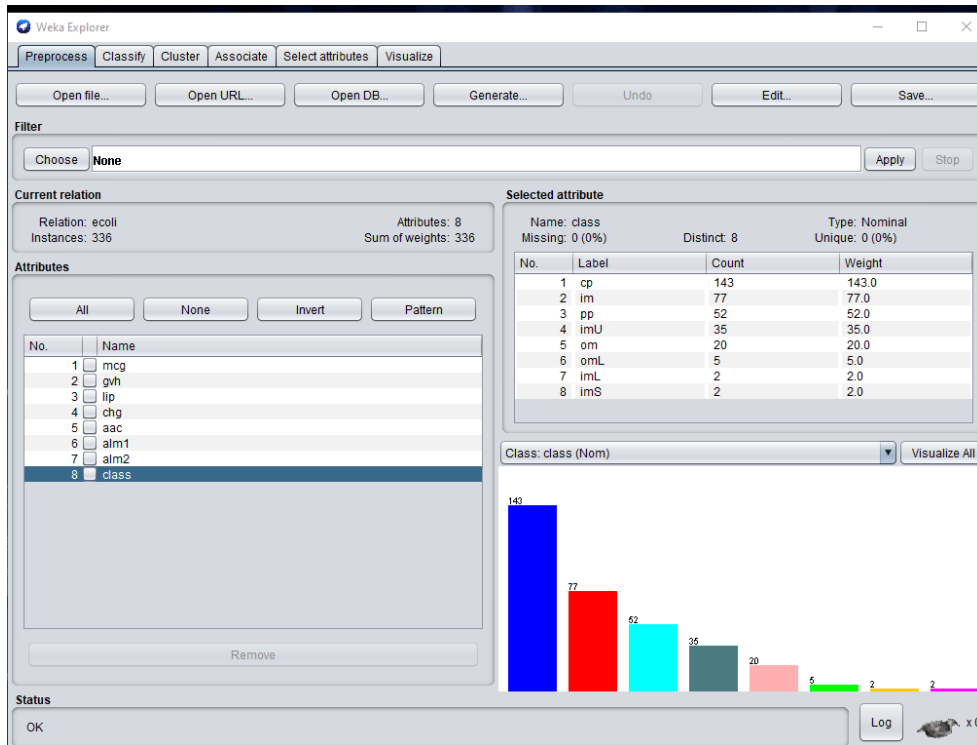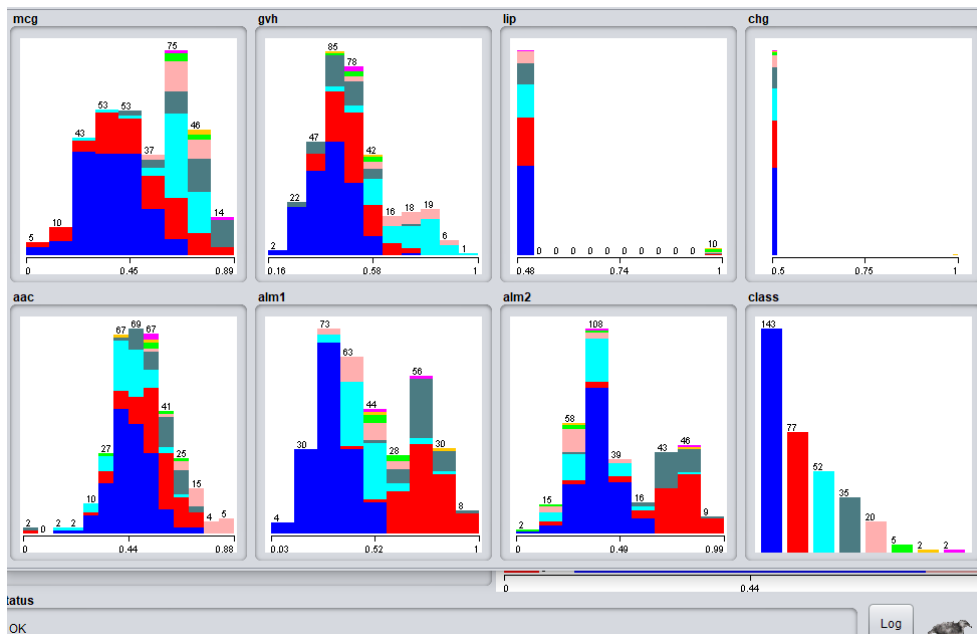# Weka Analysis Report

EECS 4412: Assignment #2 Classification by Supriyo Ghosh - 215318728

Dataset 1: Ecoli

Inspect variables. Note that the class variable has 8 values.

All the other variables are in a similar range of 0 – 1.

Most variables are normally distributed, some have a binary distribution such as lip and chg. The target class has cp has oversampled values. There is class imbalance. Hence, we apply a Resample to reduce the intensity of imbalance. After applying resampling:
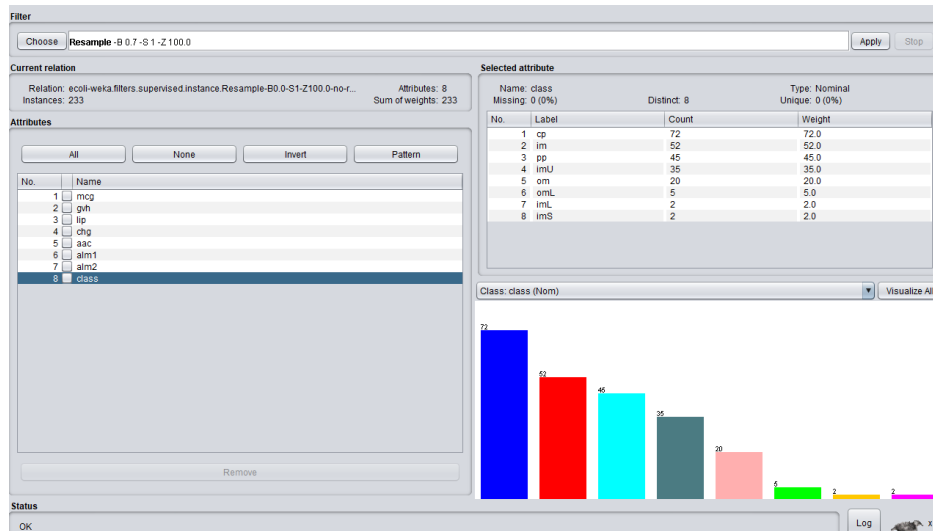
We then standardize the features.

All attributes — □ ×

mcg gvh lip chg

aac alm1 alm2 class
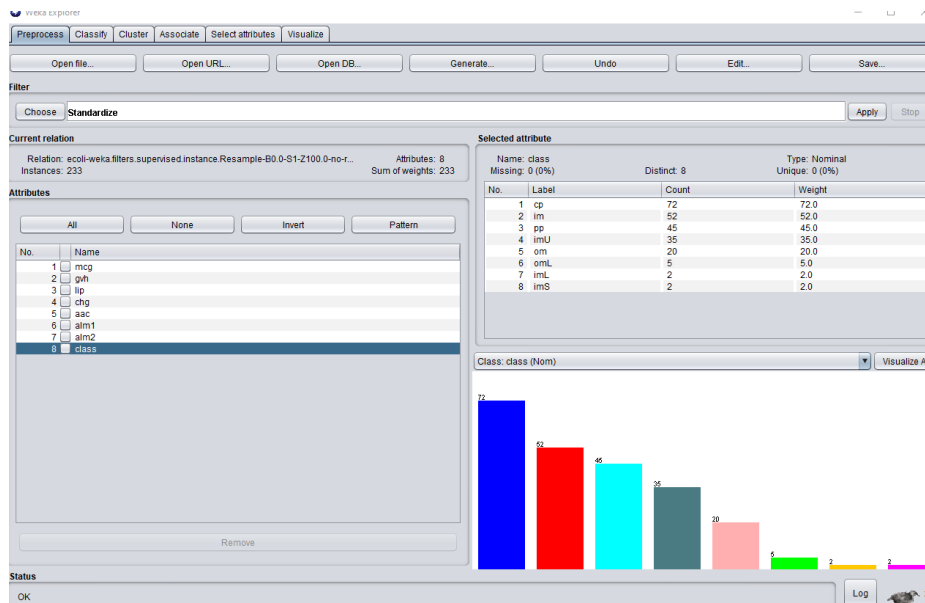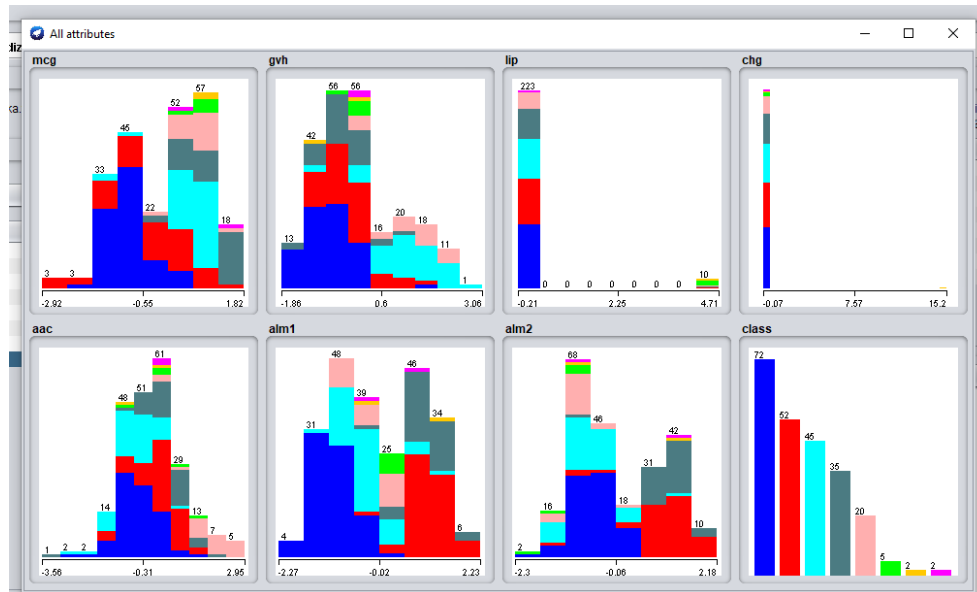
Now applying classification algos with 10-fold cross validation.

