

# CO328 - Machine Learning

## Weka Lab - Part 2

E/16/319 Rathnayake R.P.V.N

### 01. Load zoo\_train.arff

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** [Apply] [Stop]

Current relation: Relation: zoo, Instances: 81, Attributes: 18, Sum of weights: 81

Attributes: All | None | Invert | Pattern

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

Remove

Status: OK [Log] x 0

Selected attribute: Name: airborne, Missing: 0 (0%), Distinct: 2, Type: Nominal, Unique: 0 (0%)

No.	Label	Count	Weight
1	false	63	63.0
2	true	18	18.0

Class: type (Nom) [Visualize All]

Bar chart showing distribution of 'airborne' attribute: 63 (false) and 18 (true).

### 02. Build C4.5 decision tree - training set

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier: Choose **J48 - C 0.25 - M 2**

Test options: ☒ Use training set, ☐ Supplied test set, ☐ Cross-validation, ☐ Percentage split

(Nom) type: [Start] [Stop]

Result list (right-click for options): 13:05:37 - trees.J48

Classifier output:

```
=== 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  1.000  mammal
1.000  0.000  1.000  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  1.000  reptile
1.000  0.000  1.000  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  1.000  amphibian
0.833  0.000  1.000  0.833  0.909  0.907  0.992  0.910  0.910  insect
1.000  0.014  0.875  1.000  0.933  0.929  0.993  0.875  0.875  invertebrate
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
```

Status: OK [Log] x 0

### 03. 04.

The screenshot shows the Weka Explorer Classifier window. The 'Classifier' dropdown is set to 'J48 - C 0.25 - M 2'. The 'Test options' section has 'Supplied test set' selected. The 'Classifier output' pane displays the following results:

```
13:11:invertebrate,1:invertebrate,0.075
20,2:bird,2:bird,,1

=== 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
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  invertebrate
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
```

### Exercise

01

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 %

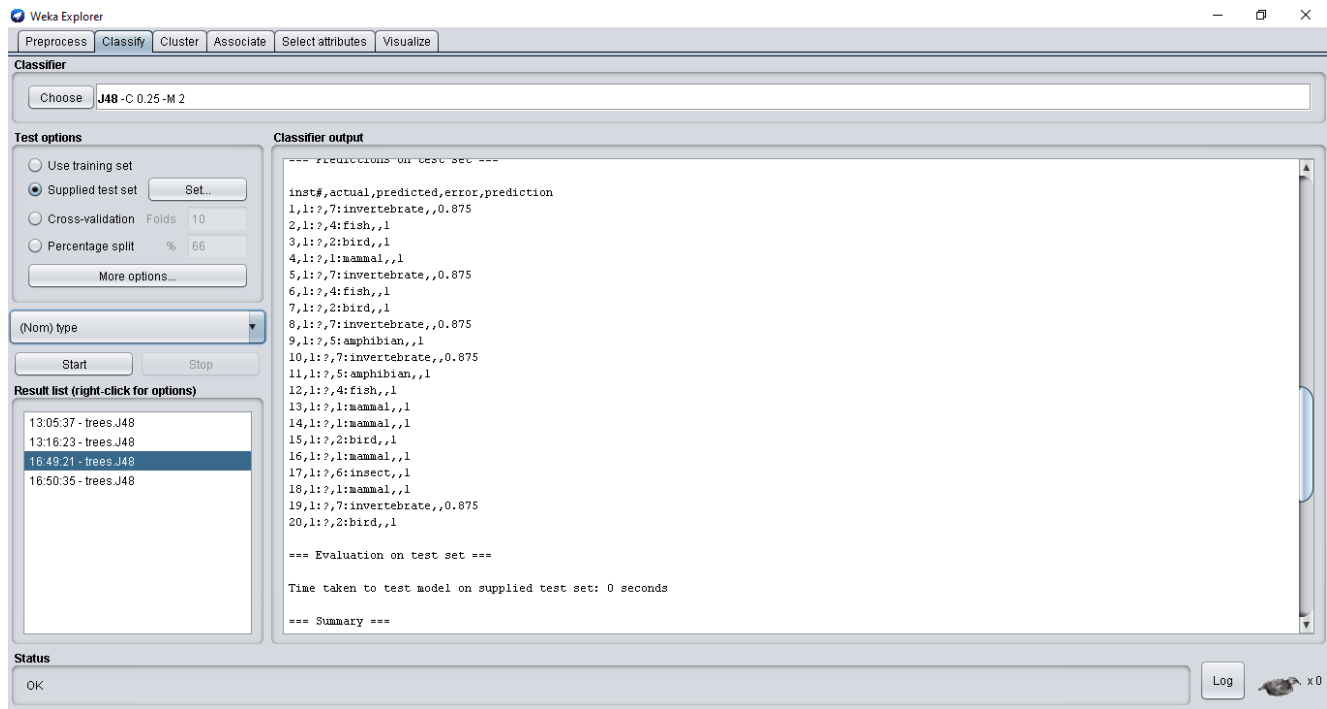
02.

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
Total Number of Instances		20

03.

The accuracy of predicting correct predict is higher for the training set than the test set model is fitted to the train set more perfectly but for the test set it shows only 85% accuracy. When we looking at the confusion matrix we can see that reptile prediction is 100% incorrect for test set.

04.



It gives no prediction this may because of missing values.