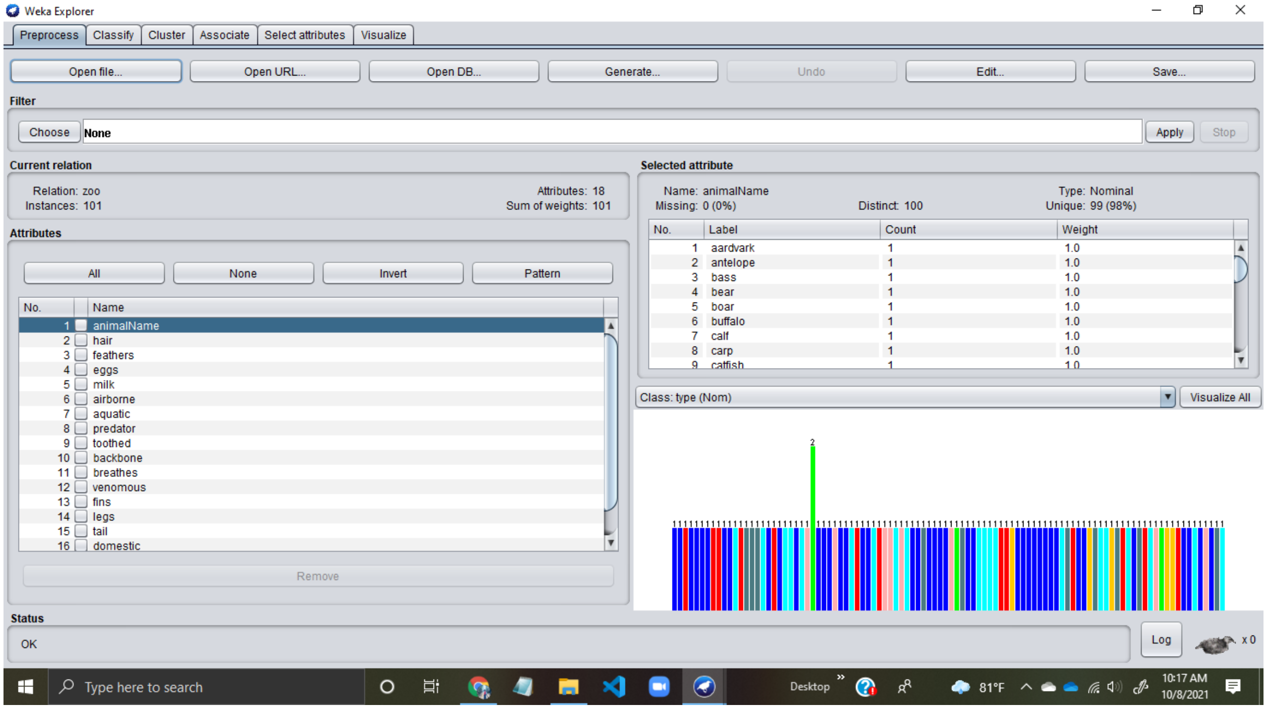
CO328 - Machine Learning
Weka Lab - Part 1
E/16/319 Rathnayake R.P.V.N

01. Load zoo data. Observe the attributes and their values



Filter: None

Current relation

Relation: zoo
Instances: 101
Attributes: 18
Sum of weights: 101

Attributes

No.	Name
1	animalName
2	hair
3	feathers
4	eggs
5	milk
6	airborne
7	aquatic
8	predator
9	toothed
10	backbone
11	breathes
12	venomous
13	fins
14	legs
15	tail
16	domestic

Selected attribute

Name: animalName
Missing: 0 (0%)
Distinct: 100
Type: Nominal
Unique: 99 (98%)

