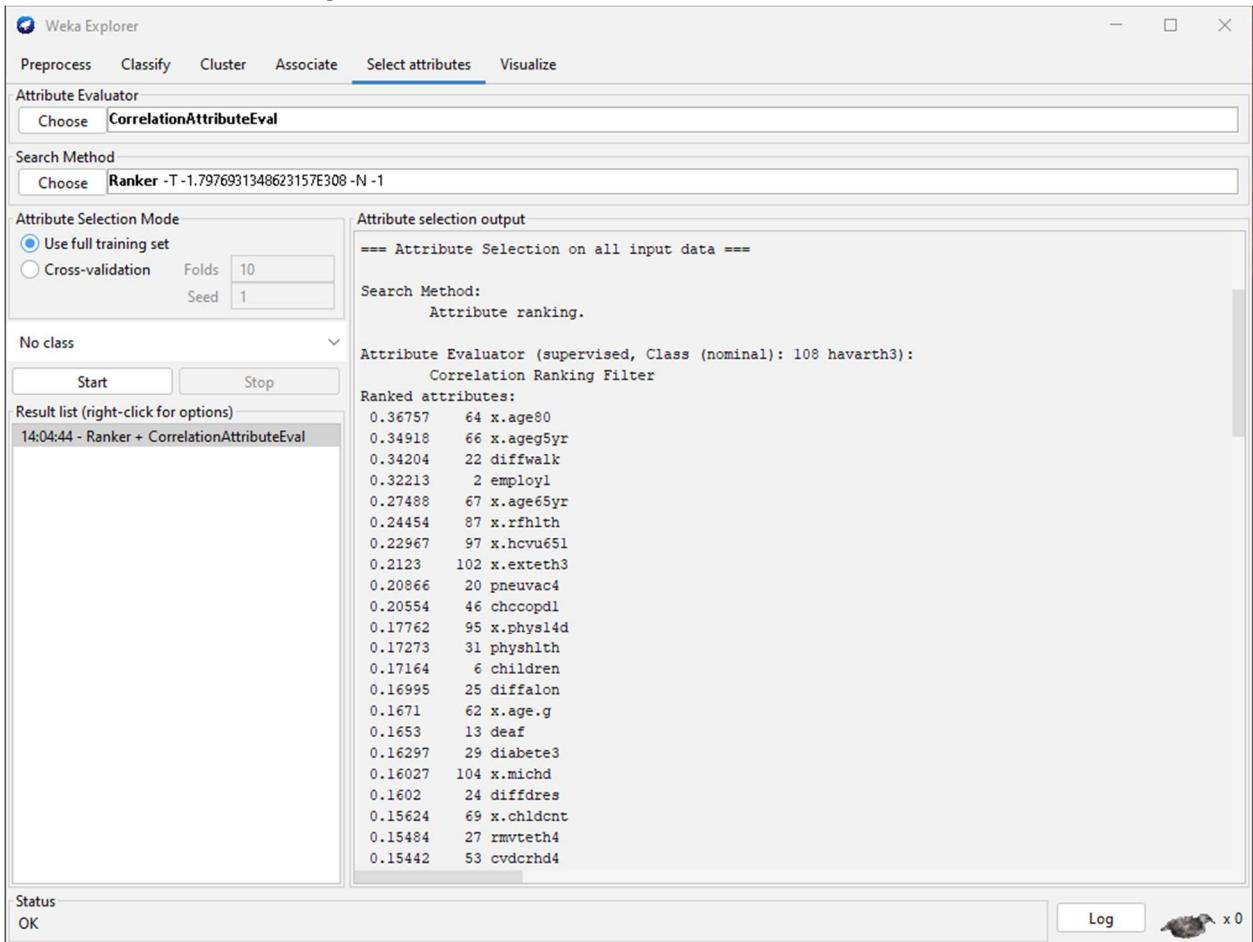
```

The 'Result list' pane shows the following entries:

- 13:55:45 - bayes.NaiveBayes
- 13:57:00 - functions.Logistic
- 13:57:51 - functions.MultilayerPerceptron
- 14:01:30 - trees.J48
- 14:01:57 - trees.RandomForest

The status bar at the bottom left shows 'Status OK' and the bottom right shows a 'Log' button.

3) Choose Attributes using CorrelationAttributeEval with Ranker search method:



Again, we choose 20 attributes as this is roughly 20% of the data.

Selected Attributes:

employ1, children, deaf, pneuvac4, diffwalk, diffdres, diffalon, diabete3, physhlth, chccopd1, x.age.g, x.age80, x.ageg5yr, x.age65yr, x.chldcnt, x.rfhlth, x.phys14d, x.hcvu651, x.exteth3, x.michd

Then run 5 classifiers:

- Naïve Bayes:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:14:29 - bayes.NaiveBayes

Classifier output

```
Time taken to build model: 0.01 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

===
Summary ===

Correctly Classified Instances      2874          70.8405 %
Incorrectly Classified Instances   1183          29.1595 %
Kappa statistic                   0.3894
Mean absolute error               0.2961
Root mean squared error           0.5043
Relative absolute error            66.205 %
Root relative squared error       106.6645 %
Total Number of Instances         4057

===
Detailed Accuracy By Class ===

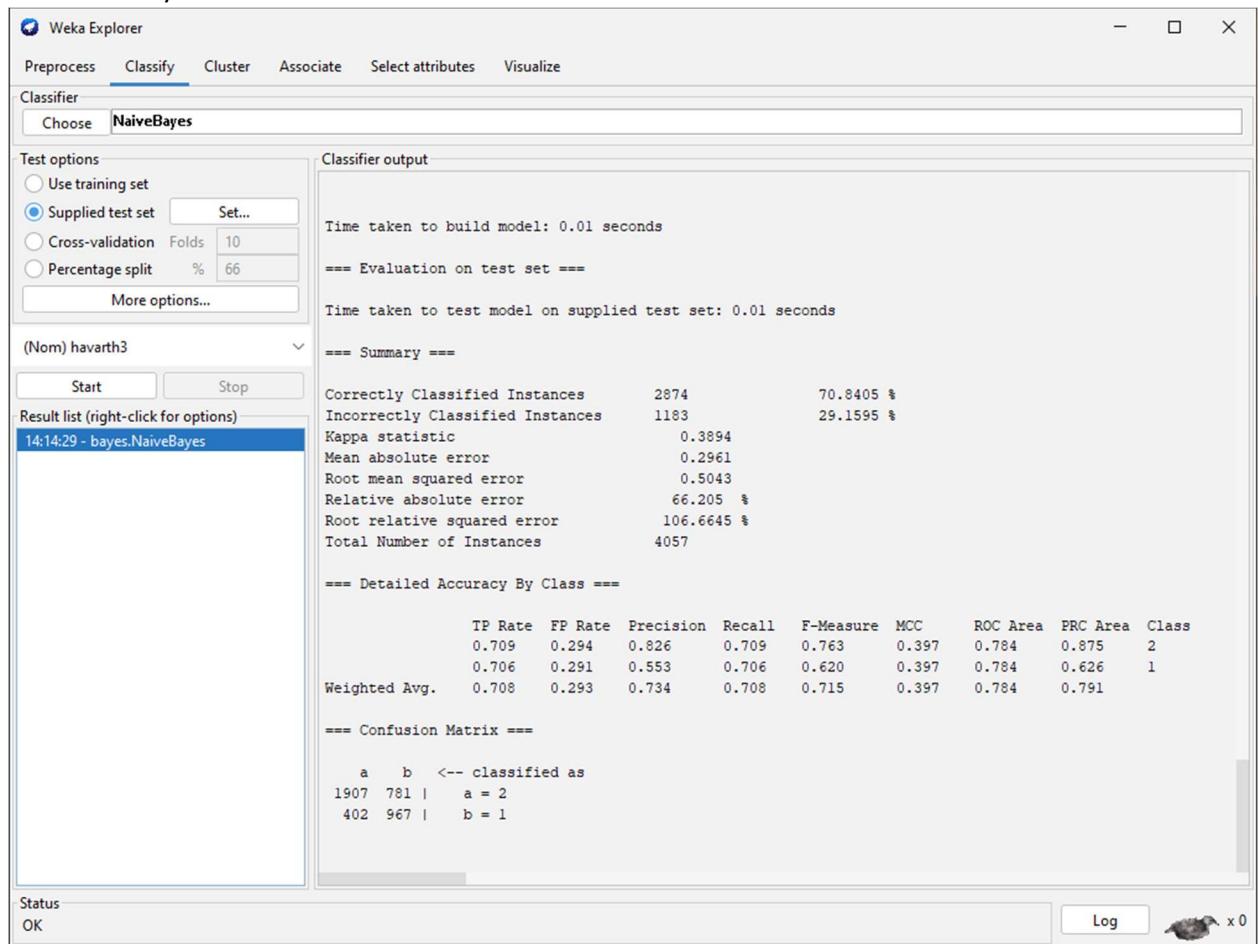
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.709   0.294    0.826     0.709    0.763     0.397   0.784    0.875     2
      0.706   0.291    0.553     0.706    0.620     0.397   0.784    0.626     1
Weighted Avg.      0.708   0.293    0.734     0.708    0.715     0.397   0.784    0.791

