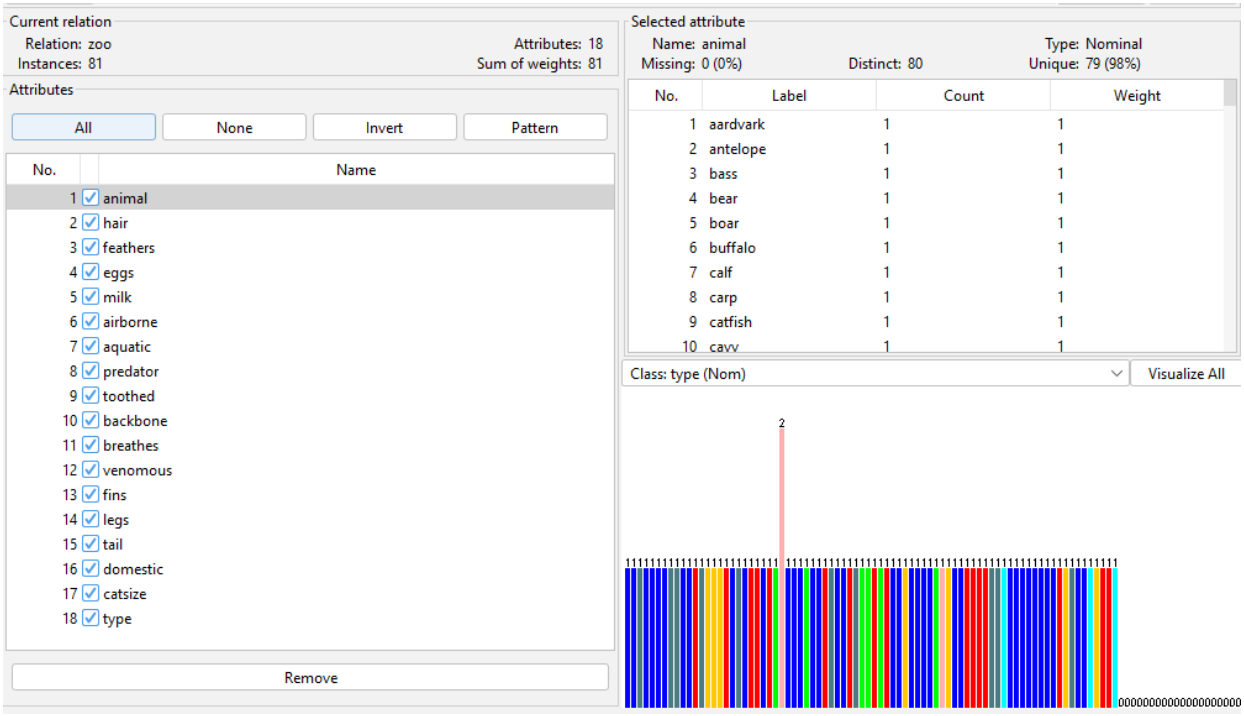
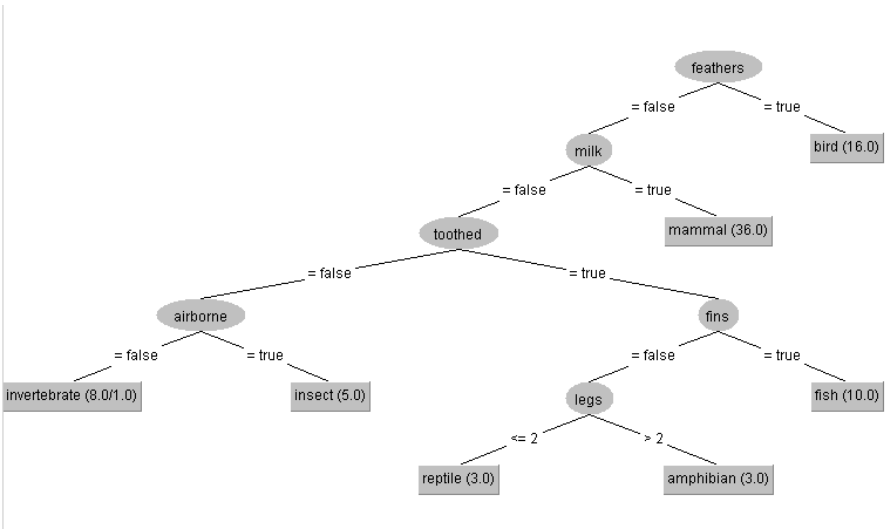
E/17/153 : Part 2 – Predicting class values

