

AI-ASSIGNMENT#2

TASK#1:

Code for the ID3 algorithm using weka api:

```
package task_1;

import java.io.BufferedReader;
import java.io.FileReader;
import java.util.Random;

//import weka.classifiers.Classifier;
import weka.classifiers.Evaluation;
import weka.classifiers.trees.Id3;
import weka.core.Instances;
//import weka.core.SerializationHelper;

//import java.lang.Object;
import weka.filters.Filter;
//import weka.filters.unsupervised.attribute.PotentialClassIgnorer;
import weka.filters.unsupervised.attribute.ReplaceMissingValues;

public class Id3_classifier
{
    public static void main(String[] args) throws Exception{
        ReplaceMissingValues m_MissingValuesFilter = null;
        BufferedReader breader=null;
```

```

        breader=new BufferedReader(new
FileReader("./train.arff"));

        Instances train=new Instances(breader);

        train.setClassIndex(train.numAttributes()-1);
        breader.close();

        //replacing missing values
        m_MissingValuesFilter = new ReplaceMissingValues();
        m_MissingValuesFilter.setInputFormat(train);
        train= Filter.useFilter(train, m_MissingValuesFilter);

        //running id3
        Id3 nb=new Id3();
        nb.buildClassifier(train);

        weka.core.SerializationHelper.write("./Id3.model", nb);

        Evaluation eval=new Evaluation(train);
        eval.crossValidateModel(nb,train, 10, new Random(1));

        System.out.println(eval.toSummaryString("*****",true));

        System.out.println(eval.fMeasure(1)+"
"+eval.precision(1)+" "+eval.recall(1));

    }
}

```

Report for training set:

=== Run information ===

Scheme: weka.classifiers.trees.Id3

Relation: mushroom-weka.filters.unsupervised.instance.Resample-S1-Z80.0-no-replacement-weka.filters.unsupervised.attribute.ReplaceMissingValues

Instances: 6499

Attributes: 23

cap-shape

cap-surface

cap-color

bruises?

odor

gill-attachment

gill-spacing

gill-size

gill-color

stalk-shape

stalk-root

stalk-surface-above-ring

stalk-surface-below-ring

stalk-color-above-ring

stalk-color-below-ring

veil-type

veil-color

ring-number

ring-type

spore-print-color

population

habitat

class

Test mode: 10-fold cross-validation

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

Id3

odor = a: e

odor = c: p

odor = f: p

odor = l: e

odor = m: p

odor = n

```

| spore-print-color = b: e
| spore-print-color = h: e
| spore-print-color = k: e
| spore-print-color = n: e
| spore-print-color = o: e
| spore-print-color = r: p
| spore-print-color = u: null
| spore-print-color = w
| | habitat = d
| | | gill-size = b: e
| | | gill-size = n: p
| | habitat = g: e
| | habitat = l
| | | cap-color = b: null
| | | cap-color = c: e
| | | cap-color = e: null
| | | cap-color = g: null
| | | cap-color = n: e
| | | cap-color = p: null
| | | cap-color = r: null
| | | cap-color = u: null
| | | cap-color = w: p
| | | cap-color = y: p
| | habitat = m: null
| | habitat = p: e
| | habitat = u: null
| | habitat = w: e
| spore-print-color = y: e
odor = p: p
odor = s: p
odor = y: p

```

Time taken to build model: 0.18 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	6498	99.9846 %
Incorrectly Classified Instances	1	0.0154 %
Kappa statistic	0.9997	
Mean absolute error	0.0002	
Root mean squared error	0.0124	
Relative absolute error	0.0308 %	
Root relative squared error	2.4819 %	
Total Number of Instances	6499	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	e
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	p
Weighted Avg.		1.000	0.000	1.000	1.000	1.000	1.000	1.000