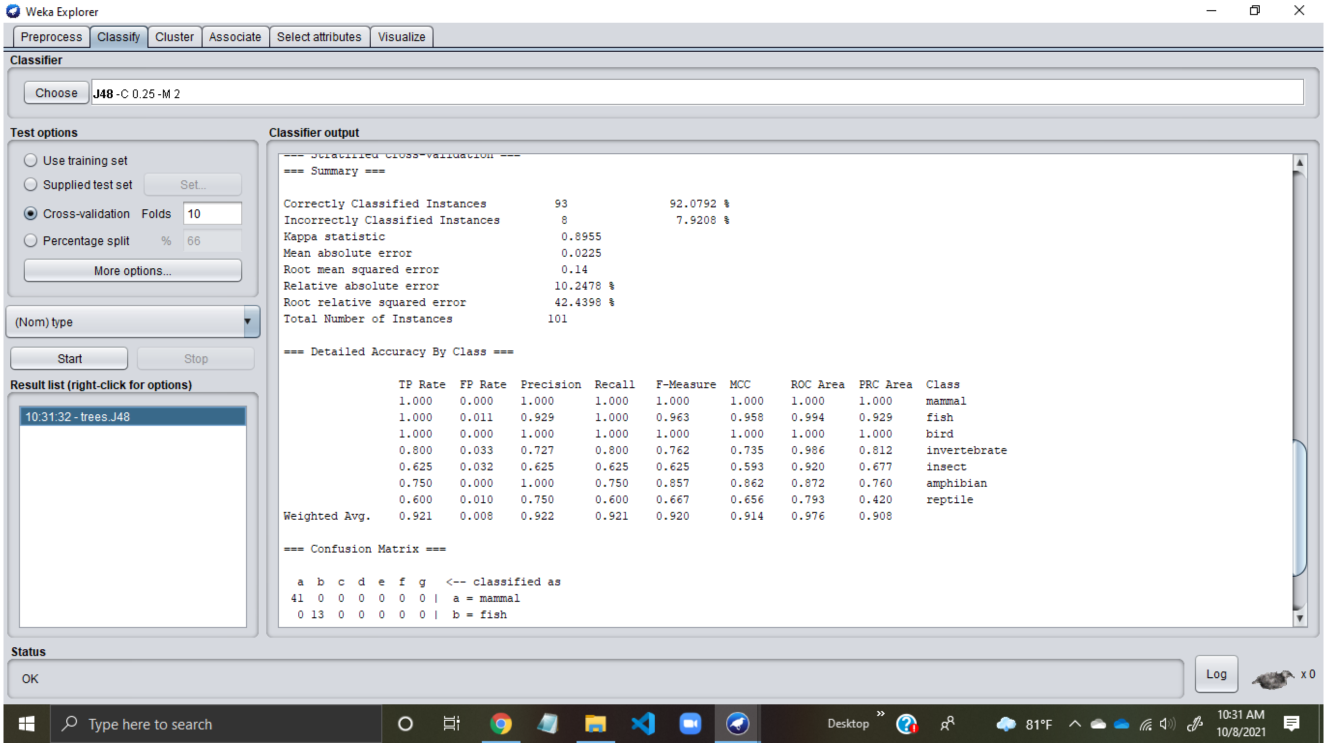
No.	Label	Count	Weight
1	aardvark	1	1.0
2	antelope	1	1.0
3	bass	1	1.0
4	bear	1	1.0
5	boar	1	1.0
6	buffalo	1	1.0
7	calf	1	1.0
8	carp	1	1.0
9	catfish	1	1.0

Class: type (Nom)

Status
OK

10:17 AM
10/8/2021

02. Build C4.5 decision tree



Classifier

J48 -C 0.25 -M 2

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) type

Result list (right-click for options)

