

1. Bagging tried with depth 3, 5 with trees 5 and 10.

```
C:\Windows\System32\cmd.exe

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>python anikamal.py bag 3 5 mushrooms
BAGGING
=====
NUMBER OF TREES USED: 5
DEPTH: 3
Accuracy: 74.5882%
Misclassification count: 540
CONFUSION MATRIX
=====
          Predict:-   Predict:+
Actual:-    1553      540
Actual:+         0       32

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>python anikamal.py bag 3 10 mushrooms
BAGGING
=====
NUMBER OF TREES USED: 10
DEPTH: 3
Accuracy: 74.5882%
Misclassification count: 540
CONFUSION MATRIX
=====
          Predict:-   Predict:+
Actual:-    1553      540
Actual:+         0       32

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>
```

Fig 1. Result for bagging with depth 3 and tree bags 5 and 10.

```
C:\Windows\System32\cmd.exe

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>python anikamal.py bag 5 5 mushrooms
BAGGING
=====
NUMBER OF TREES USED: 5
DEPTH: 5
Accuracy: 76.6588%
Misclassification count: 496
CONFUSION MATRIX
=====
          Predict:-   Predict:+
Actual:-    1597      496
Actual:+         0       32

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>python anikamal.py bag 5 10 mushrooms
BAGGING
=====
NUMBER OF TREES USED: 10
DEPTH: 5
Accuracy: 76.6588%
Misclassification count: 496
CONFUSION MATRIX
=====
          Predict:-   Predict:+
Actual:-    1597      496
Actual:+         0       32
```

Fig 2. Result for bagging with depth 5 and tree bags 5 and 10.

2) Bagging tried with depth 1, 2 with trees 5 and 10.

```
C:\Windows\System32\cmd.exe
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
self.obj[item] = s
ADABOOST
=====
NUMBER OF TREES USED: 5
DEPTH: 1
Accuracy: 83.4353%
Misclassification count: 352
CONFUSION MATRIX
=====
          Predict:- Predict:+
Actual:-    1741      352
Actual:+         0       32

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>python anikamal.py boost 1 10 mushrooms
C:\Users\anirudhkm\Anaconda3\lib\site-packages\pandas\core\indexing.py:465: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
self.obj[item] = s
ADABOOST
=====
NUMBER OF TREES USED: 10
DEPTH: 1
Accuracy: 83.4353%
Misclassification count: 352
CONFUSION MATRIX
=====
          Predict:- Predict:+
Actual:-    1741      352
Actual:+         0       32

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>
```

Fig 3. Results for adaboost with depth 1 and number of trees being 5, 10

```
C:\Windows\System32\cmd.exe
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
self.obj[item] = s
ADABOOST
=====
NUMBER OF TREES USED: 5
DEPTH: 2
Accuracy: 83.4353%
Misclassification count: 352
CONFUSION MATRIX
=====
          Predict:- Predict:+
Actual:-    1741      352
Actual:+         0       32

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>python anikamal.py boost 2 10 mushrooms
C:\Users\anirudhkm\Anaconda3\lib\site-packages\pandas\core\indexing.py:465: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
self.obj[item] = s
ADABOOST
=====
NUMBER OF TREES USED: 10
DEPTH: 2
Accuracy: 83.4353%
Misclassification count: 352
CONFUSION MATRIX
=====
          Predict:- Predict:+
Actual:-    1741      352
Actual:+         0       32

C:\Users\anirudhkm\Box Sync\Fall_2016\AML\HW\PA\PA2>
```

Fig 4. Results for adaboost with depth 2 and number of trees being 5, 10

## 3) Confusion Matrix

**BAGGING****Depth :3; Trees: 5**

	Predicted negative	Predicted positive
Actual negative	1553	540
Actual positive	0	32

**Depth: 3; Trees: 10**

	Predicted negative	Predicted positive
Actual negative	1553	540
Actual positive	0	32

**Depth: 5; Trees: 5**

	Predicted negative	Predicted positive
Actual negative	1597	496
Actual positive	0	32

