

# Machine Learning

## Classification Algorithms using WEKA

### 1. Caesarian Data set

- Applied Class Balancer in the Pre-processing step since the classes were distributed randomly and to reduce the bias.
- KNN algorithm gives more efficient result and ANN was less efficient when compared to others.

#### K Nearest Neighbor Algorithm

- When k=20, the results were more optimal.

The screenshot displays the WEKA Classifier window. The 'Classifier' dropdown is set to 'IBk -K 20 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"'. The 'Test options' section shows 'Cross-validation' selected with 'Folds' set to 10. The 'Result list' on the left shows a list of classifiers, with '22:54:44 - lazy.IBk' selected. The 'Classifier output' pane on the right displays the following results:

```
=== Classifier model (full training set) ===

IBk instance-based classifier
using 20 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      54.6803      68.3504 %
Incorrectly Classified Instances    25.3197      31.6496 %
Kappa statistic                    0.367
Mean absolute error                 0.4564
Root mean squared error             0.4679
Relative absolute error             91.1154 %
Root relative squared error         93.4046 %
Total Number of Instances          80

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               -----  -----  -
Weighted Avg.    0.684    0.316    0.699     0.684    0.677     0.382    0.708    0.719
0
0.824    0.457    0.643     0.824    0.722     0.382    0.708    0.658    0
0.543    0.176    0.755     0.543    0.632     0.382    0.708    0.781    1

=== Confusion Matrix ===

  a    b  <-- classified as
32.94  7.06 |    a = 0
18.26 21.74 |    b = 1
```

## Support Vector Machine

**Classifier**

Choose **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.Logistic -R 1.0E-8 -M -1 -nur**

**Test options**

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

(Nom) Caesarian

Start Stop

**Result list (right-click for options)**

- 11:01:58 - lazy.IBk
- 11:02:11 - lazy.IBk
- 11:04:33 - functions.SMO
- 11:06:37 - trees.J48
- 11:22:59 - trees.RandomForest
- 11:23:08 - functions.Logistic
- 11:24:38 - functions.MultilayerPerceptron
- 11:25:05 - functions.SimpleLogistic
- 11:25:49 - trees.RandomTree
- 22:53:32 - lazy.IBk
- 22:53:52 - lazy.IBk
- 22:54:06 - lazy.IBk
- 22:54:22 - lazy.IBk
- 22:54:33 - lazy.IBk
- 22:54:44 - lazy.IBk
- 22:56:28 - functions.SMO

**Classifier output**

```
Number of kernel evaluations: 2352 (87.626% cached)

Time taken to build model: 0.02 seconds

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

Correctly Classified Instances      50.2302      62.7877 %
Incorrectly Classified Instances    29.7698      37.2123 %
Kappa statistic                    0.2558
Mean absolute error                0.3721
Root mean squared error            0.61
Relative absolute error             74.2984 %
Root relative squared error        121.7858 %
Total Number of Instances         80

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.647    0.391    0.623     0.647    0.635     0.256    0.628    0.580     0
               0.609    0.353    0.633     0.609    0.621     0.256    0.628    0.581     1
Weighted Avg.   0.628    0.372    0.628     0.628    0.628     0.256    0.628    0.580

=== Confusion Matrix ===

  a    b    <-- classified as
25.88 14.12 |    a = 0
15.65 24.35 |    b = 1
```

## Decision Tree

**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) Caesarian

Start Stop

**Result list (right-click for options)**

- 11:02:11 - lazy.IBk
- 11:04:33 - functions.SMO
- 11:06:37 - trees.J48
- 11:22:59 - trees.RandomForest
- 11:23:08 - functions.Logistic
- 11:24:38 - functions.MultilayerPerceptron
- 11:25:05 - functions.SimpleLogistic
- 11:25:49 - trees.RandomTree
- 22:53:32 - lazy.IBk
- 22:53:52 - lazy.IBk
- 22:54:06 - lazy.IBk
- 22:54:22 - lazy.IBk
- 22:54:33 - lazy.IBk
- 22:54:44 - lazy.IBk
- 22:56:28 - functions.SMO
- 22:57:52 - trees.J48

