Data Analytics

Assignment 7

Logic-based approaches

Group members:

• Nanxun Xie

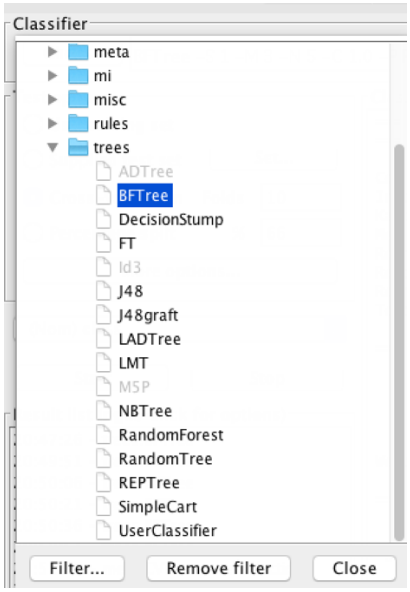
• Yifan Zhao

Using System: Weka

1. Build Classification Models for Two Data Sets

1.1 Classification Models for Iris

Firstly, open the iris.arff file in Weka, and then choose different classifier and set optimized parameters. Here is a screenshot to show the tree options.



A. BFTree

The accuracy for the default parameters is 94.67%.

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

Best-First Decision Tree

PL < 2.45: Iris_setosa(50.0/0.0)
PL >= 2.45
|  PW < 1.75
|  |  PL < 4.95: Iris_versicolor(47.0/1.0)
|  |  PL >= 4.95
|  |  |  PW < 1.55: Iris_virginica(3.0/0.0)
|  |  |  PW >= 1.55: Iris_versicolor(2.0/1.0)
|  PW >= 1.75
|  |  PL < 4.85: Iris_virginica(2.0/1.0)
|  |  PL >= 4.85: Iris_virginica(43.0/0.0)

Size of the Tree: 11

Number of Leaf Nodes: 6

Time taken to build model: 0.06 seconds

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

Correctly Classified Instances         142               94.6667 %
```

After several attempts, we found that the result would be most accurate if we change the minNumObj value from 2 to 3. The accuracy increased to

96.0%.

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

Best-First Decision Tree

PL < 2.45: Iris_setosa(50.0/0.0)
PL >= 2.45
|  PW < 1.75
|  |  PL < 4.95: Iris_versicolor(47.0/1.0)
|  |  PL >= 4.95: Iris_virginica(4.0/2.0)
|  PW >= 1.75: Iris_virginica(45.0/1.0)

Size of the Tree: 7

Number of Leaf Nodes: 4

Time taken to build model: 0 seconds

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

Correctly Classified Instances         144               96      %
```

B. DecisionStump

The accuracy for the default parameters is 66.67%. We can not set different parameters by using this approach.

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

Decision Stump

Classifications

PL <= 2.45 : Iris_setosa
PL > 2.45 : Iris_versicolor
PL is missing : Iris_setosa

Class distributions

PL <= 2.45
Iris_setosa     Iris_versicolor Iris_virginica
1.0     0.0     0.0
PL > 2.45
Iris_setosa     Iris_versicolor Iris_virginica
0.0     0.5     0.5
PL is missing
Iris_setosa     Iris_versicolor Iris_virginica
0.3333333333333333      0.3333333333333333      0.3333333333333333


Time taken to build model: 0 seconds

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

Correctly Classified Instances         100               66.6667 %
```

C. FT

The accuracy for the default parameters is 96.67%.

```
FT tree
------------------
: FT_1:15/15 (150)

Number of Leaves  :     1

Size of the Tree :      1
FT_1:
Class 0 :
23.67 +
[PL] * -7.42 +
[PW] * -5.71

Class 1 :
-5.82 +
[SL] * 1.67 +
[SW] * 0.11 +
[PL] * -0.4 +
[PW] * -1.28

Class 2 :
-31.02 +
[SL] * -0.4 +
[SW] * -3.32 +
[PL] * 5.56 +
[PW] * 9.9


Time taken to build model: 0.2 seconds

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

Correctly Classified Instances         145               96.6667 %
```

Then, we set different value of parameters. However, the result came worse.

D. J48

The accuracy for the default parameters is 96.0%.

```
J48 pruned tree
------------------

PW <= 0.6: Iris_setosa (50.0)
PW > 0.6
|   PW <= 1.7
|   |   PL <= 4.9: Iris_versicolor (48.0/1.0)
|   |   PL > 4.9
|   |   |   PW <= 1.5: Iris_virginica (3.0)
|   |   |   PW > 1.5: Iris_versicolor (3.0/1.0)
|   PW > 1.7: Iris_virginica (46.0/1.0)

Number of Leaves  :     5

Size of the tree :      9


Time taken to build model: 0 seconds

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

Correctly Classified Instances         144               96      %
```

Then, we set different value of parameters. However, the result came worse.

E. J48graft

The accuracy for the default parameters is 94.67%.

```
PW <= 0.6: Iris_setosa (50.0)
PW > 0.6
|   PW <= 1.7
|   |   PL <= 4.9: Iris_versicolor (48.0/1.0)
|   |   PL > 4.9
|   |   |   PW <= 1.5
|   |   |   |   PW <= 1.35: Iris_versicolor (0.0|28.0)
|   |   |   |   PW > 1.35
|   |   |   |   |   SL <= 5.95: Iris_versicolor (0.0|25.0)
|   |   |   |   |   SL > 5.95: Iris_virginica (3.0)
|   |   |   PW > 1.5
|   |   |   |   PL <= 5.85
|   |   |   |   |   SL <= 7.25
|   |   |   |   |   |   SW <= 2.65: Iris_virginica (0.0|7.0)
|   |   |   |   |   |   SW > 2.65
|   |   |   |   |   |   |   SW <= 3.05: Iris_versicolor (3.0/1.0)
|   |   |   |   |   |   |   SW > 3.05: Iris_virginica (0.0|19.0/2.0)
|   |   |   |   |   SL > 7.25: Iris_virginica (0.0|8.0)
|   |   |   |   PL > 5.85: Iris_virginica (0.0|13.0)
|   PW > 1.7: Iris_virginica (46.0/1.0)

Number of Leaves  :     11

Size of the tree :      21


Time taken to build model: 0.08 seconds

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

Correctly Classified Instances         142               94.6667 %
```

After several attempts, we found that the result would be most accurate if we change the minNumObj value from 2 to 3. The accuracy increased to 95.33%.