**Depth: 5; Trees: 10**

	Predicted negative	Predicted positive
Actual negative	1597	496
Actual positive	0	32

**BOOSTING****Depth 1; Trees: 5**

	Predicted negative	Predicted positive
Actual negative	1741	352
Actual positive	0	32

**Depth 1; Trees: 10**

	Predicted negative	Predicted positive
Actual negative	1741	352
Actual positive	0	32

**Depth 2; Trees: 5**

	Predicted negative	Predicted positive
Actual negative	1741	352
Actual positive	0	32

**Depth 2; Trees: 10**

	Predicted negative	Predicted positive
Actual negative	1741	352
Actual positive	0	32

#### 4) Weka results

##### Bagging Results

Iterations : 10

=== Run information ===

Scheme: weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 10 -W

weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Relation: agaricuslepiotatrain1

Instances: 8125

Attributes: 126

[list of attributes omitted]

Test mode: split 73.85% train, remainder test

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

Bagging with 10 iterations and base learner

weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Time taken to build model: 7.8 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0.1 seconds

=== Summary ===

Correctly Classified Instances	1593	74.9647 %
Incorrectly Classified Instances	532	25.0353 %
Kappa statistic	0.0812	
Mean absolute error	0.2504	
Root mean squared error	0.5004	
Relative absolute error	45.0609 %	
Root relative squared error	90.0298 %	
Total Number of Instances	2125	

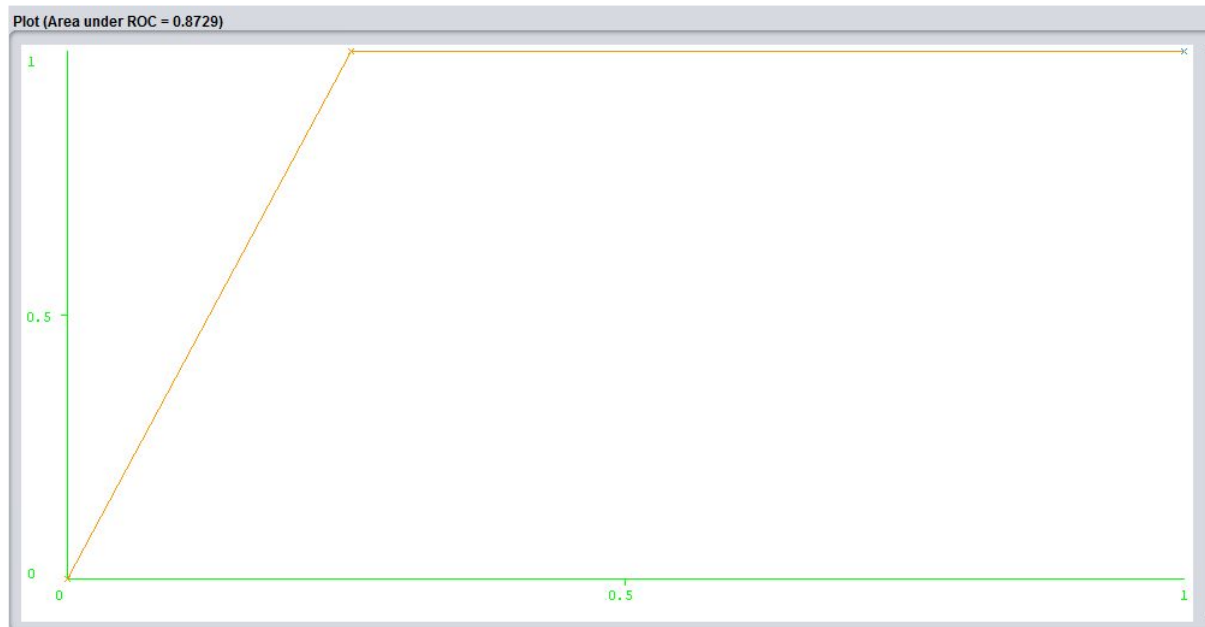
=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
1.000	0.254	0.057	1.000	0.107	0.206	0.873	0.057	Y
0.746	0.000	1.000	0.746	0.854	0.206	0.873	0.996	N
Weighted Avg.	0.750	0.004	0.986	0.750	0.843	0.206	0.873	0.982

