Adunife Kizito Okoye
100611918

Task 1(Lab 06)

• C4.5 (weka.classifier.trees.J48)

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

Correctly Classified Instances         114               77.027  %
Incorrectly Classified Instances        34               22.973  %
Kappa statistic                          0.5736
Mean absolute error                      0.1304
Root mean squared error                  0.3151
Relative absolute error                 48.619  %
Root relative squared error             86.5138 %
Total Number of Instances              148

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 1.000    0.014    0.500      1.000    0.667      0.702
0.991     0.500     normal
                 0.790    0.194    0.831      0.790    0.810      0.594
0.788     0.737     metastases
                 0.754    0.195    0.730      0.754    0.742      0.556
0.777     0.718     malign_lymph
                 0.500    0.014    0.500      0.500    0.500      0.486
0.744     0.389     fibrosis
Weighted Avg.    0.770    0.187    0.776      0.770    0.772      0.577
0.785     0.717

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  2  0  0  0 |  a = normal
  1 64 15  1 |  b = metastases
  1 13 46  1 |  c = malign_lymph
  0  0  2  2 |  d = fibrosis
```

• RIPPER (weka.classifier.rules.JRip)

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

Correctly Classified Instances         115               77.7027 %
Incorrectly Classified Instances        33               22.2973 %
Kappa statistic                          0.5725
Mean absolute error                      0.1414
Root mean squared error                  0.3108
Relative absolute error                 52.7427 %
Root relative squared error             85.3428 %
Total Number of Instances              148
```

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

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.000    0.000    ?          0.000    ?          ?
0.687     0.038     normal
                 0.827    0.254    0.798      0.827    0.812      0.576
0.805     0.808     metastases
                 0.738    0.172    0.750      0.738    0.744      0.567
0.780     0.715     malign_lymph
                 0.750    0.007    0.750      0.750    0.750      0.743
0.872     0.694     fibrosis
Weighted Avg.    0.777    0.210    ?          0.777    ?          ?
0.795     0.756

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  0  1  1  0 |  a = normal
  0 67 14  0 |  b = metastases
  0 15 45  1 |  c = malign_lymph
  0  1  0  3 |  d = fibrosis
```

Paragraph description for how both of the algorithms work.

C4.5, to put it simple it uses many different premises of information or data entropy to build decision trees from a set of training data in the same way as ID3. RIPPER(JRip) viewed as an algorithm that learns what a malicious executable is in a given set of training examples, since the algorithm is inductive by nature.

Task 2(Lab 06)

monks-3.test

• C4.5 (weka.classifier.trees.J48)

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

Correctly Classified Instances         432               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances              432

=== 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
```

```
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000
1.000     1.000

=== Confusion Matrix ===

   a   b   <-- classified as
 204   0 |   a = 0
   0 228 |   b = 1
```

• Id3 (weka.classifier.trees.Id3)

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

Correctly Classified Instances         432               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances              432

=== 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
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000
1.000     1.000

=== Confusion Matrix ===

   a   b   <-- classified as
 204   0 |   a = 0
   0 228 |   b = 1
```

• RIPPER (weka.classifier.rules.JRip)

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

Correctly Classified Instances         426               98.6111 %
Incorrectly Classified Instances         6                1.3889 %
Kappa statistic                          0.9722
Mean absolute error                      0.02
Root mean squared error                  0.112
Relative absolute error                  4.0105 %
Root relative squared error             22.4267 %
Total Number of Instances              432

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

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 1.000    0.026    0.971      1.000    0.986      0.973
0.997     0.996     0
                 0.974    0.000    1.000      0.974    0.987      0.973
0.997     0.997     1
Weighted Avg.    0.986    0.012    0.987      0.986    0.986      0.973
0.997     0.996

=== Confusion Matrix ===

   a   b   <-- classified as
 204   0 |   a = 0
   6 222 |   b = 1
```

• k-Nearest Neighbor (weka.classifiers.lazy.IBk)

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

Correctly Classified Instances         427               98.8426 %
Incorrectly Classified Instances         5                1.1574 %
Kappa statistic                          0.9768
Mean absolute error                      0.1957
Root mean squared error                  0.2172
Relative absolute error                 39.2533 %
Root relative squared error             43.4986 %
Total Number of Instances              432

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 1.000    0.022    0.976      1.000    0.988      0.977
1.000     1.000     0
                 0.978    0.000    1.000      0.978    0.989      0.977
1.000     1.000     1
Weighted Avg.    0.988    0.010    0.989      0.988    0.988      0.977
1.000     1.000