1. J48

Correctly Classified Instances   182   78.1116 %
Incorrectly Classified Instances   51   21.8884 %
Kappa statistic   0.7201
Mean absolute error   0.0672
Root mean squared error   0.2181
Relative absolute error   34.022 %
Root relative squared error   69.5081 %
Total Number of Instances   233

=== Detailed Accuracy By Class ===

                 TP Rate FP Rate Precision Recall F-Measure MCC   ROC Area PRC Area Class
                 0.944   0.050   0.895   0.944   0.919   0.882 0.950   0.839   cp
                 0.731   0.088   0.704   0.731   0.717   0.634 0.810   0.567   im
                 0.756   0.037   0.829   0.756   0.791   0.745 0.891   0.703   pp
                 0.629   0.071   0.611   0.629   0.620   0.551 0.841   0.582   imU
                 0.800   0.023   0.762   0.800   0.780   0.760 0.908   0.799   om
                 0.800   0.004   0.800   0.800   0.800   0.796 0.892   0.591   omL
                 0.000   0.000   ?   0.000   ?   ?   0.483   0.009   imL
                 0.000   0.000   ?   0.000   ?   ?   0.424   0.009   imS
Weighted Avg.   0.781   0.055   ?   0.781   ?   ?   0.878   0.691

=== Confusion Matrix ===

  a b c d e f g h   <-- classified as
 68 0 3 0 1 0 0 0 | a = cp
  2 38 1 11 0 0 0 0 | b = im
  4 4 34 0 3 0 0 0 | c = pp
  2 11 0 22 0 0 0 0 | d = imU
  0 0 3 1 16 0 0 0 | e = om
  0 0 0 0 1 4 0 0 | f = omL
  0 1 0 0 0 1 0 0 | g = imL
  0 0 0 2 0 0 0 0 | h = imS

2. Jrip

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

Correctly Classified Instances         175               75.1073 %
Incorrectly Classified Instances        58               24.8927 %
Kappa statistic                          0.6786
Mean absolute error                      0.0785
Root mean squared error                  0.2295
Relative absolute error                 39.7238 %
Root relative squared error             73.1361 %
Total Number of Instances              233

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.958	0.106	0.802	0.958	0.873	0.817	0.933	0.807	cp
	0.750	0.088	0.709	0.750	0.729	0.649	0.838	0.630	im
	0.778	0.048	0.795	0.778	0.787	0.736	0.900	0.732	pp
	0.457	0.040	0.667	0.457	0.542	0.490	0.805	0.530	imU
	0.700	0.014	0.824	0.700	0.757	0.739	0.903	0.695	om
	0.400	0.018	0.333	0.400	0.364	0.350	0.695	0.413	omL
	0.000	0.004	0.000	0.000	0.000	-0.006	0.989	0.325	imL
	0.000	0.000	?	0.000	?	?	0.281	0.009	imS
Weighted Avg.	0.751	0.069	?	0.751	?	?	0.873	0.682	

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 69  0  2  0  1  0  0  0 |  a = cp
  3 39  2  7  0  1  0  0 |  b = im
  7  2 35  0  1  0  0  0 |  c = pp
  4 13  1 16  1  0  0  0 |  d = imU
  1  0  4  0 14  1  0  0 |  e = om
  2  0  0  0  0  2  1  0 |  f = omL
  0  0  0  0  0  2  0  0 |  g = imL
  0  1  0  1  0  0  0  0 |  h = imS
```

3. Naïve bayes

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

Correctly Classified Instances         195               83.691 %
Incorrectly Classified Instances        38               16.309 %
Kappa statistic                          0.7921
Mean absolute error                      0.0493
Root mean squared error                  0.178
Relative absolute error                 24.9641 %
Root relative squared error             56.7216 %
Total Number of Instances              233

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.972	0.043	0.909	0.972	0.940	0.912	0.983	0.934	cp
	0.692	0.033	0.857	0.692	0.766	0.714	0.955	0.872	im
	0.844	0.032	0.864	0.844	0.854	0.820	0.932	0.895	pp
	0.857	0.076	0.667	0.857	0.750	0.707	0.939	0.740	imU
	0.900	0.009	0.900	0.900	0.900	0.891	0.995	0.968	om
	0.600	0.000	1.000	0.600	0.750	0.771	0.992	0.871	omL
	0.000	0.009	0.000	0.000	0.000	-0.009	0.058	0.009	imL
	0.000	0.000	?	0.000	?	?	0.106	0.007	imS
Weighted Avg.	0.837	0.039	?	0.837	?	?	0.946	0.869	

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 70  1  1  0  0  0  0  0 |  a = cp
  2 36  1 13  0  0  0  0 |  b = im
  4  1 38  0  2  0  0  0 |  c = pp
  1  4  0 30  0  0  0  0 |  d = imU
  0  0  2  0 18  0  0  0 |  e = om
  0  0  0  0  0  3  2  0 |  f = omL
  0  0  1  1  0  0  0  0 |  g = imL
  0  0  1  1  0  0  0  0 |  h = imS
```

4. BayesNet

```
Time taken to build model: 0.02 seconds

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

Correctly Classified Instances         182               78.1116 %
Incorrectly Classified Instances        51               21.8884 %
Kappa statistic                          0.7203
Mean absolute error                      0.0647
Root mean squared error                  0.1952
Relative absolute error                 32.7616 %
Root relative squared error             62.2231 %
Total Number of Instances              233

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.931	0.043	0.905	0.931	0.918	0.880	0.985	0.959	cp
	0.731	0.083	0.717	0.731	0.724	0.644	0.943	0.801	im
	0.800	0.059	0.766	0.800	0.783	0.729	0.942	0.870	pp
	0.657	0.066	0.639	0.657	0.648	0.585	0.920	0.706	imU
	0.700	0.009	0.875	0.700	0.778	0.765	0.982	0.895	om
	0.800	0.009	0.667	0.800	0.727	0.724	0.994	0.754	omL
	0.000	0.000	?	0.000	?	?	0.911	0.067	imL
	0.000	0.004	0.000	0.000	0.000	-0.006	0.617	0.051	imS
Weighted Avg.	0.781	0.054	?	0.781	?	?	0.954	0.843	

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 67  1  4  0  0  0  0  0 |  a = cp
  1 38  1 12  0  0  0  0 |  b = im
  4  3 36  0  2  0  0  0 |  c = pp
  1 10  0 23  0  0  0  1 |  d = imU
  0  0  5  0 14  1  0  0 |  e = om
  0  0  1  0  0  4  0  0 |  f = omL
  0  0  0  1  0  1  0  0 |  g = imL
  1  1  0  0  0  0  0  0 |  h = imS
```
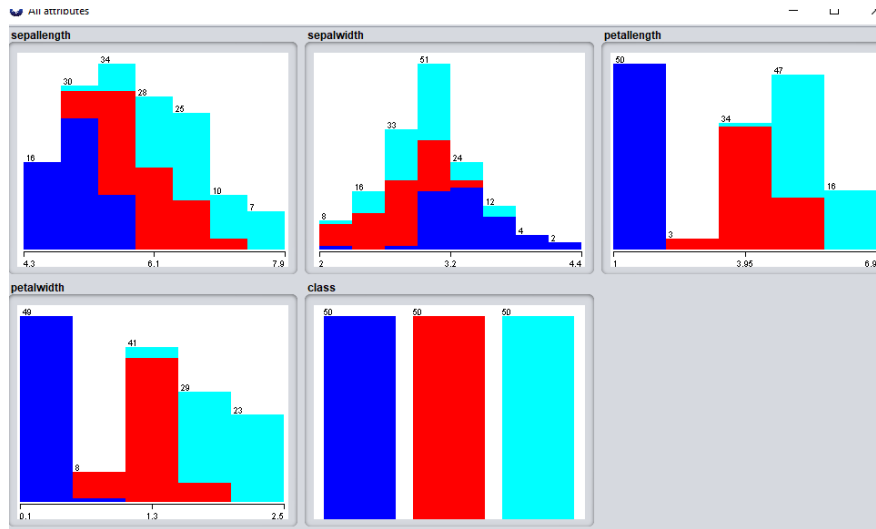
5. IBk

```
Time taken to build model: 0 seconds

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

Correctly Classified Instances         177               75.9657 %
Incorrectly Classified Instances        56               24.0343 %
Kappa statistic                          0.6935
Mean absolute error                      0.0659
Root mean squared error                  0.2409
Relative absolute error                 33.3597 %
Root relative squared error             76.7743 %
Total Number of Instances              233

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.889	0.056	0.877	0.889	0.883	0.830	0.920	0.845	cp
	0.673	0.094	0.673	0.673	0.673	0.579	0.828	0.576	im
	0.822	0.069	0.740	0.822	0.779	0.724	0.879	0.641	pp
	0.571	0.066	0.606	0.571	0.588	0.518	0.732	0.431	imU
	0.800	0.000	1.000	0.800	0.889	0.886	0.894	0.823	om
	1.000	0.004	0.833	1.000	0.909	0.911	1.000	1.000	omL
	0.000	0.004	0.000	0.000	0.000	-0.006	0.846	0.028	imL
	0.000	0.009	0.000	0.000	0.000	-0.009	0.590	0.011	imS
Weighted Avg.	0.760	0.062	0.759	0.760	0.758	0.699	0.859	0.670	

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 64  2  6  0  0  0  0  0 |  a = cp
  3 35  2 11  0  0  1  0 |  b = im
  5  2 37  0  0  0  1  1 |  c = pp
  1 12  1 20  0  0  0  1 |  d = imU
  0  0  3  1 16  0  0  0 |  e = om
  0  0  0  0  0  5  0  0 |  f = omL
  0  1  0  0  0  1  0  0 |  g = imL
  0  0  1  1  0  0  0  0 |  h = imS
```