```
PW <= 0.6: Iris_setosa (50.0)
PW > 0.6
|   PW <= 1.7
|   |   PL <= 4.9: Iris_versicolor (48.0/1.0)
|   |   PL > 4.9
|   |   |   PW <= 1.5
|   |   |   |   PW <= 1.35: Iris_versicolor (0.0|28.0)
|   |   |   |   PW > 1.35
|   |   |   |   |   SL <= 5.95: Iris_versicolor (0.0|25.0)
|   |   |   |   |   SL > 5.95: Iris_virginica (3.0)
|   |   |   PW > 1.5
|   |   |   |   PL <= 5.85
|   |   |   |   |   SL <= 7.25
|   |   |   |   |   |   SW <= 2.65: Iris_virginica (0.0|7.0)
|   |   |   |   |   |   SW > 2.65
|   |   |   |   |   |   |   SW <= 3.05: Iris_versicolor (3.0/1.0)
|   |   |   |   |   |   |   SW > 3.05: Iris_virginica (0.0|19.0/2.0)
|   |   |   |   |   SL > 7.25: Iris_virginica (0.0|8.0)
|   |   |   |   PL > 5.85: Iris_virginica (0.0|13.0)
|   PW > 1.7: Iris_virginica (46.0/1.0)

Number of Leaves  :     11

Size of the tree :      21


Time taken to build model: 0 seconds

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

Correctly Classified Instances         143               95.3333 %
```

F. LADTree

The accuracy for the default parameters is 94.0%.

```
|  (1)PW < 0.8: 2,-1,-1
|  (1)PW >= 0.8: -1,0.5,0.5
|  |  (2)PW < 1.75: -0.584,1.39,-0.805
|  |  (2)PW >= 1.75: -0.584,-0.996,1.58
|  |  (10)PL < 4.45: -0.443,1.055,-0.612
|  |  (10)PL >= 4.45: -0.457,-0.185,0.642
|  (3)PL < 4.95: 0.355,0.256,-0.611
|  |  (4)PL < 2.45: 0.931,-0.474,-0.457
|  |  (4)PL >= 2.45: -0.539,0.572,-0.033
|  |  |  (5)SL < 4.95: -0.488,-2.064,2.552
|  |  |  (5)SL >= 4.95: -0.501,0.471,0.03
|  |  |  (6)PW < 1.65: -0.433,1.23,-0.797
|  |  |  (6)PW >= 1.65: -0.462,-0.343,0.805
|  |  |  |  (7)SW < 3.1: -0.449,-0.549,0.998
|  |  |  |  (7)SW >= 3.1: -0.447,2.89,-2.443
|  (3)PL >= 4.95: -0.447,-0.858,1.305
|  |  (8)PW < 1.55: -0.476,-1.117,1.593
|  |  (8)PW >= 1.55: -0.497,0.279,0.218
|  |  |  (9)PL < 5.15: -0.472,1.239,-0.767
|  |  |  (9)PL >= 5.15: -0.439,-0.769,1.208
Legend: Iris_setosa, Iris_versicolor, Iris_virginica
#Tree size (total): 31
#Tree size (number of predictor nodes): 21
#Leaves (number of predictor nodes): 14
#Expanded nodes: 100
#Processed examples: 5966
#Ratio e/n: 59.66

Time taken to build model: 0.09 seconds

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

Correctly Classified Instances         141               94      %
```

The result would be most accurate if we change the numOfBoostingIterations value from 10 to 30. The accuracy increased to 95.33%.

```
weka.classifiers.trees.LADTree:

: 0,0,0
|  (1)PW < 0.8: 2,-1,-1
|  (1)PW >= 0.8: -1,0.5,0.5
|  |  (2)PW < 1.75: -0.584,1.39,-0.805
|  |  |  (16)PL < 5.35: -0.445,0.669,-0.224
|  |  |  (16)PL >= 5.35: -0.444,-0.448,0.892
|  |  (2)PW >= 1.75: -0.584,-0.996,1.58
|  |  (10)PL < 4.45: -0.443,1.055,-0.612
|  |  (10)PL >= 4.45: -0.457,-0.185,0.642
|  |  (13)PL < 4.85: -0.448,0.687,-0.239
|  |  (13)PL >= 4.85: -0.448,-0.089,0.537
|  |  (14)PL < 4.95: -0.89,1.287,-0.397
|  |  (14)PL >= 4.95: -0.89,-0.267,1.158
|  |  |  (15)PW < 1.55: -0.444,-0.449,0.893
|  |  |  (15)PW >= 1.55: -0.445,0.567,-0.122
|  |  |  |  (17)PW < 1.75: -0.444,0.774,-0.329
|  |  |  |  |  (18)SL < 6.95: -0.444,0.89,-0.445
|  |  |  |  |  (18)SL >= 6.95: -0.444,-0.446,0.89
|  |  |  |  (17)PW >= 1.75: -0.444,-0.445,0.889
|  (3)PL < 4.95: 0.355,0.256,-0.611
|  |  (4)PL < 2.45: 0.931,-0.474,-0.457
|  |  (4)PL >= 2.45: -0.539,0.572,-0.033
|  |  |  (5)SL < 4.95: -0.488,-2.064,2.552
|  |  |  (5)SL >= 4.95: -0.501,0.471,0.03
|  |  |  (6)PW < 1.65: -1.322,3.015,-1.694
|  |  |  (6)PW >= 1.65: -1.352,-0.911,2.263
|  |  |  |  (7)SW < 3.1: -1.782,-1.896,3.678
|  |  |  |  (7)SW >= 3.1: -1.78,5.609,-3.828
```

