

Diabetes

PREPROCESSING STEP:

Applied ClassBalancer: Number of classes unequal. May cause bias. Hence used ClassBalancer to distribute the weightage equally between the classes.

Classification Methods:

KNN:

Where k=15 , we get maximum accuracy of 73.6075%

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'IBk -K 15 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      565.3053      73.6075 %
Incorrectly Classified Instances    202.6947      26.3925 %
Kappa statistic                    0.4721
Mean absolute error                 0.3527
Root mean squared error             0.4266
Relative absolute error             70.5488 %
Root relative squared error         85.3102 %
Total Number of Instances          768

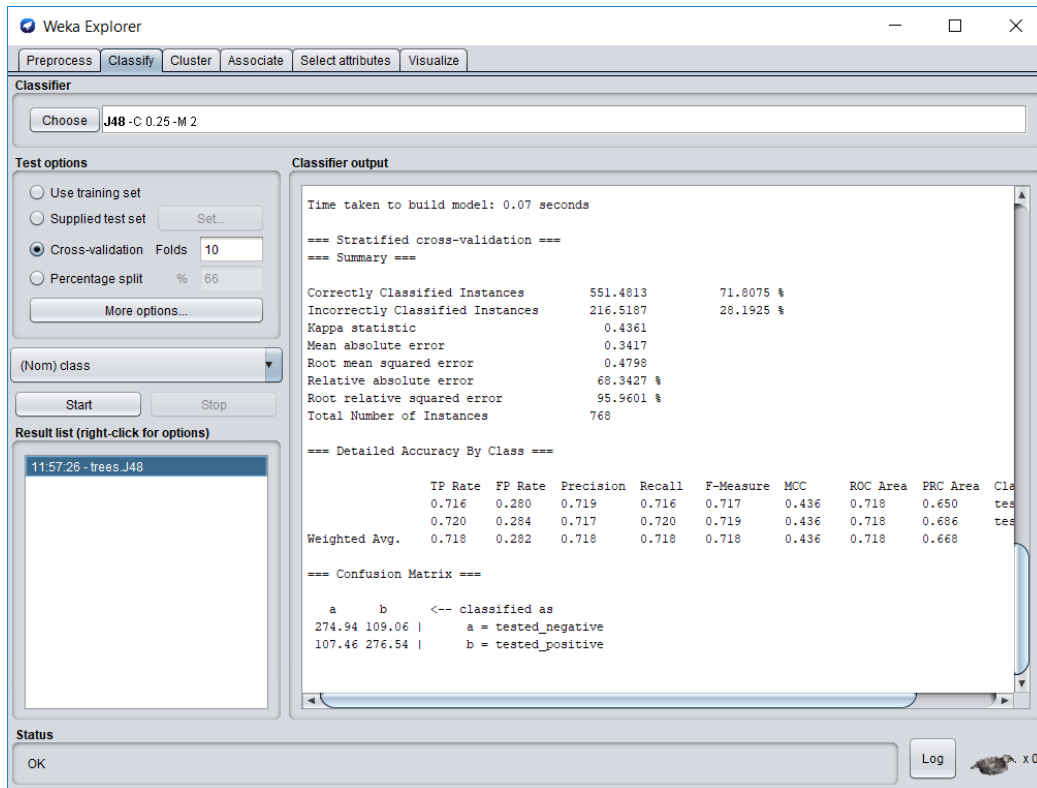
=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
Weighted Avg.   0.736   0.264   0.736     0.736   0.736     0.472   0.803    0.778   tes