=== Confusion Matrix ===

```
a  b  <-- classified as
32  0 |  a = Y
532 1561 |  b = N
```

**AUC**



**Iterations : 5**

=== Run information ===

Scheme: weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 5 -W

weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Relation: agaricuslepiotatrain1

Instances: 8125

Attributes: 126

[list of attributes omitted]

Test mode: split 73.85% train, remainder test

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

Bagging with 5 iterations and base learner

# AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Time taken to build model: 3.52 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	1593	74.9647 %
Incorrectly Classified Instances	532	25.0353 %
Kappa statistic	0.0812	
Mean absolute error	0.2504	
Root mean squared error	0.5004	
Relative absolute error	45.0609 %	
Root relative squared error	90.0298 %	
Total Number of Instances	2125	

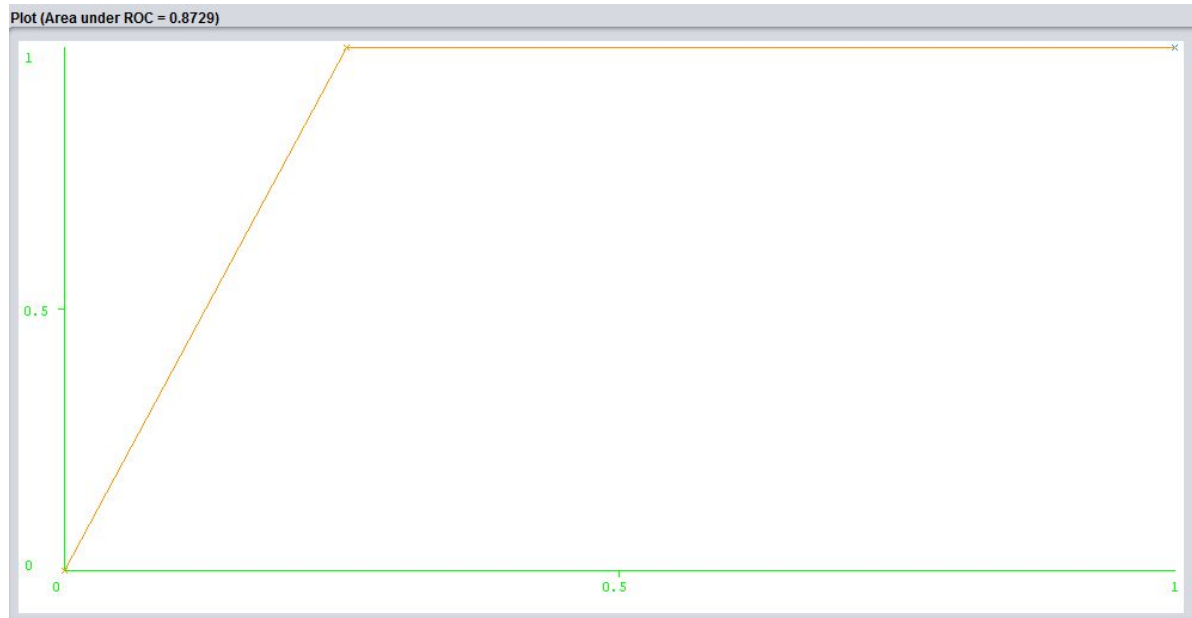
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Class								
1.000	0.254	0.057	1.000	0.107	0.206	0.873	0.057	Y
0.746	0.000	1.000	0.746	0.854	0.206	0.873	0.996	N
Weighted Avg.	0.750	0.004	0.986	0.750	0.843	0.206	0.873	0.982

=== Confusion Matrix ===

a	b	<-- classified as
32	0	a = Y
532	1561	b = N

## AUC



## Boosting

### Iterations :10

=== Run information ===

Scheme: weka.classifiers.meta.AdaBoostM1 -P 100 -S 1 -I 10 -W

weka.classifiers.trees.DecisionStump

Relation: agaricuslepiotatrain1

Instances: 8125

Attributes: 126

[list of attributes omitted]