===
Confusion Matrix ===

      a     b  <-- classified as
1907  781 |  a = 2
 402  967 |  b = 1
```

Status OK

Log x 0



- Logistic Regression:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Logistic -R 1.0E-8 -M -1 -num-decimal-places 4**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:14:29 - bayes.NaiveBayes
14:15:24 - functions.Logistic

Classifier output

```
Time taken to build model: 0.3 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

===
Summary ===

Correctly Classified Instances      3031          74.7104 %
Incorrectly Classified Instances   1026          25.2896 %
Kappa statistic                   0.3867
Mean absolute error               0.3343
Root mean squared error          0.4116
Relative absolute error           74.749 %
Root relative squared error     87.0441 %
Total Number of Instances        4057

===
Detailed Accuracy By Class ===

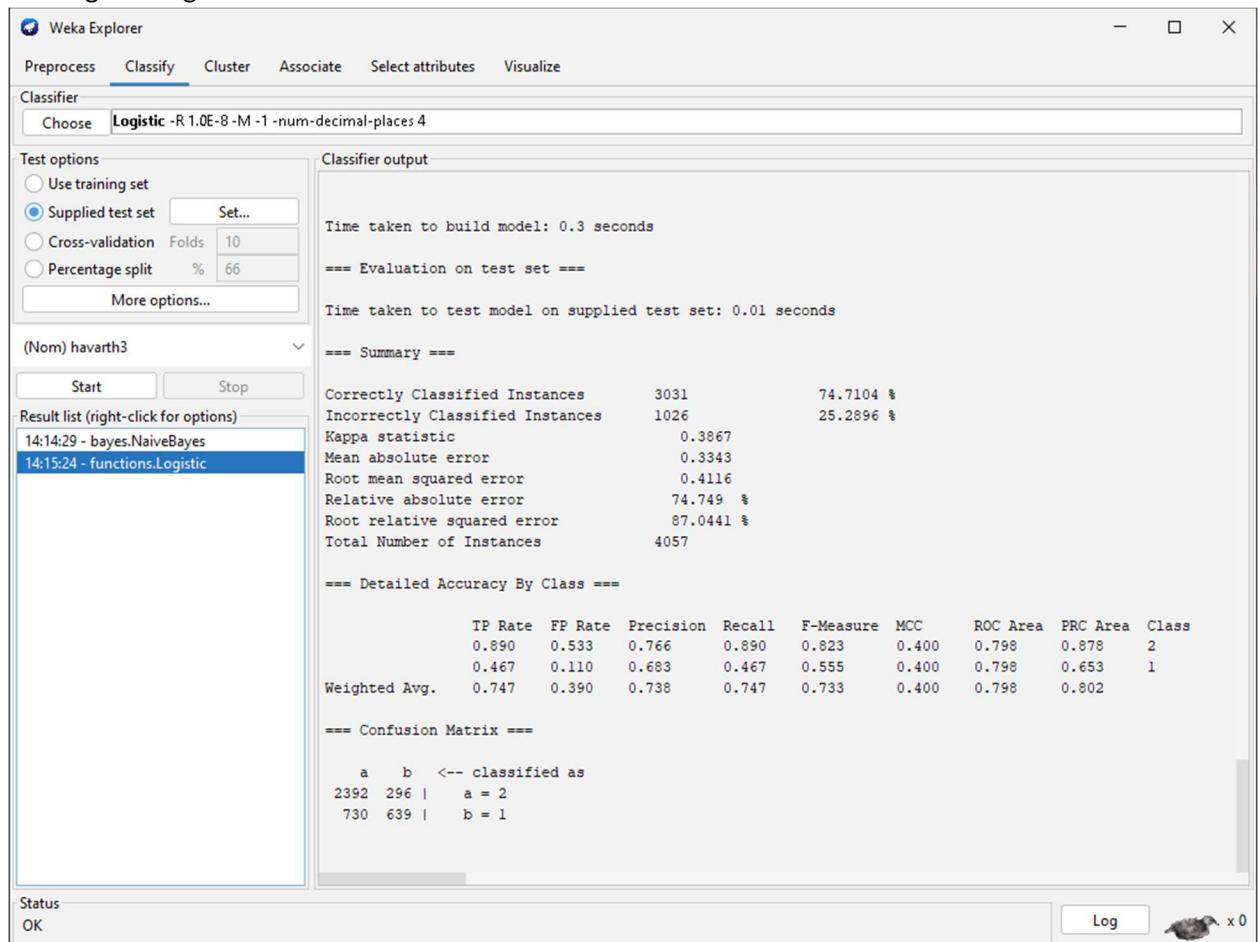
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.890   0.533    0.766    0.890    0.823     0.400   0.798    0.878    2
      0.467   0.110    0.683    0.467    0.555     0.400   0.798    0.653    1
Weighted Avg.      0.747   0.390    0.738    0.747    0.733     0.400   0.798    0.802

===
Confusion Matrix ===

      a     b  <-- classified as
2392  296 |  a = 2
730   639 |  b = 1
```

Status OK

Log x 0



- Multilayer Perceptron (Neural Network):

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:14:29 - bayes.NaiveBayes
14:15:24 - functions.Logistic
14:16:12 - functions.MultilayerPerceptron

Classifier output

```
Time taken to build model: 142.94 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.05 seconds

===
Summary ===

Correctly Classified Instances      2952           72.7631 %
Incorrectly Classified Instances   1105           27.2369 %
Kappa statistic                   0.3257
Mean absolute error               0.304
Root mean squared error          0.449
Relative absolute error           67.9907 %
Root relative squared error     94.9517 %
Total Number of Instances        4057