10:31:32 - trees.J48

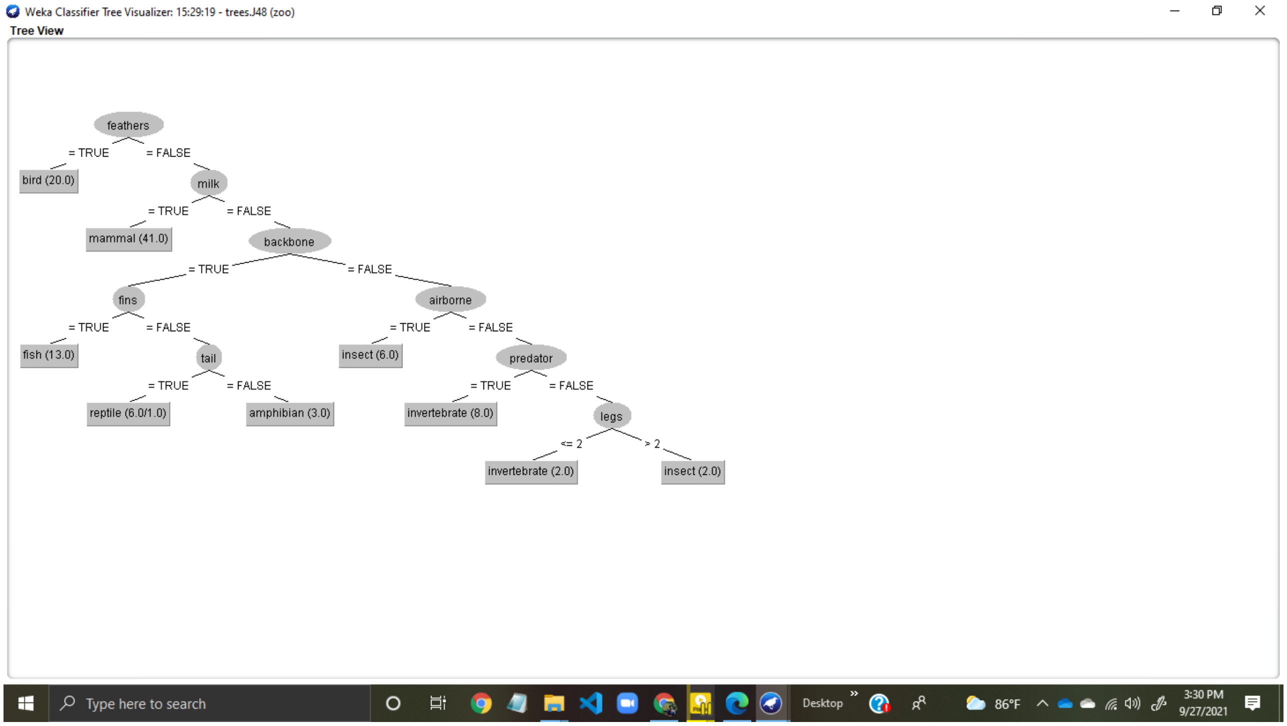
Classifier output

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

Correctly Classified Instances          93               92.0792 %
Incorrectly Classified Instances         8                7.9208 %
Kappa statistic                          0.8955
Mean absolute error                      0.0225
Root mean squared error                  0.14
Relative absolute error                 10.2478 %
Root relative squared error             42.4398 %
Total Number of Instances              101

=== 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     mammal
                 1.000    0.011    0.929      1.000    0.963      0.958    0.994     0.929     fish
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
                 0.800    0.033    0.727      0.800    0.762      0.735    0.986     0.812     invertebrate
                 0.625    0.032    0.625      0.625    0.625      0.593    0.920     0.677     insect
                 0.750    0.000    1.000      0.750    0.857      0.862    0.872     0.760     amphibian
                 0.600    0.010    0.750      0.600    0.667      0.656    0.793     0.420     reptile
Weighted Avg.    0.921    0.008    0.922      0.921    0.920      0.914    0.976     0.908

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
  0 13  0  0  0  0  0 |  b = fish
```

Status
OK

10:31 AM
10/8/2021

03. Visualize the output



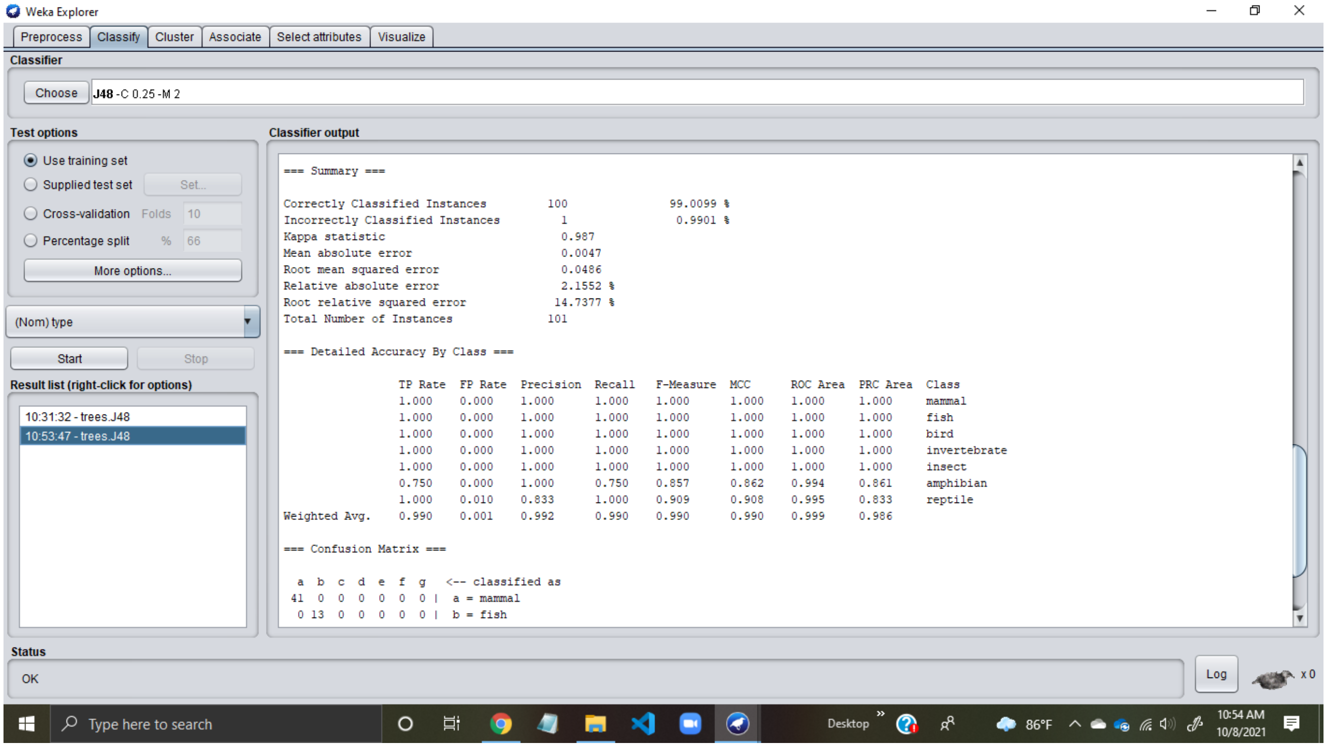
Classification accuracy : 92.079%