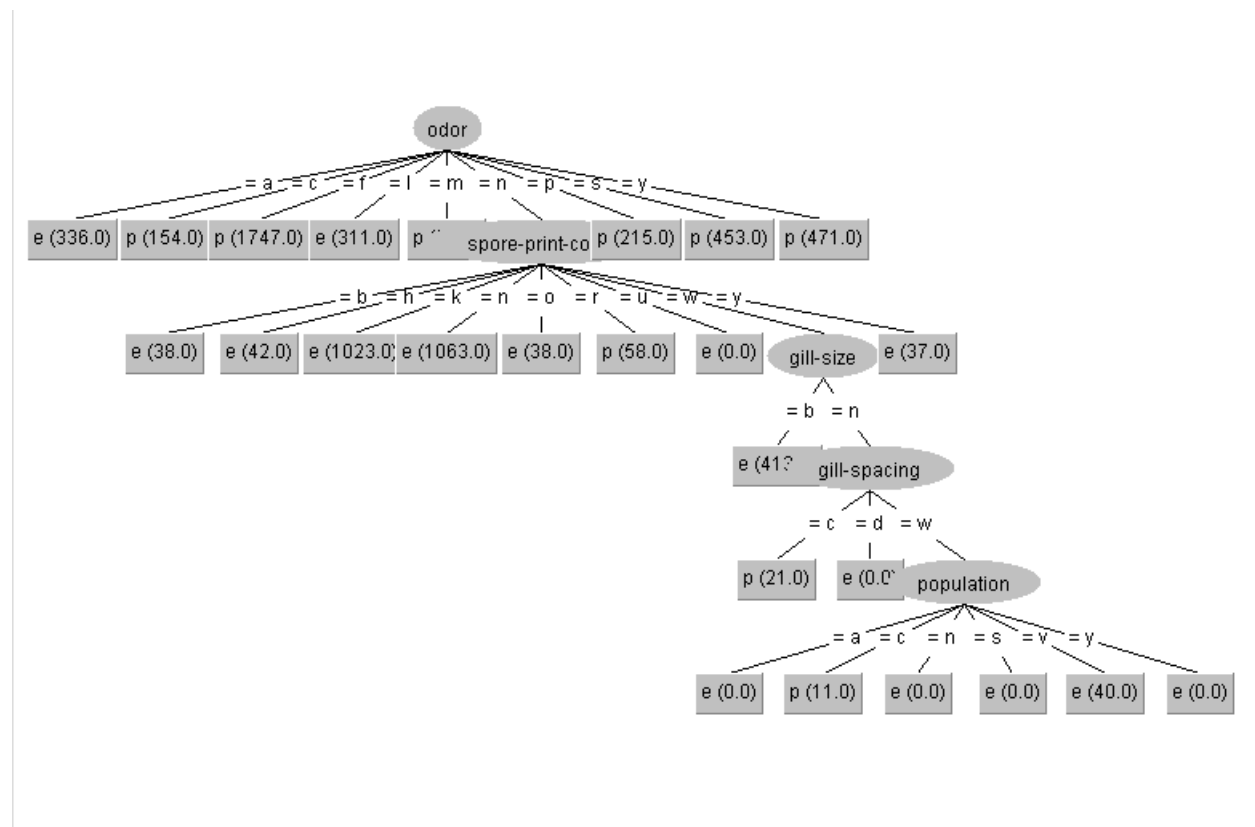
=== Confusion Matrix ===

```

a  b  <-- classified as
3340  1 |  a = e
0 3158 |  b = p

```

Decision-tree:



Code for test set accuracy report:

```

package task_1;
import java.io.BufferedReader;
import java.io.FileReader;
//import java.util.Random;

import weka.classifiers.Classifier;
import weka.classifiers.Evaluation;

```

```

import weka.core.Instances;
//import weka.core.SerializationHelper;
import weka.filters.Filter;
import weka.filters.unsupervised.attribute.ReplaceMissingValues;

//import java.lang.Object;
//import weka.filters.Filter;
//import weka.filters.unsupervised.attribute.PotentialClassIgnorer;
//import weka.filters.unsupervised.attribute.ReplaceMissingValues;

public class report_test_accuracy {
    public static void main(String args[]){
        ReplaceMissingValues m_MissingValuesFilter = null;
        try{
            Instances test = null;
            BufferedReader treader=null;
            treader=new BufferedReader(new FileReader("./test.arff"));
            test=new Instances(treader);

            test.setClassIndex(test.numAttributes()-1);

            m_MissingValuesFilter = new ReplaceMissingValues();
            m_MissingValuesFilter.setInputFormat(test);
            test= Filter.useFilter(test, m_MissingValuesFilter);

            Classifier cl1 = (Classifier) weka.core.SerializationHelper.read("./ld3.model");

            Evaluation evaluation = new Evaluation(test);
            evaluation.evaluateModel(cl1, test);
            System.out.println("Results:" + evaluation.toSummaryString());

        }catch(Exception e){
            System.out.println(e);
        }
    }
}