Test mode: split 73.85% train, remainder test

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

AdaBoostM1: Base classifiers and their weights:

Decision Stump

Classifications



## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

ring-type-pendant  $\leq 0.5$  : N  
ring-type-pendant  $> 0.5$  : Y  
ring-type-pendant is missing : N

Class distributions

ring-type-pendant  $\leq 0.5$   
Y      N  
0.04618715419773876      0.9538128458022612  
ring-type-pendant  $> 0.5$   
Y      N  
0.80241935483870960.1975806451612903  
ring-type-pendant is missing  
Y      N  
0.41550769230769230.5844923076923076

Weight: 1.99

Decision Stump

Classifications

habitat-waste  $\leq 0.5$  : N  
habitat-waste  $> 0.5$  : Y  
habitat-waste is missing : N

Class distributions

habitat-waste  $\leq 0.5$   
Y      N  
0.24698177748953432      0.7530182225104657  
habitat-waste  $> 0.5$   
Y      N  
1.00000000000000377-3.7665159879907513E-14  
habitat-waste is missing  
Y      N  
0.32104914363810520.6789508563618948

Weight: 1.25

Decision Stump

## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

### Classifications

stalk-surface-above-ring-silky  $\leq 0.5$  : Y

stalk-surface-above-ring-silky  $> 0.5$  : N

stalk-surface-above-ring-silky is missing : Y

### Class distributions

stalk-surface-above-ring-silky  $\leq 0.5$

Y      N

0.66090630132624960.33909369867375033

stalk-surface-above-ring-silky  $> 0.5$

Y      N

-4.745765883145268E-14      1.00000000000000475

stalk-surface-above-ring-silky is missing

Y      N

0.563269779365576      0.436730220634424

Weight: 0.9

### Decision Stump

### Classifications

gill-spacing-close  $\leq 0.5$  : N

gill-spacing-close  $> 0.5$  : Y

gill-spacing-close is missing : N

### Class distributions

gill-spacing-close  $\leq 0.5$

Y      N

0.046720987123564844      0.9532790128764351

gill-spacing-close  $> 0.5$

Y      N

0.50999183190734390.490008168092656

gill-spacing-close is missing

Y      N

0.39611037578889813      0.6038896242111019

Weight: 0.49

### Decision Stump

## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

### Classifications

stalk-root-missing  $\leq 0.5$  : N

stalk-root-missing  $> 0.5$  : N

stalk-root-missing is missing : N

### Class distributions

stalk-root-missing  $\leq 0.5$

Y      N

0.48564178970730215      0.5143582102926978

stalk-root-missing  $> 0.5$

Y      N

0.08914201053383776      0.9108579894661623

stalk-root-missing is missing

Y      N

0.32577310380530870.6742268961946913

Weight: 0.73

### Decision Stump

### Classifications

stalk-root-missing  $\leq 0.5$  : Y

stalk-root-missing  $> 0.5$  : N

stalk-root-missing is missing : Y

### Class distributions

stalk-root-missing  $\leq 0.5$

Y      N

0.66148455618946730.3385154438105327