```
=== 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     mammal
                 1.000    0.011    0.929      1.000    0.963      0.958    0.994     0.929     fish
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
                 0.800    0.033    0.727      0.800    0.762      0.735    0.986     0.812     invertebrate
                 0.625    0.032    0.625      0.625    0.625      0.593    0.920     0.677     insect
                 0.750    0.000    1.000      0.750    0.857      0.862    0.872     0.760     amphibian
                 0.600    0.010    0.750      0.600    0.667      0.656    0.793     0.420     reptile
Weighted Avg.    0.921    0.008    0.922      0.921    0.920      0.914    0.976     0.908

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
  0 13  0  0  0  0  0 |  b = fish
  0  0 20  0  0  0  0 |  c = bird
  0  0  0  8  2  0  0 |  d = invertebrate
  0  0  0  3  5  0  0 |  e = insect
  0  0  0  0  0  3  1 |  f = amphibian
  0  1  0  0  1  0  3 |  g = reptile
```

By looking at the confusion matrix we can see that mammals, fish, and birds are perfectly classified. But 2 invertebrates classified as insects, 3 insects classified as invertebrates. It means the identifying insects and invertebrates are not good in this model. And also 2 reptiles classified as 2 different typ so the model is unable to identify the reptiles also.

4. Evaluate the C4.5 using

a. The training set



Classifier: J48 -C 0.25 -M 2

Test options:
● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) type

Result list (right-click for options)
10:31:32 - trees.J48
10:53:47 - trees.J48

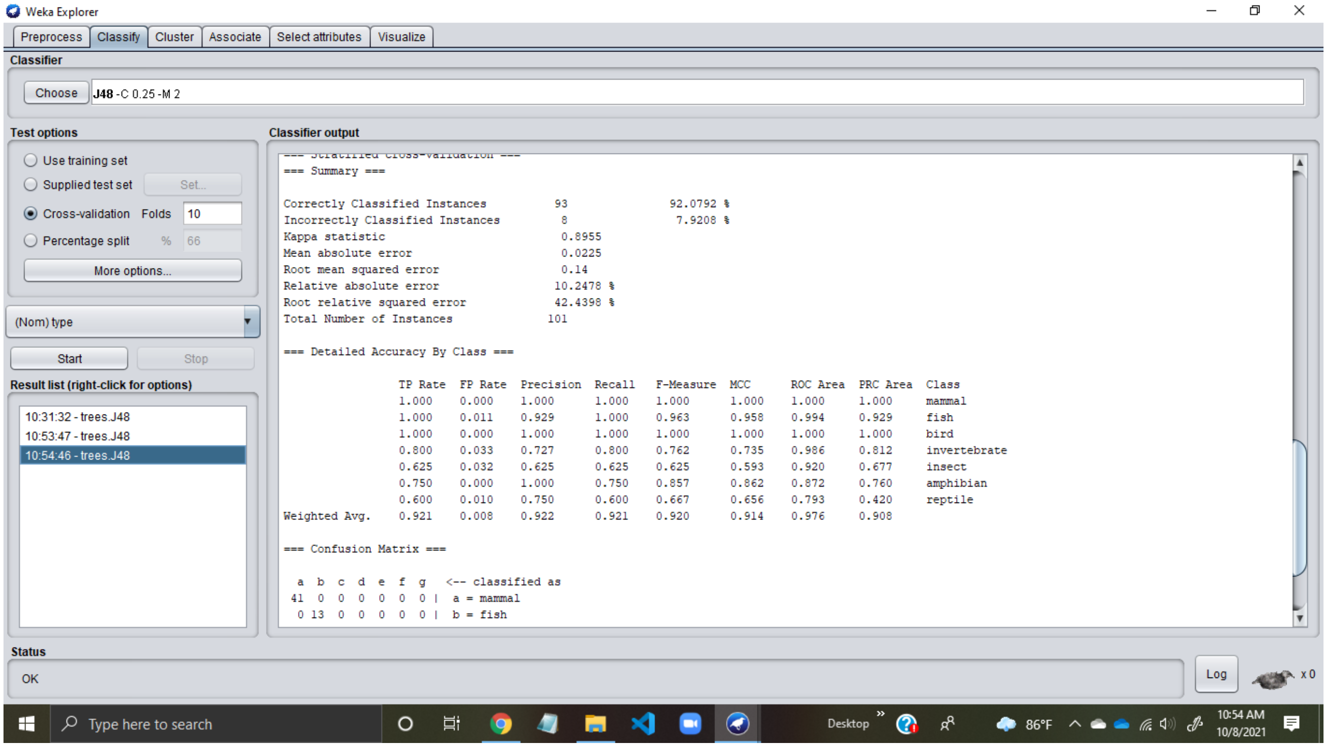
Classifier output