===
Detailed Accuracy By Class ===

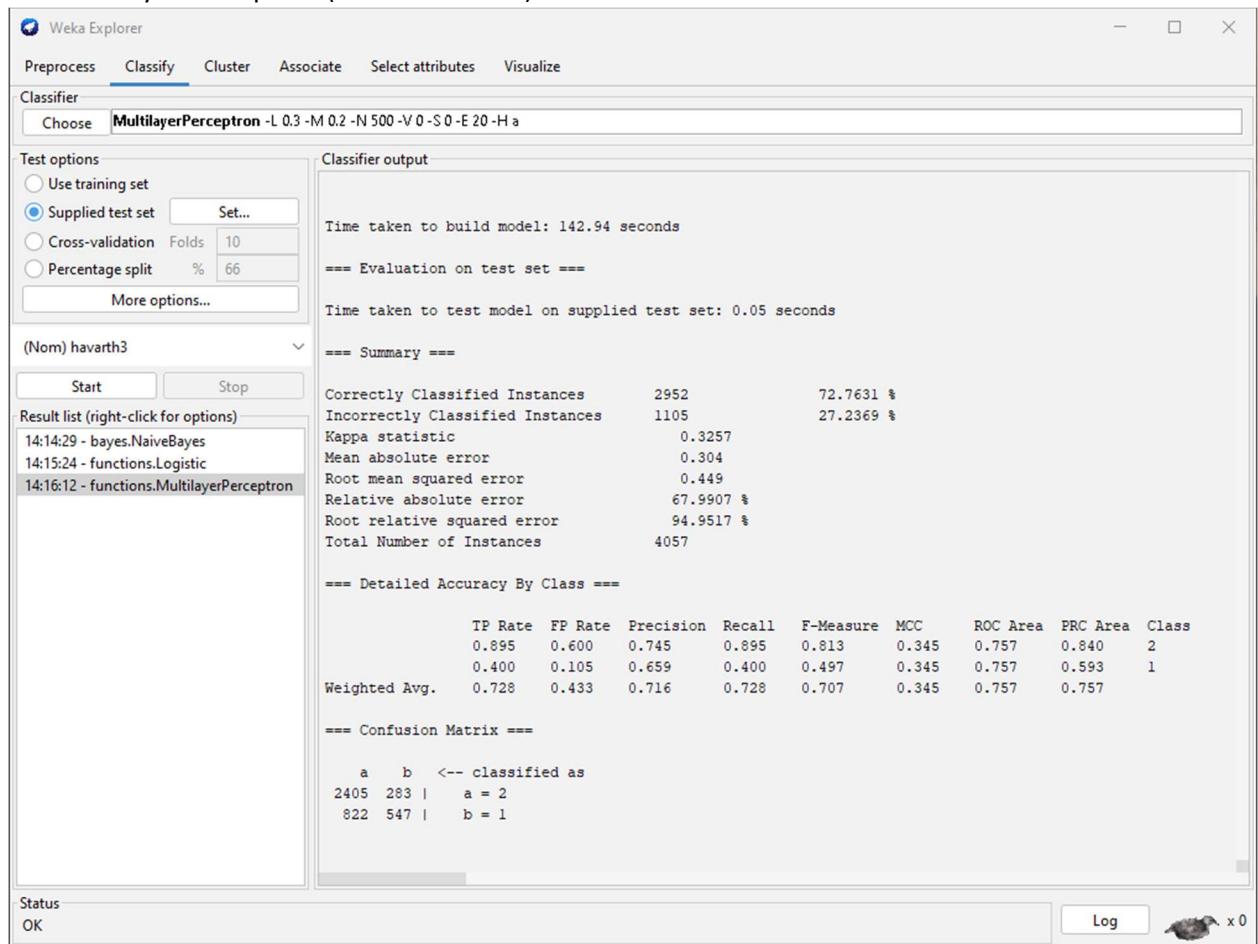
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.895   0.600    0.745    0.895    0.813     0.345   0.757    0.840    2
      0.400   0.105    0.659    0.400    0.497     0.345   0.757    0.593    1
Weighted Avg.      0.728   0.433    0.716    0.728    0.707     0.345   0.757    0.757

===
Confusion Matrix ===

      a     b  <-- classified as
2405  283 |  a = 2
 822  547 |  b = 1
```

Status OK

Log x 0



- J48:

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) havarth3

Start Stop

Result list (right-click for options)

14:14:29 - bayes.NaiveBayes
14:15:24 - functions.Logistic
14:16:12 - functions.MultilayerPerceptron
14:25:32 - trees.J48

Classifier output

```
Time taken to build model: 0.09 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

===
Summary ===

Correctly Classified Instances      3008          74.1435 %
Incorrectly Classified Instances   1049          25.8565 %
Kappa statistic                   0.3829
Mean absolute error               0.3404
Root mean squared error          0.4282
Relative absolute error           76.1205 %
Root relative squared error      90.5525 %
Total Number of Instances         4057

===
Detailed Accuracy By Class ===

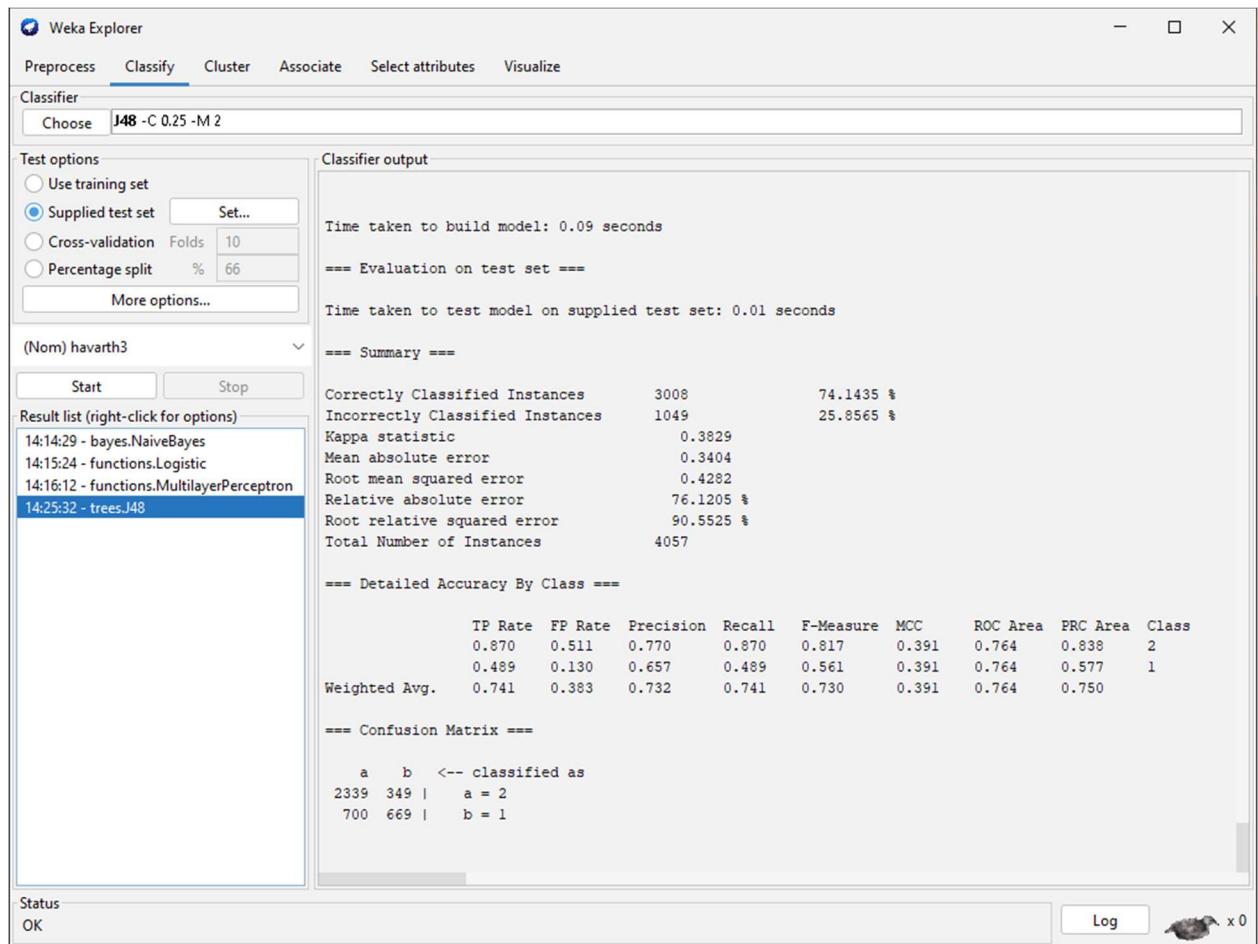
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.870   0.511    0.770     0.870    0.817     0.391   0.764    0.838    2
      0.489   0.130    0.657     0.489    0.561     0.391   0.764    0.577    1
Weighted Avg.      0.741   0.383    0.732     0.741    0.730     0.391   0.764    0.750