6. MultilayerPerceptron

```
Time taken to build model: 0.36 seconds

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

Correctly Classified Instances         190               81.5451 %
Incorrectly Classified Instances        43               18.4549 %
Kappa statistic                          0.7628
Mean absolute error                      0.0618
Root mean squared error                  0.1954
Relative absolute error                 31.2727 %
Root relative squared error             62.2912 %
Total Number of Instances              233

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC Area  Class
              0.972    0.043    0.909      0.972   0.940      0.912   0.977     0.912     cp
              0.808    0.077    0.750      0.808   0.778      0.712   0.935     0.819     im
              0.800    0.043    0.818      0.800   0.809      0.764   0.934     0.802     pp
              0.743    0.045    0.743      0.743   0.743      0.697   0.929     0.730     imU
              0.800    0.014    0.842      0.800   0.821      0.804   0.968     0.760     om
              0.000    0.009    0.000      0.000   0.000      -0.014  0.760     0.197     omL
              0.000    0.000    ?          0.000   ?          ?       0.026     0.007     imL
              0.000    0.000    ?          0.000   ?          ?       0.528     0.013     imS
Weighted Avg. 0.815    0.047    ?          0.815   ?          ?       0.935     0.799

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 70  1  1  0  0  0  0  0 |  a = cp
  1 42  1  8  0  0  0  0 |  b = im
  3  3 36  1  2  0  0  0 |  c = pp
  1  7  1 26  0  0  0  0 |  d = imU
  0  0  3  0 16  1  0  0 |  e = om
  2  1  1  0  1  0  0  0 |  f = omL
  0  1  0  0  0  1  0  0 |  g = imL
  0  1  1  0  0  0  0  0 |  h = imS
```

Summary of results

Data	Model	Classification Error
Ecoli	J48	21.9%
Ecoli	JRip	24.9%
Ecoli	NaiveBayes	16.3%
Ecoli	BayesNet	21.9%
Ecoli	IBK	24%
Ecoli	Multilayer Perceptron	18.4%

Dataset 2: Iris

Note that the class has equal balance and similar distribution.

Now applying classification algos with 10-fold cross validation.

1. J48

2. Jrip


```
Classifier
Choose   JRip -F 3 -N 2.0 -O 2 -S 1

Test options                          Classifier output
○ Use training set                                    class
○ Supplied test set    Set..          Test mode:    10-fold cross-validation
● Cross-validation  Folds  10         === Classifier model (full training set) ===
○ Percentage split    %   66
                                      JRIP rules:
       More options...               ===========

Nom) class                           (petallength >= 3.3) and (petalwidth <= 1.6) and (petallength <= 4.9) => class=Iris-versicolor (46.0/0.0)
                                     (petallength <= 1.9) => class=Iris-setosa (50.0/0.0)
    Start          Stop               => class=Iris-virginica (54.0/4.0)

sult list (right-click for options)  Number of Rules : 3

16:23:20 - trees.J48
16:24:23 - trees.J48                 Time taken to build model: 0 seconds
16:24:36 - trees.J48
16:31:12 - trees.J48                 === Stratified cross-validation ===
16:36:07 - trees.J48                 === Summary ===
16:36:58 - rules.JRip
16:37:37 - bayes.NaiveBayes          Correctly Classified Instances        143               95.3333 %
16:38:08 - bayes.BayesNet            Incorrectly Classified Instances        7                4.6667 %
16:38:39 - lazy.IBk                  Kappa statistic                        0.93
16:39:01 - functions.MultilayerPerceptron  Mean absolute error              0.0454
16:43:43 - functions.MultilayerPerceptron  Root mean squared error          0.1727
16:44:29 - functions.MultilayerPerceptron  Relative absolute error         10.2173 %
16:44:30 - functions.MultilayerPerceptron  Root relative squared error     36.6447 %
16:44:32 - functions.MultilayerPerceptron  Total Number of Instances        150
16:44:59 - trees.J48
16:45:18 - rules.JRip                === 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     Iris-setosa
                                               0.920    0.030    0.939      0.920   0.929      0.895  0.966     0.936     Iris-versicolor
                                               0.940    0.040    0.922      0.940   0.931      0.896  0.960     0.878     Iris-virginica
                                     Weighted Avg.  0.953  0.023  0.953      0.953   0.953      0.930  0.975     0.938

                                     === Confusion Matrix ===

                                      a  b  c   <-- classified as
                                     50  0  0 |  a = Iris-setosa
                                      0 46  4 |  b = Iris-versicolor
                                      0  3 47 |  c = Iris-virginica
```

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

Correctly Classified Instances        175               75.1073 %
Incorrectly Classified Instances       58               24.8927 %
Kappa statistic                        0.6786
Mean absolute error                    0.0785
Root mean squared error                0.2295
Relative absolute error               39.7238 %
Root relative squared error           73.1361 %
Total Number of Instances             233

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.958	0.106	0.802	0.958	0.873	0.817	0.933	0.807	cp
	0.750	0.088	0.709	0.750	0.729	0.649	0.838	0.630	im
	0.778	0.048	0.795	0.778	0.787	0.736	0.900	0.732	pp
	0.457	0.040	0.667	0.457	0.542	0.490	0.805	0.530	imU
	0.700	0.014	0.824	0.700	0.757	0.739	0.903	0.695	om
	0.400	0.018	0.333	0.400	0.364	0.350	0.695	0.413	omL
	0.000	0.004	0.000	0.000	0.000	-0.006	0.989	0.325	imL
	0.000	0.000	?	0.000	?	?	0.281	0.009	imS
Weighted Avg.	0.751	0.069	?	0.751	?	?	0.873	0.682	

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 69  0  2  0  1  0  0  0 |  a = cp
  3 39  2  7  0  1  0  0 |  b = im
  7  2 35  0  1  0  0  0 |  c = pp
  4 13  1 16  1  0  0  0 |  d = imU
  1  0  4  0 14  1  0  0 |  e = om
  2  0  0  0  0  2  1  0 |  f = omL
  0  0  0  0  0  2  0  0 |  g = imL
  0  1  0  1  0  0  0  0 |  h = imS
```

3. Naïve bayes

lassifier

Choose | NaiveBayes

```
est options                      Classifier output

○ Use training set               petallength
○ Supplied test set    Set...       mean              1.4694       4.2452       5.5516
                                    std. dev.         0.1782       0.4712       0.5529
● Cross-validation  Folds  10       weight sum          50           50           50
○ Percentage split    %   66        precision         0.1405       0.1405       0.1405

      More options...            petalwidth
                                    mean              0.2743       1.3097       2.0343
(Nom) class                         std. dev.         0.1096       0.1915       0.2646
                                    weight sum          50           50           50
    Start         Stop              precision         0.1143       0.1143       0.1143

esult list (right-click for options)
                                 Time taken to build model: 0 seconds
16:23:20 - trees.J48
16:24:23 - trees.J48             === Stratified cross-validation ===
16:24:36 - trees.J48             === Summary ===
16:31:12 - trees.J48
16:36:07 - trees.J48             Correctly Classified Instances      144           96      %
16:36:58 - rules.JRip            Incorrectly Classified Instances      6            4      %
16:37:37 - bayes.NaiveBayes      Kappa statistic                       0.94
16:38:08 - bayes.BayesNet        Mean absolute error                   0.0342
16:38:39 - lazy.IBk              Root mean squared error               0.155
16:39:01 - functions.MultilayerPerceptron   Relative absolute error         7.6997 %
16:43:43 - functions.MultilayerPerceptron   Root relative squared error    32.8794 %
16:44:29 - functions.MultilayerPerceptron   Total Number of Instances        150
16:44:30 - functions.MultilayerPerceptron
16:44:32 - functions.MultilayerPerceptron   === Detailed Accuracy By Class ===
16:44:59 - trees.J48
16:45:18 - rules.JRip                        TP Rate FP Rate Precision Recall F-Measure MCC   ROC Area PRC Area Class
16:45:32 - bayes.NaiveBayes                   1.000   0.000   1.000   1.000   1.000   1.000  1.000   1.000   Iris-setosa
                                              0.960   0.040   0.923   0.960   0.941   0.911  0.992   0.983   Iris-versicolor
                                              0.920   0.020   0.958   0.920   0.939   0.910  0.992   0.986   Iris-virginica
                                 Weighted Avg. 0.960  0.020   0.960   0.960   0.960   0.940  0.994   0.989

                                 === Confusion Matrix ===

                                  a  b  c   <-- classified as
                                 50  0  0 |  a = Iris-setosa
                                  0 48  2 |  b = Iris-versicolor
                                  0  4 46 |  c = Iris-virginica
```

4. BayesNet

```
class(3):
LogScore Bayes: -481.00632967833803
LogScore BDeu: -525.3834868062277
LogScore MDL: -536.5317339418378
LogScore ENTROPY: -471.39347511858665
LogScore AIC: -497.39347511858665


Time taken to build model: 0 seconds

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

Correctly Classified Instances      139           92.6667 %
Incorrectly Classified Instances     11            7.3333 %
Kappa statistic                       0.89
Mean absolute error                   0.0454
Root mean squared error               0.1828
Relative absolute error              10.2111 %
Root relative squared error          38.7793 %
Total Number of Instances            150