```
=== Summary ===

Correctly Classified Instances         100               99.0099 %
Incorrectly Classified Instances         1                0.9901 %
Kappa statistic                          0.987
Mean absolute error                      0.0047
Root mean squared error                  0.0486
Relative absolute error                  2.1552 %
Root relative squared error             14.7377 %
Total Number of Instances              101

=== 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     mammal
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     fish
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     invertebrate
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     insect
                 0.750    0.000    1.000      0.750    0.857      0.862    0.994     0.861     amphibian
                 1.000    0.010    0.833      1.000    0.909      0.908    0.995     0.833     reptile
Weighted Avg.    0.990    0.001    0.992      0.990    0.990      0.990    0.999     0.986

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
  0 13  0  0  0  0  0 |  b = fish
```

Status
OK

10:54 AM
10/8/2021

b. 10-fold cross validation



Classifier: J48 -C 0.25 -M 2

Test options:
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) type

Result list (right-click for options)
10:31:32 - trees.J48
10:53:47 - trees.J48
10:54:46 - trees.J48

Classifier output

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

Correctly Classified Instances          93               92.0792 %
Incorrectly Classified Instances         8                7.9208 %
Kappa statistic                          0.8955
Mean absolute error                      0.0225
Root mean squared error                  0.14
Relative absolute error                 10.2478 %
Root relative squared error             42.4398 %
Total Number of Instances              101

=== 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     mammal
                 1.000    0.011    0.929      1.000    0.963      0.958    0.994     0.929     fish
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
                 0.800    0.033    0.727      0.800    0.762      0.735    0.986     0.812     invertebrate
                 0.625    0.032    0.625      0.625    0.625      0.593    0.920     0.677     insect
                 0.750    0.000    1.000      0.750    0.857      0.862    0.872     0.760     amphibian
                 0.600    0.010    0.750      0.600    0.667      0.656    0.793     0.420     reptile
Weighted Avg.    0.921    0.008    0.922      0.921    0.920      0.914    0.976     0.908

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
  0 13  0  0  0  0  0 |  b = fish
```

Status
OK

10:54 AM
10/8/2021

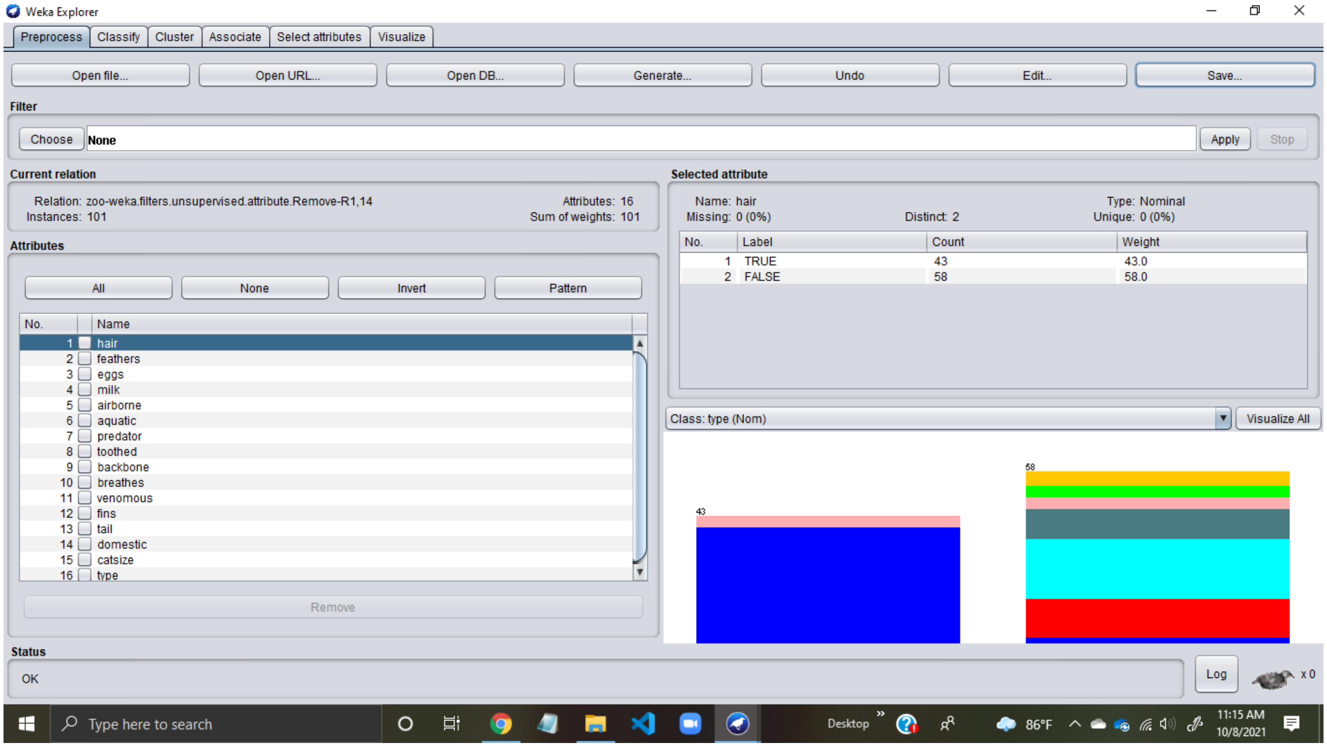
Training set accuracy : 99.008%
Cross validation accuracy : 92.07%