===
Confusion Matrix ===

      a     b  <-- classified as
2339  349 |  a = 2
 700  669 |  b = 1
```

Status OK

Log x 0



- Random Forest:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:14:29 - bayes.NaiveBayes
14:15:24 - functions.Logistic
14:16:12 - functions.MultilayerPerceptron
14:25:32 - trees.J48
14:26:00 - trees.RandomForest

Classifier output

```
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
```

Time taken to build model: 1.49 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.25 seconds

==== Summary ===

	Correctly Classified Instances	2938	72.418 %
Incorrectly Classified Instances	1119	27.582 %	
Kappa statistic	0.3601		
Mean absolute error	0.3305		
Root mean squared error	0.4354		
Relative absolute error	73.915 %		
Root relative squared error	92.0749 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

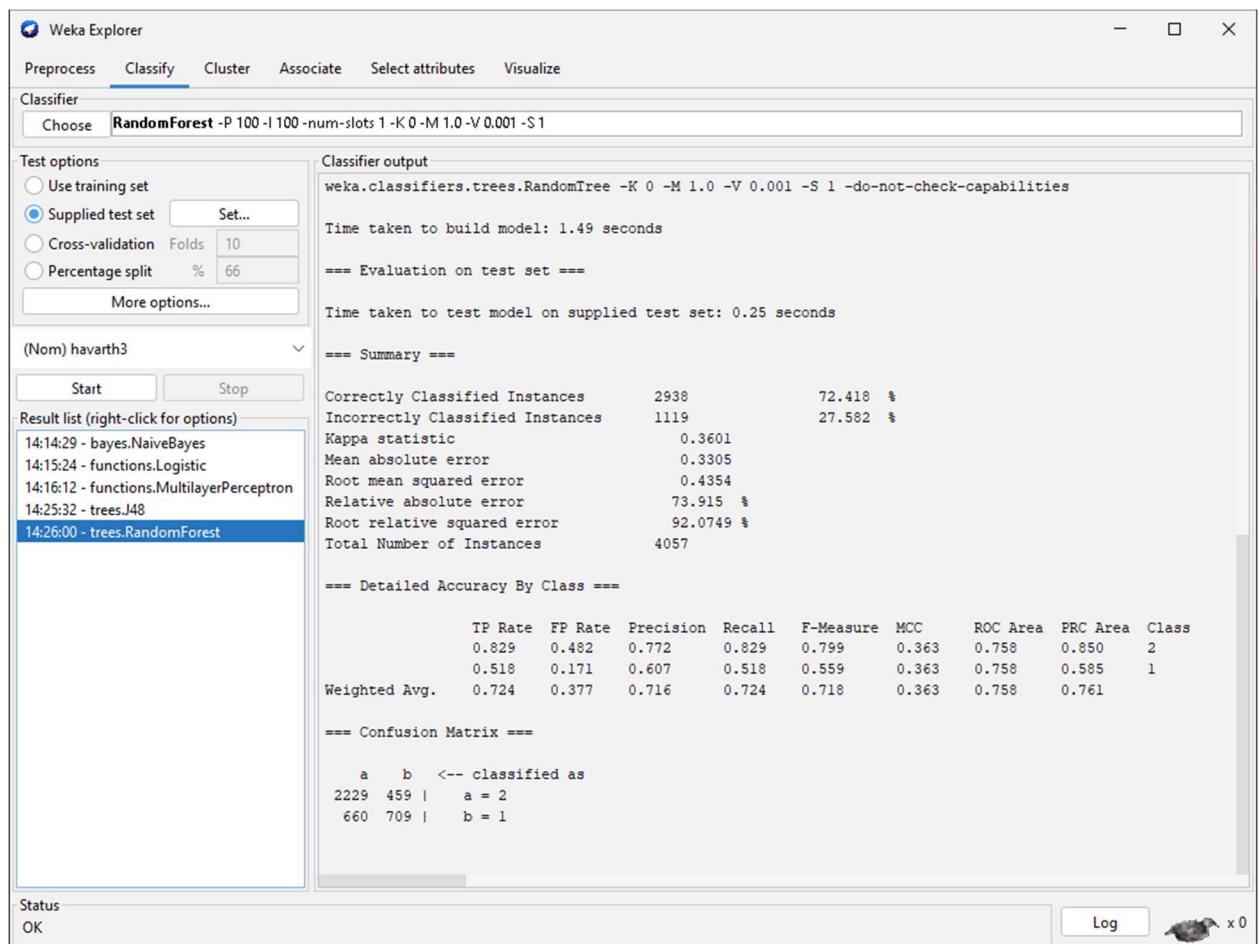
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.829	0.482	0.772	0.829	0.799	0.363	0.758	0.850	2	
0.518	0.171	0.607	0.518	0.559	0.363	0.758	0.585	1	
Weighted Avg.	0.724	0.377	0.716	0.724	0.718	0.363	0.758	0.761	

==== Confusion Matrix ===

		a b <-- classified as
		a = 2
2229	459	a = 2
660	709	b = 1

Status OK

Log x 0



4) Attributes selected with InfoGainAttributeEval using Ranker search method:

The screenshot shows the Weka Explorer interface with the following configuration:

- Preprocess**, **Classify**, **Cluster**, **Associate**, **Select attributes** (selected), **Visualize**
- Attribute Evaluator**: Choose **InfoGainAttributeEval**
- Search Method**: Choose **Ranker -T -1.7976931348623157E308 -N -1**
- Attribute Selection Mode**: Use full training set, Cross-validation (Folds: 10, Seed: 1)
- No class**: Start, Stop
- Result list (right-click for options)**: 14:28:03 - Ranker + InfoGainAttributeEval
- Attribute selection output** (Text Area):

```
==== Attribute Selection on all input data ====
Search Method:
    Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 108 havarth3):
    Information Gain Ranking Filter

Ranked attributes:
0.11056128 64 x.age80
0.11001453 66 x.ageg5yr
0.10680984 62 x.age.g
0.09485689 2 employ1
0.07832746 22 diffwalk
0.06586308 34 genhlth
0.05819653 67 x.age65yr
0.0544134 97 x.hcvu651
0.05245567 31 physhlth
0.04877981 95 x.phys14d
0.04567538 27 rmvteth4
0.04277001 20 pneuvac4
0.04152785 87 x.rfhlth
0.03578182 102 x.exteth3
0.03527668 11 marital
0.02920094 46 chccopdl
0.0287828 69 x.chldcnt
0.02671903 6 children
0.02321593 41 persdoc2
0.02268924 43 checkup1
0.02062293 29 diabete3
0.01912437 54 qstver
0.01883658 25 diffalon
0.01839936 3 income2
0.01812092 13 deaf
0.01748861 104 x.michd
0.01688165 71 x.incomg
0.01676461 24 diffdrres
0.01648448 53 cvdcrhd4
0.016113 86 x.llcpwrt
0.01566449 78 x.smoker3
0.01516519 59 x.bmi5
0.01350118 45 chcoconr
0.01343537 60 x.bmi5cat
0.01341882 26 smoke100
0.01339554 88 x.dualuse
0.01304164 44 exranv2
```
- Status**: OK
- Log**: x 0

Again, select 20 attributes as this is about 20% of the data.

Selected Attributes:

employ1, children, marital, pneuvac4, diffwalk, rmvteth4, physhlth, genhlth, persdoc2, checkup1, chccopdl, x.age.g, x.age80, x.ageg5yr, x.age65yr, x.chldcnt, x.rfhlth, x.phys14d, x.hcvu651, x.exteth3

Then run 5 classifiers on the data:

- Naïve Bayes

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **NaiveBayes**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:34:50 - bayes.NaiveBayes

Classifier output

