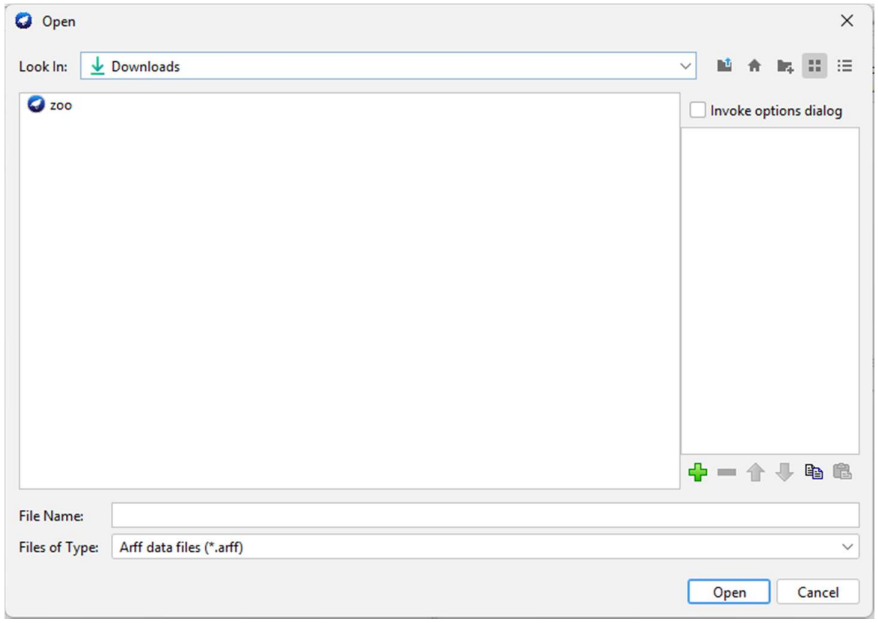
WEEK-8

Demonstration of Classification algorithm using Bayesian approach.

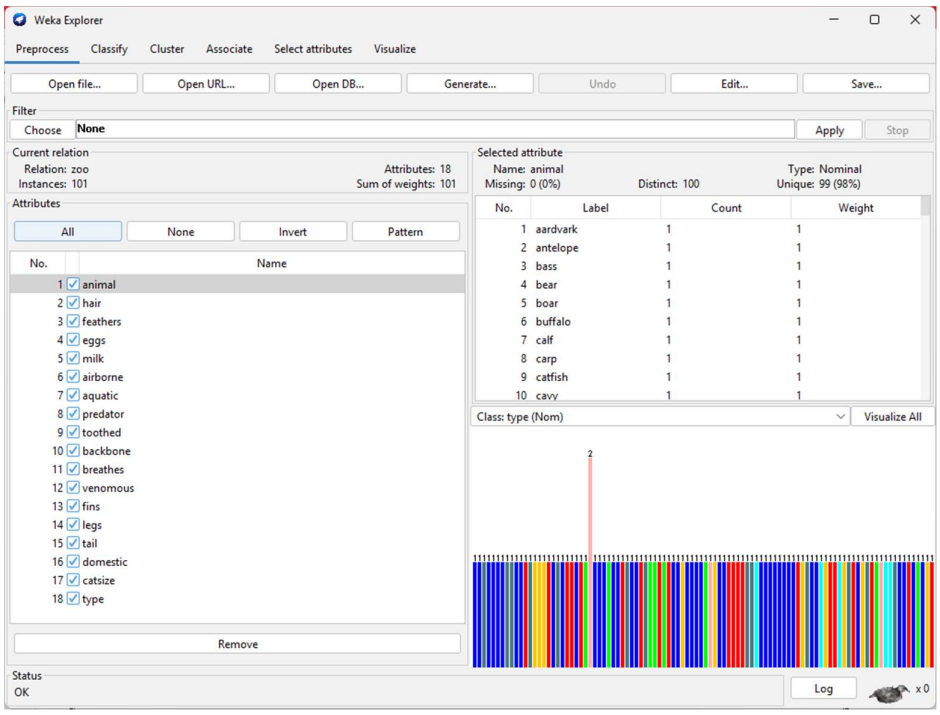
Step 1: Link to zoo.arff dataset (https://github.com/renatopp/arff-datasets/blob/master/classification/zoo.arff)