Training set accuracy is higher than the cross validation. That happen because of

05. Can we use ID3?

Can't use it because in 2 attributes there are some missing attributes. Therefore we need to remove or fill those missing values.

06, Remover 2 instances



Filter: None

Current relation

Relation: zoo-weka.filters.unsupervised.attribute.Remove-R1,14
Instances: 101
Attributes: 16
Sum of weights: 101

Selected attribute

Name: hair
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	TRUE	43	43.0
2	FALSE	58	58.0

Attributes

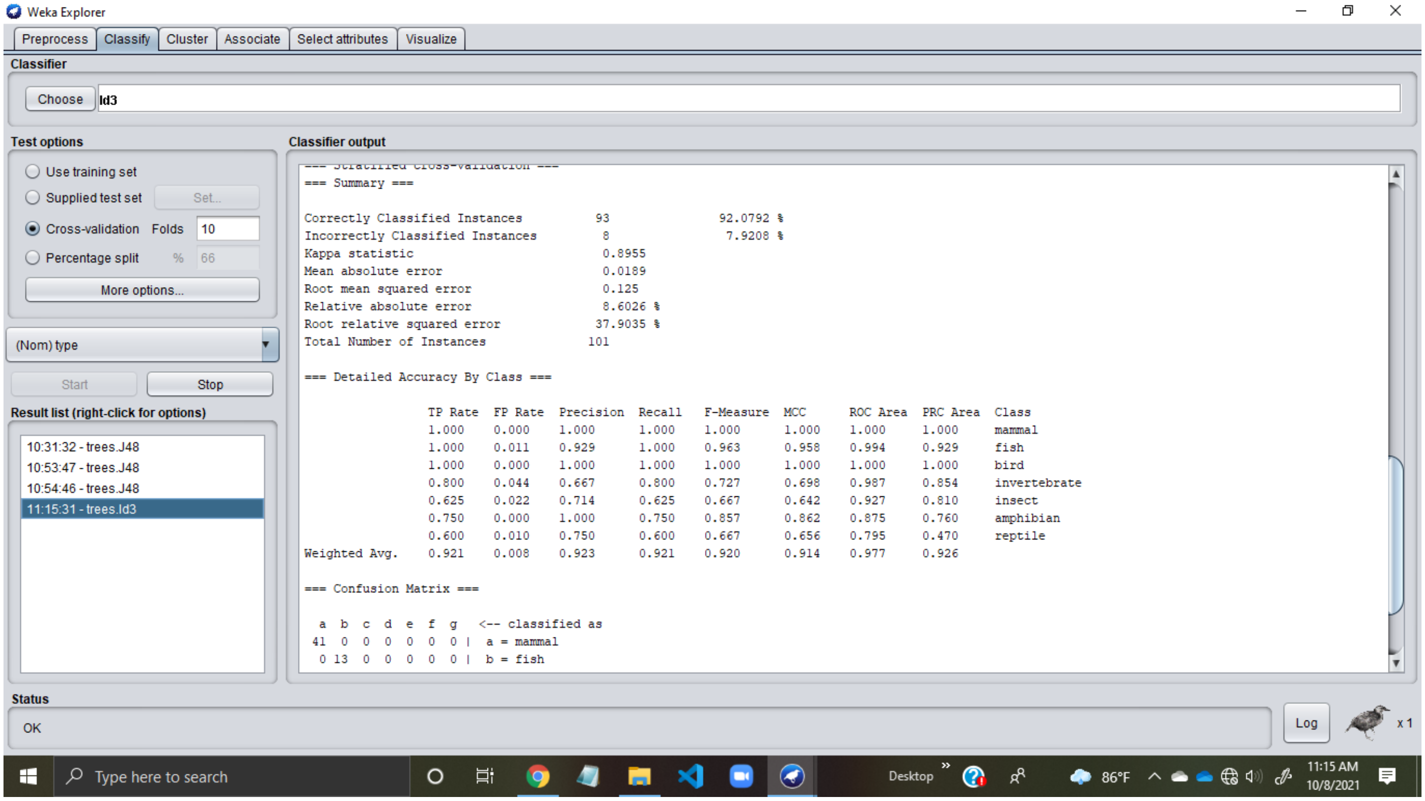
No.	Name
1	hair
2	feathers
3	eggs
4	milk
5	airborne
6	aquatic
7	predator
8	toothed
9	backbone
10	breathes
11	venomous
12	fins
13	tail
14	domestic
15	catsize
16	type