```

Report for the supplied test set accrcy:

=== Run information ===

Scheme: weka.classifiers.trees.Id3
 Relation: mushroom-weka.filters.unsupervised.instance.Resample-S1-Z80.0-no-replacement-
 weka.filters.unsupervised.attribute.ReplaceMissingValues
 Instances: 6499

Attributes: 23

- cap-shape
- cap-surface
- cap-color
- bruises?
- odor
- gill-attachment
- gill-spacing
- gill-size
- gill-color
- stalk-shape
- stalk-root
- stalk-surface-above-ring
- stalk-surface-below-ring
- stalk-color-above-ring
- stalk-color-below-ring
- veil-type
- veil-color
- ring-number
- ring-type
- spore-print-color
- population
- habitat
- class

Test mode: 10-fold cross-validation

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

Id3

- odor = a: e
- odor = c: p
- odor = f: p
- odor = l: e
- odor = m: p
- odor = n
 - | spore-print-color = b: e
 - | spore-print-color = h: e
 - | spore-print-color = k: e
 - | spore-print-color = n: e
 - | spore-print-color = o: e
 - | spore-print-color = r: p
 - | spore-print-color = u: null
 - | spore-print-color = w
 - | | habitat = d
 - | | | gill-size = b: e
 - | | | gill-size = n: p

```

| | habitat = g: e
| | habitat = l
| | | cap-color = b: null
| | | cap-color = c: e
| | | cap-color = e: null
| | | cap-color = g: null
| | | cap-color = n: e
| | | cap-color = p: null
| | | cap-color = r: null
| | | cap-color = u: null
| | | cap-color = w: p
| | | cap-color = y: p
| | habitat = m: null
| | habitat = p: e
| | habitat = u: null
| | habitat = w: e
| spore-print-color = y: e
odor = p: p
odor = s: p
odor = y: p

```

Time taken to build model: 0.18 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	6498	99.9846 %
Incorrectly Classified Instances	1	0.0154 %
Kappa statistic	0.9997	
Mean absolute error	0.0002	
Root mean squared error	0.0124	
Relative absolute error	0.0308 %	
Root relative squared error	2.4819 %	
Total Number of Instances	6499	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	e
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	p
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

```

a  b  <-- classified as
3340  1 | a = e
0 3158 | b = p

```


=== Re-evaluation on test set ===

User supplied test set

Relation: mushroom-weka.filters.unsupervised.instance.Resample-S1-Z80.0-no-replacement-V

Instances: unknown (yet). Reading incrementally

Attributes: 23

=== Summary ===

Correctly Classified Instances	1110	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Total Number of Instances	1110		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	e
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	p
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	<-- classified as
713	0	a = e
0	397	b = p

TASK#2:

Code for the naïve-bayes classifier:

```
package task_2;
import java.io.BufferedReader;
import java.io.FileReader;
import java.util.Random;

//import weka.classifiers.Classifier;
import weka.classifiers.Evaluation;
import weka.classifiers.bayes.NaiveBayes;
import weka.core.Instances;
//import weka.core.SerializationHelper;

public class NB_classifier
{
    public static void main(String[] args) throws Exception{
        BufferedReader breader=null;
        breader=new BufferedReader(new FileReader("./train.arff"));
        Instances train=new Instances(breader);

        train.setClassIndex(train.numAttributes()-1);
        breader.close();
        NaiveBayes nb=new NaiveBayes();
        nb.buildClassifier(train);
        weka.core.SerializationHelper.write("./nb.model", nb);
        Evaluation eval=new Evaluation(train);
        eval.crossValidateModel(nb,train, 10, new Random(1));
        System.out.println(eval.toSummaryString("*****",true));
        System.out.println(eval.fMeasure(1)+" "+eval.precision(1)+" "+eval.recall(1));

    }
}
```

Report for the naïve bayes classification:

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes
Relation: mushroom-weka.filters.unsupervised.instance.Resample-S1-Z80.0-no-replacement-weka.filters.unsupervised.attribute.ReplaceMissingValues
Instances: 6499
Attributes: 23

cap-shape
 cap-surface
 cap-color
 bruises?
 odor
 gill-attachment
 gill-spacing
 gill-size
 gill-color
 stalk-shape
 stalk-root
 stalk-surface-above-ring
 stalk-surface-below-ring
 stalk-color-above-ring
 stalk-color-below-ring
 veil-type
 veil-color
 ring-number
 ring-type
 spore-print-color
 population
 habitat
 class