=== Confusion Matrix ===

   a   b   <-- classified as
 204   0 |   a = 0
   5 223 |   b = 1
```

• Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes)

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

Correctly Classified Instances         420               97.2222 %
Incorrectly Classified Instances        12                2.7778 %
Kappa statistic                          0.9444
Mean absolute error                      0.1381
Root mean squared error                  0.1859
Relative absolute error                 27.7123 %
Root relative squared error             37.2363 %
```

```
Total Number of Instances              432

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 1.000    0.053    0.944      1.000    0.971      0.946
0.998     0.997     0
                 0.947    0.000    1.000      0.947    0.973      0.946
0.998     0.998     1
Weighted Avg.    0.972    0.025    0.974      0.972    0.972      0.946
0.998     0.998

=== Confusion Matrix ===

   a   b   <-- classified as
 204   0 |   a = 0
  12 216 |   b = 1
```

• Neural Networks (weka.classifiers.functions.MultilayerPerceptron)

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

Correctly Classified Instances         432               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.0021
Root mean squared error                  0.0033
Relative absolute error                  0.4127 %
Root relative squared error              0.6643 %
Total Number of Instances              432

=== 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
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000
1.000     1.000

=== Confusion Matrix ===

   a   b   <-- classified as
 204   0 |   a = 0
   0 228 |   b = 1
```

monks-3-train

• C4.5 (weka.classifier.trees.J48)

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

```
Correctly Classified Instances         114               93.4426 %
Incorrectly Classified Instances         8                6.5574 %
Kappa statistic                          0.8689
Mean absolute error                      0.12
Root mean squared error                  0.2577
Relative absolute error                 24.0001 %
Root relative squared error             51.531  %
Total Number of Instances              122

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.919    0.050    0.950      0.919    0.934      0.869
0.936     0.933     0
                 0.950    0.081    0.919      0.950    0.934      0.869
0.936     0.915     1
Weighted Avg.    0.934    0.065    0.935      0.934    0.934      0.869
0.936     0.924

=== Confusion Matrix ===

  a  b   <-- classified as
 57  5 |  a = 0
  3 57 |  b = 1
```

• Id3 (weka.classifier.trees.Id3)

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

Correctly Classified Instances         105               86.0656 %
Incorrectly Classified Instances         9                7.377  %
Kappa statistic                          0.8415
Mean absolute error                      0.0789
Root mean squared error                  0.281
Relative absolute error                 16.9135 %
Root relative squared error             58.1805 %
UnClassified Instances                   8                6.5574 %
Total Number of Instances              122

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.918    0.075    0.933      0.918    0.926      0.842
0.918     0.892     0
                 0.925    0.082    0.907      0.925    0.916      0.842
0.868     0.831     1
Weighted Avg.    0.921    0.078    0.921      0.921    0.921      0.842
0.895     0.864

=== Confusion Matrix ===

  a  b   <-- classified as
```

```
 56  5 |  a = 0
  4 49 |  b = 1
```

• RIPPER (weka.classifier.rules.JRip)

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

Correctly Classified Instances         106               86.8852 %
Incorrectly Classified Instances        16               13.1148 %
Kappa statistic                          0.7382
Mean absolute error                      0.1806
Root mean squared error                  0.3277
Relative absolute error                 36.1236 %
Root relative squared error             65.5376 %
Total Number of Instances              122

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.806    0.067    0.926      0.806    0.862      0.745
0.902     0.880     0
                 0.933    0.194    0.824      0.933    0.875      0.745
0.902     0.891     1
Weighted Avg.    0.869    0.129    0.876      0.869    0.868      0.745
0.902     0.885

=== Confusion Matrix ===

  a  b   <-- classified as
 50 12 |  a = 0
  4 56 |  b = 1
```

• k-Nearest Neighbor (weka.classifiers.lazy.IBk)

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

Correctly Classified Instances          97               79.5082 %
Incorrectly Classified Instances        25               20.4918 %
Kappa statistic                          0.5891
Mean absolute error                      0.2368
Root mean squared error                  0.3826
Relative absolute error                 47.3616 %
Root relative squared error             76.5254 %
Total Number of Instances              122

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.871    0.283    0.761      0.871    0.812      0.596
0.876     0.864     0
                 0.717    0.129    0.843      0.717    0.775      0.596
0.876     0.854     1
Weighted Avg.    0.795    0.207    0.801      0.795    0.794      0.596
0.876     0.859
```

```
=== Confusion Matrix ===

  a  b   <-- classified as
 54  8 |  a = 0
 17 43 |  b = 1
```

• Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes)