**Classifier output**

```
Number of Leaves :      2

Size of the tree :      3

Time taken to build model: 0 seconds

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

Correctly Classified Instances      49.4629      61.8286 %
Incorrectly Classified Instances    30.5371      38.1714 %
Kappa statistic                    0.2366
Mean absolute error                0.4558
Root mean squared error            0.4875
Relative absolute error            91.0056 %
Root relative squared error        97.3227 %
Total Number of Instances         80

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.824    0.587    0.584     0.824    0.683     0.259    0.573    0.529     0
               0.413    0.176    0.701     0.413    0.520     0.259    0.573    0.574     1
Weighted Avg.   0.618    0.382    0.642     0.618    0.601     0.259    0.573    0.552

=== Confusion Matrix ===

  a    b    <-- classified as
32.94  7.06 |    a = 0
23.48 16.52 |    b = 1
```

Artificial Neural Network

Classifier

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

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Set...

Folds10

%66

More options...

(Nom) Caesarian

Start

Stop

Result list (right-click for options)

11:04:33 - functions.SMO

11:06:37 - trees.J48

11:22:59 - trees.RandomForest

11:23:08 - functions.Logistic

11:24:38 - functions.MultilayerPerceptron

11:25:05 - functions.SimpleLogistic

11:25:49 - trees.RandomTree

22:53:32 - lazy.IBk

22:53:52 - lazy.IBk

22:54:06 - lazy.IBk

22:54:22 - lazy.IBk

22:54:33 - lazy.IBk

22:54:44 - lazy.IBk

22:56:28 - functions.SMO

22:57:52 - trees.J48

22:59:30 - functions.MultilayerPerceptron

Classifier output

Node 0

Class 1

Input

Node 1

Time taken to build model: 0.38 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances

41.1253

51.4066 %

Incorrectly Classified Instances

38.8747

48.5934 %

Kappa statistic

0.0281

Mean absolute error

0.4705

Root mean squared error

0.6317

Relative absolute error

93.9428 %

Root relative squared error

126.111 %

Total Number of Instances

80

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.441	0.413	0.516	0.441	0.476	0.028	0.550	0.563	0
	0.587	0.559	0.512	0.587	0.547	0.028	0.550	0.571	1
Weighted Avg.	0.514	0.486	0.514	0.514	0.511	0.028	0.550	0.567	

=== Confusion Matrix ===

a

b

<-- classified as

17.65 22.35 | a = 0

16.52 23.48 | b = 1

## 2. Diabetics Dataset

- Applied Class Balancer in the Pre-processing step since the classes were distributed randomly and to reduce the bias.
- SVM Algorithm gives more efficient and Decision Tree gives less efficient results when compared to others.

### K Nearest Neighbor

- When k=20, the results were more optimal.

**Classifier**

Choose **IBk -K 20 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

**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)**

- 23:18:09 - lazy.IBk
- 23:18:37 - lazy.IBk
- 23:32:26 - lazy.IBk
- 23:32:37 - lazy.IBk
- 23:32:50 - lazy.IBk
- 23:33:15 - lazy.IBk**

**Classifier output**

```
=== Classifier model (full training set) ===

IB1 instance-based classifier
using 20 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      566.3713      73.7463 %
Incorrectly Classified Instances    201.6287      26.2537 %
Kappa statistic                    0.4749
Mean absolute error                 0.361
Root mean squared error             0.4285
Relative absolute error             72.2076 %
Root relative squared error         85.7044 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MDC     ROC Area  PRC Area  Class
               0.710    0.235    0.751    0.710    0.730    0.476    0.799    0.792    tested_negative
               0.765    0.290    0.725    0.765    0.744    0.476    0.798    0.769    tested_positive
Weighted Avg.   0.737    0.263    0.738    0.737    0.737    0.476    0.799    0.780

=== Confusion Matrix ===

  a    b  <-- classified as
272.64 111.36 |   a = tested_negative
 90.27 293.73 |   b = tested_positive
```