```
|  (3)PL >= 4.95: -0.447,-0.858,1.305
|  |  (8)PW < 1.55: -0.476,-1.117,1.593
|  |  (8)PW >= 1.55: -0.497,0.279,0.218
|  |  |  (9)PL < 5.15: -0.917,1.91,-0.992
|  |  |  |  (11)PW < 1.75: -1.335,2.925,-1.589
|  |  |  |  (11)PW >= 1.75: -1.334,-1.358,2.692
|  |  |  (9)PL >= 5.15: -0.883,-1.232,2.115
|  (12)PL < 2.45: 2.681,-1.343,-1.338
|  (12)PL >= 2.45: -1.352,0.406,0.946
|  (19)PW < 1.35: 0.875,-0.417,-0.458
|  (19)PW >= 1.35: -0.445,-0.076,0.52
Legend: Iris_setosa, Iris_versicolor, Iris_virginica
#Tree size (total): 58
#Tree size (number of predictor nodes): 39
#Leaves (number of predictor nodes): 27
#Expanded nodes: 690
#Processed examples: 31732
#Ratio e/n: 45.98840579710145

Time taken to build model: 0.04 seconds

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

Correctly Classified Instances         143               95.3333 %
```

G. LMT

The accuracy for the default parameters is 94.0%.

```
Logistic model tree
------------------
: LM_1:18/18 (150)

Number of Leaves  :     1

Size of the Tree :      1
LM_1:
Class 0 :
29.99 +
[PL] * -9.96 +
[PW] * -5.71

Class 1 :
-6.15 +
[SL] * 1.67 +
[SW] * 0.82 +
[PL] * -0.74 +
[PW] * -1.28

Class 2 :
-34.94 +
[SL] * -0.4 +
[SW] * -3.76 +
[PL] * 6.27 +
[PW] * 10.89


Time taken to build model: 0.31 seconds

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

Correctly Classified Instances         141               94       %
```

The result would be most accurate if we change the numOfBoostingIterations value from -1 to 15 and the minNumInstances value from 15 to 39. The accuracy increased to 98.0%.

```
: LM_1:15/15 (150)

Number of Leaves  :     1

Size of the Tree :      1
LM_1:
Class 0 :
23.67 +
[PL] * -7.42 +
[PW] * -5.71

Class 1 :
-5.82 +
[SL] * 1.67 +
[SW] * 0.11 +
[PL] * -0.4 +
[PW] * -1.28

Class 2 :
-31.02 +
[SL] * -0.4 +
[SW] * -3.32 +
[PL] * 5.56 +
[PW] * 9.9


Time taken to build model: 0.05 seconds

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

Correctly Classified Instances         147               98      %
```

H. NBTree

The accuracy for the default parameters is 94.67%. We can not set different parameters by using this approach.

```
NBTree
------------------

SW <= 3.35
|   SW <= 2.95
|   |   PL <= 4.75: NB 3
|   |   PL > 4.75: NB 4
|   SW > 2.95: NB 5
SW > 3.35
|   SL <= 5.9: NB 7
|   SL > 5.9: NB 8

Leaf number: 3 Naive Bayes Classifier

                          Class
Attribute            Iris_setosa Iris_versicolor  Iris_virginica
                         (0.08)          (0.86)          (0.05)
===============================================================
SL
  'All'                     3.0            32.0             2.0
  [total]                   3.0            32.0             2.0

SW
  'All'                     3.0            32.0             2.0
  [total]                   3.0            32.0             2.0

PL
  '(-inf-2.2]'              3.0             1.0             1.0
  '(2.2-inf)'               1.0            32.0             2.0
  [total]                   4.0            33.0             3.0
```

```
Number of Leaves  :     5

Size of the tree :      9


Time taken to build model: 0.24 seconds

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

Correctly Classified Instances         142               94.6667 %
```

I. RandomForest

The accuracy for the default parameters is 95.33%.

```
Random forest of 100 trees, each constructed while considering 3 random features.
Out of bag error: 0.0467



Time taken to build model: 0.15 seconds

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

Correctly Classified Instances         143               95.3333 %
```

Then, we set different value of parameters. However, the result did not turn out to be better.

J. RandomTree

The accuracy for the default parameters is 91.33%.

```
RandomTree
==========

PL < 2.45 : Iris_setosa (50/0)
PL >= 2.45
|   PW < 1.75
|   |   PL < 4.95
|   |   |   PW < 1.65 : Iris_versicolor (47/0)
|   |   |   PW >= 1.65 : Iris_virginica (1/0)
|   |   PL >= 4.95
|   |   |   PW < 1.55 : Iris_virginica (3/0)
|   |   |   PW >= 1.55
|   |   |   |   SL < 6.95 : Iris_versicolor (2/0)
|   |   |   |   SL >= 6.95 : Iris_virginica (1/0)
|   PW >= 1.75
|   |   PL < 4.85
|   |   |   SL < 5.95 : Iris_versicolor (1/0)
|   |   |   SL >= 5.95 : Iris_virginica (2/0)
|   |   PL >= 4.85 : Iris_virginica (43/0)

Size of the tree : 17

Time taken to build model: 0 seconds

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

Correctly Classified Instances         137               91.3333 %
```

The result would be most accurate if we change the maxDepth value from 0 to 10 and the minNum value from 1 to 8 and the seed value from 1 to 10. The accuracy increased to 95.33%.

```
RandomTree
==========

PL < 2.45 : Iris_setosa (50/0)
PL >= 2.45
|   PW < 1.75
|   |   PL < 4.95
|   |   |   PW < 1.65 : Iris_versicolor (47/0)
|   |   |   PW >= 1.65 : Iris_virginica (1/0)
|   |   PL >= 4.95 : Iris_virginica (6/2)
|   PW >= 1.75
|   |   PL < 4.85 : Iris_virginica (3/1)
|   |   PL >= 4.85 : Iris_virginica (43/0)

Size of the tree : 11
Max depth of tree: 10

Time taken to build model: 0 seconds

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

Correctly Classified Instances         143               95.3333 %
```

K. REPTree

The accuracy for the default parameters is 94.0%.