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

Correctly Classified Instances         114               93.4426 %
Incorrectly Classified Instances         8                6.5574 %
Kappa statistic                          0.8689
Mean absolute error                      0.2192
Root mean squared error                  0.2895
Relative absolute error                 43.8555 %
Root relative squared error             57.9011 %
Total Number of Instances              122

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.919    0.050    0.950      0.919    0.934      0.869
0.911     0.879     0
                 0.950    0.081    0.919      0.950    0.934      0.869
0.911     0.895     1
Weighted Avg.    0.934    0.065    0.935      0.934    0.934      0.869
0.911     0.887

=== Confusion Matrix ===

  a  b   <-- classified as
 57  5 |  a = 0
  3 57 |  b = 1
```

• Neural Networks (weka.classifiers.functions.MultilayerPerceptron)

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

Correctly Classified Instances         107               87.7049 %
Incorrectly Classified Instances        15               12.2951 %
Kappa statistic                          0.7537
Mean absolute error                      0.1271
Root mean squared error                  0.3224
Relative absolute error                 25.4232 %
Root relative squared error             64.4867 %
Total Number of Instances              122

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.919    0.167    0.851      0.919    0.884      0.756
0.912     0.859     0
```

```
                 0.833    0.081    0.909      0.833    0.870      0.756
0.912     0.909     1
Weighted Avg.    0.877    0.124    0.879      0.877    0.877      0.756
0.912     0.883

=== Confusion Matrix ===

  a  b   <-- classified as
 57  5 |  a = 0
 10 50 |  b = 1
```

Paragraph description for how the algorithms work.

- C4.5 (weka.classifier.trees.J48) – The C4.5 algorithm works by checking for the bases case. Then it loops each attribute *x* while finding the normalized information gain from splitting on *x.* After which let there be a best *x* attribute which is the highest normalized information gain. Then Create a decision node that splits on the best *x.* Recurse it on the sub-lists obtained by splitting on the best *x*, and adding those nodes as children of node.
- Id3 (weka.classifier.trees.Id3) – The Id3 algorithm works by recursing through each subset considering only attributes never selected before.
- RIPPER (weka.classifier.rules.JRip) – The JRip algorithm works by using propositional learning, then repeated incremental running in order to produce error reduction. During the grow phase it uses p(log(p/t)-log(P/T) to find the value of attribute with highest information gain the it moves to the pruning and optimization stages.
- k-Nearest Neighbor (weka.classifiers.lazy.IBk) – The IBk algorithm works by predicting the class of the single nearest training instance for each test instance.
- Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes) – The NaiveBayes uses estimator classes, hence numeric estimator precision values are chosen based on analysis of the training data
- Neural Networks (weka.classifiers.functions.MultilayerPerceptron) – The Neural Networks being an acyclic network. Backpropagation is used to classify instances. In this classifier networks can be created via hand, created by an algorithm or both. Hence enabling the network to be not only monitored but as well modified during training time.

Task 3(Lab 07)

• C4.5 (weka.classifier.trees.J48)

```
=== Summary ===

Correctly Classified Instances         178               85.9903 %
Incorrectly Classified Instances        29               14.0097 %
Kappa statistic                          0.7168
Mean absolute error                      0.1958
Root mean squared error                  0.3288
Relative absolute error                 39.4502 %
Root relative squared error             65.6306 %
Total Number of Instances              207

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
```

```
                 0.776    0.064    0.916      0.776    0.840      0.725
0.901     0.857     +
                 0.936    0.224    0.823      0.936    0.876      0.725
0.901     0.872     -
Weighted Avg.    0.860    0.149    0.867      0.860    0.859      0.725
0.901     0.865

=== Confusion Matrix ===

   a   b   <-- classified as
  76  22 |   a = +
   7 102 |   b = -
```

• Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes)

```
=== Summary ===

Correctly Classified Instances         156               75.3623 %
Incorrectly Classified Instances        51               24.6377 %
Kappa statistic                          0.4968
Mean absolute error                      0.2468
Root mean squared error                  0.4633
Relative absolute error                 49.7186 %
Root relative squared error             92.494  %
Total Number of Instances              207

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.561    0.073    0.873      0.561    0.683      0.529
0.880     0.869     +
                 0.927    0.439    0.701      0.927    0.798      0.529
0.880     0.887     -
Weighted Avg.    0.754    0.266    0.783      0.754    0.744      0.529
0.880     0.879

=== Confusion Matrix ===

   a   b   <-- classified as
  55  43 |   a = +
   8 101 |   b = -
```