stalk-root-missing  $> 0.5$

Y      N

0.16843063382215528      0.8315693661778447

stalk-root-missing is missing

Y      N

0.50000000000000720.4999999999999928

Weight: 0.93

## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

Decision Stump

Classifications

gill-size-broad  $\leq 0.5$  : N

gill-size-broad  $> 0.5$  : Y

gill-size-broad is missing : N

Class distributions

gill-size-broad  $\leq 0.5$

Y      N

0.11302960855589052      0.8869703914441095

gill-size-broad  $> 0.5$

Y      N

0.57256094725394280.42743905274605715

gill-size-broad is missing

Y      N

0.40765282033491984      0.5923471796650801

Weight: 0.78

Decision Stump

Classifications

habitat-waste  $\leq 0.5$  : N

habitat-waste  $> 0.5$  : Y

habitat-waste is missing : N

Class distributions

habitat-waste  $\leq 0.5$

Y      N

0.28111316013139853      0.7188868398686015

habitat-waste  $> 0.5$

Y      N

1.00000000000000189-1.8799107911556663E-14

habitat-waste is missing

Y      N

0.33226088234711140.6677391176528886

Weight: 1.04

## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

Decision Stump

Classifications

stalk-surface-below-ring-silky  $\leq 0.5$  : Y

stalk-surface-below-ring-silky  $> 0.5$  : N

stalk-surface-below-ring-silky is missing : Y

Class distributions

stalk-surface-below-ring-silky  $\leq 0.5$

Y      N

0.62126352328467320.3787364767153269

stalk-surface-below-ring-silky  $> 0.5$

Y      N

1.1178292244425353E-15    0.9999999999999989

stalk-surface-below-ring-silky is missing

Y      N

0.54814568594415860.4518543140558415

Weight: 0.69

Decision Stump

Classifications

cap-surface-scaly  $\leq 0.5$  : N

cap-surface-scaly  $> 0.5$  : Y

cap-surface-scaly is missing : N

Class distributions

cap-surface-scaly  $\leq 0.5$

Y      N

0.30357127922197170.6964287207780283

cap-surface-scaly  $> 0.5$

Y      N

0.735468789109858    0.264531210890142

cap-surface-scaly is missing

Y      N

0.41162101308705380.5883789869129462

# AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

Weight: 0.88

Number of performed Iterations: 10

Time taken to build model: 5.62 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0.08 seconds

=== Summary ===

Correctly Classified Instances	1611	75.8118 %
Incorrectly Classified Instances	514	24.1882 %
Kappa statistic	0.0847	
Mean absolute error	0.2452	
Root mean squared error	0.4889	
Relative absolute error	44.1395 %	
Root relative squared error	87.9716 %	
Total Number of Instances	2125	

=== Detailed Accuracy By Class ===

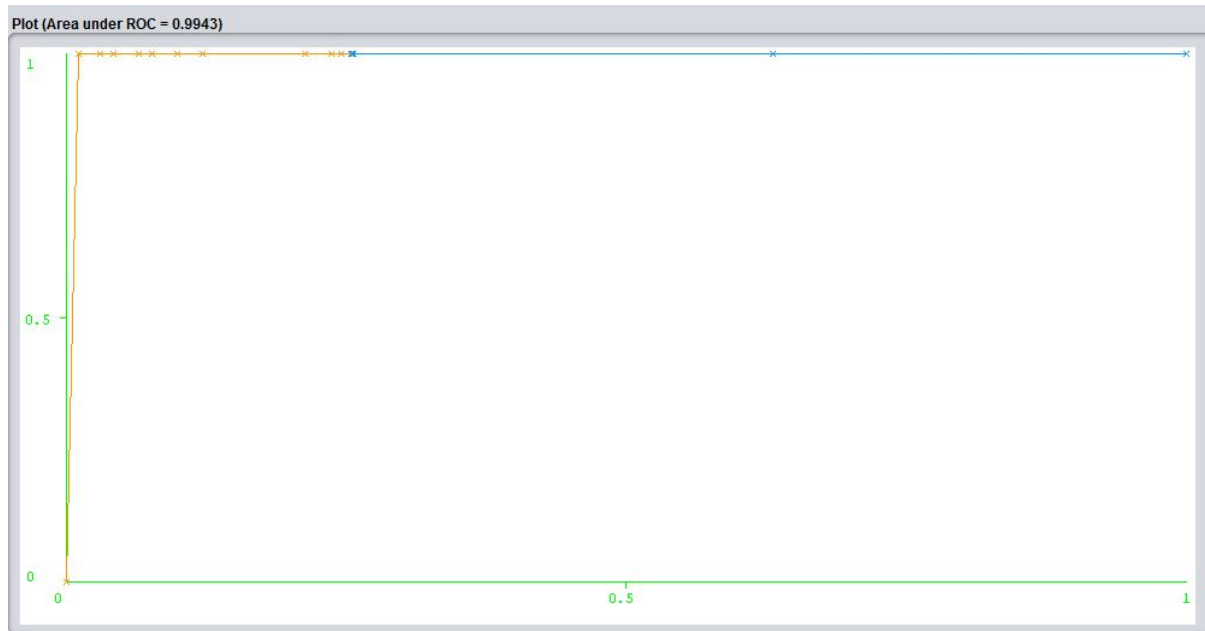
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Class								
	1.000	0.246	0.059	1.000	0.111	0.210	0.994	0.571
	0.754	0.000	1.000	0.754	0.860	0.210	0.994	1.000
Weighted Avg.	0.758	0.004	0.986	0.758	0.849	0.210	0.994	0.993