Test mode: 10-fold cross-validation

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

Naive Bayes Classifier

Attribute	Class	
	e	p
	(0.51)	(0.49)
=====		
cap-shape		
b	320.0	38.0
c	1.0	5.0
f	1263.0	1250.0
k	179.0	463.0
s	27.0	1.0
x	1557.0	1407.0
[total]	3347.0	3164.0
cap-surface		
f	1261.0	605.0
g	1.0	4.0
s	891.0	1144.0
y	1192.0	1409.0
[total]	3345.0	3162.0

cap-color		
b	39.0	91.0
c	25.0	11.0
e	490.0	697.0
g	812.0	662.0
n	1011.0	812.0
p	41.0	70.0
r	13.0	1.0
u	16.0	1.0
w	580.0	272.0
y	324.0	551.0
[total]	3351.0	3168.0

bruises?		
f	1151.0	2647.0
t	2192.0	513.0
[total]	3343.0	3160.0

odor		
a	337.0	1.0
c	1.0	155.0
f	1.0	1748.0
l	312.0	1.0
m	1.0	29.0
n	2695.0	91.0
p	1.0	216.0
s	1.0	454.0
y	1.0	472.0
[total]	3350.0	3167.0

gill-attachment		
a	147.0	16.0
d	1.0	1.0
f	3196.0	3144.0
n	1.0	1.0
[total]	3345.0	3162.0

gill-spacing		
c	2390.0	3072.0
d	1.0	1.0
w	953.0	88.0
[total]	3344.0	3161.0

gill-size		
b	3107.0	1380.0
n	236.0	1780.0
[total]	3343.0	3160.0

gill-color

b	1.0	1379.0
e	73.0	1.0
g	197.0	417.0
h	162.0	431.0
k	268.0	53.0
n	744.0	99.0
o	53.0	1.0
p	703.0	515.0
r	1.0	23.0
u	347.0	39.0
w	756.0	194.0
y	48.0	18.0
[total]	3353.0	3170.0

stalk-shape

e	1288.0	1549.0
t	2055.0	1611.0
[total]	3343.0	3160.0

stalk-root

b	2091.0	2912.0
c	412.0	33.0
e	682.0	216.0
r	160.0	1.0
u	1.0	1.0
z	1.0	1.0
[total]	3347.0	3164.0

stalk-surface-above-ring

f	317.0	116.0
k	115.0	1801.0
s	2901.0	1240.0
y	12.0	5.0
[total]	3345.0	3162.0

stalk-surface-below-ring

f	366.0	117.0
k	107.0	1745.0
s	2701.0	1246.0
y	171.0	54.0
[total]	3345.0	3162.0

stalk-color-above-ring

b	1.0	358.0
c	1.0	29.0
e	70.0	1.0

g	465.0	1.0
n	12.0	353.0
o	147.0	1.0
p	447.0	1040.0
w	2206.0	1379.0
y	1.0	5.0
[total]	3350.0	3167.0

stalk-color-below-ring

b	1.0	360.0
c	1.0	29.0
e	75.0	1.0
g	455.0	1.0
n	52.0	366.0
o	147.0	1.0
p	442.0	1036.0
w	2176.0	1357.0
y	1.0	16.0
[total]	3350.0	3167.0

veil-type

p	3342.0	3159.0
u	1.0	1.0
[total]	3343.0	3160.0

veil-color

n	76.0	1.0
o	72.0	1.0
w	3196.0	3155.0
y	1.0	5.0
[total]	3345.0	3162.0

ring-number

n	1.0	29.0
o	2929.0	3073.0
t	414.0	59.0
[total]	3344.0	3161.0

ring-type

c	1.0	1.0
e	794.0	1404.0
f	43.0	1.0
l	1.0	1062.0
n	1.0	29.0
p	2507.0	667.0
s	1.0	1.0
z	1.0	1.0
[total]	3349.0	3166.0

spore-print-color

b	39.0	1.0
h	43.0	1294.0
k	1312.0	182.0
n	1385.0	189.0
o	39.0	1.0
r	1.0	59.0
u	39.0	1.0
w	454.0	1439.0
y	38.0	1.0
[total]	3350.0	3167.0

population

a	300.0	1.0
c	222.0	40.0
n	324.0	1.0
s	706.0	306.0
v	965.0	2296.0
y	830.0	520.0
[total]	3347.0	3164.0

habitat

d	1496.0	995.0
g	1120.0	605.0
l	187.0	482.0
m	206.0	31.0
p	113.0	827.0
u	77.0	224.0
w	149.0	1.0
[total]	3348.0	3165.0