=== 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   Iris-setosa
               0.880   0.050   0.898   0.880   0.889   0.834  0.971   0.906   Iris-versicolor
               0.900   0.060   0.882   0.900   0.891   0.836  0.970   0.919   Iris-virginica
Weighted Avg.  0.927   0.037   0.927   0.927   0.927   0.890  0.980   0.942

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 44  6 |  b = Iris-versicolor
  0  5 45 |  c = Iris-virginica
```

5. IBk

Choose   **IBk** -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options

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

More options...

(Nom) class

Start   Stop

Result list (right-click for options)

- 16:23:20 - trees.J48
- 16:24:23 - trees.J48
- 16:24:36 - trees.J48
- 16:31:12 - trees.J48
- 16:36:07 - trees.J48
- 16:36:58 - rules.JRip
- 16:37:37 - bayes.NaiveBayes
- 16:38:08 - bayes.BayesNet
- 16:38:39 - lazy.IBk
- 16:39:01 - functions.MultilayerPerceptron
- 16:43:43 - functions.MultilayerPerceptron
- 16:44:29 - functions.MultilayerPerceptron
- 16:44:30 - functions.MultilayerPerceptron
- 16:44:32 - functions.MultilayerPerceptron
- 16:44:59 - trees.J48
- 16:45:18 - rules.JRip
- 16:45:32 - bayes.NaiveBayes
- 16:45:47 - bayes.BayesNet
- 16:46:02 - lazy.IBk

Classifier output

```
Instances:    150
Attributes:   5
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
Test mode:    10-fold cross-validation

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

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances       143               95.3333 %
Incorrectly Classified Instances       7                4.6667 %
Kappa statistic                        0.93
Mean absolute error                    0.0399
Root mean squared error                0.1747
Relative absolute error                8.9763 %
Root relative squared error           37.0695 %
Total Number of Instances            150

=== 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     Iris-setosa
                 0.940    0.040    0.922      0.940   0.931      0.896  0.952     0.887     Iris-versicolor
                 0.920    0.030    0.939      0.920   0.929      0.895  0.947     0.894     Iris-virginica
Weighted Avg.    0.953    0.023    0.953      0.953   0.953      0.930  0.966     0.927

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 47  3 |  b = Iris-versicolor
  0  4 46 |  c = Iris-virginica
```

6. MultilayerPerceptron

Classifier

Choose   **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

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

More options...

(Nom) class

Start   Stop

Result list (right-click for options)

- 16:23:20 - trees.J48
- 16:24:23 - trees.J48
- 16:24:36 - trees.J48
- 16:31:12 - trees.J48
- 16:36:07 - trees.J48
- 16:36:58 - rules.JRip
- 16:37:37 - bayes.NaiveBayes
- 16:38:08 - bayes.BayesNet
- 16:38:39 - lazy.IBk
- 16:39:01 - functions.MultilayerPerceptron
- 16:43:43 - functions.MultilayerPerceptron
- 16:44:29 - functions.MultilayerPerceptron
- 16:44:30 - functions.MultilayerPerceptron
- 16:44:32 - functions.MultilayerPerceptron
- 16:44:59 - trees.J48
- 16:45:18 - rules.JRip
- 16:45:32 - bayes.NaiveBayes
- 16:45:47 - bayes.BayesNet
- 16:46:02 - lazy.IBk
- 16:46:16 - functions.MultilayerPerceptron

Classifier output

```
    Attrib sepallength   -1.2198676622036794
    Attrib sepalwidth    -3.5332821317534946
    Attrib petallength    8.401834252274107
    Attrib petalwidth     9.460215580472836
Class Iris-setosa
    Input
    Node 0
Class Iris-versicolor
    Input
    Node 1
Class Iris-virginica
    Input
    Node 2

Time taken to build model: 0.06 seconds

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

Correctly Classified Instances       146               97.3333 %
Incorrectly Classified Instances       4                2.6667 %
Kappa statistic                        0.96
Mean absolute error                    0.0327
Root mean squared error                0.1291
Relative absolute error                7.3555 %
Root relative squared error           27.3796 %
Total Number of Instances            150

=== 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     Iris-setosa
                 0.960    0.020    0.960      0.960   0.960      0.940  0.996     0.993     Iris-versicolor
                 0.960    0.020    0.960      0.960   0.960      0.940  0.996     0.993     Iris-virginica
Weighted Avg.    0.973    0.013    0.973      0.973   0.973      0.960  0.998     0.995

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 48  2 |  b = Iris-versicolor
  0  2 48 |  c = Iris-virginica
```
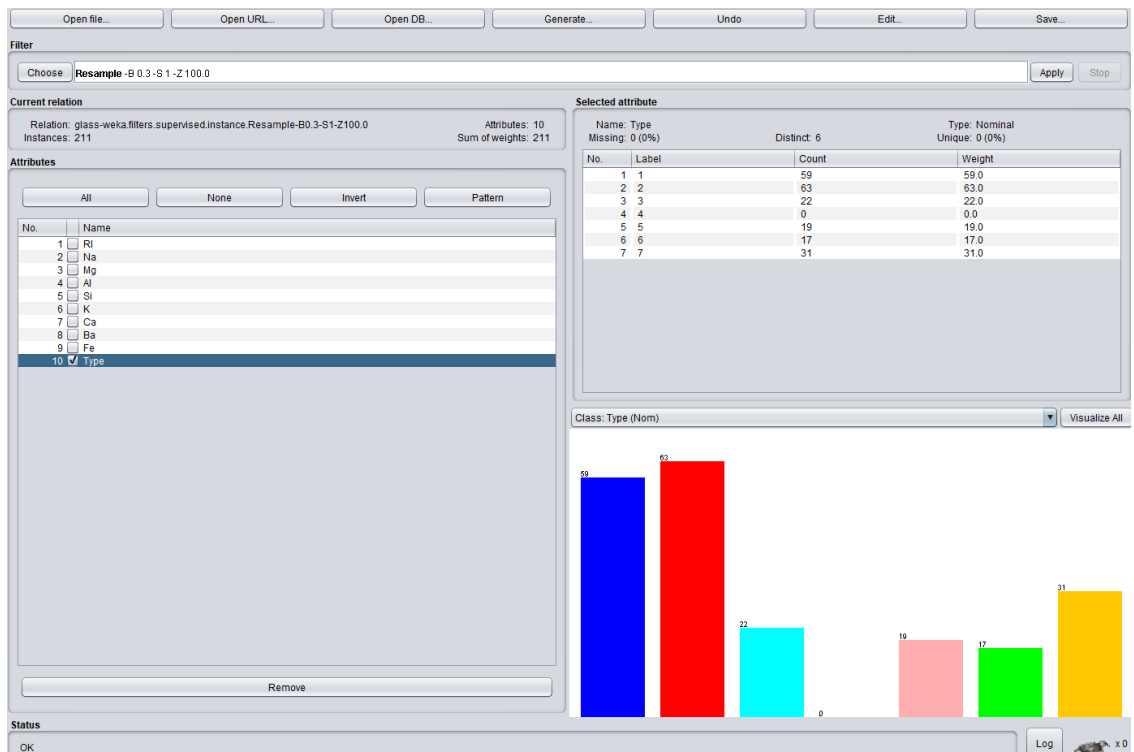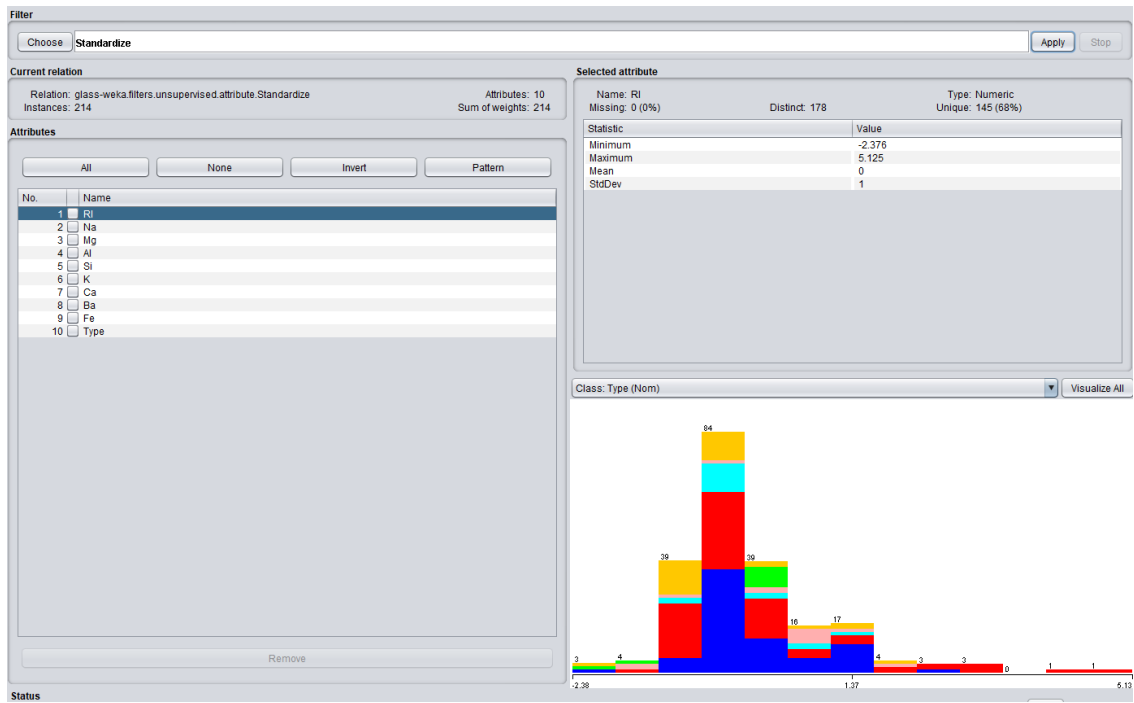
Summary of results

Data	Model	Classification Error
Iris	J48	4%
Iris	JRip	4.7%
Iris	NaiveBayes	4%
Iris	BayesNet	7.3%
Iris	IBK	4.7%
Iris	Multilayer Perceptron	2.7%

Dataset 3: Glass

We need to apply class balancing (resample) and normalize the variables (Standardize)as shown below:

Now applying classification algos with 10-fold cross validation:

1. J48

2. Jrip