Class: type (Nom)

Status

OK

07. Build ID3



Classifier: Id3

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) type

Result list (right-click for options)

10:31:32 - trees.J48
10:53:47 - trees.J48
10:54:46 - trees.J48
11:15:31 - trees.Id3

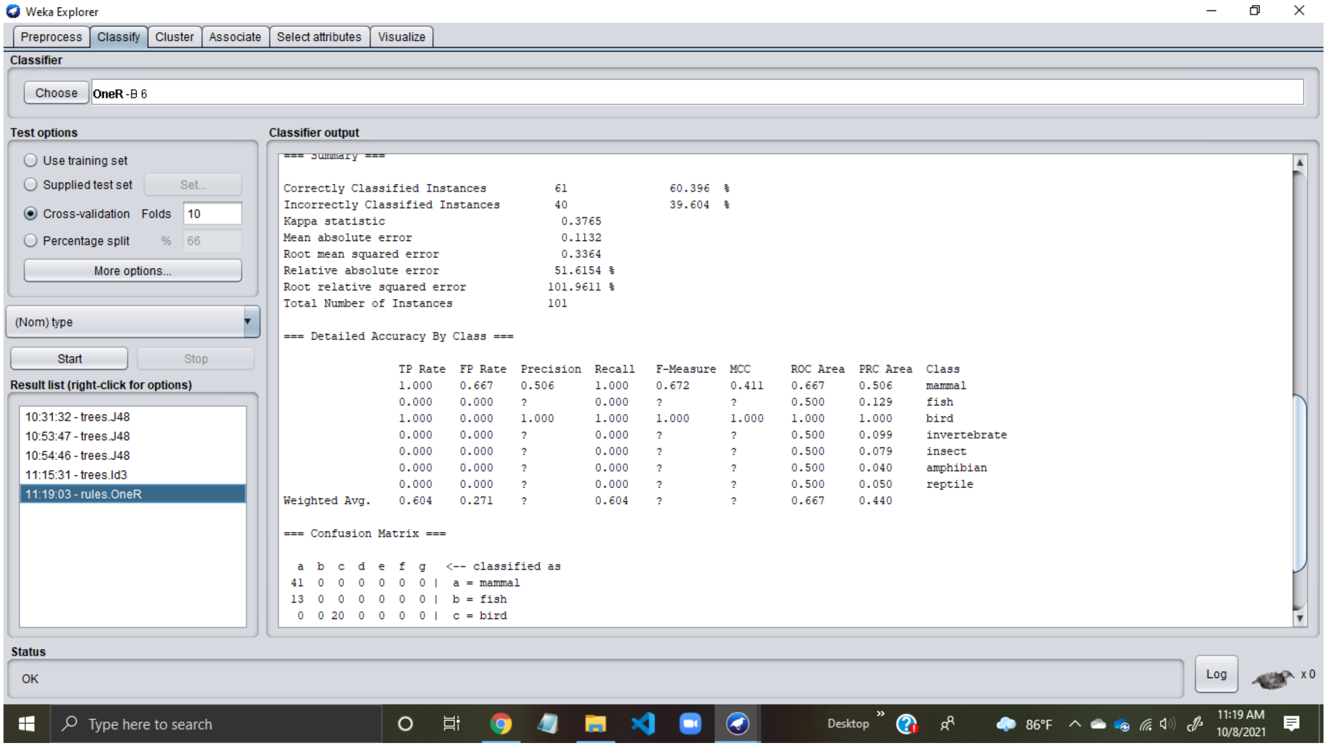
Classifier output

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

Correctly Classified Instances          93               92.0792 %
Incorrectly Classified Instances         8                7.9208 %
Kappa statistic                          0.8955
Mean absolute error                      0.0189
Root mean squared error                  0.125
Relative absolute error                  8.6026 %
Root relative squared error             37.9035 %
Total Number of Instances              101

=== 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     mammal
                 1.000    0.011    0.929      1.000    0.963      0.958    0.994     0.929     fish
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
                 0.800    0.044    0.667      0.800    0.727      0.698    0.987     0.854     invertebrate
                 0.625    0.022    0.714      0.625    0.667      0.642    0.927     0.810     insect
                 0.750    0.000    1.000      0.750    0.857      0.862    0.875     0.760     amphibian
                 0.600    0.010    0.750      0.600    0.667      0.656    0.795     0.470     reptile
Weighted Avg.    0.921    0.008    0.923      0.921    0.920      0.914    0.977     0.926