• Neural Networks (weka.classifiers.functions.MultilayerPerceptron)

```
=== Summary ===

Correctly Classified Instances         160               77.2947 %
Incorrectly Classified Instances        47               22.7053 %
Kappa statistic                          0.5401
Mean absolute error                      0.2173
Root mean squared error                  0.4352
Relative absolute error                 43.7768 %
Root relative squared error             86.8833 %
Total Number of Instances              207

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

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.663    0.128    0.823      0.663    0.734      0.550
0.869     0.864     +
                 0.872    0.337    0.742      0.872    0.802      0.550
0.869     0.840     -
Weighted Avg.    0.773    0.238    0.780      0.773    0.770      0.550
0.869     0.851

=== Confusion Matrix ===

  a  b   <-- classified as
 65 33 |  a = +
 14 95 |  b = -
```

Task 4(Lab 07)

ecoli

• C4.5 (weka.classifier.trees.J48)

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

Correctly Classified Instances         283               84.2262 %
Incorrectly Classified Instances        53               15.7738 %
Kappa statistic                          0.7824
Mean absolute error                      0.0486
Root mean squared error                  0.1851
Relative absolute error                 26.5877 %
Root relative squared error             61.3413 %
Total Number of Instances              336

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.951    0.036    0.951      0.951    0.951      0.915
0.962     0.915     cp
                 0.844    0.066    0.793      0.844    0.818      0.762
0.907     0.784     im
                 0.865    0.032    0.833      0.865    0.849      0.821
0.904     0.669     pp
                 0.571    0.030    0.690      0.571    0.625      0.589
0.855     0.635     imU
                 0.700    0.028    0.609      0.700    0.651      0.629
0.890     0.655     om
```

```
                 0.600    0.006    0.600      0.600    0.600      0.594
0.993     0.604     omL
                 0.000    0.000    ?          0.000    ?          ?
0.490     0.006     imL
                 0.000    0.000    ?          0.000    ?          ?
0.479     0.006     imS
Weighted Avg.    0.842    0.040    ?          0.842    ?          ?
0.920     0.787

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 136   0   4   0   3   0   0   0 |   a = cp
   2  65   0   8   2   0   0   0 |   b = im
   4   2  45   0   1   0   0   0 |   c = pp
   1  12   1  20   1   0   0   0 |   d = imU
   0   3   3   0  14   0   0   0 |   e = om
   0   0   0   0   2   3   0   0 |   f = omL
   0   0   0   0   0   2   0   0 |   g = imL
   0   0   1   1   0   0   0   0 |   h = imS
```

• RIPPER (weka.classifier.rules.JRip)

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

Correctly Classified Instances         271               80.6548 %
Incorrectly Classified Instances        65               19.3452 %
Kappa statistic                          0.7311
Mean absolute error                      0.0608
Root mean squared error                  0.2013
Relative absolute error                 33.2586 %
Root relative squared error             66.7354 %
Total Number of Instances              336

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.951    0.088    0.889      0.951    0.919      0.857
0.943     0.882     cp
                 0.766    0.054    0.808      0.766    0.787      0.726
0.928     0.821     im
                 0.788    0.025    0.854      0.788    0.820      0.789
0.924     0.751     pp
                 0.514    0.060    0.500      0.514    0.507      0.449
0.852     0.435     imU
                 0.750    0.013    0.789      0.750    0.769      0.755
0.874     0.602     om
                 0.400    0.015    0.286      0.400    0.333      0.326
0.767     0.165     omL
                 0.000    0.000    ?          0.000    ?          ?
0.708     0.086     imL
                 0.000    0.000    ?          0.000    ?          ?
0.380     0.006     imS
```

```
Weighted Avg.    0.807    0.061    ?          0.807    ?          ?
0.916     0.764

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 136   0   3   2   2   0   0   0 |   a = cp
   2  59   0  14   1   1   0   0 |   b = im
   8   1  41   2   0   0   0   0 |   c = pp
   3  12   0  18   1   1   0   0 |   d = imU
   1   0   3   0  15   1   0   0 |   e = om
   2   0   1   0   0   2   0   0 |   f = omL
   0   0   0   0   0   2   0   0 |   g = imL
   1   1   0   0   0   0   0   0 |   h = imS
```

• k-Nearest Neighbor (weka.classifiers.lazy.IBk)

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

Correctly Classified Instances         270               80.3571 %
Incorrectly Classified Instances        66               19.6429 %
Kappa statistic                          0.7295
Mean absolute error                      0.0535
Root mean squared error                  0.2189
Relative absolute error                 29.238  %
Root relative squared error             72.5574 %
Total Number of Instances              336