### Support Vector Machine

**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)**

- 23:18:09 - lazy.IBk
- 23:18:37 - lazy.IBk
- 23:32:26 - lazy.IBk
- 23:32:37 - lazy.IBk
- 23:32:50 - lazy.IBk
- 23:33:15 - lazy.IBk
- 23:36:32 - functions.SMO**
- 23:36:51 - functions.MultilayerPerceptron
- 23:37:12 - trees.J48

**Classifier output**

```
Number of kernel evaluations: 21182 (67.241% cached)

Time taken to build model: 0.05 seconds

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

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  MDC     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    0.691

=== Confusion Matrix ===

  a    b  <-- classified as
297.22  86.78 |   a = tested_negative
110.33 273.67 |   b = tested_positive
```

# Artificial Neural Network

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Percentage split

Set...

Folds 10

% 66

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

23:18:09 - lazy.IBk

23:18:37 - lazy.IBk

23:32:26 - lazy.IBk

23:32:37 - lazy.IBk

23:32:50 - lazy.IBk

23:33:15 - lazy.IBk

23:36:32 - functions.SMO

23:36:51 - functions.MultilayerPerceptron

23:37:12 - trees.J48

Classifier output

Node 0

Class tested\_positive

Input

Node 1

Time taken to build model: 0.46 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	Class
	0.736	0.276	0.727	0.736	0.732	0.460	0.793	0.776	tested_negative
	0.724	0.264	0.733	0.724	0.728	0.460	0.793	0.777	tested_positive
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

# Decision Tree

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Percentage split

Set...

Folds 10

% 66

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

23:18:09 - lazy.IBk

23:18:37 - lazy.IBk

23:32:26 - lazy.IBk

23:32:37 - lazy.IBk

23:32:50 - lazy.IBk

23:33:15 - lazy.IBk

23:36:32 - functions.SMO

23:36:51 - functions.MultilayerPerceptron

23:37:12 - trees.J48

Classifier output

Number of Leaves : 24

Size of the tree : 47

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

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	0.687	tested_positive
Weighted Avg.	0.718	0.282	0.718	0.718	0.718	0.436	0.718	0.668	

=== Confusion Matrix ===

a

b

<-- classified as

274.94

109.06

|

a = tested\_negative

107.46

276.54

|

b = tested\_positive

### 3. Heart Disease Dataset

- Applied Class Balancer in the Pre-processing step since the classes were distributed randomly and to reduce the bias.
- KNN Algorithm given more efficient result and Decision Tree algorithm gives less efficient results when compared with others.

#### KNN

When k=1, the results were more optimal.

Classifier

Choose J48 -C 0.25 -M 2

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Folds 10

% 66

More options...

(Nom) Class

Start

Stop

Result list (right-click for options)

23:40:22 - lazy Jk

23:41:01 - lazy Jk

23:41:14 - lazy Jk

23:41:24 - lazy Jk

23:59:06 - functions SMO

23:59:13 - functions MultilayerPerceptron

23:59:34 - trees J48

Classifier output

=== Classifier model (full training set) ===  
  
IB1 instance-based classifier  
using 1 nearest neighbour(s) for classification  
  
Time taken to build model: 0 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  
  
Correctly Classified Instances 187.6212 96.216 %  
Incorrectly Classified Instances 7.3788 3.784 %  
Kappa statistic 0.9243  
Mean absolute error 0.043  
Root mean squared error 0.1935  
Relative absolute error 8.6043 %  
Root relative squared error 38.6842 %  
Total Number of Instances 195  
  
=== Detailed Accuracy By Class ===  

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.958	0.034	0.966	0.958	0.962	0.924	0.967	0.943	1
	0.966	0.042	0.959	0.966	0.962	0.924	0.967	0.963	2
Weighted Avg.	0.962	0.038	0.962	0.962	0.962	0.924	0.967	0.953	

  
=== Confusion Matrix ===  

a	b	<-- classified as	
93.44	4.06	a = 1	
3.32	94.18	b = 2	

#### Support Vector Machine

Classifier

Choose J48 -C 0.25 -M 2

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Folds 10

% 66

More options...