Procedure for applying Bayesian approach for zoo.arff

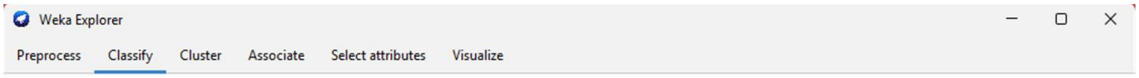
Step 1: Load the **zoo.arff** data file



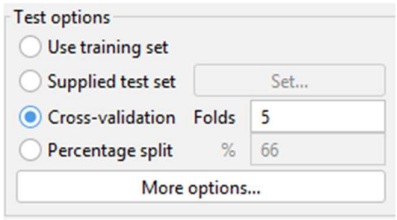
Step 2: Select all the attributes



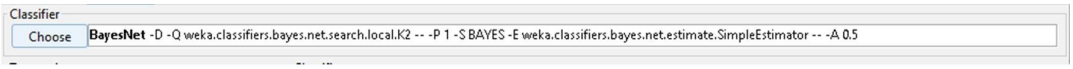
Step 3: Go to classify tab



Under the test options, change the **folds to 5**



Then click on choose, under the classifier, and select **BayesNet**



Click on the start. **(Output for 5-fold)**

Classifier output

```
Correctly Classified Instances          94               93.0693 %
Incorrectly Classified Instances         7                6.9307 %
Kappa statistic                          0.9089
Mean absolute error                      0.02
Root mean squared error                  0.1105
Relative absolute error                  9.1047 %
Root relative squared error             33.5013 %
Total Number of Instances              101

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.976    0.000    1.000      0.976    0.988      0.980    1.000     1.000     mammal
               1.000    0.012    0.952      1.000    0.976      0.970    1.000     1.000     bird
               0.600    0.021    0.600      0.600    0.600      0.579    0.988     0.839     reptile
               1.000    0.023    0.867      1.000    0.929      0.920    1.000     1.000     fish
               0.750    0.000    1.000      0.750    0.857      0.862    1.000     1.000     amphibian
               1.000    0.022    0.800      1.000    0.889      0.885    0.995     0.947     insect
               0.700    0.000    1.000      0.700    0.824      0.823    0.989     0.923     invertebrate
Weighted Avg.  0.931    0.008    0.938      0.931    0.929      0.923    0.998     0.980

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 40  0  0  1  0  0  0 |  a = mammal
  0 20  0  0  0  0  0 |  b = bird
  0  1  3  1  0  0  0 |  c = reptile
  0  0  0 13  0  0  0 |  d = fish
  0  0  1  0  3  0  0 |  e = amphibian
  0  0  0  0  0  8  0 |  f = insect
  0  0  1  0  0  2  7 |  g = invertebrate
```

Output for 10-fold validation

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

Correctly Classified Instances          95               94.0594 %
Incorrectly Classified Instances         6                5.9406 %
Kappa statistic                          0.9216
Mean absolute error                      0.0179
Root mean squared error                  0.0934
Relative absolute error                  8.1655 %
Root relative squared error             28.3188 %
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    0.012    0.952      1.000    0.976      0.970    1.000     1.000     bird
               0.600    0.021    0.600      0.600    0.600      0.579    0.988     0.844     reptile
               1.000    0.011    0.929      1.000    0.963      0.958    1.000     1.000     fish
               0.750    0.000    1.000      0.750    0.857      0.862    1.000     1.000     amphibian
               1.000    0.022    0.800      1.000    0.889      0.885    1.000     1.000     insect
               0.700    0.000    1.000      0.700    0.824      0.823    0.998     0.983     invertebrate
Weighted Avg.  0.941    0.007    0.946      0.941    0.939      0.936    0.999     0.991