```
x.hcvu651
1      3323.0  1081.0
9      1418.0  1466.0
2      480.0   114.0
[total] 5221.0  2661.0

x.exteth3
1      3068.0  941.0
2      2063.0  1641.0
9      90.0    79.0
[total] 5221.0  2661.0
```

Time taken to build model: 0.01 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

	Correctly Classified Instances	2883	71.0624 %
Incorrectly Classified Instances	1174	28.9376 %	
Kappa statistic	0.3969		
Mean absolute error	0.2908		
Root mean squared error	0.5006		
Relative absolute error	65.0283 %		
Root relative squared error	105.8769 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.706	0.280	0.832	0.706	0.764	0.406	0.790	0.877	2	
0.720	0.294	0.555	0.720	0.627	0.406	0.790	0.630	1	
Weighted Avg.	0.711	0.285	0.738	0.711	0.718	0.406	0.790	0.794	

==== Confusion Matrix ===

a	b	<- classified as
1897	791	a = 2
383	986	b = 1

Status OK Log x 0

- Logistic Regression:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. A classifier named 'Logistic - R 1.0E-8 -M -1 -num-decimal-places 4' is chosen. The 'Test options' panel indicates a 'Supplied test set' has been selected. The 'Classifier output' panel displays the following results:

```

x.rfhlth=1      1.1236
x.rfhlth=2      0.8845
x.rfhlth=9      1.3886
x.phys14d=1      1.2956
x.phys14d=2      0.8899
x.phys14d=9      0.8901
x.phys14d=3      0.7192
x.hcvu65l=1      1.0678
x.hcvu65l=9      0.909
x.hcvu65l=2      1.0898
x.exteth3=1      1.0368
x.exteth3=2      0.9648
x.exteth3=9      0.9948

```

Below the classifier output, the terminal logs show:

```

Time taken to build model: 0.42 seconds
==== Evaluation on test set ====
Time taken to test model on supplied test set: 0.01 seconds
==== Summary ====
Correctly Classified Instances      3051      75.2034 %
Incorrectly Classified Instances   1006      24.7966 %
Kappa statistic                   0.4052
Mean absolute error               0.3298
Root mean squared error           0.4081
Relative absolute error           73.7548 %
Root relative squared error      86.3168 %
Total Number of Instances        4057
==== Detailed Accuracy By Class ====
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.883    0.505    0.774    0.883    0.825    0.416    0.804    0.880    2
          0.495    0.117    0.683    0.495    0.574    0.416    0.804    0.662    1
Weighted Avg.    0.752    0.374    0.744    0.752    0.740    0.416    0.804    0.807
==== Confusion Matrix ====
      a     b  <- classified as
2374  314 |     a = 2
 692  677 |     b = 1

```

The status bar at the bottom left shows 'Status OK' and the log icon at the bottom right shows 'Log x 0'.

- Multilayer Perceptron (Neural Network):

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a'. The 'Test options' section indicates a 'Supplied test set' was used. The 'Classifier output' pane displays the following details:

- Attrib** x.phys14d=3 7.785037280118111
- Attrib x.hcvu65l=1 -3.96400828957967
- Attrib x.hcvu65l=9 0.4791917696345055
- Attrib x.hcvu65l=2 6.180010934295802
- Attrib x.exteth3=1 6.246556594364493
- Attrib x.exteth3=2 -6.208488079590952
- Attrib x.exteth3=9 2.710075220207287

Class 2

- Input
- Node 0

Class 1

- Input
- Node 1

Time taken to build model: 206.93 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.07 seconds

==== Summary ===

	Correctly Classified Instances	2870	70.7419 %
Incorrectly Classified Instances	1187	29.2581 %	
Kappa statistic	0.3123		
Mean absolute error	0.3089		
Root mean squared error	0.4747		
Relative absolute error	69.0853 %		
Root relative squared error	100.3896 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.830	0.533	0.754	0.830	0.790	0.316	0.740	0.843	2	
0.467	0.170	0.583	0.467	0.519	0.316	0.740	0.569	1	
Weighted Avg.	0.707	0.410	0.696	0.707	0.698	0.316	0.740	0.750	

==== Confusion Matrix ===

		a b <-- classified as
2230	458	a = 2
729	640	b = 1

Status: OK Log x 0

- J48:

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) havarth3

Start Stop

Result list (right-click for options)

14:34:50 - bayes.NaiveBayes
14:35:37 - functions.Logistic
14:36:14 - functions.MultilayerPerceptron
14:41:03 - trees.J48

Classifier output

```
| diffwalk = 7
|   | chccopdl = 2
|   |   | x.rfhlth = 1: 1 (10.35/2.26)
|   |   | x.rfhlth = 2: 2 (5.11/1.06)
|   |   | x.rfhlth = 9: 2 (1.0)
|   | chccopdl = 1: 1 (3.03/0.01)
|   | chccopdl = 7: 1 (0.0)
|   | chccopdl = 9: 2 (0.0)
| diffwalk = 9: 2 (6.16/2.05)

Number of Leaves : 344

Size of the tree : 463

Time taken to build model: 0.08 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances      3002          73.9956 %
Incorrectly Classified Instances   1055          26.0044 %
Kappa statistic                   0.3913
Mean absolute error               0.3366
Root mean squared error           0.4294
Relative absolute error           75.282 %
Root relative squared error      90.8035 %
Total Number of Instances        4057

==== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
      0.850    0.476    0.778    0.850    0.812    0.395    0.761    0.834    2
      0.524    0.150    0.640    0.524    0.576    0.395    0.761    0.585    1
Weighted Avg.      0.740    0.366    0.731    0.740    0.733    0.395    0.761    0.750

==== Confusion Matrix ====

      a     b  <- classified as
2285  403 |     a = 2
       652  717 |     b = 1
```

Status OK Log x 0

- Random Forest

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. A 'Classifier' panel on the left displays the configuration: 'Choose' set to 'RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section shows 'Supplied test set' selected. The 'Classifier output' panel on the right contains the following text:

```
x.rfhlth
x.phys14d
x.hcvu651
x.exteth3
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

== Classifier model (full training set) ==

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 1.13 seconds

== Evaluation on test set ==

Time taken to test model on supplied test set: 0.22 seconds

== Summary ==

Correctly Classified Instances      2918          71.9251 %
Incorrectly Classified Instances   1139          28.0749 %
Kappa statistic                   0.3496
Mean absolute error               0.3354
Root mean squared error          0.4344
Relative absolute error           75.0012 %
Root relative squared error     91.8631 %
Total Number of Instances        4057