(Nom) Class

Start

Stop

Result list (right-click for options)

23:40:22 - lazy Jk

23:41:01 - lazy Jk

23:41:14 - lazy Jk

23:41:24 - lazy Jk

23:59:06 - functions SMO

23:59:13 - functions MultilayerPerceptron

23:59:34 - trees J48

Classifier output

Number of kernel evaluations: 2516 (71.96% cached)  
  
Time taken to build model: 0.03 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  
  
Correctly Classified Instances 156.8622 80.4422 %  
Incorrectly Classified Instances 38.1378 19.5578 %  
Kappa statistic 0.6088  
Mean absolute error 0.1956  
Root mean squared error 0.4422  
Relative absolute error 39.1054 %  
Root relative squared error 88.424 %  
Total Number of Instances 195  
  
=== Detailed Accuracy By Class ===  

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.833	0.224	0.788	0.833	0.810	0.610	0.804	0.740	1
	0.776	0.167	0.823	0.776	0.799	0.610	0.804	0.751	2
Weighted Avg.	0.804	0.196	0.805	0.804	0.804	0.610	0.804	0.745	

  
=== Confusion Matrix ===  

a	b	<-- classified as	
81.25	16.25	a = 1	
21.89	75.61	b = 2	

# Artificial Neural Network

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Folds 10

Percentage split % 66

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

23:40:22 - lazy.JBk

23:41:01 - lazy.JBk

23:41:14 - lazy.JBk

23:41:24 - lazy.JBk

23:59:06 - functions.SMO

23:59:13 - functions.MultilayerPerceptron

23:59:34 - trees.J48

Classifier output

Node 0

Class 2

Input

Node 1

Time taken to build model: 0.45 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances177.547891.0502 %

Incorrectly Classified Instances17.45228.9498 %

Kappa statistic0.821

Mean absolute error0.118

Root mean squared error0.2769

Relative absolute error23.5868 %

Root relative squared error55.3672 %

Total Number of Instances195

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.896	0.075	0.923	0.896	0.909	0.821	0.953	0.943	1
	0.925	0.104	0.899	0.925	0.912	0.821	0.953	0.958	2
Weighted Avg.	0.911	0.089	0.911	0.911	0.910	0.821	0.953	0.950	

=== Confusion Matrix ===

a b <-- classified as

87.34 10.16 | a = 1

7.3 90.2 | b = 2

# Decision Tree

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Folds 10

Percentage split % 66

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

23:40:22 - lazy.JBk

23:41:01 - lazy.JBk

23:41:14 - lazy.JBk

23:41:24 - lazy.JBk

23:59:06 - functions.SMO

23:59:13 - functions.MultilayerPerceptron

23:59:34 - trees.J48

Classifier output

Number of Leaves : 14

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances151.348977.6148 %

Incorrectly Classified Instances43.651122.3852 %

Kappa statistic0.5523

Mean absolute error0.2346

Root mean squared error0.4608

Relative absolute error46.9011 %

Root relative squared error92.1327 %

Total Number of Instances195

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.729	0.177	0.805	0.729	0.765	0.555	0.792	0.763	1
	0.823	0.271	0.752	0.823	0.786	0.555	0.791	0.727	2
Weighted Avg.	0.776	0.224	0.779	0.776	0.776	0.555	0.791	0.745	

=== Confusion Matrix ===

a b <-- classified as

71.09 26.41 | a = 1

17.24 80.26 | b = 2

## 4. Fertility Dataset

- Applied Class Balancer in the Pre-processing step since the classes were distributed randomly and to reduce the bias.
- KNN Algorithm shows more efficient result and ANN Algorithm gives less efficient results when compared with others.

### K Nearest Neighbor

When k= 5, the results were more optimal.