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.930    0.052    0.930      0.930    0.930      0.878
0.942     0.900     cp
                 0.727    0.081    0.727      0.727    0.727      0.646
0.814     0.609     im
                 0.846    0.046    0.772      0.846    0.807      0.771
0.903     0.695     pp
                 0.486    0.056    0.500      0.486    0.493      0.435
0.713     0.304     imU
                 0.750    0.006    0.882      0.750    0.811      0.803
0.896     0.680     om
                 1.000    0.003    0.833      1.000    0.909      0.911
0.999     0.867     omL
                 0.000    0.006    0.000      0.000    0.000      -0.006
0.695     0.010     imL
                 0.000    0.000    ?          0.000    ?          ?
0.698     0.010     imS
Weighted Avg.    0.804    0.054    ?          0.804    ?          ?
0.878     0.715

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 133   4   6   0   0   0   0   0 |   a = cp
```

```
   4  56   1  15   0   0   1   0 |   b = im
   5   1  44   0   2   0   0   0 |   c = pp
   1  15   1  17   0   0   1   0 |   d = imU
   0   0   4   1  15   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
```

• Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes)

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

Correctly Classified Instances         287               85.4167 %
Incorrectly Classified Instances        49               14.5833 %
Kappa statistic                          0.8002
Mean absolute error                      0.0429
Root mean squared error                  0.1639
Relative absolute error                 23.461  %
Root relative squared error             54.3314 %
Total Number of Instances              336

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.958    0.041    0.945      0.958    0.951      0.915
0.986     0.973     cp
                 0.727    0.031    0.875      0.727    0.794      0.745
0.966     0.904     im
                 0.846    0.032    0.830      0.846    0.838      0.808
0.945     0.901     pp
                 0.829    0.060    0.617      0.829    0.707      0.677
0.937     0.630     imU
                 0.900    0.009    0.857      0.900    0.878      0.870
0.996     0.964     om
                 0.600    0.000    1.000      0.600    0.750      0.772
0.996     0.883     omL
                 0.000    0.006    0.000      0.000    0.000      -0.006
0.060     0.006     imL
                 0.000    0.003    0.000      0.000    0.000      -0.004
0.148     0.005     imS
Weighted Avg.    0.854    0.036    0.861      0.854    0.854      0.819
0.960     0.897

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 137   2   4   0   0   0   0   0 |   a = cp
   3  56   1  16   0   0   0   1 |   b = im
   4   1  44   0   3   0   0   0 |   c = pp
   1   5   0  29   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
```

• Neural Networks (weka.classifiers.functions.MultilayerPerceptron)

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

Correctly Classified Instances         289               86.0119 %
Incorrectly Classified Instances        47               13.9881 %
Kappa statistic                          0.8066
Mean absolute error                      0.0484
Root mean squared error                  0.1704
Relative absolute error                 26.479  %
Root relative squared error             56.4913 %
Total Number of Instances              336

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.965    0.036    0.952      0.965    0.958      0.927
0.980     0.962     cp
                 0.831    0.062    0.800      0.831    0.815      0.759
0.951     0.870     im
                 0.846    0.032    0.830      0.846    0.838      0.808
0.952     0.806     pp
                 0.629    0.037    0.667      0.629    0.647      0.608
0.935     0.580     imU
                 0.850    0.009    0.850      0.850    0.850      0.841
0.977     0.887     om
                 0.800    0.003    0.800      0.800    0.800      0.797
0.997     0.786     omL
                 0.000    0.000    ?          0.000    ?          ?
0.187     0.005     imL
                 0.000    0.000    ?          0.000    ?          ?
0.340     0.007     imS
Weighted Avg.    0.860    0.039    ?          0.860    ?          ?
0.956     0.859

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 138   1   4   0   0   0   0   0 |   a = cp
   2  64   1  10   0   0   0   0 |   b = im
   4   2  44   0   2   0   0   0 |   c = pp
   1  12   0  22   0   0   0   0 |   d = imU
   0   0   3   0  17   0   0   0 |   e = om
   0   0   0   0   1   4   0   0 |   f = omL
   0   0   0   1   0   1   0   0 |   g = imL
   0   1   1   0   0   0   0   0 |   h = imS
```

glass

• C4.5 (weka.classifier.trees.J48)

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

```
Correctly Classified Instances         141               65.8879 %
Incorrectly Classified Instances        73               34.1121 %
Kappa statistic                          0.5412
Mean absolute error                      0.1059
Root mean squared error                  0.2928
Relative absolute error                 50.0098 %
Root relative squared error             90.2088 %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.714    0.181    0.658      0.714    0.685      0.523
0.792     0.625     1
                 0.566    0.167    0.652      0.566    0.606      0.414
0.756     0.617     2
                 0.294    0.061    0.294      0.294    0.294      0.233
0.722     0.186     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.846    0.025    0.688      0.846    0.759      0.746
0.944     0.590     5
                 0.889    0.015    0.727      0.889    0.800      0.795
0.938     0.690     6
                 0.828    0.022    0.857      0.828    0.842      0.818
0.884     0.716     7
Weighted Avg.    0.659    0.128    0.658      0.659    0.656      0.526
0.801     0.600