```
Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         162               76.7773 %
Incorrectly Classified Instances        49               23.2227 %
Kappa statistic                          0.7016
Mean absolute error                      0.0724
Root mean squared error                  0.2343
Relative absolute error                 32.1592 %
Root relative squared error             69.9299 %
Total Number of Instances              211

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.746	0.118	0.710	0.746	0.727	0.618	0.902	0.728	1
	0.730	0.149	0.676	0.730	0.702	0.569	0.879	0.714	2
	0.591	0.021	0.765	0.591	0.667	0.640	0.842	0.673	3
	?	0.000	?	?	?	?	?	?	4
	0.789	0.010	0.882	0.789	0.833	0.819	0.977	0.794	5
	0.941	0.000	1.000	0.941	0.970	0.968	0.969	0.946	6
	0.903	0.017	0.903	0.903	0.903	0.887	0.980	0.951	7
Weighted Avg.	0.768	0.083	0.773	0.768	0.768	0.692	0.912	0.774	

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 44 13  2  0  0  0  0 |  a = 1
 13 46  0  0  1  0  3 |  b = 2
  4  5 13  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  4  0  0 15  0  0 |  e = 5
  0  0  0  0  1 16  0 |  f = 6
  1  0  2  0  0  0 28 |  g = 7
```

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

Correctly Classified Instances         175               75.1073 %
Incorrectly Classified Instances        58               24.8927 %
Kappa statistic                          0.6786
Mean absolute error                      0.0785
Root mean squared error                  0.2295
Relative absolute error                 39.7238 %
Root relative squared error             73.1361 %
Total Number of Instances              233

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.958	0.106	0.802	0.958	0.873	0.817	0.933	0.807	cp
	0.750	0.088	0.709	0.750	0.729	0.649	0.838	0.630	im
	0.778	0.048	0.795	0.778	0.787	0.736	0.900	0.732	pp
	0.457	0.040	0.667	0.457	0.542	0.490	0.805	0.530	imU
	0.700	0.014	0.824	0.700	0.757	0.739	0.903	0.695	om
	0.400	0.018	0.333	0.400	0.364	0.350	0.695	0.413	omL
	0.000	0.004	0.000	0.000	0.000	-0.006	0.989	0.325	imL
	0.000	0.000	?	0.000	?	?	0.281	0.009	imS
Weighted Avg.	0.751	0.069	?	0.751	?	?	0.873	0.682	

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 69  0  2  0  1  0  0  0 |  a = cp
  3 39  2  7  0  1  0  0 |  b = im
  7  2 35  0  1  0  0  0 |  c = pp
  4 13  1 16  1  0  0  0 |  d = imU
  1  0  4  0 14  1  0  0 |  e = om
  2  0  0  0  0  2  1  0 |  f = omL
  0  0  0  0  0  2  0  0 |  g = imL
  0  1  0  1  0  0  0  0 |  h = imS
```

3. Naïve bayes

4. BayesNet

5. IBk


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

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

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

Correctly Classified Instances       180               85.3081 %
Incorrectly Classified Instances      31               14.6919 %
Kappa statistic                        0.8127
Mean absolute error                    0.0475
Root mean squared error                0.2023
Relative absolute error               21.104  %
Root relative squared error           60.3651 %
Total Number of Instances            211

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                 0.864    0.086    0.797      0.864   0.829      0.760  0.915     0.760     1
                 0.810    0.061    0.850      0.810   0.829      0.760  0.874     0.784     2
                 0.682    0.026    0.750      0.682   0.714      0.684  0.830     0.595     3
                 ?        0.000    ?          ?       ?          ?      ?         ?         4
                 0.842    0.010    0.889      0.842   0.865      0.852  0.884     0.705     5
                 0.941    0.010    0.889      0.941   0.914      0.907  0.935     0.803     6
                 1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     7
Weighted Avg.    0.853    0.047    0.853      0.853   0.852      0.807  0.905     0.784

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 51  5  3  0  0  0  0 |  a = 1
  9 51  2  0  1  0  0 |  b = 2
  4  3 15  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  1  0  0 16  2  0 |  e = 5
  0  0  0  0  1 16  0 |  f = 6
  0  0  0  0  0  0 31 |  g = 7
```

6. MultilayerPerceptron


```
      Input
        Node 5
Class 7
      Input
        Node 6

Time taken to build model: 0.29 seconds

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

Correctly Classified Instances       164               77.7251 %
Incorrectly Classified Instances      47               22.2749 %
Kappa statistic                        0.715
Mean absolute error                    0.0814
Root mean squared error                0.2172
Relative absolute error               36.1272 %
Root relative squared error           64.8177 %
Total Number of Instances            211

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                 0.729    0.138    0.672      0.729   0.699      0.577  0.884     0.735     1
                 0.714    0.122    0.714      0.714   0.714      0.593  0.889     0.719     2
                 0.455    0.037    0.588      0.455   0.513      0.469  0.832     0.474     3
                 ?        0.000    ?          ?       ?          ?      ?         ?         4
                 0.947    0.005    0.947      0.947   0.947      0.942  0.973     0.956     5
                 1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     6
                 1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     7
Weighted Avg.    0.777    0.079    0.775      0.777   0.775      0.699  0.914     0.783

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 43 11  5  0  0  0  0 |  a = 1
 15 45  2  0  1  0  0 |  b = 2
  6  6 10  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  1  0  0 18  0  0 |  e = 5
  0  0  0  0  0 17  0 |  f = 6
  0  0  0  0  0  0 31 |  g = 7
```

Summary of results

Data	Model	Classification Error
Glass	J48	20.3%
Glass	JRip	23.2%
Glass	NaiveBayes	43.6%
Glass	BayesNet	27.5%
Glass	IBK	14.7%
Glass	Multilayer Perceptron	22.3%

Dataset 4: Yeast
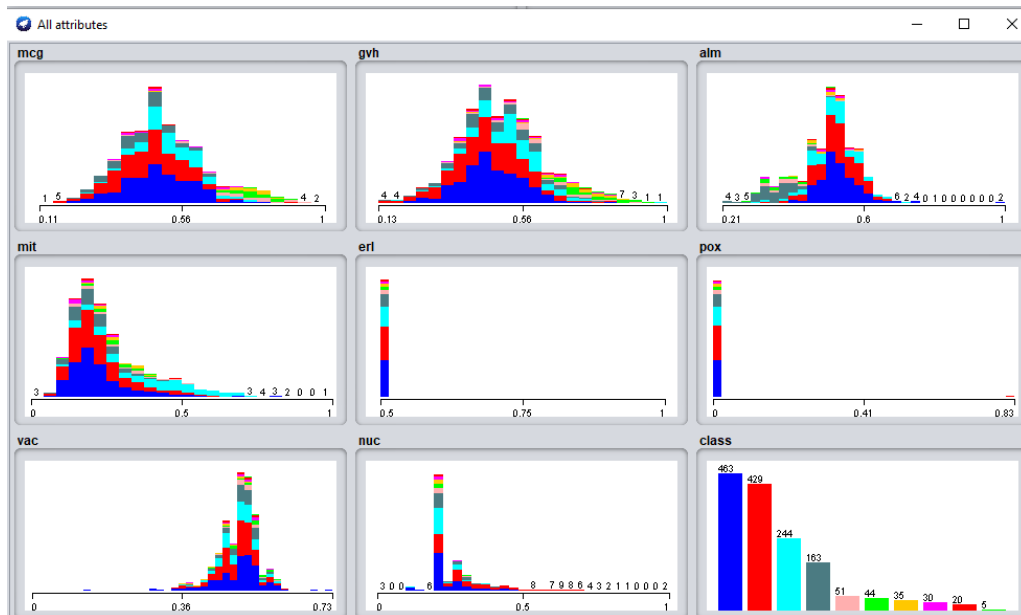
We need to apply following preprocessing.

1. Remove outliers – using interquartile range.
2. Standardize
3. Resample the class

Outlier catches 82 outliers.

We use RemoveWithValues as follows to remove outliers.

Now we have 1402 instances which are not outlier.

Now Standardize and Resample.

Remove the outlier and extreme value attributes using remove option.

Choose | Resample -B 0.3 -S 1 -Z 100.0

Current relation   Sele

Relation: yeast-weka.filters.unsupervised.attribute.InterquartileRange-Rfirst-last-O3.0-E6....   Attributes: 11
Instances: 1401   Sum of weights: 1401   M

Ne

Attributes

| All | None | Invert | Pattern |

No.		Name
1	☐	mcg
2	☐	gvh
3	☐	alm
4	☐	mit
5	☐	erl
6	☐	pox
7	☐	vac
8	☐	nuc
9	☐	class
10	☐	Outlier
11	☑	ExtremeValue

Clas

Remove

Status

Now applying classification algos with 10-fold cross validation.

1. J48


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

Classifier

Choose   J48 -C 0.25 -M 2

Test options                          Classifier output
 Use training set                     Time taken to build model: 0.03 seconds
 Supplied test set    Set..
                                      === Stratified cross-validation ===
 Cross-validation  Folds  10          === Summary ===
 Percentage split    %  66
                                      Correctly Classified Instances      1067            76.1599 %
         More options...              Incorrectly Classified Instances     334            23.8401 %
                                      Kappa statistic                     0.7027
Nom) class                            Mean absolute error                 0.0545
                                      Root mean squared error             0.2047
     Start          Stop              Relative absolute error            33.8095 %
                                      Root relative squared error        72.153  %
esult list (right-click for options)  Total Number of Instances           1401

16:24:36 - trees.J48                  === Detailed Accuracy By Class ===
16:31:12 - trees.J48
16:36:07 - trees.J48                               TP Rate FP Rate Precision Recall F-Measure MCC   ROC Area PRC Area Class
16:36:58 - rules.JRip                              0.782   0.120   0.729    0.782  0.755    0.649  0.869    0.713    CYT
16:37:37 - bayes.NaiveBayes                        0.720   0.082   0.748    0.720  0.734    0.646  0.855    0.688    NUC
16:38:08 - bayes.BayesNet                          0.711   0.055   0.726    0.711  0.719    0.662  0.877    0.651    MIT
16:38:39 - lazy.IBk                                0.829   0.021   0.811    0.829  0.820    0.800  0.934    0.726    ME3
16:39:01 - functions.MultilayerPerceptron          0.692   0.016   0.621    0.692  0.655    0.642  0.911    0.582    ME2
16:43:43 - functions.MultilayerPerceptron          0.854   0.001   0.972    0.854  0.909    0.909  0.949    0.873    ME1
16:44:29 - functions.MultilayerPerceptron          0.784   0.003   0.879    0.784  0.829    0.826  0.912    0.732    EXC
16:44:30 - functions.MultilayerPerceptron          0.053   0.003   0.200    0.053  0.083    0.096  0.688    0.055    VAC
16:44:32 - functions.MultilayerPerceptron          0.934   0.005   0.910    0.934  0.922    0.918  0.992    0.942    POX
16:44:59 - trees.J48                               1.000   0.000   1.000    1.000  1.000    1.000  1.000    1.000    ERL
16:45:18 - rules.JRip                 Weighted Avg. 0.762   0.068   0.758    0.762  0.758    0.693  0.886    0.708
16:45:32 - bayes.NaiveBayes
16:45:47 - bayes.BayesNet             === Confusion Matrix ===
16:46:02 - lazy.IBk
16:46:16 - functions.MultilayerPerceptron   a   b   c   d   e   f   g   h   i   j   <-- classified as
16:58:24 - trees.J48                      320  51  29   2   3   0   0   1   3   0 |  a = CYT
16:59:00 - rules.JRip                      61 255  22   8   5   0   1   1   1   0 |  b = NUC
16:59:22 - bayes.NaiveBayes                37  18 170   8   4   0   0   0   2   0 |  c = MIT
16:59:38 - bayes.BayesNet                  12   8   3 116   1   0   0   0   0   0 |  d = ME3
16:59:56 - lazy.IBk                         0   3   6   4  36   1   1   1   0   0 |  e = ME2
17:00:15 - functions.MultilayerPerceptron   0   0   0   0   3  35   1   1   1   0 |  f = ME1
17:26:12 - trees.J48                        2   1   3   0   2   0  29   0   0   0 |  g = EXC
                                            5   4   1   5   2   0   1   1   0   0 |  h = VAC
                                            2   1   0   0   2   0   0   0  71   0 |  i = POX
                                            0   0   0   0   0   0   0   0   0  34 |  j = ERL

tatus
```