== Detailed Accuracy By Class ==

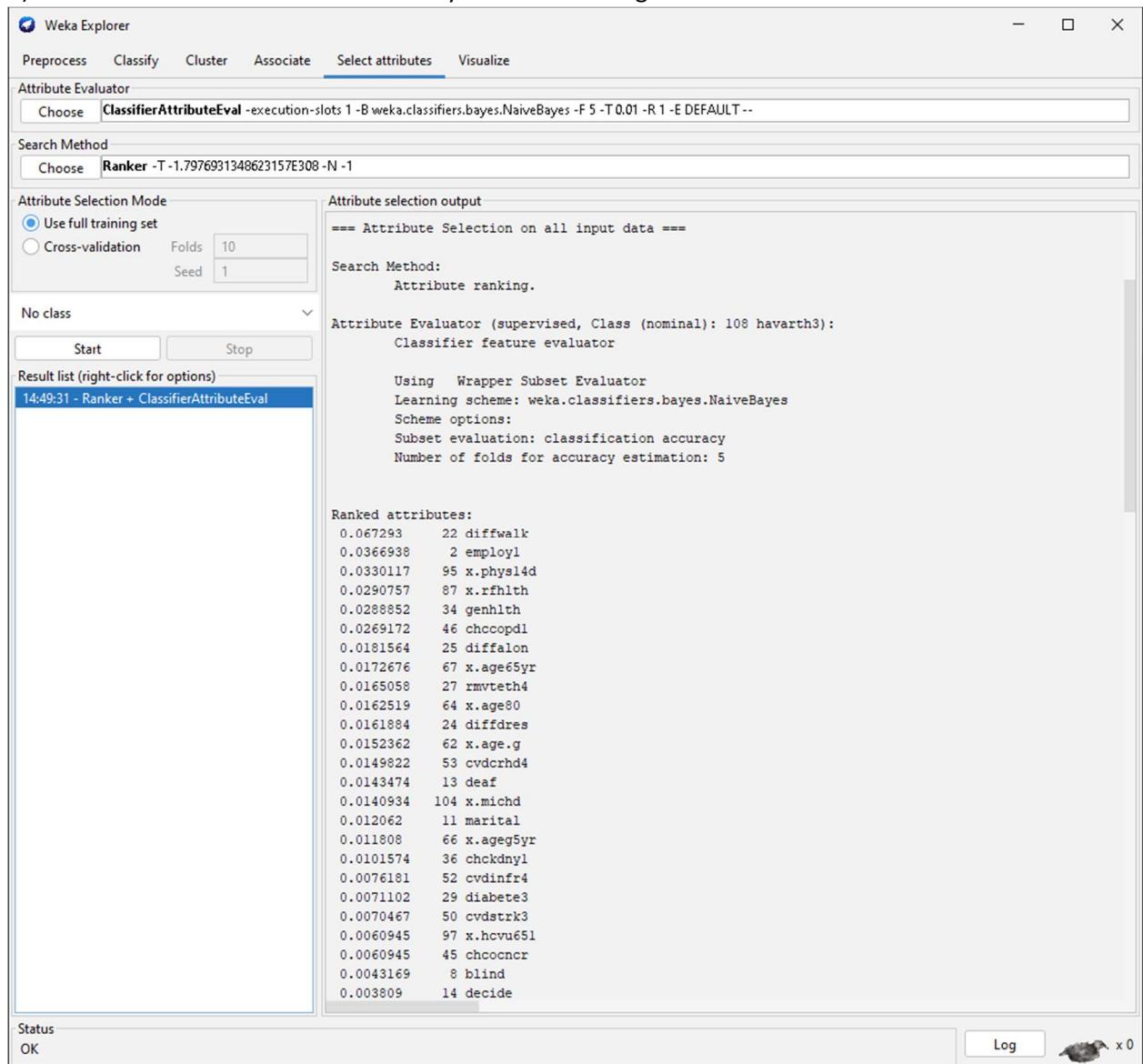
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.824   0.486   0.769   0.824   0.795   0.352   0.759   0.851   2
          0.514   0.176   0.598   0.514   0.552   0.352   0.759   0.583   1
Weighted Avg.      0.719   0.382   0.711   0.719   0.713   0.352   0.759   0.760

== Confusion Matrix ==

      a     b  <- classified as
2215  473 |     a = 2
 666  703 |     b = 1
```

The status bar at the bottom left shows 'Status OK' and the bottom right has a 'Log' button.

5) ClassifierAttributeEval with Naïve Bayes Classifier using Ranker search method



Again, select top 20 attributes for roughly 20% of the data.

Selected Attributes:

employ1, marital, deaf, diffwalk, diffdres, diffalon, rmvteth4, diabete3, genhlth, chckdny1, chccopdl, cvdinfr4, cvdinfr4, cvdcrhd4, x.age.g, x.age80, x.ageg5yr, x.age65yr, x.rfhlth, x.phys14d, x.michd

Then run 5 classifiers:

- Naïve Bayes:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:58:52 - bayes.NaiveBayes

Classifier output

```
Time taken to build model: 0 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

===
Summary ===

Correctly Classified Instances      2915          71.8511 %
Incorrectly Classified Instances   1142          28.1489 %
Kappa statistic                   0.4014
Mean absolute error               0.2875
Root mean squared error          0.4818
Relative absolute error           64.2914 %
Root relative squared error     101.9033 %
Total Number of Instances        4057

===
Detailed Accuracy By Class ===

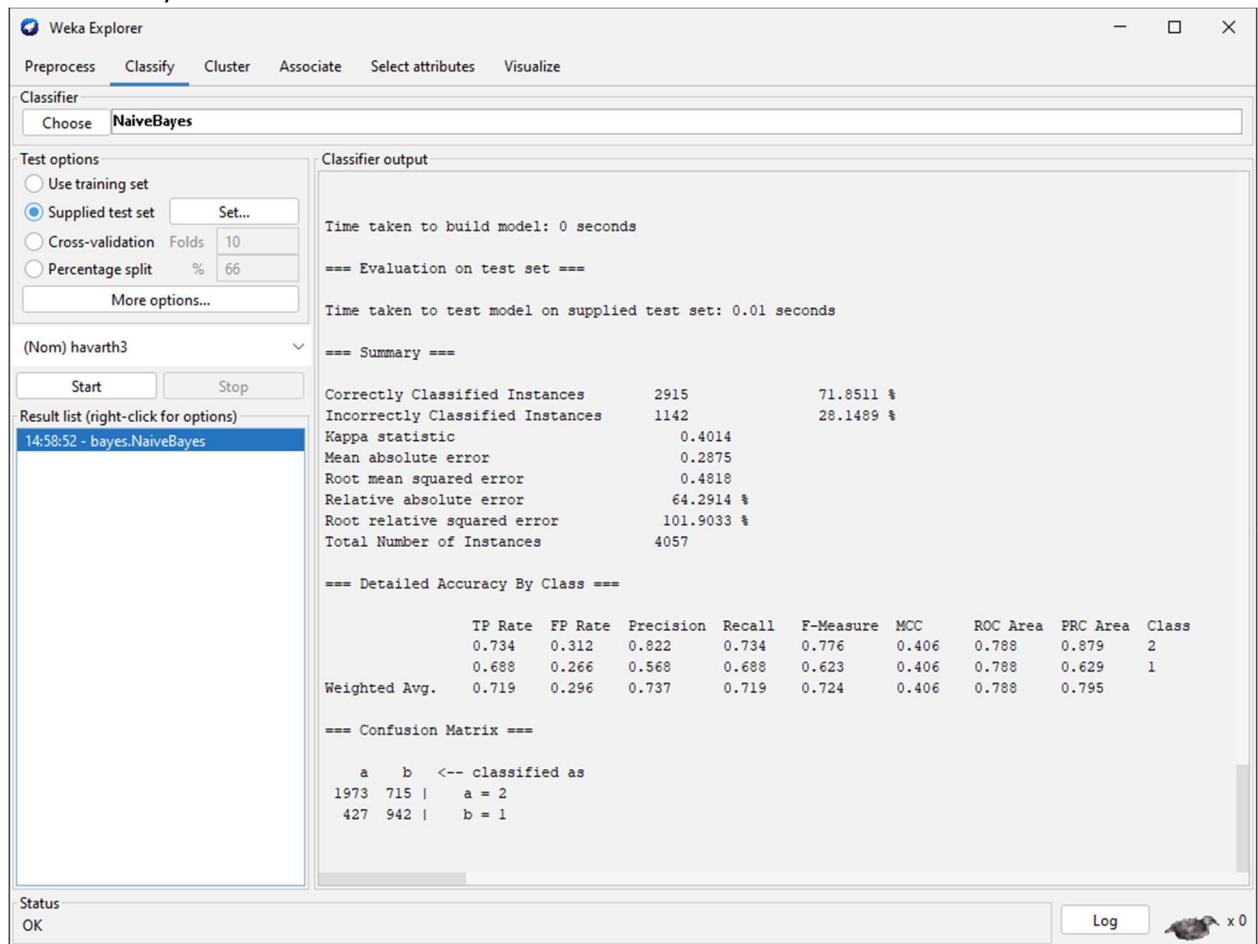
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.734    0.312    0.822     0.734    0.776     0.406   0.788    0.879     2
      0.688    0.266    0.568     0.688    0.623     0.406   0.788    0.629     1
Weighted Avg.      0.719    0.296    0.737     0.719    0.724     0.406   0.788    0.795