```
REPTree
============

PL < 2.5 : Iris_setosa (33/0) [17/0]
PL >= 2.5
|   PW < 1.75 : Iris_versicolor (36/3) [18/2]
|   PW >= 1.75 : Iris_virginica (31/1) [15/0]

Size of the tree : 5

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         141               94      %
```

Then, we set different value of parameters. However, the result did not turn out to be better.

L. SimpleCart

The accuracy for the default parameters is 95.33%.

```
CART Decision Tree

PL < 2.45: Iris_setosa(50.0/0.0)
PL >= 2.45
|  PW < 1.75
|  |  PL < 4.95: Iris_versicolor(47.0/1.0)
|  |  PL >= 4.95
|  |  |  PW < 1.55: Iris_virginica(3.0/0.0)
|  |  |  PW >= 1.55: Iris_versicolor(2.0/1.0)
|  PW >= 1.75: Iris_virginica(45.0/1.0)

Number of Leaf Nodes: 5

Size of the Tree: 9

Time taken to build model: 0.03 seconds

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

Correctly Classified Instances         143               95.3333 %
```

Then, we set different value of parameters. However, the result did not turn out to be better.

1.2 Classification Models for Congressional Voting Records

Firstly, open the house-votes-84.arff file in Weka, and then choose different classifier, choose the (nom)party option and set optimized parameters. Here is a screenshot to show the tree options.

A. ADTree

The accuracy for the default parameters is 95.86%.

```
Alternating decision tree:

: 0.231
|  (1)physician_fee_freeze = y: -1.417
|  |  (5)adoption_of_the_budget_resolution = n: -0.558
|  |  |  (8)superfund_right_to_sue = y: -0.62
|  |  |  (8)superfund_right_to_sue != y: 0.59
|  |  (5)adoption_of_the_budget_resolution != n: 0.518
|  |  |  (7)nti_satellite_test_ban = n: 0.984
|  |  |  (7)nti_satellite_test_ban != n: -0.981
|  |  (6)immigration = n: 0.442
|  |  (6)immigration != n: -0.829
|  (1)physician_fee_freeze != y: 1.66
|  |  (2)adoption_of_the_budget_resolution = y: 1.775
|  |  (2)adoption_of_the_budget_resolution != y: -0.924
|  (3)synfuels_corporation_cutback = y: 0.914
|  |  (9)export_administration_act_sa = y: 0.011
|  |  (9)export_administration_act_sa != y: 0.791
|  (3)synfuels_corporation_cutback != y: -0.566
|  (4)education_spending = n: 0.684
|  (4)education_spending != n: -0.346
|  |  (10)physician_fee_freeze = n: 0.614
|  |  (10)physician_fee_freeze != n: -0.381
Legend: -ve = republican, +ve = democrat
Tree size (total number of nodes): 31
Leaves (number of predictor nodes): 21

Time taken to build model: 0.03 seconds

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

Correctly Classified Instances         417               95.8621 %
```

The result would be most accurate if we change the numOfBoostingIterations value from 10 to 25. The accuracy increased to 96.32%.

```
|  |  (9)export_administration_act_sa = y: 0.011
|  |  (9)export_administration_act_sa != y: 0.791
|  (3)synfuels_corporation_cutback != y: -0.566
|  |  (21)water_project_cost_sharing = n: -0.363
|  |  (21)water_project_cost_sharing != n: 0.161
|  (4)education_spending = n: 0.684
|  |  (11)duty_free_exports = n: -0.314
|  |  (11)duty_free_exports != n: 0.558
|  |  (15)immigration = y: -0.139
|  |  (15)immigration != y: 0.514
|  |  |  (18)aid_to_nicaraguan_contras = w: -0.229
|  |  |  (18)aid_to_nicaraguan_contras != w: 0.526
|  (4)education_spending != n: -0.346
|  |  (10)physician_fee_freeze = n: 0.614
|  |  |  (14)education_spending = y: 0.682
|  |  |  (14)education_spending != y: -0.392
|  |  (10)physician_fee_freeze != n: -0.381
|  (22)mx_missile = y: 0.344
|  |  (24)duty_free_exports = n: -0.249
|  |  (24)duty_free_exports != n: 0.554
|  (22)mx_missile != y: -0.119
|  |  (23)education_spending = w: -0.619
|  |  (23)education_spending != w: 0.187
Legend: -ve = republican, +ve = democrat
Tree size (total number of nodes): 73
Leaves (number of predictor nodes): 49

Time taken to build model: 0.02 seconds

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

Correctly Classified Instances         419               96.3218 %
```

B. BFTree

The accuracy for the default parameters is 95.17%.

```
physician_fee_freeze=(n)|(w)
|  adoption_of_the_budget_resolution=(y)|(n)
|  |  adoption_of_the_budget_resolution=(y)|(w): democrat(224.0/0.0)
|  |  adoption_of_the_budget_resolution!=(y)|(w): democrat(23.0/2.0)
|  adoption_of_the_budget_resolution!=(y)|(n)
|  |  mx_missile=(n)|(y)
|  |  |  education_spending=(n)|(w): democrat(5.0/0.0)
|  |  |  education_spending!=(n)|(w): republican(1.0/1.0)
|  |  mx_missile!=(n)|(y): republican(2.0/0.0)
physician_fee_freeze!=(n)|(w)
|  synfuels_corporation_cutback=(y)
|  |  adoption_of_the_budget_resolution=(w)|(y)
|  |  |  nti_satellite_test_ban=(n)|(w): democrat(6.0/0.0)
|  |  |  nti_satellite_test_ban!=(n)|(w): republican(3.0/0.0)
|  |  adoption_of_the_budget_resolution!=(w)|(y)
|  |  |  el_salvador_aid=(n): democrat(2.0/0.0)
|  |  |  el_salvador_aid!=(n)
|  |  |  |  export_administration_act_sa=(w)|(n): republican(8.0/3.0)
|  |  |  |  export_administration_act_sa!=(w)|(n): republican(10.0/0.0)
|  synfuels_corporation_cutback!=(y)
|  |  duty_free_exports=(y)
|  |  |  immigration=(n): republican(2.0/2.0)
|  |  |  immigration!=(n): republican(9.0/0.0)
|  |  duty_free_exports!=(y)
|  |  |  adoption_of_the_budget_resolution=(y)
|  |  |  |  export_administration_act_sa=(w): republican(1.0/1.0)
|  |  |  |  export_administration_act_sa!=(w): republican(12.0/0.0)
|  |  |  adoption_of_the_budget_resolution!=(y): republican(118.0/0.0)

Size of the Tree: 29

Number of Leaf Nodes: 15

Time taken to build model: 0.09 seconds
```

The result would be most accurate if we change the minNumObj value from 2 to 5 and seed value from 1 to 2. The accuracy increased to 95.86%.