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
  0 13  0  0  0  0  0 |  b = fish
```

Status

OK

08. One R algorithm



Classifier

Choose OneR -B 6

Test options

○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66
More options...

(Nom) type

Start Stop

Result list (right-click for options)

10:31:32 - trees.J48
10:53:47 - trees.J48
10:54:46 - trees.J48
11:15:31 - trees.Id3
11:19:03 - rules.OneR

Classifier output

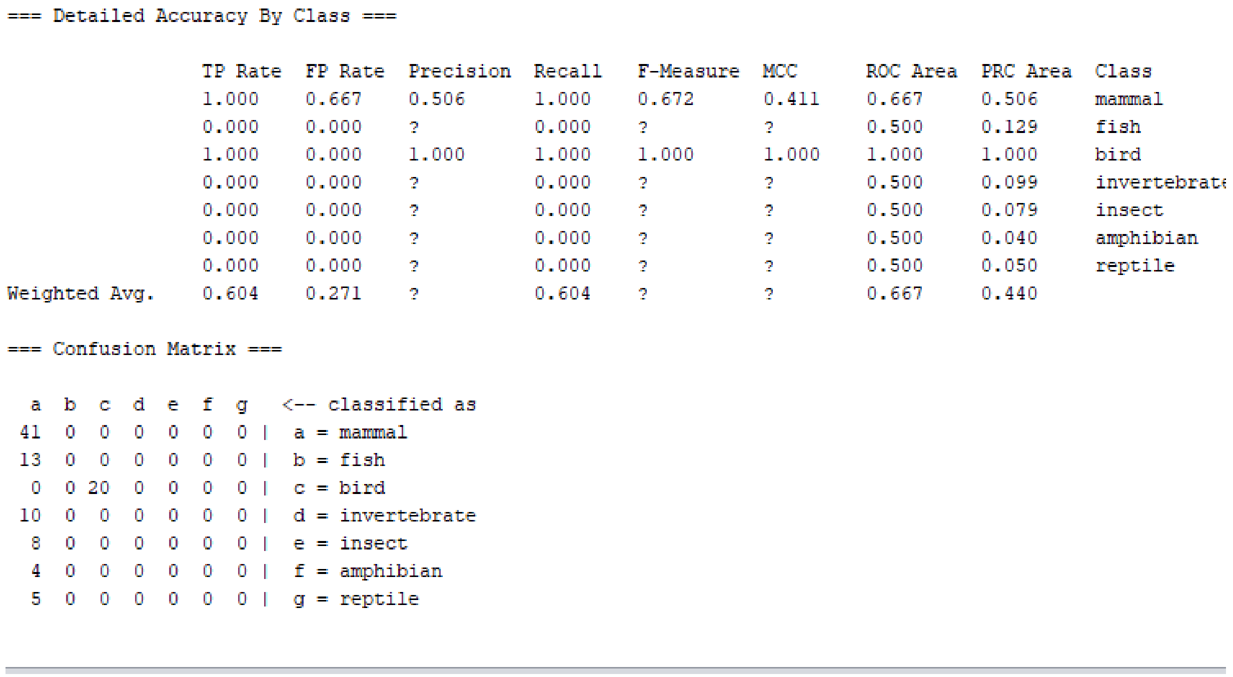
```
=== Summary ===

Correctly Classified Instances          61               60.396 %
Incorrectly Classified Instances        40               39.604 %
Kappa statistic                          0.3765
Mean absolute error                      0.1132
Root mean squared error                  0.3364
Relative absolute error                 51.6154 %
Root relative squared error            101.9611 %
Total Number of Instances              101

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.667    0.506      1.000    0.672      0.411    0.667     0.506     mammal
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.129     fish
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.099     invertebrate
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.079     insect
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.040     amphibian
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.050     reptile
Weighted Avg.    0.604    0.271    ?          0.604    ?          ?        0.667     0.440

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
 13  0  0  0  0  0  0 |  b = fish
  0  0 20  0  0  0  0 |  c = bird
```

Status

OK

Log x 0

In the oneR algorithm accuracy is less which is 60%. Only birds and mammals are correctly classified. And all others were not classified at all. Most of the ones classified as mammals with the whole category. It means that accuracy is 60% because of having 41 mammals. If not the accuracy may fall in to very less.



```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.667    0.506      1.000    0.672      0.411    0.667     0.506     mammal
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.129     fish
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.099     invertebrate
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.079     insect
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.040     amphibian
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.050     reptile
Weighted Avg.    0.604    0.271    ?          0.604    ?          ?        0.667     0.440

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
 13  0  0  0  0  0  0 |  b = fish
  0  0 20  0  0  0  0 |  c = bird
 10  0  0  0  0  0  0 |  d = invertebrate
  8  0  0  0  0  0  0 |  e = insect
  4  0  0  0  0  0  0 |  f = amphibian
  5  0  0  0  0  0  0 |  g = reptile
```