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 50 14  4  0  0  1  1 |  a = 1
 16 43  8  0  5  2  2 |  b = 2
  7  5  5  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  1  0  0 11  0  1 |  e = 5
  1  0  0  0  0  8  0 |  f = 6
  2  3  0  0  0  0 24 |  g = 7
```

- **RIPPER (weka.classifier.rules.JRip)**

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

Correctly Classified Instances         149               69.6262 %
Incorrectly Classified Instances        65               30.3738 %
Kappa statistic                          0.5741
Mean absolute error                      0.1139
Root mean squared error                  0.2657
Relative absolute error                 53.8052 %
Root relative squared error             81.8743 %
Total Number of Instances              214

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

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.714    0.167    0.676      0.714    0.694      0.540
0.816     0.655     1
                 0.789    0.225    0.659      0.789    0.719      0.547
0.818     0.649     2
                 0.118    0.010    0.500      0.118    0.190      0.215
0.655     0.182     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.615    0.015    0.727      0.615    0.667      0.650
0.803     0.475     5
                 0.556    0.015    0.625      0.556    0.588      0.572
0.805     0.387     6
                 0.828    0.011    0.923      0.828    0.873      0.856
0.906     0.797     7
Weighted Avg.    0.696    0.138    0.690      0.696    0.681      0.567
0.815     0.613

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 50 17  1  0  0  1  1 |  a = 1
 14 60  1  0  0  1  0 |  b = 2
  9  5  2  0  0  1  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  4  0  0  8  0  1 |  e = 5
  0  4  0  0  0  5  0 |  f = 6
  1  1  0  0  3  0 24 |  g = 7
```

• k-Nearest Neighbor (weka.classifiers.lazy.IBk)

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

Correctly Classified Instances         151               70.5607 %
Incorrectly Classified Instances        63               29.4393 %
Kappa statistic                          0.6017
Mean absolute error                      0.0897
Root mean squared error                  0.2852
Relative absolute error                 42.3765 %
Root relative squared error             87.8768 %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.771    0.139    0.730      0.771    0.750      0.624
0.823     0.660     1
                 0.671    0.145    0.718      0.671    0.694      0.535
0.749     0.605     2
                 0.353    0.061    0.333      0.353    0.343      0.285
0.670     0.182     3
```

```
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.769    0.030    0.625      0.769    0.690      0.671
0.824     0.483     5
                 0.667    0.010    0.750      0.667    0.706      0.695
0.819     0.518     6
                 0.828    0.016    0.889      0.828    0.857      0.836
0.903     0.736     7
Weighted Avg.    0.706    0.106    0.710      0.706    0.707      0.600
0.795     0.596

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 54  9  7  0  0  0  0 |  a = 1
 14 51  5  0  4  1  1 |  b = 2
  6  5  6  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  2  0  0 10  0  1 |  e = 5
  0  1  0  0  1  6  1 |  f = 6
  0  3  0  0  1  1 24 |  g = 7
```

• Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes)

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

Correctly Classified Instances         106               49.5327 %
Incorrectly Classified Instances       108               50.4673 %
Kappa statistic                          0.334
Mean absolute error                      0.1521
Root mean squared error                  0.3343
Relative absolute error                 71.8506 %
Root relative squared error            102.9939 %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.714    0.396    0.467      0.714    0.565      0.299
0.718     0.473     1
                 0.197    0.087    0.556      0.197    0.291      0.159
0.717     0.506     2
                 0.353    0.112    0.214      0.353    0.267      0.193
0.699     0.196     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.231    0.045    0.250      0.231    0.240      0.193
0.835     0.342     5
                 0.889    0.020    0.667      0.889    0.762      0.758
0.985     0.776     6
                 0.828    0.022    0.857      0.828    0.842      0.818
0.941     0.817     7
Weighted Avg.    0.495    0.176    0.527      0.495    0.470      0.324
0.765     0.514
```

