

# CSCI4030U: Big Data Analytics

## Lab07

Syed Naqvi  
100590852

April 8, 2024

1.

(a) **C4.5 (weka.classifier.trees.J48)**

Classification Accuracy: 85.9903%

Sumamry and Confusion matrix:

```
=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds

=== 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 = -
```

(b) **Naive Bayesian Classification (weka.classifiers.bayes.NaiveBayes)**

Classification Accuracy: 75.3623%

Sumamry and Confusion matrix:

```
=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds

=== 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 = -
```

(c) **Neural Networks (weka.classifiers.functions.MultilayerPerceptron)**

Classification Accuracy: 77.2947%

Summary and Confusion matrix:

```
=== Evaluation on test split ===  
  
Time taken to test model on test split: 0 seconds  
  
=== 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 = -
```