```
Best-First Decision Tree

physician_fee_freeze=(n)|(w): democrat(253.0/5.0)
physician_fee_freeze!=(n)|(w): republican(163.0/14.0)

Size of the Tree: 3

Number of Leaf Nodes: 2

Time taken to build model: 0.05 seconds

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

Correctly Classified Instances         417               95.8621 %
```

C. DecisionStump

The accuracy for the default parameters is 95.63%. We can not set different parameters by using this approach.

```
Decision Stump

Classifications

physician_fee_freeze = y : republican
physician_fee_freeze != y : democrat
physician_fee_freeze is missing : democrat

Class distributions

physician_fee_freeze = y
republican      democrat
0.9209039548022598      0.07909604519774012
physician_fee_freeze != y
republican      democrat
0.01937984496124031     0.9806201550387597
physician_fee_freeze is missing
republican      democrat
0.38620689655172413     0.6137931034482759


Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         416               95.6322 %
```

D. FT

The accuracy for the default parameters is 96.78%.

```
FT tree
------------------

N0#1 <= 0.595535: FT_1:15/30 (268)
N0#1 > 0.595535: FT_2:15/30 (167)

Number of Leaves  :     2

Size of the Tree :      3
FT_N0#1:
Class 0 :
0.61 +
[water_project_cost_sharing=y] * -0.51 +
[adoption_of_the_budget_resolution=n] * 0.38 +
[adoption_of_the_budget_resolution=y] * -0.85 +
[physician_fee_freeze=y] * 1.8  +
[physician_fee_freeze=n] * -1.58 +
[nti_satellite_test_ban=n] * -0.38 +
[mx_missile=y] * -0.54 +
[immigration=n] * -0.59 +
[synfuels_corporation_cutback=n] * 0.46 +
[synfuels_corporation_cutback=y] * -0.8 +
[education_spending=n] * -0.58 +
[duty_free_exports=y] * -0.41 +
[export_administration_act_sa=y] * 0.41
```

Then, we set different value of parameters. However, the result did not turn out to be better.

E. Id3

The accuracy for the default parameters is 94.25%.

```
Id3


physician_fee_freeze = y
|  synfuels_corporation_cutback = w: republican
|  synfuels_corporation_cutback = n
|  |  duty_free_exports = n
|  |  |  adoption_of_the_budget_resolution = n: republican
|  |  |  adoption_of_the_budget_resolution = y
|  |  |  |  export_administration_act_sa = y: republican
|  |  |  |  export_administration_act_sa = w
|  |  |  |  |  handicapped_infants = n: democrat
|  |  |  |  |  handicapped_infants = w: null
|  |  |  |  |  handicapped_infants = y: republican
|  |  |  |  export_administration_act_sa = n: null
|  |  |  adoption_of_the_budget_resolution = w: null
|  |  duty_free_exports = y
|  |  |  immigration = y: republican
|  |  |  immigration = n
|  |  |  |  export_administration_act_sa = y: democrat
|  |  |  |  export_administration_act_sa = w
|  |  |  |  |  water_project_cost_sharing = y: republican
|  |  |  |  |  water_project_cost_sharing = n: democrat
|  |  |  |  |  water_project_cost_sharing = w: null
|  |  |  |  export_administration_act_sa = n: republican
|  |  |  immigration = w: null
|  |  duty_free_exports = w: republican
|  synfuels_corporation_cutback = y
|  |  adoption_of_the_budget_resolution = n
|  |  |  el_salvador_aid = y
|  |  |  |  export_administration_act_sa = y: republican
|  |  |  |  export_administration_act_sa = w
```

Then, we set different value of parameters. However, the result did not turn out to be better.

F. J48

The accuracy for the default parameters is 94.94%.

```
J48 pruned tree
------------------

physician_fee_freeze = y
|   synfuels_corporation_cutback = w: republican (7.0)
|   synfuels_corporation_cutback = n: republican (138.0/3.0)
|   synfuels_corporation_cutback = y
|   |   mx_missile = n
|   |   |   adoption_of_the_budget_resolution = n: republican (21.0/3.0)
|   |   |   adoption_of_the_budget_resolution = y
|   |   |   |   water_project_cost_sharing = y: democrat (4.0)
|   |   |   |   water_project_cost_sharing = n: republican (2.0)
|   |   |   |   water_project_cost_sharing = w: democrat (0.0)
|   |   |   adoption_of_the_budget_resolution = w: republican (0.0)
|   |   mx_missile = y: democrat (5.0/1.0)
|   |   mx_missile = w: republican (0.0)
physician_fee_freeze = w: democrat (11.0/3.0)
physician_fee_freeze = n: democrat (247.0/2.0)

Number of Leaves  :     11

Size of the tree :      16


Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         413               94.9425 %
```

The result would be most accurate if we change the minNumObj value from 2 to 7 and confidenceFactor value from 0.25 to 1. The accuracy increased to 96.32%.