Time taken to build model: 0.11 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	6192	95.2762 %
Incorrectly Classified Instances	307	4.7238 %
Kappa statistic	0.9052	
Mean absolute error	0.0471	
Root mean squared error	0.1895	
Relative absolute error	9.4214 %	
Root relative squared error	37.9104 %	
Total Number of Instances	6499	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.994	0.091	0.921	0.994	0.956	0.908	0.998	0.998	e
	0.909	0.006	0.993	0.909	0.949	0.908	0.998	0.998	p
Weighted Avg.	0.953	0.050	0.956	0.953	0.953	0.908	0.998	0.998	

=== Confusion Matrix ===

```
a  b  <-- classified as
3320 21 | a = e
286 2872 | b = p
```

Code for the supplied test set accuracy report:

```
package task_2;
import java.io.BufferedReader;
import java.io.FileReader;
//import java.util.Random;

import weka.classifiers.Classifier;
import weka.classifiers.Evaluation;
import weka.core.Instances;
//import weka.core.SerializationHelper;

public class report_accuracy {
    public static void main(String args[]){
        try{
            Instances test = null;
            BufferedReader treader=null;
            treader=new BufferedReader(new FileReader("./test.arff"));
            test=new Instances(treader);

            test.setClassIndex(test.numAttributes()-1);

            Classifier cl1 = (Classifier) weka.core.SerializationHelper.read("./nb.model");

            Evaluation evaluation = new Evaluation(test);
            evaluation.evaluateModel(cl1, test);
            System.out.println("Results:" + evaluation.toSummaryString());

        }catch(Exception e){
            System.out.println(e);
        }
    }
}
```


Accuracy report for the supplied test:

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: mushroom-weka.filters.unsupervised.instance.Resample-S1-Z80.0-no-replacement-weka.filters.unsupervised.attribute.ReplaceMissingValues

Instances: 6499

Attributes: 23

cap-shape
cap-surface
cap-color
bruises?
odor
gill-attachment
gill-spacing
gill-size
gill-color
stalk-shape
stalk-root
stalk-surface-above-ring
stalk-surface-below-ring
stalk-color-above-ring
stalk-color-below-ring
veil-type
veil-color
ring-number
ring-type
spore-print-color
population
habitat
class

Test mode: 10-fold cross-validation

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

Naive Bayes Classifier

Attribute	Class	e	p
		(0.51)	(0.49)
=====			
cap-shape			
b	320.0	38.0	
c	1.0	5.0	
f	1263.0	1250.0	
k	179.0	463.0	
s	27.0	1.0	

x	1557.0	1407.0
[total]	3347.0	3164.0

cap-surface

f	1261.0	605.0
g	1.0	4.0
s	891.0	1144.0
y	1192.0	1409.0
[total]	3345.0	3162.0

cap-color

b	39.0	91.0
c	25.0	11.0
e	490.0	697.0
g	812.0	662.0
n	1011.0	812.0
p	41.0	70.0
r	13.0	1.0
u	16.0	1.0
w	580.0	272.0
y	324.0	551.0
[total]	3351.0	3168.0

bruises?

f	1151.0	2647.0
t	2192.0	513.0
[total]	3343.0	3160.0

odor

a	337.0	1.0
c	1.0	155.0
f	1.0	1748.0
l	312.0	1.0
m	1.0	29.0
n	2695.0	91.0
p	1.0	216.0
s	1.0	454.0
y	1.0	472.0
[total]	3350.0	3167.0

gill-attachment

a	147.0	16.0
d	1.0	1.0
f	3196.0	3144.0
n	1.0	1.0
[total]	3345.0	3162.0

gill-spacing

c	2390.0	3072.0
d	1.0	1.0
w	953.0	88.0
[total]	3344.0	3161.0

gill-size		
b	3107.0	1380.0
n	236.0	1780.0
[total]	3343.0	3160.0

gill-color		
b	1.0	1379.0
e	73.0	1.0
g	197.0	417.0
h	162.0	431.0
k	268.0	53.0
n	744.0	99.0
o	53.0	1.0
p	703.0	515.0
r	1.0	23.0
u	347.0	39.0
w	756.0	194.0
y	48.0	18.0
[total]	3353.0	3170.0