The screenshot shows the Weka Classifier window with the 'IBk' classifier selected. The 'Test options' tab is active, showing 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
=== Classifier model (full training set) ===
IB1 instance-based classifier
using 5 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      74.053      74.053 %
Incorrectly Classified Instances    25.947      25.947 %
Kappa statistic                    0.4811
Mean absolute error                 0.319
Root mean squared error             0.4558
Relative absolute error             63.5455 %
Root relative squared error         90.7713 %
Total Number of Instances          100

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.648    0.167    0.795     0.648    0.714      0.490    0.761    0.716     1
          0.833    0.352    0.703     0.833    0.763      0.490    0.761    0.704     2
Weighted Avg.   0.741    0.255    0.749     0.741    0.738      0.490    0.761    0.710

=== Confusion Matrix ===

  a    b  <-- classified as
32.39 17.61 |   a = 1
 8.33 41.67 |   b = 2
```

### Support Vector Machine

The screenshot shows the Weka Classifier window with the 'J48' classifier selected. The 'Test options' tab is active, showing 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
Number of kernel evaluations: 1662 (81.073% cached)

Time taken to build model: 0.02 seconds

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

Correctly Classified Instances      57.7652      57.7652 %
Incorrectly Classified Instances    42.2348      42.2348 %
Kappa statistic                    0.1553
Mean absolute error                 0.4223
Root mean squared error             0.6499
Relative absolute error             84.1337 %
Root relative squared error         129.4259 %
Total Number of Instances          100

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.739    0.583    0.559     0.739    0.636      0.164    0.578    0.543     1
          0.417    0.261    0.615     0.417    0.497      0.164    0.578    0.548     2
Weighted Avg.   0.578    0.422    0.587     0.578    0.566      0.164    0.578    0.546

=== Confusion Matrix ===

  a    b  <-- classified as
36.93 13.07 |   a = 1
29.17 20.83 |   b = 2
```



## Decision Tree

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)

00:03:20 - lazy.JBk

00:03:31 - lazy.JBk

00:06:23 - lazy.JBk

00:06:34 - lazy.JBk

00:07:35 - functions.SMO

00:19:21 - functions.MultilayerPerceptron

00:19:31 - trees.J48

Classifier output

Number of Leaves : 10

Size of the tree : 19

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	55.8712	55.8712 %
Incorrectly Classified Instances	44.1288	44.1288 %
Kappa statistic	0.1174	
Mean absolute error	0.4334	
Root mean squared error	0.646	
Relative absolute error	86.3431 %	
Root relative squared error	128.6432 %	
Total Number of Instances	100	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.784	0.667	0.540	0.784	0.640	0.132	0.589	0.546	1
	0.333	0.216	0.607	0.333	0.430	0.132	0.589	0.634	2
Weighted Avg.	0.559	0.441	0.574	0.559	0.535	0.132	0.589	0.590	

=== Confusion Matrix ===

a b <-- classified as

39.2	10.8		a = 1
33.33	16.67		b = 2

## Artificial Neural Network

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)

00:03:20 - lazy.JBk

00:03:31 - lazy.JBk

00:06:23 - lazy.JBk

00:06:34 - lazy.JBk

00:07:35 - functions.SMO

00:19:21 - functions.MultilayerPerceptron

00:19:31 - trees.J48

Classifier output

Node 0

Class 2

Input

Node 1

Time taken to build model: 0.08 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	55.6818	55.6818 %
Incorrectly Classified Instances	44.3182	44.3182 %
Kappa statistic	0.1136	
Mean absolute error	0.4218	
Root mean squared error	0.5967	
Relative absolute error	84.0318 %	
Root relative squared error	118.8391 %	
Total Number of Instances	100	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.864	0.750	0.535	0.864	0.661	0.144	0.667	0.657	1
	0.250	0.136	0.647	0.250	0.361	0.144	0.667	0.682	2
Weighted Avg.	0.557	0.443	0.591	0.557	0.511	0.144	0.667	0.670	

=== Confusion Matrix ===

a b <-- classified as

43.18	6.82		a = 1
37.5	12.5		b = 2

## 5. Breast Cancer Dataset

- Applied Class Balancer in the Pre-processing step since the classes were distributed randomly and to reduce the bias.
- SVM Algorithm is more efficient out of all and Decision Tree Algorithm is least efficient.

### K Nearest Neighbor

When k=20, the results were more optimal.