```
J48 pruned tree
------------------

physician_fee_freeze = y
|   synfuels_corporation_cutback = w: republican (7.0)
|   synfuels_corporation_cutback = n: republican (138.0/3.0)
|   synfuels_corporation_cutback = y
|   |   adoption_of_the_budget_resolution = n: republican (23.0/5.0)
|   |   adoption_of_the_budget_resolution = y: democrat (8.0/3.0)
|   |   adoption_of_the_budget_resolution = w: democrat (1.0)
physician_fee_freeze = w: democrat (11.0/3.0)
physician_fee_freeze = n
|   adoption_of_the_budget_resolution = n
|   |   education_spending = y: democrat (10.0)
|   |   education_spending = n: democrat (14.0/1.0)
|   |   education_spending = w: republican (1.0)
|   adoption_of_the_budget_resolution = y: democrat (219.0)
|   adoption_of_the_budget_resolution = w: democrat (3.0)

Number of Leaves  :     11

Size of the tree :      16


Time taken to build model: 0.05 seconds

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

Correctly Classified Instances         419               96.3218 %
```

G. J48graft

The accuracy for the default parameters is 95.40%.

```
J48graft pruned tree
------------------

physician_fee_freeze = y
|   synfuels_corporation_cutback = w: republican (7.0)
|   synfuels_corporation_cutback = n: republican (138.0/3.0)
|   synfuels_corporation_cutback = y
|   |   mx_missile = n
|   |   |   adoption_of_the_budget_resolution = n: republican (21.0/3.0)
|   |   |   adoption_of_the_budget_resolution = y
|   |   |   |   water_project_cost_sharing = y: democrat (4.0)
|   |   |   |   water_project_cost_sharing = n: republican (2.0)
|   |   |   |   water_project_cost_sharing = w: democrat (0.0)
|   |   |   adoption_of_the_budget_resolution = w: republican (0.0)
|   |   mx_missile = y: democrat (5.0/1.0)
|   |   mx_missile = w: republican (0.0)
physician_fee_freeze = w
|   adoption_of_the_budget_resolution = n: republican (0.0|171.0/29.0)
|   adoption_of_the_budget_resolution != n: democrat (11.0/3.0)
physician_fee_freeze = n: democrat (247.0/2.0)

Number of Leaves  :     12

Size of the tree :      18


Time taken to build model: 0.02 seconds

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

Correctly Classified Instances         415               95.4023 %
```

The result would be most accurate if we change the minNumObj value from 2 to 7 and confidenceFactor value from 0.25 to 1. The accuracy increased to 96.32%.

```
physician_fee_freeze = w
|   adoption_of_the_budget_resolution = n: republican (0.0|171.0/29.0)
|   adoption_of_the_budget_resolution != n: democrat (11.0/3.0)
physician_fee_freeze = n
|   adoption_of_the_budget_resolution = n
|   |   education_spending = y: democrat (10.0)
|   |   education_spending = n: democrat (14.0/1.0)
|   |   education_spending = w
|   |   |   crime = n: democrat (0.0|15.0)
|   |   |   crime != n
|   |   |   |   aid_to_nicaraguan_contras = y: democrat (0.0|24.0/1.0)
|   |   |   |   aid_to_nicaraguan_contras != y
|   |   |   |   |   el_salvador_aid = n: democrat (0.0|23.0/1.0)
|   |   |   |   |   el_salvador_aid != n
|   |   |   |   |   |   synfuels_corporation_cutback = y: democrat (0.0|22.0/1.0)
|   |   |   |   |   |   synfuels_corporation_cutback != y
|   |   |   |   |   |   |   mx_missile = y: democrat (0.0|21.0/1.0)
|   |   |   |   |   |   |   mx_missile != y
|   |   |   |   |   |   |   |   handicapped_infants = y: democrat (0.0|21.0/2.0)
|   |   |   |   |   |   |   |   handicapped_infants != y: republican (1.0)
|   adoption_of_the_budget_resolution = y: democrat (219.0)
|   adoption_of_the_budget_resolution = w: democrat (3.0)

Number of Leaves  :     18

Size of the tree :      30


Time taken to build model: 0.07 seconds

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

Correctly Classified Instances         419               96.3218 %
```

H. LADTree

The accuracy for the default parameters is 94.94%.

```
: 0,0
|   (1)physician_fee_freeze = y: 0.842,-0.842
|   |   (6)immigration = n: -0.34,0.34
|   |   |   (9)mx_missile = n: 0.201,-0.201
|   |   |   |   (10)superfund_right_to_sue = y: 0.236,-0.236
|   |   |   |   (10)superfund_right_to_sue != y: -0.987,0.987
|   |   |   (9)mx_missile != n: -0.847,0.847
|   |   (6)immigration != n: 0.532,-0.532
|   (1)physician_fee_freeze != y: -0.961,0.961
|   (2)adoption_of_the_budget_resolution = y: -0.529,0.529
|   |   (7)nti_satellite_test_ban = y: 0.071,-0.071
|   |   |   (8)physician_fee_freeze = y: 0.714,-0.714
|   |   |   (8)physician_fee_freeze != y: -0.512,0.512
|   |   (7)nti_satellite_test_ban != y: -0.878,0.878
|   (2)adoption_of_the_budget_resolution != y: 0.345,-0.345
|   (3)synfuels_corporation_cutback = y: -0.61,0.61
|   (3)synfuels_corporation_cutback != y: 0.217,-0.217
|   |   (4)physician_fee_freeze = n: -0.491,0.491
|   |   (4)physician_fee_freeze != n: 0.452,-0.452
|   (5)education_spending = n: -0.435,0.435
|   (5)education_spending != n: 0.191,-0.191
Legend: republican, democrat
#Tree size (total): 31
#Tree size (number of predictor nodes): 21
#Leaves (number of predictor nodes): 14
#Expanded nodes: 100
#Processed examples: 20626
#Ratio e/n: 206.26

Time taken to build model: 0.13 seconds

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

Correctly Classified Instances         413               94.9425 %
```

The result would be most accurate if we change the numOfBoostingIterations value from 10 to 20. The accuracy increased to 96.32%.

