

Decision Tree

- I) Vote.arff
 - a) Without removing attributes

Classifier

Choose RandomTree -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) Class

Start Stop

Result list (right-click for options)

12:12:53 - trees.RandomTree

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      407          93.5632 %
Incorrectly Classified Instances    28           6.4368 %
Kappa statistic                    0.8636
Mean absolute error                0.0699
Root mean squared error            0.2379
Relative absolute error            14.7341 %
Root relative squared error        48.8605 %
Total Number of Instances         435

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.955   0.095   0.941     0.955   0.948     0.864   0.966    0.971    democrat
               0.905   0.045   0.927     0.905   0.916     0.864   0.967    0.937    republican
Weighted Avg.   0.936   0.076   0.936     0.936   0.935     0.864   0.966    0.958

=== Confusion Matrix ===
  a  b  <-- classified as
255 12 | a = democrat
 16 152 | b = republican
  
```

Accuracy = 93.56 %

Precision = $255 / (255 + 16) = 0.94$

Recall = $255 / (255 + 12) = 0.96$

- b) After removing three attributes (mix-missive, immigration, crime)

Classifier

Choose RandomTree -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) Class

Start Stop

Result list (right-click for options)

12:12:53 - trees.RandomTree

12:19:47 - trees.RandomTree

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      409          94.023 %
Incorrectly Classified Instances    26           5.977 %
Kappa statistic                    0.8739
Mean absolute error                0.0668
Root mean squared error            0.22
Relative absolute error            14.0774 %
Root relative squared error        45.1933 %
Total Number of Instances         435

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.951   0.077   0.951     0.951   0.951     0.874   0.969    0.973    democrat
               0.923   0.049   0.923     0.923   0.923     0.874   0.968    0.953    republican
Weighted Avg.   0.940   0.066   0.940     0.940   0.940     0.874   0.969    0.965

=== Confusion Matrix ===
  a  b  <-- classified as
254 13 | a = democrat
 13 155 | b = republican
  
```

Accuracy = 94.023%

Precision = $254 / (254 + 13) = 0.95$

Recall = $254 / (254 + 13) = 0.95$

II) Diabetes.arff
a) Without removing attributes

Classifier
Choose RandomTree -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) class
 Start Stop

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      523           68.099 %
Incorrectly Classified Instances    245           31.901 %
Kappa statistic                    0.3033
Mean absolute error                0.319
Root mean squared error            0.5648
Relative absolute error            70.1883 %
Root relative squared error        118.4973 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.746   0.440   0.760     0.746   0.753     0.303   0.653    0.732   tested_negative
               0.560   0.254   0.542     0.560   0.550     0.303   0.653    0.457   tested_positive
Weighted Avg.   0.681   0.375   0.684     0.681   0.682     0.303   0.653    0.636

=== Confusion Matrix ===
  a  b  <-- classified as
373 127 | a = tested_negative
118 150 | b = tested_positive
  
```

Result list (right-click for options)
 12:23:04 - trees.RandomTree

Accuracy = 68.099%

Precision = $373/(373+118) = 0.76$

Recall = $373/(373+127) = 0.746$

b) After removing three attributes (press, skin, mass)

Classifier
Choose RandomTree -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) class
 Start Stop

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      531           69.1406 %
Incorrectly Classified Instances    237           30.8594 %
Kappa statistic                    0.3249
Mean absolute error                0.3086
Root mean squared error            0.5555
Relative absolute error            67.8965 %
Root relative squared error        116.5466 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.756   0.429   0.767     0.756   0.761     0.325   0.663    0.739   tested_negative
               0.571   0.244   0.556     0.571   0.564     0.325   0.663    0.467   tested_positive
Weighted Avg.   0.691   0.365   0.693     0.691   0.692     0.325   0.663    0.644

=== Confusion Matrix ===
  a  b  <-- classified as
378 122 | a = tested_negative
115 153 | b = tested_positive
  
```

Result list (right-click for options)
 12:23:04 - trees.RandomTree
 12:27:51 - trees.RandomTree

Accuracy = 69.14%

Precision = $378/(378+115) = 0.77$

Recall = $378/(378+122) = 0.76$

Random Forest

- I) Vote.arff
 - a) Without removing attributes

The Classifier window displays the following information:

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) Class: (Nom) Class

Start **Stop**

Result list (right-click for options):

- 12:31:59 -trees RandomForest

Classifier output:

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      418          96.092 %
Incorrectly Classified Instances    17           3.908 %
Kappa statistic                    0.9175
Mean absolute error                 0.0714
Root mean squared error             0.1742
Relative absolute error             15.0587 %
Root relative squared error         35.7776 %
Total Number of Instances          435

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.970    0.054    0.966      0.970    0.968      0.917    0.993    0.996    democrat
          0.946    0.030    0.952      0.946    0.949      0.917    0.993    0.988    republican
Weighted Avg.   0.961    0.044    0.961      0.961    0.961      0.917    0.993    0.993

=== Confusion Matrix ===
  a  b  <-- classified as
259  8  | a = democrat
  9 159 | b = republican
  
```

Accuracy = 96.092 %

Precision = $259 / (259 + 9) = 0.97$

Recall = $259 / (259 + 8) = 0.97$

- c) After removing three attributes (education-spending, superfund-right-to-sue, crime)

The Classifier window displays the following information:

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) Class: (Nom) Class

Start **Stop**

Result list (right-click for options):

- 12:31:59 -trees RandomForest
- 12:37:57 -trees RandomForest

Classifier output:

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      419          96.3218 %
Incorrectly Classified Instances    16           3.6782 %
Kappa statistic                    0.9224
Mean absolute error                 0.0693
Root mean squared error             0.1734
Relative absolute error             14.6171 %
Root relative squared error         35.6039 %
Total Number of Instances          435

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.970    0.048    0.970      0.970    0.970      0.922    0.994    0.996    democrat
          0.952    0.030    0.952      0.952    0.952      0.922    0.994    0.990    republican
Weighted Avg.   0.963    0.041    0.963      0.963    0.963      0.922    0.994    0.994

=== Confusion Matrix ===
  a  b  <-- classified as
259  8  | a = democrat
  8 160 | b = republican
  
```

Accuracy = 96.3218%

Precision = $259 / (259 + 8) = 0.97$

Recall = $259 / (259 + 8) = 0.97$

II) Diabetes.arff
a) Without removing attributes

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
☒ Cross-validation Folds **10**
☐ Percentage split % 66

 (Nom) class

Result list (right-click for options)
 12:41:29 - trees.RandomForest

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      582           75.7813 %
Incorrectly Classified Instances    186           24.2188 %
Kappa statistic                    0.4566
Mean absolute error                 0.3106
Root mean squared error             0.4031
Relative absolute error             68.3405 %
Root relative squared error        84.5604 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.836   0.388   0.801    0.836   0.818     0.458   0.820    0.886   tested_negative
               0.612   0.164   0.667    0.612   0.638     0.458   0.820    0.679   tested_positive
Weighted Avg.   0.758   0.310   0.754    0.758   0.755     0.458   0.820    0.814

=== Confusion Matrix ===
  a  b  <-- classified as
418 82 | a = tested_negative
104 164 | b = tested_positive
  
```

Accuracy = 75.78%

Precision = $418 / (418 + 108) = 0.795$

Recall = $418 / (418 + 82) = 0.836$

b) After removing two attributes (preg, insu)

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
☒ Cross-validation Folds **10**
☐ Percentage split % 66

 (Nom) class

Result list (right-click for options)
 12:41:29 - trees.RandomForest
 12:56:35 - trees.RandomForest

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      587           76.4323 %
Incorrectly Classified Instances    181           23.5677 %
Kappa statistic                    0.4717
Mean absolute error                 0.31
Root mean squared error             0.405
Relative absolute error             68.2059 %
Root relative squared error        84.9767 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.840   0.377   0.806    0.840   0.823     0.473   0.820    0.893   tested_negative
               0.623   0.160   0.676    0.623   0.649     0.473   0.820    0.673   tested_positive
Weighted Avg.   0.764   0.301   0.761    0.764   0.762     0.473   0.820    0.816

=== Confusion Matrix ===
  a  b  <-- classified as
420 80 | a = tested_negative
101 167 | b = tested_positive
  
```

Accuracy = 76.43%

Precision = $420 / (420 + 101) = 0.81$

Recall = $420 / (420 + 80) = 0.84$

Naïve Bayes

- I) Vote.arff
- a) Without removing attributes

Classifier

Choose **NaiveBayes**

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)

- 12:35:16 - bayes.NaiveBayes

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      392          90.1149 %
Incorrectly Classified Instances    43           9.8851 %
Kappa statistic                    0.7949
Mean absolute error                 0.0995
Root mean squared error             0.2977
Relative absolute error             20.9815 %
Root relative squared error         61.1406 %
Total Number of Instances          435