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 50  4 14  0  0  2  0 |  a = 1
 46 15  8  0  5  1  1 |  b = 2
 10  0  6  0  0  1  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  8  0  0  3  0  2 |  e = 5
  0  0  0  0  0  8  1 |  f = 6
  1  0  0  0  4  0 24 |  g = 7
```

• Neural Networks (weka.classifiers.functions.MultilayerPerceptron)

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

Correctly Classified Instances         148               69.1589 %
Incorrectly Classified Instances        66               30.8411 %
Kappa statistic                          0.5677
Mean absolute error                      0.1067
Root mean squared error                  0.2471
Relative absolute error                 50.3806 %
Root relative squared error             76.124  %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.814    0.229    0.633      0.814    0.713      0.556
0.881     0.747     1
                 0.697    0.174    0.688      0.697    0.693      0.522
0.856     0.758     2
                 0.059    0.005    0.500      0.059    0.105      0.151
0.732     0.192     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.692    0.020    0.692      0.692    0.692      0.672
0.981     0.843     5
                 0.556    0.015    0.625      0.556    0.588      0.572
0.933     0.613     6
                 0.793    0.005    0.958      0.793    0.868      0.854
0.908     0.801     7
Weighted Avg.    0.692    0.140    0.690      0.692    0.672      0.560
0.872     0.715

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 57 12  1  0  0  0  0 |  a = 1
 20 53  0  0  1  2  0 |  b = 2
 10  6  1  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  3  0  0  9  0  1 |  e = 5
  2  1  0  0  1  5  0 |  f = 6
```

```
  1   2   0   0   2   1 23 |  g = 7
```

image

• C4.5 (weka.classifier.trees.J48)

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

Correctly Classified Instances         141               65.8879 %
Incorrectly Classified Instances        73               34.1121 %
Kappa statistic                          0.5412
Mean absolute error                      0.1059
Root mean squared error                  0.2928
Relative absolute error                 50.0098 %
Root relative squared error             90.2088 %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.714    0.181    0.658      0.714    0.685      0.523
0.792     0.625     1
                 0.566    0.167    0.652      0.566    0.606      0.414
0.756     0.617     2
                 0.294    0.061    0.294      0.294    0.294      0.233
0.722     0.186     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.846    0.025    0.688      0.846    0.759      0.746
0.944     0.590     5
                 0.889    0.015    0.727      0.889    0.800      0.795
0.938     0.690     6
                 0.828    0.022    0.857      0.828    0.842      0.818
0.884     0.716     7
Weighted Avg.    0.659    0.128    0.658      0.659    0.656      0.526
0.801     0.600

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 50 14  4  0  0  1  1 |  a = 1
 16 43  8  0  5  2  2 |  b = 2
  7  5  5  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  1  0  0 11  0  1 |  e = 5
  1  0  0  0  0  8  0 |  f = 6
  2  3  0  0  0  0 24 |  g = 7
```

• RIPPER (weka.classifier.rules.JRip)

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

```
Correctly Classified Instances         149               69.6262 %
Incorrectly Classified Instances        65               30.3738 %
Kappa statistic                          0.5741
Mean absolute error                      0.1139
Root mean squared error                  0.2657
Relative absolute error                 53.8052 %
Root relative squared error             81.8743 %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.714    0.167    0.676      0.714    0.694      0.540
0.816     0.655     1
                 0.789    0.225    0.659      0.789    0.719      0.547
0.818     0.649     2
                 0.118    0.010    0.500      0.118    0.190      0.215
0.655     0.182     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.615    0.015    0.727      0.615    0.667      0.650
0.803     0.475     5
                 0.556    0.015    0.625      0.556    0.588      0.572
0.805     0.387     6
                 0.828    0.011    0.923      0.828    0.873      0.856
0.906     0.797     7
Weighted Avg.    0.696    0.138    0.690      0.696    0.681      0.567
0.815     0.613

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 50 17  1  0  0  1  1 |  a = 1
 14 60  1  0  0  1  0 |  b = 2
  9  5  2  0  0  1  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  4  0  0  8  0  1 |  e = 5
  0  4  0  0  0  5  0 |  f = 6
  1  1  0  0  3  0 24 |  g = 7
```

• k-Nearest Neighbor (weka.classifiers.lazy.IBk)

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

Correctly Classified Instances         151               70.5607 %
Incorrectly Classified Instances        63               29.4393 %
Kappa statistic                          0.6017
Mean absolute error                      0.0897
Root mean squared error                  0.2852
Relative absolute error                 42.3765 %
Root relative squared error             87.8768 %
Total Number of Instances              214
```

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

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.771    0.139    0.730      0.771    0.750      0.624
0.823     0.660     1
                 0.671    0.145    0.718      0.671    0.694      0.535
0.749     0.605     2
                 0.353    0.061    0.333      0.353    0.343      0.285
0.670     0.182     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.769    0.030    0.625      0.769    0.690      0.671
0.824     0.483     5
                 0.667    0.010    0.750      0.667    0.706      0.695
0.819     0.518     6
                 0.828    0.016    0.889      0.828    0.857      0.836
0.903     0.736     7
Weighted Avg.    0.706    0.106    0.710      0.706    0.707      0.600
0.795     0.596