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 41  0  0  0  0  0  0 |  a = mammal
  0 20  0  0  0  0  0 |  b = bird
  0  1  3  1  0  0  0 |  c = reptile
  0  0  0 13  0  0  0 |  d = fish
  0  0  1  0  3  0  0 |  e = amphibian
  0  0  0  0  0  8  0 |  f = insect
  0  0  1  0  0  2  7 |  g = invertebrate
```

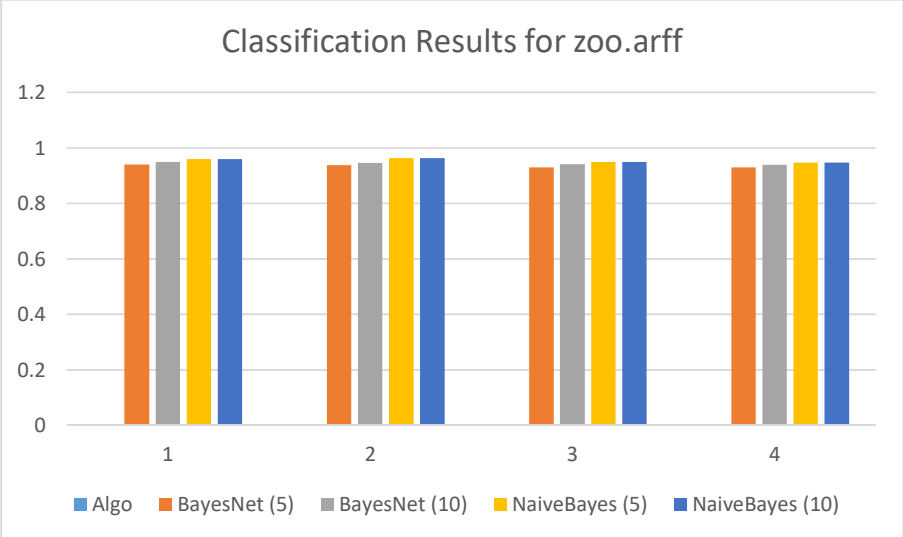
NaiveBayes (5-fold)

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

Correctly Classified Instances          96               95.0495 %
Incorrectly Classified Instances         5                4.9505 %
Kappa statistic                          0.9352
Mean absolute error                      0.0167
Root mean squared error                  0.1004
Relative absolute error                  7.5962 %
Root relative squared error             30.4618 %
Total Number of Instances              101

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.951    0.000    1.000      0.951    0.975      0.959    1.000     1.000     mammal
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
               0.400    0.000    1.000      0.400    0.571      0.623    0.985     0.825     reptile
               1.000    0.034    0.813      1.000    0.897      0.886    1.000     1.000     fish
               1.000    0.021    0.667      1.000    0.800      0.808    1.000     1.000     amphibian
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     insect
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     invertebrate
Weighted Avg.  0.950    0.005    0.963      0.950    0.947      0.943    0.999     0.991

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 39  0  0  2  0  0  0 |  a = mammal
  0 20  0  0  0  0  0 |  b = bird
  0  0  2  1  2  0  0 |  c = reptile
  0  0  0 13  0  0  0 |  d = fish
  0  0  0  0  4  0  0 |  e = amphibian
  0  0  0  0  0  8  0 |  f = insect
  0  0  0  0  0  0 10 |  g = invertebrate
```

NaiveBayes (10-fold)

```
Time taken to build model: 0 seconds

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

Correctly Classified Instances          96               95.0495 %
Incorrectly Classified Instances         5                4.9505 %
Kappa statistic                          0.9352
Mean absolute error                      0.0153
Root mean squared error                  0.098
Relative absolute error                  6.9784 %
Root relative squared error             29.693  %
Total Number of Instances              101

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.951    0.000    1.000      0.951    0.975      0.959    1.000     1.000     mammal
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     bird
               0.400    0.000    1.000      0.400    0.571      0.623    0.994     0.925     reptile
               1.000    0.034    0.813      1.000    0.897      0.886    1.000     1.000     fish
               1.000    0.021    0.667      1.000    0.800      0.808    1.000     1.000     amphibian
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     insect
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     invertebrate
Weighted Avg.  0.950    0.005    0.963      0.950    0.947      0.943    1.000     0.996

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 39  0  0  2  0  0  0 |  a = mammal
  0 20  0  0  0  0  0 |  b = bird
  0  0  2  1  2  0  0 |  c = reptile
  0  0  0 13  0  0  0 |  d = fish
  0  0  0  0  4  0  0 |  e = amphibian
  0  0  0  0  0  8  0 |  f = insect
  0  0  0  0  0  0 10 |  g = invertebrate
```

Visualization



Algo	Accuracy	Precision	Recall	F1 score
BayesNet (5)	0.94	0.938	0.931	0.929
BayesNet (10)	0.95	0.946	0.941	0.939
NaiveBayes (5)	0.96	0.963	0.95	0.947
NaiveBayes (10)	0.96	0.963	0.95	0.947

Applying Bayesian approach for AER_Credit_Card_Data.csv

Output:

BayesNet (5-Fold)

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

Correctly Classified Instances        1295               98.1804 %
Incorrectly Classified Instances        24                1.8196 %
Kappa statistic                          0.9491
Mean absolute error                      0.0185
Root mean squared error                  0.1343
Relative absolute error                  5.3044 %
Root relative squared error             32.1898 %
Total Number of Instances             1319

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.978    0.003    0.999      0.978    0.988      0.950    0.994     0.998     yes
               0.997    0.022    0.928      0.997    0.961      0.950    0.994     0.975     no
Weighted Avg.  0.982    0.008    0.983      0.982    0.982      0.950    0.994     0.993