stalk-shape		
e	1288.0	1549.0
t	2055.0	1611.0
[total]	3343.0	3160.0

stalk-root		
b	2091.0	2912.0
c	412.0	33.0
e	682.0	216.0
r	160.0	1.0
u	1.0	1.0
z	1.0	1.0
[total]	3347.0	3164.0

stalk-surface-above-ring		
f	317.0	116.0
k	115.0	1801.0
s	2901.0	1240.0
y	12.0	5.0
[total]	3345.0	3162.0

stalk-surface-below-ring		
f	366.0	117.0

k	107.0	1745.0
s	2701.0	1246.0
y	171.0	54.0
[total]	3345.0	3162.0

stalk-color-above-ring

b	1.0	358.0
c	1.0	29.0
e	70.0	1.0
g	465.0	1.0
n	12.0	353.0
o	147.0	1.0
p	447.0	1040.0
w	2206.0	1379.0
y	1.0	5.0
[total]	3350.0	3167.0

stalk-color-below-ring

b	1.0	360.0
c	1.0	29.0
e	75.0	1.0
g	455.0	1.0
n	52.0	366.0
o	147.0	1.0
p	442.0	1036.0
w	2176.0	1357.0
y	1.0	16.0
[total]	3350.0	3167.0

veil-type

p	3342.0	3159.0
u	1.0	1.0
[total]	3343.0	3160.0

veil-color

n	76.0	1.0
o	72.0	1.0
w	3196.0	3155.0
y	1.0	5.0
[total]	3345.0	3162.0

ring-number

n	1.0	29.0
o	2929.0	3073.0
t	414.0	59.0
[total]	3344.0	3161.0

ring-type

c	1.0	1.0
e	794.0	1404.0
f	43.0	1.0
l	1.0	1062.0
n	1.0	29.0
p	2507.0	667.0
s	1.0	1.0
z	1.0	1.0
[total]	3349.0	3166.0

spore-print-color

b	39.0	1.0
h	43.0	1294.0
k	1312.0	182.0
n	1385.0	189.0
o	39.0	1.0
r	1.0	59.0
u	39.0	1.0
w	454.0	1439.0
y	38.0	1.0
[total]	3350.0	3167.0

population

a	300.0	1.0
c	222.0	40.0
n	324.0	1.0
s	706.0	306.0
v	965.0	2296.0
y	830.0	520.0
[total]	3347.0	3164.0

habitat

d	1496.0	995.0
g	1120.0	605.0
l	187.0	482.0
m	206.0	31.0
p	113.0	827.0
u	77.0	224.0
w	149.0	1.0
[total]	3348.0	3165.0

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	6192	95.2762 %
Incorrectly Classified Instances	307	4.7238 %
Kappa statistic	0.9052	
Mean absolute error	0.0471	
Root mean squared error	0.1895	
Relative absolute error	9.4214 %	
Root relative squared error	37.9104 %	
Total Number of Instances	6499	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.994	0.091	0.921	0.994	0.956	0.908	0.998	0.998	e
	0.909	0.006	0.993	0.909	0.949	0.908	0.998	0.998	p
Weighted Avg.	0.953	0.050	0.956	0.953	0.953	0.908	0.998	0.998	

=== Confusion Matrix ===

```

a   b  <-- classified as
3320 21 |  a = e
286 2872 |  b = p

```

=== Re-evaluation on test set ===

User supplied test set

Relation: mushroom-weka.filters.unsupervised.instance.Resample-S1-Z80.0-no-replacement-V

Instances: unknown (yet). Reading incrementally

Attributes: 23

=== Summary ===

Correctly Classified Instances	1563	96.1846 %
Incorrectly Classified Instances	62	3.8154 %
Kappa statistic	0.923	
Mean absolute error	0.0387	
Root mean squared error	0.1701	
Total Number of Instances	1625	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.995	0.077	0.937	0.995	0.965	0.925	0.999	0.999	e
	0.923	0.005	0.994	0.923	0.958	0.925	0.999	0.998	p
Weighted Avg.	0.962	0.043	0.964	0.962	0.962	0.925	0.999	0.999	

=== Confusion Matrix ===

```
a b <-- classified as  
863 4 | a = e  
58 700 | b = p
```