===
Confusion Matrix ===

      a     b  <-- classified as
1973  715 |  a = 2
427   942 |  b = 1
```

Status OK

Log x 0



- Logistic Regression

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Logistic -R 1.0E-8 -M -1 -num-decimal-places 4**

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:58:52 - bayes.NaiveBayes
14:59:56 - functions.Logistic

Classifier output

```
Time taken to build model: 0.38 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

===
Summary ===

Correctly Classified Instances      3040          74.9322 %
Incorrectly Classified Instances   1017          25.0678 %
Kappa statistic                   0.3976
Mean absolute error               0.3325
Root mean squared error           0.4098
Relative absolute error            74.3626 %
Root relative squared error       86.6734 %
Total Number of Instances         4057

===
Detailed Accuracy By Class ===

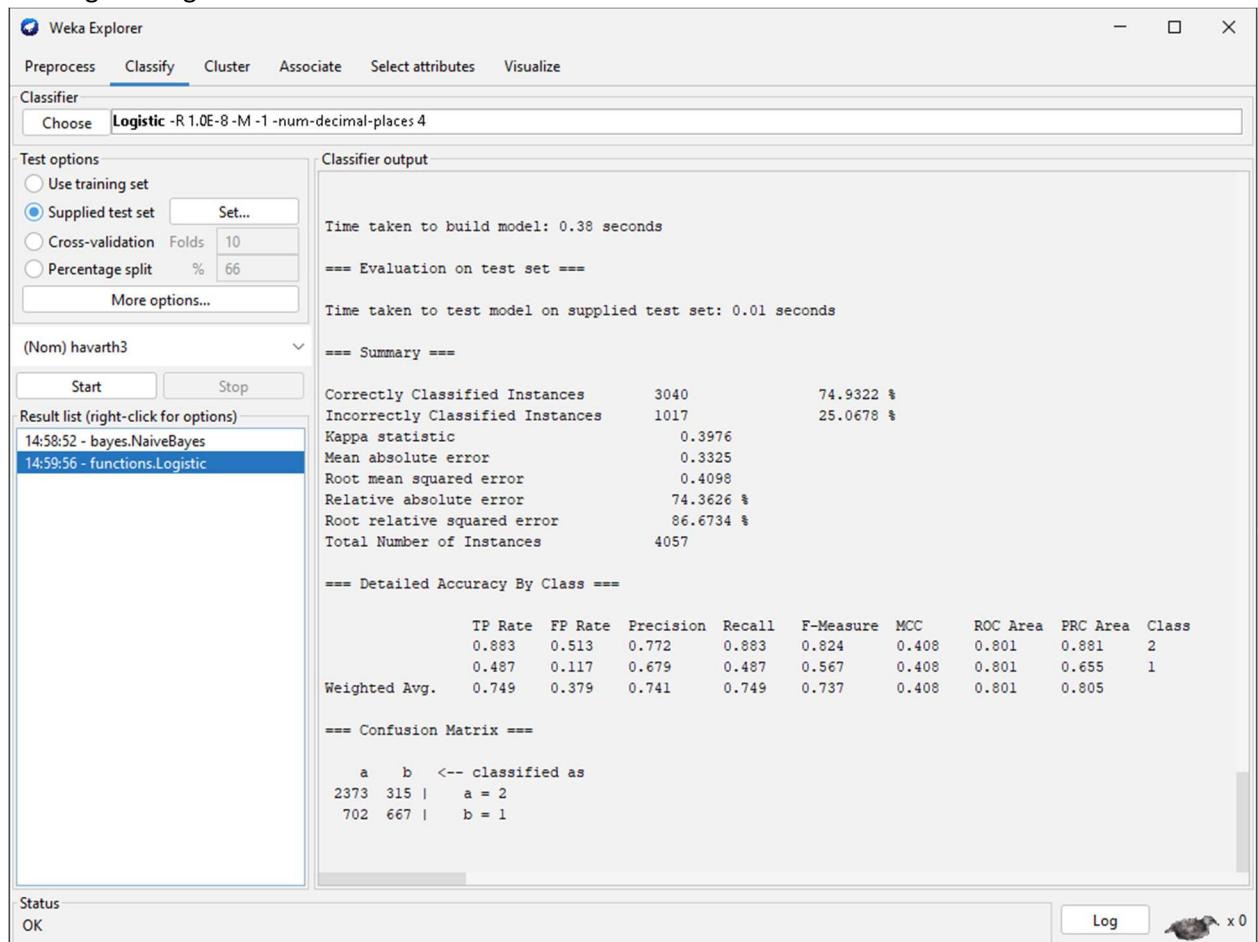
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.883    0.513    0.772     0.883    0.824     0.408   0.801    0.881    2
      0.487    0.117    0.679     0.487    0.567     0.408   0.801    0.655    1
Weighted Avg.                      0.749    0.379    0.741     0.749    0.737     0.408   0.801    0.805

===
Confusion Matrix ===

      a     b  <-- classified as
2373  315 |  a = 2
702   667 |  b = 1
```

Status OK

Log x 0



- Multilayer Perceptron (Neural Network):

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:58:52 - bayes.NaiveBayes
14:59:56 - functions.Logistic
15:00:18 - functions.MultilayerPerceptron

Classifier output

```
Time taken to build model: 205.4 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.07 seconds

===
Summary ===

Correctly Classified Instances      2902          71.5307 %
Incorrectly Classified Instances   1155          28.4693 %
Kappa statistic                   0.3234
Mean absolute error               0.3153
Root mean squared error          0.4734
Relative absolute error           70.5122 %
Root relative squared error     100.1235 %
Total Number of Instances        4057

===
Detailed Accuracy By Class ===

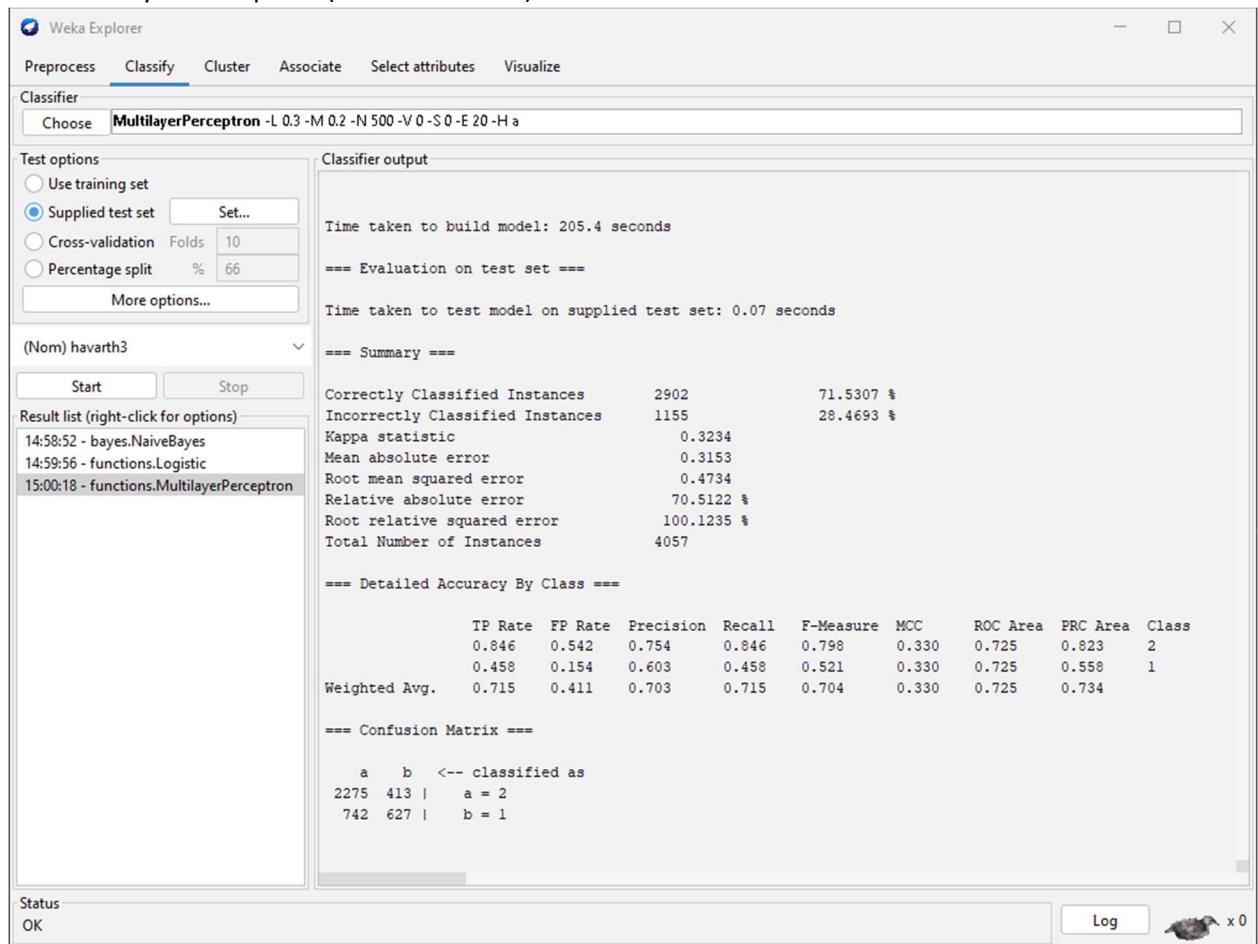
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.846    0.542    0.754     0.846    0.798     0.330   0.725    0.823    2
      0.458    0.154    0.603     0.458    0.521     0.330   0.725    0.558    1
Weighted Avg.      0.715    0.411    0.703     0.715    0.704     0.330   0.725    0.734