=== Confusion Matrix ===

   a   b   <-- classified as
1000  23 |   a = yes
   1 295 |   b = no
```

BayesNet (10-Fold)

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

Correctly Classified Instances        1296               98.2563 %
Incorrectly Classified Instances        23                1.7437 %
Kappa statistic                          0.9511
Mean absolute error                      0.0177
Root mean squared error                  0.1315
Relative absolute error                  5.0951 %
Root relative squared error             31.5192 %
Total Number of Instances             1319

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.978    0.003    0.999      0.978    0.989      0.952    0.994     0.999     yes
               0.997    0.022    0.931      0.997    0.962      0.952    0.994     0.977     no
Weighted Avg.  0.983    0.007    0.984      0.983    0.983      0.952    0.994     0.994

=== Confusion Matrix ===

    a    b   <-- classified as
 1001   22 |    a = yes
    1  295 |    b = no
```

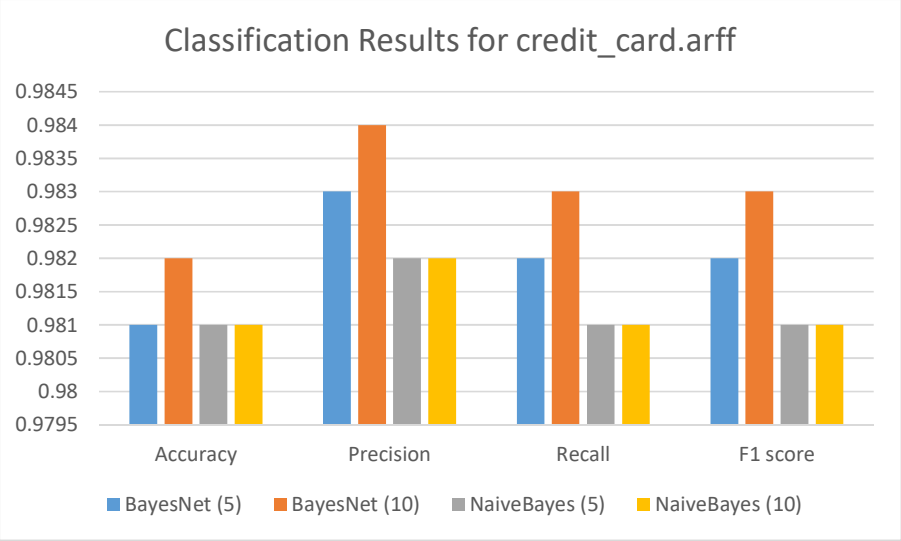
NaiveBayes (5-Fold)

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

Correctly Classified Instances        1294               98.1046 %
Incorrectly Classified Instances        25                1.8954 %
Kappa statistic                          0.947
Mean absolute error                      0.0195
Root mean squared error                  0.1368
Relative absolute error                  5.5907 %
Root relative squared error             32.7868 %
Total Number of Instances             1319

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.977    0.003    0.999      0.977    0.988      0.948    0.995     0.999     yes
               0.997    0.023    0.925      0.997    0.959      0.948    0.995     0.980     no
Weighted Avg.  0.981    0.008    0.982      0.981    0.981      0.948    0.995     0.994

=== Confusion Matrix ===

   a   b   <-- classified as
 999  24 |   a = yes
   1 295 |   b = no
```

NaiveBayes (10-Fold)

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

Correctly Classified Instances        1294               98.1046 %
Incorrectly Classified Instances        25                1.8954 %
Kappa statistic                          0.947
Mean absolute error                      0.0197
Root mean squared error                  0.1371
Relative absolute error                  5.6542 %
Root relative squared error             32.8573 %
Total Number of Instances             1319

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.977    0.003    0.999      0.977    0.988      0.948    0.996     0.999     yes
               0.997    0.023    0.925      0.997    0.959      0.948    0.996     0.982     no
Weighted Avg.  0.981    0.008    0.982      0.981    0.981      0.948    0.996     0.995

=== Confusion Matrix ===

   a   b   <-- classified as
 999  24 |   a = yes
   1 295 |   b = no
```

Visualization



Algo	Accuracy	Precision	Recall	F1 score
BayesNet (5)	0.981	0.983	0.982	0.982
BayesNet (10)	0.982	0.984	0.983	0.983
NaiveBayes (5)	0.981	0.982	0.981	0.981
NaiveBayes (10)	0.981	0.982	0.981	0.981