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 54  9  7  0  0  0  0 |  a = 1
 14 51  5  0  4  1  1 |  b = 2
  6  5  6  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  2  0  0 10  0  1 |  e = 5
  0  1  0  0  1  6  1 |  f = 6
  0  3  0  0  1  1 24 |  g = 7
```

• Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes)

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

Correctly Classified Instances         106               49.5327 %
Incorrectly Classified Instances       108               50.4673 %
Kappa statistic                          0.334
Mean absolute error                      0.1521
Root mean squared error                  0.3343
Relative absolute error                 71.8506 %
Root relative squared error            102.9939 %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.714    0.396    0.467      0.714    0.565      0.299
0.718     0.473     1
                 0.197    0.087    0.556      0.197    0.291      0.159
0.717     0.506     2
                 0.353    0.112    0.214      0.353    0.267      0.193
0.699     0.196     3
```

```
                 ?         0.000    ?          ?         ?          ?
?         ?         4
                 0.231     0.045    0.250      0.231     0.240      0.193
0.835     0.342     5
                 0.889     0.020    0.667      0.889     0.762      0.758
0.985     0.776     6
                 0.828     0.022    0.857      0.828     0.842      0.818
0.941     0.817     7
Weighted Avg.    0.495     0.176    0.527      0.495     0.470      0.324
0.765     0.514

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 50  4 14  0  0  2  0 |  a = 1
 46 15  8  0  5  1  1 |  b = 2
 10  0  6  0  0  1  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  8  0  0  3  0  2 |  e = 5
  0  0  0  0  0  8  1 |  f = 6
  1  0  0  0  4  0 24 |  g = 7
```

• Neural Networks (weka.classifiers.functions.MultilayerPerceptron)

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

Correctly Classified Instances         148               69.1589 %
Incorrectly Classified Instances        66               30.8411 %
Kappa statistic                          0.5677
Mean absolute error                      0.1067
Root mean squared error                  0.2471
Relative absolute error                 50.3806 %
Root relative squared error             76.124  %
Total Number of Instances              214

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
ROC Area  PRC Area  Class
                 0.814    0.229    0.633      0.814    0.713      0.556
0.881     0.747     1
                 0.697    0.174    0.688      0.697    0.693      0.522
0.856     0.758     2
                 0.059    0.005    0.500      0.059    0.105      0.151
0.732     0.192     3
                 ?        0.000    ?          ?        ?          ?
?         ?         4
                 0.692    0.020    0.692      0.692    0.692      0.672
0.981     0.843     5
                 0.556    0.015    0.625      0.556    0.588      0.572
0.933     0.613     6
                 0.793    0.005    0.958      0.793    0.868      0.854
0.908     0.801     7
Weighted Avg.    0.692    0.140    0.690      0.692    0.672      0.560
0.872     0.715
```

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 57 12  1  0  0  0  0 |  a = 1
 20 53  0  0  1  2  0 |  b = 2
 10  6  1  0  0  0  0 |  c = 3
  0  0  0  0  0  0  0 |  d = 4
  0  3  0  0  9  0  1 |  e = 5
  2  1  0  0  1  5  0 |  f = 6
  1  2  0  0  2  1 23 |  g = 7
```

Discuss the results regarding whether there is an overall winner and whether the misclassification rates for the algorithms are significantly different.
The overall winning classifier is the k-Nearest Neighbor (weka.classifiers.lazy.IBk). It is very clear the that the misclassification rates for the algorithms are significantly different regardless of how they varied from dataset to classifier.

Task 5(Lab 07-08)

```
=== Summary ===

Correlation coefficient                  0.1292
Mean absolute error                      0.5834
Root mean squared error                  1.0143
Relative absolute error                 94.4768 %
Root relative squared error            129.689  %
Total Number of Instances             4229
Ignored Class Unknown Instances               4771
```

Task 6(Lab 08-09)

```
# Perceptron Algorithm
# initialize θ and θ0 with 0
θ = 0  (vector)
θ0 = 0 (scalar)# totally T epoches to iterate
for t = 1 .. T do
    # totally m data points
    for i = 1 .. m do
        # misclassify data points
        if y(i)(θ · x(i) + θ0) ≦ 0
        then
            θ  = θ + y(i) · x(i)
            θ0 = θ0 + y(i)return θ, θ0
```