The screenshot shows the RStudio Classifier window with the following details:

- Classifier:** MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
- Test options:** Cross-validation, Folds: 10
- Classifier output:**
  - Classifier model (full training set) ===
  - IB1 instance-based classifier using 20 nearest neighbour(s) for classification
  - Time taken to build model: 0 seconds
  - Stratified cross-validation summary:

	Correctly Classified Instances	Incorrectly Classified Instances
Summary	185.3936	100.6064
  - Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.826	0.529	0.609	0.826	0.701	0.317	0.696	0.670	no-recurrence-events
	0.471	0.174	0.730	0.471	0.572	0.317	0.696	0.729	recurrence-events
  - Confusion Matrix:

a \ b	no-recurrence-events	recurrence-events
no-recurrence-events	118.1	24.9
recurrence-events	75.71	67.29

### Support Vector Machine

The screenshot shows the RStudio Classifier window with the following details:

- Classifier:** MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
- Test options:** Cross-validation, Folds: 10
- Classifier output:**
  - Number of kernel evaluations: 34884 (93.07% cached)
  - Time taken to build model: 0.09 seconds
  - Stratified cross-validation summary:

	Correctly Classified Instances	Incorrectly Classified Instances
Summary	185.854	100.146
  - Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.711	0.412	0.633	0.711	0.670	0.302	0.650	0.595	no-recurrence-events
	0.588	0.289	0.671	0.588	0.627	0.302	0.650	0.601	recurrence-events
  - Confusion Matrix:

a \ b	no-recurrence-events	recurrence-events
no-recurrence-events	101.74	41.26
recurrence-events	58.88	84.12

## Artificial Neural Network

**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   
☐ Percentage split %   
More options...  
(Nom) Class  
Start Stop

**Result list (right-click for options)**  
23:08:04 - lazy.IBk  
23:08:36 - lazy.IBk  
23:08:47 - lazy.IBk  
23:09:29 - functions.SMO  
23:16:57 - trees.J48  
23:17:05 - functions.MultilayerPerceptron

**Classifier output**  
Node 0  
Class recurrence-events  
Input  
Node 1  
  
Time taken to build model: 2.78 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  

Correctly Classified Instances	176.4043	61.6798 %
Incorrectly Classified Instances	109.5957	38.3202 %
Kappa statistic	0.2336	
Mean absolute error	0.3825	
Root mean squared error	0.5838	
Relative absolute error	76.4855 %	
Root relative squared error	116.7402 %	
Total Number of Instances	286	

  
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.751	0.518	0.592	0.751	0.662	0.243	0.646	0.637	no-recurrence-events
	0.482	0.249	0.660	0.482	0.557	0.243	0.646	0.657	recurrence-events
Weighted Avg.	0.617	0.383	0.626	0.617	0.610	0.243	0.646	0.647	

  
=== Confusion Matrix ===  
a b <-- classified as  
107.43 35.57 | a = no-recurrence-events  
74.02 68.98 | b = recurrence-events

## Decision Tree

**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   
☐ Percentage split %   
More options...  
(Nom) Class  
Start Stop

**Result list (right-click for options)**  
23:08:04 - lazy.IBk  
23:08:36 - lazy.IBk  
23:08:47 - lazy.IBk  
23:09:29 - functions.SMO  
23:16:57 - trees.J48  
23:17:05 - functions.MultilayerPerceptron

**Classifier output**  
Number of Leaves : 27  
Size of the tree : 31  
  
Time taken to build model: 0.01 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  

Correctly Classified Instances	169.7419	59.3503 %
Incorrectly Classified Instances	116.2581	40.6497 %
Kappa statistic	0.187	
Mean absolute error	0.4728	
Root mean squared error	0.5192	
Relative absolute error	94.5445 %	
Root relative squared error	103.8343 %	
Total Number of Instances	286	

  
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.716	0.529	0.575	0.716	0.638	0.193	0.565	0.524	no-recurrence-events
	0.471	0.284	0.624	0.471	0.537	0.193	0.565	0.559	recurrence-events
Weighted Avg.	0.594	0.406	0.600	0.594	0.587	0.193	0.565	0.541	

  
=== Confusion Matrix ===  
a b <-- classified as  
102.45 40.55 | a = no-recurrence-events  
75.71 67.29 | b = recurrence-events