2. Jrip


```
Weka Explorer                                                                                          _

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

Classifier

Choose   JRip -F 3 -N 2.0 -O 2 -S 1

est options                          Classifier output
 Use training set                    Time taken to build model: 0.32 seconds
 Supplied test set    Set..
                                     === Stratified cross-validation ===
 Cross-validation  Folds  10         === Summary ===
 Percentage split    %  66
                                     Correctly Classified Instances       976            69.6645 %
         More options...             Incorrectly Classified Instances     425            30.3355 %
                                     Kappa statistic                     0.6196
(Nom) class                          Mean absolute error                 0.0771
                                     Root mean squared error             0.216
     Start          Stop             Relative absolute error            47.84   %
                                     Root relative squared error        76.1166 %
esult list (right-click for options) Total Number of Instances           1401

16:31:12 - trees.J48                 === Detailed Accuracy By Class ===
16:36:07 - trees.J48
16:36:58 - rules.JRip                              TP Rate FP Rate Precision Recall F-Measure MCC   ROC Area PRC Area Class
16:37:37 - bayes.NaiveBayes                        0.743   0.210   0.594    0.743  0.660    0.504  0.830    0.587    CYT
16:38:08 - bayes.BayesNet                          0.576   0.089   0.687    0.576  0.627    0.518  0.832    0.658    NUC
16:38:39 - lazy.IBk                                0.636   0.048   0.731    0.636  0.680    0.622  0.860    0.612    MIT
16:39:01 - functions.MultilayerPerceptron          0.836   0.026   0.780    0.836  0.807    0.795  0.931    0.828    ME3
16:43:43 - functions.MultilayerPerceptron          0.635   0.011   0.688    0.635  0.660    0.648  0.875    0.575    ME2
16:44:29 - functions.MultilayerPerceptron          0.878   0.005   0.837    0.878  0.857    0.853  0.946    0.755    ME1
16:44:30 - functions.MultilayerPerceptron          0.730   0.003   0.871    0.730  0.794    0.792  0.911    0.715    EXC
16:44:32 - functions.MultilayerPerceptron          0.053   0.002   0.250    0.053  0.087    0.109  0.638    0.047    VAC
16:44:59 - trees.J48                               0.895   0.005   0.919    0.895  0.907    0.901  0.989    0.884    POX
16:45:18 - rules.JRip                              1.000   0.000   1.000    1.000  1.000    1.000  1.000    1.000    ERL
16:45:32 - bayes.NaiveBayes          Weighted Avg. 0.697   0.095   0.700    0.697  0.693    0.607  0.863    0.660
16:45:47 - bayes.BayesNet
16:46:02 - lazy.IBk                  === Confusion Matrix ===
16:46:16 - functions.MultilayerPerceptron
16:58:24 - trees.J48                   a   b   c   d   e   f   g   h   i   j   <-- classified as
16:59:00 - rules.JRip                304  70  28   5   0   0   0   1   1   0 |  a = CYT
16:59:22 - bayes.NaiveBayes          116 204  18   8   2   0   1   1   4   0 |  b = NUC
16:59:38 - bayes.BayesNet             61  12 152   9   4   1   0   0   0   0 |  c = MIT
16:59:56 - lazy.IBk                    9   6   2 117   6   0   0   0   0   0 |  d = ME3
17:00:15 - functions.MultilayerPerceptron  5   1   3   6  33   2   2   0   0   0 |  e = ME2
17:26:12 - trees.J48                   0   0   0   1   2  36   1   1   0   0 |  f = ME1
                                       5   2   1   0   0   1  27   0   1   0 |  g = EXC
                                       7   1   3   4   0   3   0   1   0   0 |  h = VAC
                                       5   1   1   0   1   0   0   0  68   0 |  i = POX
                                       0   0   0   0   0   0   0   0   0  34 |  j = ERL
```

3. Naïve bayes


```
Classifier

Choose    NaiveBayes

Test options                              Classifier output
○ Use training set                        Time taken to build model: 0 seconds
○ Supplied test set      Set..
● Cross-validation  Folds  10             === Stratified cross-validation ===
○ Percentage split    %   66              === Summary ===

          More options...                 Correctly Classified Instances         805              57.459 %
                                          Incorrectly Classified Instances       596              42.541 %
(Nom) class                               Kappa statistic                          0.4702
                                          Mean absolute error                      0.0989
      Start            Stop               Root mean squared error                  0.2423
                                          Relative absolute error                 61.3483 %
Result list (right-click for options)     Root relative squared error             85.4015 %
                                          Total Number of Instances              1401
16:36:07 - trees.J48
16:36:58 - rules.JRip                     === Detailed Accuracy By Class ===
16:37:37 - bayes.NaiveBayes
16:38:08 - bayes.BayesNet                              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC Area  Class
16:38:39 - lazy.IBk                                    0.687    0.272    0.510      0.687   0.585      0.386   0.767     0.531     CYT
16:39:01 - functions.MultilayerPerceptron              0.305    0.078    0.568      0.305   0.397      0.288   0.805     0.537     NUC
16:43:43 - functions.MultilayerPerceptron              0.640    0.094    0.584      0.640   0.611      0.527   0.852     0.629     MIT
16:44:29 - functions.MultilayerPerceptron              0.707    0.022    0.780      0.707   0.742      0.715   0.981     0.801     ME3
16:44:30 - functions.MultilayerPerceptron              0.423    0.027    0.379      0.423   0.400      0.376   0.908     0.416     ME2
16:44:32 - functions.MultilayerPerceptron              0.829    0.013    0.667      0.829   0.739      0.735   0.989     0.679     ME1
16:44:59 - trees.J48                                   0.703    0.024    0.441      0.703   0.542      0.542   0.960     0.685     EXC
16:45:18 - rules.JRip                                  0.000    0.002    0.000      0.000   0.000     -0.005   0.679     0.025     VAC
16:45:32 - bayes.NaiveBayes                            0.684    0.014    0.743      0.684   0.712      0.697   0.959     0.847     POX
16:45:47 - bayes.BayesNet                              0.882    0.000    1.000      0.882   0.938      0.938   1.000     1.000     ERL
16:46:02 - lazy.IBk               Weighted Avg.        0.575    0.120    0.580      0.575   0.562      0.457   0.844     0.602
16:46:16 - functions.MultilayerPerceptron
16:58:24 - trees.J48                      === Confusion Matrix ===
16:59:00 - rules.JRip
16:59:22 - bayes.NaiveBayes                 a   b   c   d   e   f   g   h   i   j   <-- classified as
16:59:38 - bayes.BayesNet                 281  57  51   1   3   0  10   0   6   0 |   a = CYT
16:59:56 - lazy.IBk                       192 108  30  13   5   0   6   0   0   0 |   b = NUC
17:00:15 - functions.MultilayerPerceptron  42  11 153   6   4   2   9   0  12   0 |   c = MIT
17:26:12 - trees.J48                       12   8   2  99  16   0   0   3   0   0 |   d = ME3
17:26:34 - rules.JRip                      11   0   3   3  22  12   1   0   0   0 |   e = ME2
                                            0   0   0   0   4  34   3   0   0   0 |   f = ME1
                                            1   4   3   0   0   3  26   0   0   0 |   g = EXC
                                            8   2   2   5   0   0   2   0   0   0 |   h = VAC
                                            4   0  18   0   0   0   2   0  52   0 |   i = POX
                                            0   0   0   0   4   0   0   0   0  30 |   j = ERL
```

