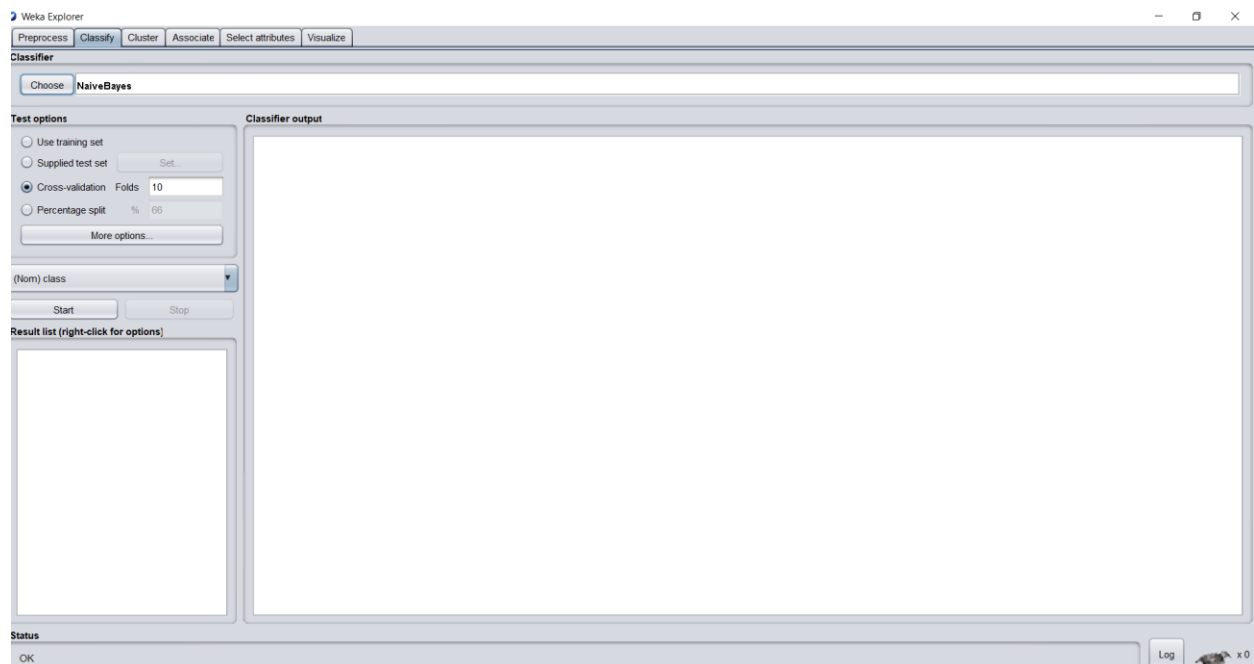


SIT719 Security and Privacy Issues in Analytics

PASS TASK 4.1: ATTACK CLASSIFICATION USING NAÏVE BAYES ALGORITHM

Now apply “Naïve Bayes” classification algorithm from the “Classify” tab



Successfully selected the “Naïve Bayes” classification algorithm from the “Classify” tab.

Check the results with a 10-fold cross validation.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

09:21:18 - bayes.NaiveBayes

Classifier output

```

std. dev.          0.1922      0.4034
weight sum        67343      58630
precision          0.01        0.01

Time taken to build model: 0.87 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   113956      90.3813 %
Incorrectly Classified Instances 12117      9.6187 %
Kappa statistic                 0.8059
Mean absolute error             0.0965
Root mean squared error         0.3058
Relative absolute error         19.3981 %
Root relative squared error      61.312 %
Total Number of Instances      125973

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MDC   ROC Area  PRC Area  Class
0.936   0.134   0.890   0.936   0.912   0.807   0.967   0.964   normal
0.866   0.064   0.922   0.866   0.893   0.807   0.965   0.949   anomaly
Weighted Avg.  0.904   0.101   0.905   0.904   0.904   0.807   0.966   0.957

=== Confusion Matrix ===
      a    b  <-- classified as
63058 4285 |  a = normal
 7832 50798 |  b = anomaly
  
```

Status

OK Log x 0

Successfully made the cross-validation with 10 folds and got the above results with correctly classified instances – 90.3813% and incorrectly classified instances – 9.6187%. Time taken to build the model is 0.87 seconds.

Now, upload the test dataset and check the classification results.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☒ Supplied test set

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

09:21:18 - bayes.NaiveBayes

09:26:15 - bayes.NaiveBayes

Classifier output

```

Time taken to build model: 0.9 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.34 seconds

=== Summary ===
Correctly Classified Instances   17161      76.1222 %
Incorrectly Classified Instances 5383      23.8778 %
Kappa statistic                 0.5366
Mean absolute error             0.2386
Root mean squared error         0.4862
Relative absolute error         47.2755 %
Root relative squared error      96.0968 %
Total Number of Instances      22544

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MDC   ROC Area  PRC Area  Class
0.931   0.367   0.657   0.931   0.771   0.572   0.895   0.844   normal
0.633   0.069   0.924   0.633   0.751   0.572   0.917   0.911   anomaly
Weighted Avg.  0.761   0.197   0.809   0.761   0.759   0.572   0.908   0.882

=== Confusion Matrix ===
      a    b  <-- classified as
9041  670 |  a = normal
4713 8120 |  b = anomaly
  
```

Status

OK Log x 0

Successfully uploaded the test data and got the above results with correctly classified instances – 76.1222% and incorrectly classified instances – 23.8778%. Time taken to build the model is 0.9 seconds.

Compare the results between 10-fold cross validation and the one obtained using the test dataset. Use confusion matrix to explain the results.

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

```
      a      b  <-- classified as
63058  4285 |      a = normal
 7832 50798 |      b = anomaly
```

Confusion matrix for 10 folds

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

```
      a      b  <-- classified as
 9041   670 |      a = normal
 4713 8120 |      b = anomaly
```

Confusion matrix for test data

On comparing both confusion matrix the matrix resulted using 10 folds has a high true positive value and which rated the correctly classified instances – 90.3813%. On comparing the false positive values, the matrix using 10 folds resulted the value of 4285 whereas the test data resulted 670. And the values of false negative are 7832 for 10 folds and 4713 for test data. The values of true negative are 50798 for 10 folds and 8120 for test data.

Finally, the output result obtained from the “Naïve Bayes” cross validation test option using 10 folds resulted in the high accuracy whereas the resulted output using test data is bit low accuracy on comparing to the cross validation.