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
               0.891   0.083   0.944    0.891   0.917     0.797  0.973    0.984   democrat
               0.917   0.109   0.842    0.917   0.877     0.797  0.973    0.957   republican
Weighted Avg.   0.901   0.093   0.905    0.901   0.902     0.797  0.973    0.973

=== Confusion Matrix ===
  a  b  <-- classified as
238 29 | a = democrat
 14 154 | b = republican
    
```

Accuracy = 90.1149 %

Precision = $238 / (238 + 14) = 0.94$

Recall = $238 / (238 + 29) = 0.89$

- b) After removing three attributes (adoption-of-the-budget-resolution, religious-groups-in-school, mx-missile)

Classifier

Choose **NaiveBayes**

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)

- 12:35:16 - bayes.NaiveBayes
- 12:43:21 - bayes.NaiveBayes

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      394          90.5747 %
Incorrectly Classified Instances    41           9.4253 %
Kappa statistic                    0.804
Mean absolute error                 0.0941
Root mean squared error             0.2846
Relative absolute error             19.8394 %
Root relative squared error         58.4448 %
Total Number of Instances          435

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
               0.899   0.083   0.945    0.899   0.921     0.806  0.973    0.984   democrat
               0.917   0.101   0.851    0.917   0.883     0.806  0.973    0.955   republican
Weighted Avg.   0.906   0.090   0.909    0.906   0.906     0.806  0.973    0.973

=== Confusion Matrix ===
  a  b  <-- classified as
240 27 | a = democrat
 14 154 | b = republican
    
```

Accuracy = 90.5747%

Precision = $240 / (240 + 14) = 0.95$

Recall = $240 / (240 + 27) = 0.90$

- II) Diabetes.arff
- a) Without removing attributes

The Classifier window shows the NaiveBayes classifier results for the Diabetes.arff dataset. The test options are set to Cross-validation with 10 folds. The classifier output displays the following summary and detailed accuracy by class.

Summary:

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 586 | 76.3021 % |
| Incorrectly Classified Instances | 182 | 23.6979 % |
| Kappa statistic | 0.4664 | |
| Mean absolute error | 0.2841 | |
| Root mean squared error | 0.4168 | |
| Relative absolute error | 62.5028 % | |
| Root relative squared error | 87.4349 % | |
| Total Number of Instances | 768 | |

Detailed Accuracy By Class:

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-----------------|
| Weighted Avg. | 0.844 | 0.388 | 0.802 | 0.844 | 0.823 | 0.468 | 0.819 | 0.892 | tested_negative |
| | 0.612 | 0.156 | 0.678 | 0.612 | 0.643 | 0.468 | 0.819 | 0.671 | tested_positive |

Confusion Matrix:

```

a  b  <-- classified as
422 78 | a = tested_negative
104 164 | b = tested_positive
  
```

Accuracy = 76.3021%

Precision = $422 / (422 + 104) = 0.802$

Recall = $422 / (422 + 78) = 0.844$

- b) After removing two attributes (preg, insu)

The Classifier window shows the NaiveBayes classifier results for the Diabetes.arff dataset after removing two attributes (preg, insu). The test options are set to Cross-validation with 10 folds. The classifier output displays the following summary and detailed accuracy by class.

Summary:

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 590 | 76.8229 % |
| Incorrectly Classified Instances | 178 | 23.1771 % |
| Kappa statistic | 0.4707 | |
| Mean absolute error | 0.2942 | |
| Root mean squared error | 0.4034 | |
| Relative absolute error | 64.7195 % | |
| Root relative squared error | 84.632 % | |
| Total Number of Instances | 768 | |

Detailed Accuracy By Class:

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-----------------|
| Weighted Avg. | 0.864 | 0.410 | 0.797 | 0.864 | 0.829 | 0.474 | 0.826 | 0.888 | tested_negative |
| | 0.590 | 0.136 | 0.699 | 0.590 | 0.640 | 0.474 | 0.826 | 0.696 | tested_positive |

Confusion Matrix:

```

a  b  <-- classified as
432 68 | a = tested_negative
110 158 | b = tested_positive
  
```

Accuracy = 76.8229%

Precision = $432 / (432 + 110) = 0.797$

Recall = $432 / (432 + 68) = 0.864$