**Project 2**

**Task 1:**

**J48 Tree**

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

J48 pruned tree

------------------

lym\_nodes\_dimin <= 1

| changes\_in\_node = no

| | defect\_in\_node = no: normal (3.0/1.0)

| | defect\_in\_node = lacunar: malign\_lymph (2.0)

| | defect\_in\_node = lac\_margin: normal (0.0)

| | defect\_in\_node = lac\_central: normal (0.0)

| changes\_in\_node = lacunar

| | exclusion\_of\_no = no: metastases (10.0/1.0)

| | exclusion\_of\_no = yes

| | | special\_forms = no: metastases (3.0/1.0)

| | | special\_forms = chalices

| | | | lym\_nodes\_enlar <= 2: malign\_lymph (3.0)

| | | | lym\_nodes\_enlar > 2: metastases (2.0)

| | | special\_forms = vesicles: malign\_lymph (19.0/1.0)

| changes\_in\_node = lac\_margin

| | block\_of\_affere = no

| | | extravasates = no

| | | | lymphatics = normal: metastases (0.0)

| | | | lymphatics = arched

| | | | | early\_uptake\_in = no: metastases (5.0/1.0)

| | | | | early\_uptake\_in = yes: malign\_lymph (4.0/1.0)

| | | | lymphatics = deformed: metastases (5.0)

| | | | lymphatics = displaced: malign\_lymph (1.0)

| | | extravasates = yes: malign\_lymph (4.0)

| | block\_of\_affere = yes: metastases (56.0/3.0)

| changes\_in\_node = lac\_central

| | no\_of\_nodes\_in <= 1

| | | block\_of\_affere = no: malign\_lymph (3.0)

| | | block\_of\_affere = yes: metastases (2.0)

| | no\_of\_nodes\_in > 1: malign\_lymph (20.0)

lym\_nodes\_dimin > 1

| by\_pass = no: metastases (2.0/1.0)

| by\_pass = yes: fibrosis (4.0)

Number of Leaves : 21

Size of the tree : 34

Time taken to build model: 0.01 seconds

**JRIP**

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

JRIP rules:

===========

(lymphatics = normal) => class=normal (2.0/0.0)

(lym\_nodes\_dimin >= 2) and (by\_pass = yes) => class=fibrosis (4.0/0.0)

(no\_of\_nodes\_in >= 3) and (special\_forms = vesicles) => class=malign\_lymph (41.0/5.0)

(block\_of\_affere = no) and (extravasates = yes) => class=malign\_lymph (8.0/0.0)

(changes\_in\_node = lac\_central) => class=malign\_lymph (8.0/2.0)

=> class=metastases (85.0/11.0)

Number of Rules : 6

Time taken to build model: 0.02 seconds

**Task 2:**

**J48**

=== Summary ===

Correctly Classified Instances 420 97.2222 %

Incorrectly Classified Instances 12 2.7778 %

Kappa statistic 0.9444

Mean absolute error 0.0892

Root mean squared error 0.1831

Relative absolute error 17.8311 %

Root relative squared error 36.5759 %

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 0.983 0.964 0

0.947 0.000 1.000 0.947 0.973 0.946 0.983 0.981 1

Weighted Avg. 0.972 0.025 0.974 0.972 0.972 0.946 0.983 0.973

=== Confusion Matrix ===

a b <-- classified as

204 0 | a = 0

12 216 | b = 1

**JRIP**

=== Summary ===

Correctly Classified Instances 390 90.2778 %

Incorrectly Classified Instances 42 9.7222 %

Kappa statistic 0.8053

Mean absolute error 0.1314

Root mean squared error 0.277

Relative absolute error 26.2643 %

Root relative squared error 55.3461 %

Total Number of Instances 432

=== Detailed Accuracy By Class ===

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

0.912 0.105 0.886 0.912 0.899 0.806 0.938 0.879 0

0.895 0.088 0.919 0.895 0.907 0.806 0.938 0.942 1

Weighted Avg. 0.903 0.096 0.903 0.903 0.903 0.806 0.938 0.912

=== Confusion Matrix ===

a b <-- classified as

186 18 | a = 0

24 204 | b = 1

**IBk**

=== Summary ===

Correctly Classified Instances 378 87.5 %

Incorrectly Classified Instances 54 12.5 %

Kappa statistic 0.7512

Mean absolute error 0.191

Root mean squared error 0.3228

Relative absolute error 38.1693 %

Root relative squared error 64.5029 %

Total Number of Instances 432

=== Detailed Accuracy By Class ===

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

0.941 0.184 0.821 0.941 0.877 0.758 0.933 0.902 0

0.816 0.059 0.939 0.816 0.873 0.758 0.933 0.935 1

Weighted Avg. 0.875 0.118 0.883 0.875 0.875 0.758 0.933 0.919

=== Confusion Matrix ===

a b <-- classified as

192 12 | a = 0

42 186 | b = 1

**NaiveBayes**

=== Summary ===

Correctly Classified Instances 420 97.2222 %

Incorrectly Classified Instances 12 2.7778 %

Kappa statistic 0.9444

Mean absolute error 0.1863

Root mean squared error 0.2323

Relative absolute error 37.2363 %

Root relative squared error 46.4131 %

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 0.975 0.961 0

0.947 0.000 1.000 0.947 0.973 0.946 0.975 0.985 1

Weighted Avg. 0.972 0.025 0.974 0.972 0.972 0.946 0.975 0.973

=== Confusion Matrix ===

a b <-- classified as

204 0 | a = 0

12 216 | b = 1

**MultiLayer Perception**

=== Summary ===

Correctly Classified Instances 404 93.5185 %

Incorrectly Classified Instances 28 6.4815 %

Kappa statistic 0.8709

Mean absolute error 0.068

Root mean squared error 0.2322

Relative absolute error 13.5875 %

Root relative squared error 46.3993 %

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 0.123 0.879 1.000 0.936 0.878 0.967 0.941 0

0.877 0.000 1.000 0.877 0.935 0.878 0.967 0.981 1

Weighted Avg. 0.935 0.058 0.943 0.935 0.935 0.878 0.967 0.962

=== Confusion Matrix ===

a b <-- classified as

204 0 | a = 0

28 200 | b = 1

Based the outputs above, J48 and NaiveBayes have the highest rates of correctly classified instance at 97.222%, while IBK has the lowest at 87.5%

**Task 3:**

**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 = -

**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 = -

**MultiLayer Perception**

=== 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:**

Time

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | J48 | JRIP | NaiveBayes | IBK | MultiLayerPerception |
| Ecoli | 0 | 0.03 | 0 | 0 | 0.35 |
| Glass | 0 | 0.01 | 0 | 0 | 0.26 |
| Image | 0 | 0.02 | 0 | 0 | 0.55 |

Correctly Classified Percentage

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | J48 | JRIP | NaiveBayes | IBK | MultiLayerPerception |
| Ecoli | 84.2262 % | 80.6548 % | 85.4167 % | 80.3571 % | 86.0119 % |
| Glass | 65.8879 % | 69.6262 % | 49.5327 % | 70.5607 % | 69.1589 % |
| Image | 89.0476 % | 79.5238 % | 77.619 % | 87.1429 % | 88.5714 % |

In terms of time, MultiLayer Perception was the longest time wise followed by JRIP. IBK is showing a better consistency than other methods in correct classification percentages.