```
: 0,0
|  (1)physician_fee_freeze = y: 0.977,-0.977
|  |  (6)immigration = n: -0.34,0.34
|  |  |  (9)mx_missile = n: 0.201,-0.201
|  |  |  |  (10)superfund_right_to_sue = y: 0.236,-0.236
|  |  |  |  |  (13)export_administration_act_sa = y: 0.571,-0.571
|  |  |  |  |  (13)export_administration_act_sa != y: -0.284,0.284
|  |  |  |  (10)superfund_right_to_sue != y: -0.987,0.987
|  |  |  |  |  (15)export_administration_act_sa = w: 0.588,-0.588
|  |  |  |  |  (15)export_administration_act_sa != w: -0.613,0.613
|  |  |  (9)mx_missile != n: -0.847,0.847
|  |  (6)immigration != n: 0.532,-0.532
|  (1)physician_fee_freeze != y: -1.236,1.236
|  (2)adoption_of_the_budget_resolution = y: -0.529,0.529
|  |  (7)nti_satellite_test_ban = y: 0.071,-0.071
|  |  |  (8)physician_fee_freeze = y: 1.403,-1.403
|  |  |  (8)physician_fee_freeze != y: -1.016,1.016
|  |  (7)nti_satellite_test_ban != y: -0.878,0.878
|  |  |  (11)water_project_cost_sharing = y: -0.679,0.679
|  |  |  (11)water_project_cost_sharing != y: 0.632,-0.632
|  (2)adoption_of_the_budget_resolution != y: 0.345,-0.345
|  (3)synfuels_corporation_cutback = y: -0.61,0.61
|  (3)synfuels_corporation_cutback != y: 0.217,-0.217
|  |  (4)physician_fee_freeze = n: -0.491,0.491
|  |  |  (12)superfund_right_to_sue = n: 0.573,-0.573
|  |  |  |  (16)adoption_of_the_budget_resolution = n: 0.716,-0.716
|  |  |  |  (16)adoption_of_the_budget_resolution != n: -0.512,0.512
|  |  |  (12)superfund_right_to_sue != n: -0.605,0.605
|  |  (4)physician_fee_freeze != n: 0.452,-0.452
|  |  (14)education_spending = n: -0.183,0.183
|  |  |  (17)religious_groups_in_schools = n: 0.465,-0.465
|  |  |  |  (18)crime = y: 0.837,-0.837
|  |  |  |  (18)crime != y: -0.786,0.786
```

I. LMT

The accuracy for the default parameters is 96.78%.

```
Number of Leaves  :     1

Size of the Tree :      1
LM_1:
Class 0 :
0.31 +
[water_project_cost_sharing=y] * -0.51 +
[adoption_of_the_budget_resolution=n] * 0.38 +
[adoption_of_the_budget_resolution=y] * -0.85 +
[physician_fee_freeze=y] * 1.8  +
[physician_fee_freeze=n] * -1.58 +
[nti_satellite_test_ban=n] * -0.38 +
[nti_satellite_test_ban=y] * 0.32 +
[aid_to_nicaraguan_contras=w] * 0.44 +
[mx_missile=y] * -0.54 +
[immigration=n] * -0.59 +
[synfuels_corporation_cutback=n] * 0.71 +
[synfuels_corporation_cutback=y] * -0.8 +
[education_spending=n] * -0.58 +
[education_spending=w] * 0.35 +
[superfund_right_to_sue=w] * -0.46 +
[duty_free_exports=y] * -0.41 +
[export_administration_act_sa=y] * 0.41

Class 1 :
-0.31 +
[water_project_cost_sharing=y] * 0.51 +
[adoption_of_the_budget_resolution=n] * -0.38 +
[adoption_of_the_budget_resolution=y] * 0.85 +
[physician_fee_freeze=y] * -1.8 +
[physician_fee_freeze=n] * 1.58 +
[nti_satellite_test_ban=n] * 0.38 +
[nti_satellite_test_ban=y] * -0.32 +
```

Then, we set different value of parameters. However, the result did not turn out to be better.

J. NBTree

The accuracy for the default parameters is 94.71%. We can not set different parameters by using this approach.

```
NBTree
------------------

nti_satellite_test_ban = n
|   aid_to_nicaraguan_contras = n: NB 2
|   aid_to_nicaraguan_contras = y: NB 3
|   aid_to_nicaraguan_contras = w: NB 4
nti_satellite_test_ban = y
|   superfund_right_to_sue = y: NB 6
|   superfund_right_to_sue = w: NB 7
|   superfund_right_to_sue = n
|   |   physician_fee_freeze = y: NB 9
|   |   physician_fee_freeze = w: NB 10
|   |   physician_fee_freeze = n
|   |   |   synfuels_corporation_cutback = w: NB 12
|   |   |   synfuels_corporation_cutback = n: NB 13
|   |   |   synfuels_corporation_cutback = y: NB 14
nti_satellite_test_ban = w: NB 15

Leaf number: 2 Naive Bayes Classifier
```

K. RandomForest

The accuracy for the default parameters is 95.86%.

```
Random forest of 100 trees, each constructed while considering 5 random features.
Out of bag error: 0.0414


Time taken to build model: 0.15 seconds

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

Correctly Classified Instances         417               95.8621 %
```

Then, we set different value of parameters. However, the result did not turn out to be better.

L. RandomTree

The accuracy for the default parameters is 93.56%.