=== Confusion Matrix ===
      a      b  <-- classified as
288.77  95.23 |  a = tested_negative
107.46 276.54 |  b = tested_positive
```

The 'Result list (right-click for options)' on the left shows a list of test results, with '01:04:52 - lazy.IBk' selected. The 'Status' bar at the bottom indicates 'OK'.

Decision Tree:

Accuracy : 71.8075%



The screenshot shows the Weka Explorer interface with the Classifier tab selected. The classifier chosen is J48 - C 0.25 - M 2. The test options are set to Cross-validation with 10 folds. The classifier output window displays the following results:

Time taken to build model: 0.07 seconds

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

Metric	Value	Percentage
Correctly Classified Instances	551.4813	71.8075 %
Incorrectly Classified Instances	216.5187	28.1925 %
Kappa statistic	0.4361	
Mean absolute error	0.3417	
Root mean squared error	0.4798	
Relative absolute error	68.3427 %	
Root relative squared error	95.9601 %	
Total Number of Instances	768	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.716	0.280	0.719	0.716	0.717	0.436	0.718	0.650	tested_negative
	0.720	0.284	0.717	0.720	0.719	0.436	0.718	0.686	tested_positive
Weighted Avg.	0.718	0.282	0.718	0.718	0.718	0.436	0.718	0.668	

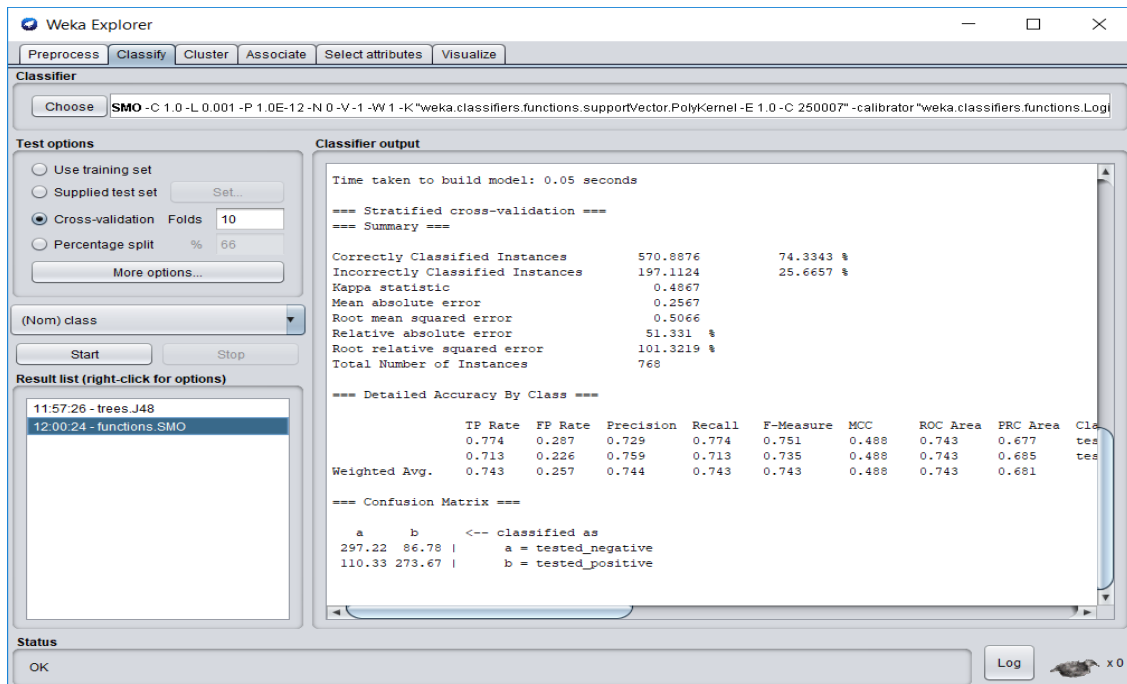
=== Confusion Matrix ===

a \ b	tested_negative	tested_positive
tested_negative	274.94	109.06
tested_positive	107.46	276.54

The result list on the left shows the file 11:57:26 - trees.J48.

SVM:

Accuracy :74.3343%



The screenshot shows the Weka Explorer interface with the Classifier tab selected. The classifier chosen is SMO - C 1.0 - L 0.001 - P 1.0E-12 - N 0 - V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel" - E 1.0 - C 250007 - calibrator "weka.classifiers.functions.Logit". The test options are set to Cross-validation with 10 folds. The classifier output window displays the following results:

Time taken to build model: 0.05 seconds

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

Metric	Value	Percentage
Correctly Classified Instances	570.8876	74.3343 %
Incorrectly Classified Instances	197.1124	25.6657 %
Kappa statistic	0.4867	
Mean absolute error	0.2567	
Root mean squared error	0.5066	
Relative absolute error	51.331 %	
Root relative squared error	101.3219 %	
Total Number of Instances	768	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.774	0.287	0.729	0.774	0.751	0.488	0.743	0.677	tested_negative
	0.713	0.226	0.759	0.713	0.735	0.488	0.743	0.685	tested_positive
Weighted Avg.	0.743	0.257	0.744	0.743	0.743	0.488	0.743	0.681	

=== Confusion Matrix ===

a \ b	tested_negative	tested_positive
tested_negative	297.22	86.78
tested_positive	110.33	273.67

The result list on the left shows the file 11:57:26 - trees.J48 and 12:00:24 - functions.SMO.

ANN:

Accuracy: 72.994%

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) class

Start Stop

Result list (right-click for options)

- 11:57:26 - trees.J48
- 12:00:24 - functions.SMO
- 12:01:45 - functions.MultilayerPerceptron**

Classifier output

```
Time taken to build model: 0.69 seconds

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

Correctly Classified Instances      560.5941      72.994 %
Incorrectly Classified Instances    207.4059      27.006 %
Kappa statistic                     0.4599
Mean absolute error                  0.3157
Root mean squared error              0.445
Relative absolute error              63.1457 %
Root relative squared error          88.9998 %
Total Number of Instances           768

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
          0.736   0.276   0.727     0.736   0.732     0.460   0.793    0.776   tes
          0.724   0.264   0.733     0.724   0.728     0.460   0.793    0.777   tes

Weighted Avg.   0.730   0.270   0.730     0.730   0.730     0.460   0.793    0.777

=== Confusion Matrix ===

  a    b  <-- classified as
282.62 101.38 |    a = tested_negative
106.03 277.97 |    b = tested_positive
```

Status

OK Log x0

SVM has better accuracy than all of these models.