4. BayesNet


```
Classifier

Choose    BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Test options                              Classifier output
○ Use training set                        Time taken to build model: 0.02 seconds
○ Supplied test set      Set..
● Cross-validation  Folds  10             === Stratified cross-validation ===
○ Percentage split    %   66              === Summary ===

          More options...                 Correctly Classified Instances         887              63.3119 %
                                          Incorrectly Classified Instances       514              36.6881 %
(Nom) class                               Kappa statistic                          0.5425
                                          Mean absolute error                      0.0946
      Start            Stop               Root mean squared error                  0.2194
                                          Relative absolute error                 58.6975 %
Result list (right-click for options)     Root relative squared error             77.3205 %
                                          Total Number of Instances              1401

                                          === Detailed Accuracy By Class ===

                                                       TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC Area  Class
                                                       0.577    0.205    0.538      0.577   0.557      0.365   0.802     0.611     CYT
                                                       0.545    0.141    0.566      0.545   0.555      0.409   0.820     0.566     NUC
                                                       0.619    0.071    0.643      0.619   0.631      0.557   0.877     0.671     MIT
                                                       0.857    0.024    0.800      0.857   0.828      0.808   0.982     0.843     ME3
                                                       0.481    0.010    0.641      0.481   0.549      0.541   0.937     0.628     ME2
                                                       0.854    0.007    0.778      0.854   0.814      0.809   0.992     0.891     ME1
                                                       0.784    0.007    0.744      0.784   0.763      0.757   0.977     0.715     EXC
                                                       0.053    0.004    0.167      0.053   0.080      0.087   0.674     0.090     VAC
                                                       0.868    0.009    0.846      0.868   0.857      0.849   0.968     0.906     POX
                                                       1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     ERL
                                          Weighted Avg. 0.633   0.111    0.628      0.633   0.629      0.521   0.865     0.663

                                          === Confusion Matrix ===

                                            a   b   c   d   e   f   g   h   i   j   <-- classified as
                                          236 121  44   0   2   0   1   2   3   0 |   a = CYT
                                          113 193  28  15   1   0   2   2   0   0 |   b = NUC
                                           51  15 148   8   4   1   2   1   9   0 |   c = MIT
                                           12   5   1 120   2   0   0   0   0   0 |   d = ME3
                                            7   4   4   2  25   8   2   0   0   0 |   e = ME2
                                            1   0   0   0   3  35   2   0   0   0 |   f = ME1
                                            5   1   2   0   0   0  29   0   0   0 |   g = EXC
                                            7   2   2   5   0   1   1   1   0   0 |   h = VAC
                                            7   0   1   0   2   0   0   0  66   0 |   i = POX
                                            0   0   0   0   0   0   0   0   0  34 |   j = ERL
```

5. IBk

6. MultilayerPerceptron

Summary of results

Data	Model	Classification Error
Yeast	J48	23.9%
Yeast	JRip	30.3%
Yeast	NaiveBayes	42.5%
Yeast	BayesNet	36.7%
Yeast	IBK	17.5%
Yeast	Multilayer Perceptron	34.4%

Dataset 5: Gait

Note that the class variable has 8 values.

Now applying classification algos with 10-fold cross validation.

1. J48

```
Time taken to build model: 0.08 seconds

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

Correctly Classified Instances          19               39.5833 %
Incorrectly Classified Instances        29               60.4167 %
Kappa statistic                          0.3556
Mean absolute error                      0.0755
Root mean squared error                  0.2748
Relative absolute error                 63.6438 %
Root relative squared error            112.0714 %
Total Number of Instances               48

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.044	0.000	0.000	0.000	-0.054	0.478	0.063	0
	0.667	0.022	0.667	0.667	0.667	0.644	0.822	0.465	1
	0.667	0.067	0.400	0.667	0.500	0.475	0.800	0.288	2
	0.333	0.044	0.333	0.333	0.333	0.289	0.644	0.153	3
	0.667	0.044	0.500	0.667	0.571	0.545	0.811	0.354	4
	0.000	0.067	0.000	0.000	0.000	-0.067	0.467	0.063	5
	0.333	0.044	0.333	0.333	0.333	0.289	0.644	0.153	6
	0.000	0.044	0.000	0.000	0.000	-0.054	0.478	0.063	7
	0.000	0.067	0.000	0.000	0.000	-0.067	0.467	0.063	8
	0.667	0.022	0.667	0.667	0.667	0.644	0.822	0.465	9
	0.000	0.044	0.000	0.000	0.000	-0.054	0.478	0.063	10
	0.667	0.000	1.000	0.667	0.800	0.808	0.833	0.688	11
	0.667	0.000	1.000	0.667	0.800	0.808	0.833	0.688	12
	0.333	0.089	0.200	0.333	0.250	0.194	0.622	0.108	13
	0.667	0.044	0.500	0.667	0.571	0.545	0.811	0.354	14
	0.667	0.000	1.000	0.667	0.800	0.808	0.833	0.688	15
Weighted Avg.	0.396	0.040	0.412	0.396	0.393	0.360	0.678	0.295	

```
=== Confusion Matrix ===
```

2. Jrip

```
=== Summary ===

Correctly Classified Instances          20               41.6667 %
Incorrectly Classified Instances        28               58.3333 %
Kappa statistic                          0.3778
Mean absolute error                      0.0698
Root mean squared error                  0.2224
Relative absolute error                 58.792 %
Root relative squared error             90.6831 %
Total Number of Instances               48

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.667	0.200	0.182	0.667	0.286	0.269	0.770	0.354	0
	0.333	0.067	0.250	0.333	0.286	0.234	0.781	0.283	1
	0.000	0.089	0.000	0.000	0.000	-0.078	0.459	0.063	2
	0.000	0.089	0.000	0.000	0.000	-0.078	0.400	0.063	3
	0.667	0.022	0.667	0.667	0.667	0.644	0.804	0.465	4
	0.667	0.022	0.667	0.667	0.667	0.644	0.981	0.644	5
	0.667	0.000	1.000	0.667	0.800	0.808	0.826	0.688	6
	0.333	0.044	0.333	0.333	0.333	0.289	0.607	0.375	7
	0.667	0.000	1.000	0.667	0.800	0.808	0.978	0.792	8
	0.667	0.000	1.000	0.667	0.800	0.808	1.000	1.000	9
	0.000	0.022	0.000	0.000	0.000	-0.038	0.593	0.125	10
	0.667	0.044	0.500	0.667	0.571	0.545	0.826	0.688	11
	0.000	0.000	?	0.000	?	?	0.570	0.079	12
	0.667	0.000	1.000	0.667	0.800	0.808	0.796	0.688	13
	0.333	0.000	1.000	0.333	0.500	0.565	0.644	0.375	14
	0.333	0.022	0.500	0.333	0.400	0.377	0.915	0.345	15
Weighted Avg.	0.417	0.039	?	0.417	?	?	0.747	0.439	

```
=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p   <-- classified as
 2 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 | a = 0
 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 | b = 1
 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 | c = 2
```

3. Naïve bayes

```
Time taken to build model: 0.01 seconds

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

Correctly Classified Instances           9               18.75   %
Incorrectly Classified Instances        39               81.25   %
Kappa statistic                          0.1333
Mean absolute error                      0.1016
Root mean squared error                  0.3187
Relative absolute error                 85.5899 %
Root relative squared error            129.9655 %
Total Number of Instances               48

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
              0.000    0.133    0.000      0.000   0.000      -0.098   0.433     0.063     0
              0.000    0.000    ?          0.000   ?          ?        0.500     0.063     1
              0.333    0.022    0.500      0.333   0.400      0.377    0.656     0.208     2
              0.333    0.511    0.042      0.333   0.074      -0.086   0.411     0.056     3
              0.333    0.000    1.000      0.333   0.500      0.565    0.667     0.375     4
              0.333    0.044    0.333      0.333   0.333      0.289    0.644     0.153     5
              0.000    0.000    ?          0.000   ?          ?        0.500     0.063     6
              0.333    0.022    0.500      0.333   0.400      0.377    0.656     0.208     7
              0.000    0.044    0.000      0.000   0.000      -0.054   0.478     0.063     8
              0.333    0.000    1.000      0.333   0.500      0.565    0.667     0.375     9
              0.000    0.089    0.000      0.000   0.000      -0.078   0.444     0.063     10
              0.000    0.000    ?          0.000   ?          ?        0.500     0.063     11
              0.000    0.000    ?          0.000   ?          ?        0.500     0.063     12
              0.333    0.000    1.000      0.333   0.500      0.565    0.667     0.375     13
              0.000    0.000    ?          0.000   ?          ?        0.500     0.063     14
              0.667    0.000    1.000      0.667   0.800      0.808    0.833     0.688     15
Weighted Avg. 0.188    0.054    ?          0.188   ?          ?        0.566     0.184

=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p   <-- classified as
 0 0 0 2 0 0 0 1 0 0 0 0 0 0 0 0 | a = 0
 0 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 | b = 1
 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 | c = 2
 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 | d = 3
 0 0 0 2 1 0 0 0 0 0 0 0 0 0 0 0 | e = 4
 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 | f = 5
 0 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 | g = 6
```