===
Confusion Matrix ===

      a     b  <-- classified as
2275  413 |  a = 2
742   627 |  b = 1
```

Status OK

Log x 0



- J48:

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) havarth3

Start Stop

Result list (right-click for options)

14:58:52 - bayes.NaiveBayes
14:59:56 - functions.Logistic
15:00:18 - functions.MultilayerPerceptron
15:54:05 - trees.J48

Classifier output

```
Time taken to build model: 0.08 seconds
===
Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

===
Summary ===

Correctly Classified Instances      3018          74.3899 %
Incorrectly Classified Instances   1039          25.6101 %
Kappa statistic                   0.3874
Mean absolute error               0.3365
Root mean squared error          0.4262
Relative absolute error           75.252 %
Root relative squared error     90.1368 %
Total Number of Instances        4057

===
Detailed Accuracy By Class ===

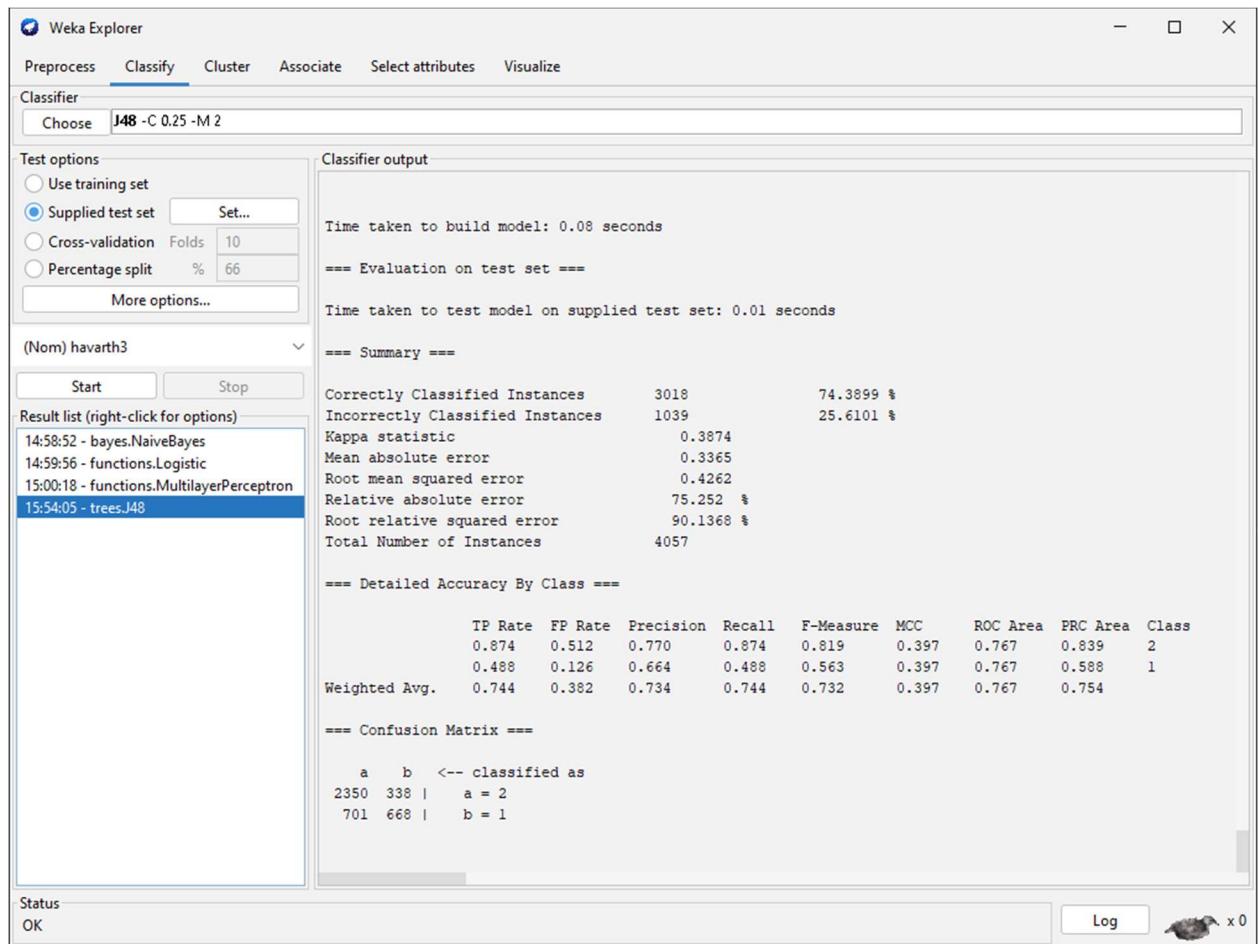
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.874    0.512    0.770     0.874    0.819     0.397   0.767    0.839    2
      0.488    0.126    0.664     0.488    0.563     0.397   0.767    0.588    1
Weighted Avg.      0.744    0.382    0.734     0.744    0.732     0.397   0.767    0.754

===
Confusion Matrix ===

      a     b  <- classified as
2350  338 |  a = 2
 701  668 |  b = 1
```

Status OK

Log x 0



- Random Forest:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

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

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:58:52 - bayes.NaiveBayes
14:59:56 - functions.Logistic
15:00:18 - functions.MultilayerPerceptron
15:54:05 - trees.J48
15:54:42 - trees.RandomForest

Classifier output

```
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
```

Time taken to build model: 1.19 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.23 seconds

==== Summary ===

	Correctly Classified Instances	71.161 %
Incorrectly Classified Instances	1170	28.839 %
Kappa statistic	0.3363	
Mean absolute error	0.3356	
Root mean squared error	0.4392	
Relative absolute error	75.0558 %	
Root relative squared error	92.879 %	
Total Number of Instances	4057	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.812	0.485	0.767	0.812	0.789	0.338	0.753	0.853	2	
0.515	0.188	0.582	0.515	0.547	0.338	0.753	0.564	1	
Weighted Avg.	0.712	0.385	0.704	0.712	0.707	0.338	0.753	0.756	

==== Confusion Matrix ===

		a b <-- classified as
2182 506		a = 2
664 705		b = 1

Status OK

Log x 0