```
crime = y
|   synfuels_corporation_cutback = w
|   |   el_salvador_aid = y
|   |   |   export_administration_act_sa = y
|   |   |   |   education_spending = y : republican (2/0)
|   |   |   |   education_spending = n : democrat (1/0)
|   |   |   |   education_spending = w : republican (1/0)
|   |   |   export_administration_act_sa = w : republican (1/0)
|   |   |   export_administration_act_sa = n
|   |   |   |   education_spending = y : republican (1/0)
|   |   |   |   education_spending = n : democrat (1/0)
|   |   |   |   education_spending = w : republican (1/0)
|   |   el_salvador_aid = w : republican (1/0)
|   |   el_salvador_aid = n : democrat (1/0)
|   synfuels_corporation_cutback = n
|   |   physician_fee_freeze = y
|   |   |   export_administration_act_sa = y
|   |   |   |   nti_satellite_test_ban = n
|   |   |   |   |   water_project_cost_sharing = y
|   |   |   |   |   |   adoption_of_the_budget_resolution = n
|   |   |   |   |   |   |   superfund_right_to_sue = y : republican (21/0)
|   |   |   |   |   |   |   superfund_right_to_sue = w : democrat (1/0)
|   |   |   |   |   |   |   superfund_right_to_sue = n : republican (0/0)
|   |   |   |   |   |   adoption_of_the_budget_resolution = y : republican (1/0)
|   |   |   |   |   |   adoption_of_the_budget_resolution = w : republican (0/0)
|   |   |   |   |   water_project_cost_sharing = n : republican (18/0)
|   |   |   |   |   water_project_cost_sharing = w : republican (9/0)
|   |   |   |   nti_satellite_test_ban = y : republican (27/0)
|   |   |   |   nti_satellite_test_ban = w : republican (1/0)
|   |   |   export_administration_act_sa = w
```

The result would be most accurate if we change the seed value from 1 to 15. The accuracy increased to 94.25%.

```
aid_to_nicaraguan_contras = n
|   education_spending = y
|   |   synfuels_corporation_cutback = w : republican (3/0)
|   |   synfuels_corporation_cutback = n
|   |   |   physician_fee_freeze = y
|   |   |   |   adoption_of_the_budget_resolution = n : republican (90/0)
|   |   |   |   adoption_of_the_budget_resolution = y
|   |   |   |   |   water_project_cost_sharing = y
|   |   |   |   |   |   export_administration_act_sa = y : republican (3/0)
|   |   |   |   |   |   export_administration_act_sa = w : democrat (1/0)
|   |   |   |   |   |   export_administration_act_sa = n : republican (0/0)
|   |   |   |   |   water_project_cost_sharing = n : republican (3/0)
|   |   |   |   |   water_project_cost_sharing = w : republican (0/0)
|   |   |   |   adoption_of_the_budget_resolution = w : republican (1/0)
|   |   |   physician_fee_freeze = w : republican (0/0)
|   |   |   physician_fee_freeze = n : democrat (3/0)
|   |   synfuels_corporation_cutback = y
|   |   |   export_administration_act_sa = y
|   |   |   |   adoption_of_the_budget_resolution = n
|   |   |   |   |   superfund_right_to_sue = y
|   |   |   |   |   |   immigration = y
|   |   |   |   |   |   |   water_project_cost_sharing = y : republican (3/0)
|   |   |   |   |   |   |   water_project_cost_sharing = n : democrat (1/0)
|   |   |   |   |   |   |   water_project_cost_sharing = w : republican (0/0)
|   |   |   |   |   |   immigration = n : republican (4/0)
|   |   |   |   |   |   immigration = w : republican (0/0)
|   |   |   |   |   superfund_right_to_sue = w : republican (0/0)
|   |   |   |   |   superfund_right_to_sue = n
|   |   |   |   |   |   handicapped_infants = n : republican (1/0)
|   |   |   |   |   |   handicapped_infants = w : republican (0/0)
|   |   |   |   |   |   handicapped_infants = y : democrat (1/0)
```

M. REPTree

The accuracy for the default parameters is 94.71%.

```
REPTree
============

physician_fee_freeze = y
|   synfuels_corporation_cutback = w : republican (5/0) [2/0]
|   synfuels_corporation_cutback = n : republican (87/1) [51/2]
|   synfuels_corporation_cutback = y
|   |   water_project_cost_sharing = y
|   |   |   adoption_of_the_budget_resolution = n : republican (13/4) [4/0]
|   |   |   adoption_of_the_budget_resolution = y : democrat (5/1) [1/0]
|   |   |   adoption_of_the_budget_resolution = w : democrat (1/0) [0/0]
|   |   water_project_cost_sharing = n : republican (6/0) [1/1]
|   |   water_project_cost_sharing = w : republican (1/0) [0/0]
physician_fee_freeze = w : democrat (7/2) [4/1]
physician_fee_freeze = n : democrat (165/2) [82/0]

Size of the tree : 13

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         412               94.7126 %
```

The result would be most accurate if we change the minNum value from 2.0 to 1.0, the numFolds value from 3 to 2 and the seed value from 1 to 15 . The accuracy increased to 95.86%.

N. SimpleCart

The accuracy for the default parameters is 95.63%.

```
CART Decision Tree

physician_fee_freeze=(n)|(w)
|  adoption_of_the_budget_resolution=(y)|(n): democrat(247.0/2.0)
|  adoption_of_the_budget_resolution!=(y)|(n)
|  |  mx_missile=(n)|(y): democrat(6.0/1.0)
|  |  mx_missile!=(n)|(y): republican(2.0/0.0)
physician_fee_freeze!=(n)|(w)
|  synfuels_corporation_cutback=(y)
|  |  adoption_of_the_budget_resolution=(w)|(y)
|  |  |  nti_satellite_test_ban=(n)|(w): democrat(6.0/0.0)
|  |  |  nti_satellite_test_ban!=(n)|(w): republican(3.0/0.0)
|  |  adoption_of_the_budget_resolution!=(w)|(y)
|  |  |  el_salvador_aid=(n): democrat(2.0/0.0)
|  |  |  el_salvador_aid!=(n): republican(18.0/3.0)
|  synfuels_corporation_cutback!=(y): republican(142.0/3.0)

Number of Leaf Nodes: 8

Size of the Tree: 15

Time taken to build model: 0.21 seconds

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

Correctly Classified Instances         416               95.6322 %
```

Then, we set different value of parameters. However, the result did not turn out to be better.

2. Observation and Conclusion

For the “iris.arff” file, the LMT classifier has best accuracy which is 98.0%. So it seems that LMT classifier is the best method for iris classification. For the “house-votes-84.arff” file, the FT classifier and LMT classifier have best accuracy which is 96.78%. Therefore, the LMT classifier and FT classifier are both the best classifiers for congressmen classification. Taken together, the LMT method is a better classifier than others