1. As it shows there are 18 attributes and 81 instances in the ‘zoo_train’ dataset



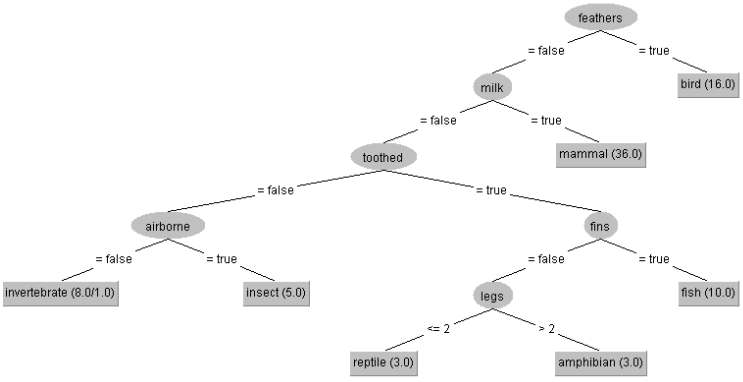
2. C4.5 decision tree with ‘Use training test’ option

```
=== Summary ===

Correctly Classified Instances          80               98.7654 %
Incorrectly Classified Instances         1                1.2346 %
Kappa statistic                          0.9831
Mean absolute error                      0.0062
Root mean squared error                  0.0556
Relative absolute error                  2.9101 %
Root relative squared error             17.1647 %
Total Number of Instances               81

=== 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     bird
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     reptile
               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     amphibian
               0.833    0.000    1.000      0.833    0.909      0.907    0.992     0.910     insect
               1.000    0.014    0.875      1.000    0.933      0.929    0.993     0.875     invertebra
Weighted Avg.  0.988    0.001    0.989      0.988    0.988      0.987    0.999     0.983

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 36  0  0  0  0  0  0 |  a = mammal
  0 16  0  0  0  0  0 |  b = bird
  0  0  3  0  0  0  0 |  c = reptile
  0  0  0 10  0  0  0 |  d = fish
  0  0  0  0  3  0  0 |  e = amphibian
  0  0  0  0  0  5  1 |  f = insect
  0  0  0  0  0  0  7 |  g = invertebrate
```



3. C4.5 decision tree with ‘Supplied test set’ option

```
=== Summary ===

Correctly Classified Instances          17               85      %
Incorrectly Classified Instances         3               15      %
Kappa statistic                          0.8187
Mean absolute error                      0.0464
Root mean squared error                  0.1965
Relative absolute error                 20.0843 %
Root relative squared error             55.849  %
Total Number of Instances               20

=== 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     bird
               0.000    0.000    ?          0.000    ?          ?        0.500     0.100     reptile
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     fish
               1.000    0.053    0.500      1.000    0.667      0.688    0.974     0.500     amphibian
               0.500    0.000    1.000      0.500    0.667      0.688    0.944     0.667     insect
               1.000    0.118    0.600      1.000    0.750      0.728    0.941     0.600     invertebra
Weighted Avg.  0.850    0.020    ?          0.850    ?          ?        0.934     0.792

=== Confusion Matrix ===

 a b c d e f g   <-- classified as
 5 0 0 0 0 0 0 | a = mammal
 0 4 0 0 0 0 0 | b = bird
 0 0 0 0 1 0 1 | c = reptile
 0 0 0 3 0 0 0 | d = fish
 0 0 0 0 1 0 0 | e = amphibian
 0 0 0 0 0 1 1 | f = insect
 0 0 0 0 0 0 3 | g = invertebrate
```