4. BayesNet

```
=== Summary ===

Correctly Classified Instances          47               97.9167 %
Incorrectly Classified Instances         1                2.0833 %
Kappa statistic                          0.9778
Mean absolute error                      0.0026
Root mean squared error                  0.051
Relative absolute error                  2.2012 %
Root relative squared error             20.8112 %
Total Number of Instances               48

=== 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     0
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     1
              1.000    0.022    0.750      1.000   0.857      0.856   1.000     1.000     2
              0.667    0.000    1.000      0.667   0.800      0.808   0.993     0.917     3
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     4
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     5
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     6
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     7
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     8
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     9
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     10
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     11
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     12
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     13
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     14
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     15
Weighted Avg. 0.979    0.001    0.984      0.979   0.979      0.979   1.000     0.995

=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p   <-- classified as
 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | a = 0
 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = 1
 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 | c = 2
 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 | d = 3
 0 0 0 0 3 0 0 0 0 0 0 0 0 0 0 0 | e = 4
 0 0 0 0 0 3 0 0 0 0 0 0 0 0 0 0 | f = 5
 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 0 | g = 6
 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 | h = 7
 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 | i = 8
 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0 | j = 9
 0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 | k = 10
```

Classifier output

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

m) attribute_321

Start   Stop

lt list (right-click for options)

:40:06 - rules.ZeroR
:40:16 - trees.J48
:40:36 - rules.JRip
:41:16 - bayes.NaiveBayes
:41:29 - bayes.BayesNet
:41:41 - lazy.IBk
:41:51 - functions.MultilayerPerceptron

5. IBk

st options

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

(lom) attribute_321 ▼

Start	Stop

sult list (right-click for options)

17:40:06 - rules.ZeroR
17:40:16 - trees.J48
17:40:36 - rules.JRip
17:41:16 - bayes.NaiveBayes
17:41:29 - bayes.BayesNet
17:41:41 - lazy.IBk
17:41:51 - functions.MultilayerPerceptron

Classifier output

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

Correctly Classified Instances         41               85.4167 %
Incorrectly Classified Instances        7               14.5833 %
Kappa statistic                          0.8444
Mean absolute error                      0.045
Root mean squared error                  0.1326
Relative absolute error                 37.9036 %
Root relative squared error             54.058  %
Total Number of Instances               48

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.667	0.000	1.000	0.667	0.800	0.808	0.859	0.691	0
	0.667	0.000	1.000	0.667	0.800	0.808	0.863	0.692	1
	1.000	0.067	0.500	1.000	0.667	0.683	0.967	0.500	2
	0.667	0.022	0.667	0.667	0.667	0.644	0.852	0.469	3
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	4
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	5
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	6
	1.000	0.022	0.750	1.000	0.857	0.856	0.989	0.750	7
	0.333	0.000	1.000	0.333	0.500	0.565	0.726	0.383	8
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	9
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	10
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	11
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	12
	0.667	0.000	1.000	0.667	0.800	0.808	0.863	0.692	13
	0.667	0.022	0.667	0.667	0.667	0.644	0.852	0.580	14
	1.000	0.022	0.750	1.000	0.857	0.856	0.989	0.750	15
Weighted Avg.	0.854	0.010	0.896	0.854	0.851	0.855	0.935	0.782	

```
=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p   <-- classified as
 2 0 0 0 0 0 0 0 0 0 0 0 0 1 0 | a = 0
 0 2 0 0 0 0 1 0 0 0 0 0 0 0 0 | b = 1
 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 | c = 2
 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 | d = 3
 0 0 0 0 3 0 0 0 0 0 0 0 0 0 0 | e = 4
 0 0 0 0 0 3 0 0 0 0 0 0 0 0 0 | f = 5
 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 | g = 6
 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 | h = 7
 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0 | i = 8
 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 | j = 9
```

atus

6. MultilayerPerceptron

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

Correctly Classified Instances         34               70.8333 %
Incorrectly Classified Instances        14               29.1667 %
Kappa statistic                          0.6889
Mean absolute error                      0.0466
Root mean squared error                  0.1563
Relative absolute error                 39.2662 %
Root relative squared error             63.7559 %
Total Number of Instances               48

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.667	0.022	0.667	0.667	0.667	0.644	0.993	0.917	0
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1
	1.000	0.044	0.600	1.000	0.750	0.757	1.000	1.000	2
	0.333	0.000	1.000	0.333	0.500	0.565	0.926	0.744	3
	0.667	0.000	1.000	0.667	0.800	0.808	0.889	0.722	4
	0.667	0.000	1.000	0.667	0.800	0.808	0.993	0.917	5
	0.333	0.000	1.000	0.333	0.500	0.565	0.896	0.618	6
	0.667	0.000	1.000	0.667	0.800	0.808	1.000	1.000	7
	0.333	0.067	0.250	0.333	0.286	0.234	0.948	0.610	8
	0.333	0.022	0.500	0.333	0.400	0.377	0.911	0.577	9
	1.000	0.044	0.600	1.000	0.750	0.757	0.956	0.478	10
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	11
	0.667	0.000	1.000	0.667	0.800	0.808	0.933	0.750	12
	0.667	0.022	0.667	0.667	0.667	0.644	0.970	0.810	13
	1.000	0.089	0.429	1.000	0.600	0.625	1.000	1.000	14
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	15
Weighted Avg.	0.708	0.019	0.794	0.708	0.707	0.712	0.963	0.821	

```
=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p   <-- classified as
 2 0 0 0 0 0 0 0 0 0 0 0 0 1 0 | a = 0
 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = 1
 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 | c = 2
 0 0 0 1 0 0 0 0 2 0 0 0 0 0 0 | d = 3
 1 0 0 0 2 0 0 0 0 0 0 0 0 0 0 | e = 4
 0 0 0 0 0 2 0 0 0 0 0 0 0 1 0 | f = 5
 0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 | g = 6
```

Summary of results

Data	Model	Classification Error
Gait	J48	60.4%
Gait	JRip	58.3%
Gait	NaiveBayes	81.2%
Gait	BayesNet	2.1%
Gait	IBK	14.6%
Gait	Multilayer Perceptron	29.2.%

All Results Summary

Data	Model	Classification Error
Ecoli	J48	21.90%
Ecoli	JRip	24.90%
Ecoli	NaiveBayes	16.30%
Ecoli	BayesNet	21.90%
Ecoli	IBK	24%
Ecoli	Multilayer Perceptron	18.40%
Iris	J48	4%
Iris	JRip	4.70%
Iris	NaiveBayes	4%
Iris	BayesNet	7.30%
Iris	IBK	4.70%
Iris	Multilayer Perceptron	2.70%
Glass	J48	20.30%
Glass	JRip	23.20%
Glass	NaiveBayes	43.60%

Glass	BayesNet	27.50%
Glass	IBK	14.70%
Glass	Multilayer Perceptron	22.30%
Yeast	J48	23.90%
Yeast	JRip	30.30%
Yeast	NaiveBayes	42.50%
Yeast	BayesNet	36.70%
Yeast	IBK	17.50%
Yeast	Multilayer Perceptron	34.40%
Gait	J48	60.40%
Gait	JRip	58.30%
Gait	NaiveBayes	81.20%
Gait	BayesNet	2.10%
Gait	IBK	14.60%
Gait	Multilayer Perceptron	29.2.%

1. **Cleaning**

The following cleaning methods have been used.
Resampling – Observed that there was class imbalance in most of the datasets. Therefore, Weka's Resample filter is used to reduce the impact of different number of classes.
Outlier removal - Used the Interquartile Range filter to identify and remove 82 outliers in yeast data.
Standardization – Used to standardize filter to normalize features to have 0 mean and 1 standard deviation.

2. **Method Utility**

a. Classification-error rate has been reported on the above dataset. Following are the average classification error rates:

Model	Average Classification Error
BayesNet	19.1
IBK	15.1
J48	26.1
JRip	28.28
Multilayer Perceptron	21.4
NaiveBayes	37.52

b. Ranked

Model	Average Classification Error
IBK	15.1
BayesNet	19.1
Multilayer Perceptron	21.4
J48	26.1
JRip	28.28
NaiveBayes	37.52

b. i. The error rate has an average difference of 4% for two methods. It is not significantly difficult but might vary from data to data. What we observe is that each model is best on a separate dataset/problem. For each problem, we need to find the best model on it.
b. ii. There is a significant difference between methods with lowest and highest error rate. Interesting to note that each model still is specialized on their own dataset. Though on an average IBK is better model. Also, it is important to note, a lot of datasets were imbalanced and hence models behave differently.

3. **Dataset Differences**

a. Classification error varies very significantly across datasets. This is based on how the quality of features, the outliers in data, the class imbalance, the correlation of input with target, the number of attributes and most importantly number of instances (size of data). To give a real-world example, an 80% accurate recommendation engine for amazon is good enough. However, an 80% accurate cancer detection model is very poor and risky.

b. Yes, each model is specialized on certain problems. Also model generally need tuning. Multilayer perceptron is specialist for large size data, Decision tree are specialized when there are large number of categorical values with good split points.

Accuracy Improvement

The accuracy for Iris dataset has been improved using Jrip. We reduced error from 4.7% to 2.7% by setting value of optimizations parameter to 5. As the algorithm will get chance to perform more optimizations, we expected the accuracy increase.