=== Confusion Matrix ===

```

a  b  <-- classified as
32  0 |  a = Y
514 1579 |  b = N
    
```

## AUC



## Iterations: 5

=== Run information ===

Scheme: weka.classifiers.meta.AdaBoostM1 -P 100 -S 1 -I 5 -W

weka.classifiers.trees.DecisionStump

Relation: agaricuslepiotatrain1

Instances: 8125

Attributes: 126

[list of attributes omitted]

Test mode: split 73.85% train, remainder test

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

AdaBoostM1: Base classifiers and their weights:

Decision Stump

Classifications

## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

ring-type-pendant  $\leq 0.5$  : N  
ring-type-pendant  $> 0.5$  : Y  
ring-type-pendant is missing : N

Class distributions

ring-type-pendant  $\leq 0.5$   
Y      N  
0.04618715419773876      0.9538128458022612  
ring-type-pendant  $> 0.5$   
Y      N  
0.80241935483870960.1975806451612903  
ring-type-pendant is missing  
Y      N  
0.41550769230769230.5844923076923076

Weight: 1.99

Decision Stump

Classifications

habitat-waste  $\leq 0.5$  : N  
habitat-waste  $> 0.5$  : Y  
habitat-waste is missing : N

Class distributions

habitat-waste  $\leq 0.5$   
Y      N  
0.24698177748953432      0.7530182225104657  
habitat-waste  $> 0.5$   
Y      N  
1.00000000000000377-3.7665159879907513E-14  
habitat-waste is missing  
Y      N  
0.32104914363810520.6789508563618948

Weight: 1.25

Decision Stump

Classifications



## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

stalk-surface-above-ring-silky  $\leq 0.5$  : Y  
stalk-surface-above-ring-silky  $> 0.5$  : N  
stalk-surface-above-ring-silky is missing : Y

Class distributions

stalk-surface-above-ring-silky  $\leq 0.5$   
Y      N  
0.6609063013262496 0.33909369867375033  
stalk-surface-above-ring-silky  $> 0.5$   
Y      N  
-4.745765883145268E-14    1.00000000000000475  
stalk-surface-above-ring-silky is missing  
Y      N  
0.563269779365576    0.436730220634424

Weight: 0.9

Decision Stump

Classifications

gill-spacing-close  $\leq 0.5$  : N  
gill-spacing-close  $> 0.5$  : Y  
gill-spacing-close is missing : N

Class distributions

gill-spacing-close  $\leq 0.5$   
Y      N  
0.046720987123564844    0.9532790128764351  
gill-spacing-close  $> 0.5$   
Y      N  
0.50999183190734390.490008168092656  
gill-spacing-close is missing  
Y      N  
0.39611037578889813    0.6038896242111019

Weight: 0.49

Decision Stump

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
-------	---------	---------	-----------	--------	-----------	-----	----------	----------

## AML - PROGRAMMING ASSIGNMENT - II

ANIRUDH K MURALIDHAR

[anikamal@iu.edu](mailto:anikamal@iu.edu)

ARUN RAM SANKARANARAYANAN

[arunsank@iu.edu](mailto:arunsank@iu.edu)

	1.000	0.237	0.061	1.000	0.114	0.215	0.966	0.182	Y
	0.763	0.000	1.000	0.763	0.866	0.215	0.966	0.999	N
Weighted Avg.	0.767	0.004	0.986	0.767	0.854	0.215	0.966	0.987	

=== Confusion Matrix ===

```
a  b  <-- classified as
32  0 |  a = Y
496 1597 |  b = N
```

### AUC