4. Output predictions

```
=== Predictions on test set ===

    inst#     actual  predicted error prediction
        1 7:invertebrate 7:invertebrate       0.875
        2     4:fish     4:fish       1
        3     2:bird     2:bird       1
        4   1:mammal   1:mammal       1
        5 7:invertebrate 7:invertebrate       0.875
        6     4:fish     4:fish       1
        7     2:bird     2:bird       1
        8   6:insect 7:invertebrate   +   0.875
        9 5:amphibian 5:amphibian       1
       10  3:reptile 7:invertebrate   +   0.875
       11  3:reptile 5:amphibian   +   1
       12     4:fish     4:fish       1
       13   1:mammal   1:mammal       1
       14   1:mammal   1:mammal       1
       15     2:bird     2:bird       1
       16   1:mammal   1:mammal       1
       17   6:insect   6:insect       1
       18   1:mammal   1:mammal       1
       19 7:invertebrate 7:invertebrate       0.875
       20     2:bird     2:bird       1
```

5. Re-evaluation

There is no difference in previous output and re-evaluated output. But if there is a large data set it might show a difference.

```
=== Predictions on user test set ===

inst#,actual,predicted,error,prediction
1,7:invertebrate,7:invertebrate,,0.875
2,4:fish,4:fish,,1
3,2:bird,2:bird,,1
4,1:mammal,1:mammal,,1
5,7:invertebrate,7:invertebrate,,0.875
6,4:fish,4:fish,,1
7,2:bird,2:bird,,1
8,6:insect,7:invertebrate,+,0.875
9,5:amphibian,5:amphibian,,1
10,3:reptile,7:invertebrate,+,0.875
11,3:reptile,5:amphibian,+,1
12,4:fish,4:fish,,1
13,1:mammal,1:mammal,,1
14,1:mammal,1:mammal,,1
15,2:bird,2:bird,,1
16,1:mammal,1:mammal,,1
17,6:insect,6:insect,,1
18,1:mammal,1:mammal,,1
19,7:invertebrate,7:invertebrate,,0.875
20,2:bird,2:bird,,1
```

Exercise:

1.Training set

```
=== Summary ===

Correctly Classified Instances          80               98.7654 %
Incorrectly Classified Instances         1                1.2346 %
Kappa statistic                          0.9831
Mean absolute error                      0.0062
Root mean squared error                  0.0556
Relative absolute error                  2.9101 %
Root relative squared error             17.1647 %
Total Number of Instances               81
```

2.Test set

```
=== Summary ===

Correctly Classified Instances          17               85      %
Incorrectly Classified Instances         3               15      %
Kappa statistic                          0.8187
Mean absolute error                      0.0464
Root mean squared error                  0.1965
Relative absolute error                 20.0843 %
Root relative squared error             55.849  %
Total Number of Instances               20
```

3.The percentage of correctly classified instances is higher in training data set than the test data set. This might be because of the number of total instances is comparably higher in the training set.

4. No result was obtained. Since the attributes are not the same in the test and train data sets it might be not able to perform the prediction using C4.5 model.

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

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 ?        ?        ?          ?        ?          ?        ?         ?         mammal
                 ?        ?        ?          ?        ?          ?        ?         ?         bird
                 ?        ?        ?          ?        ?          ?        ?         ?         reptile
                 ?        ?        ?          ?        ?          ?        ?         ?         fish
                 ?        ?        ?          ?        ?          ?        ?         ?         amphibian
                 ?        ?        ?          ?        ?          ?        ?         ?         insect
                 ?        ?        ?          ?        ?          ?        ?         ?         invertebra
Weighted Avg.    ?        ?        ?          ?        ?          ?        ?         ?

=== Confusion Matrix ===

 a b c d e f g   <-- classified as
 0 0 0 0 0 0 0 | a = mammal
 0 0 0 0 0 0 0 | b = bird
 0 0 0 0 0 0 0 | c = reptile
 0 0 0 0 0 0 0 | d = fish
 0 0 0 0 0 0 0 | e = amphibian
 0 0 0 0 0 0 0 | f = insect
 0 0 0 0 0 0 0 | g = invertebrate
```