Ahsanullah University of Science and Technology

Department of Computer Science and Engineering
Fall2023

Program: Bachelor of Science in Computer Science and Engineering

Course No: CSE4142
Course Title: Data Warehousing and Mining Lab

Assignment No: 04
Date of Submission: 07/07/2024

Submitted to

Mr Saha Reno
Assistant Professor, Department of CSE, AUST.

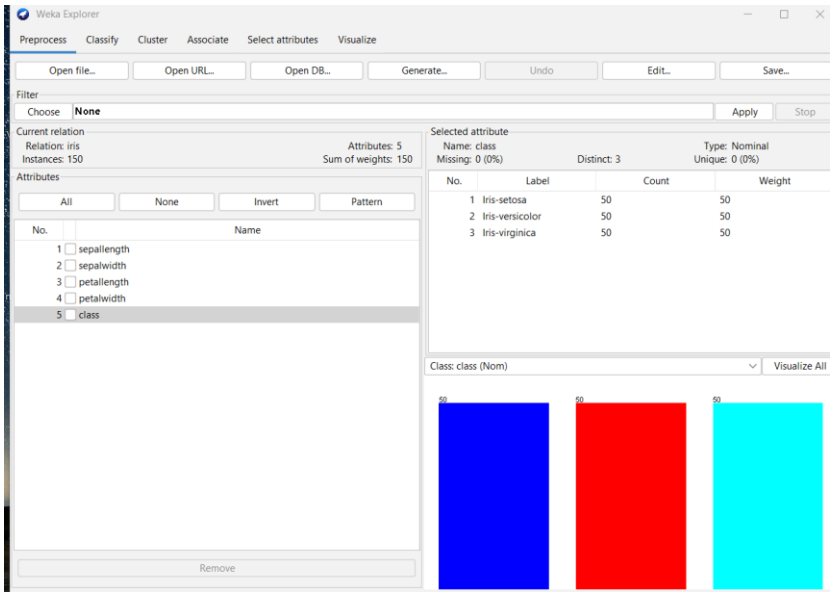
Mr Raiyan Jahangir
Lecturer, Department of CSE,AUST

Submitted by

Name: Amit Karmakar
Student ID: 20200204056

1)Take any dataset from Weka Repository or Online (Must be in ARFF format),

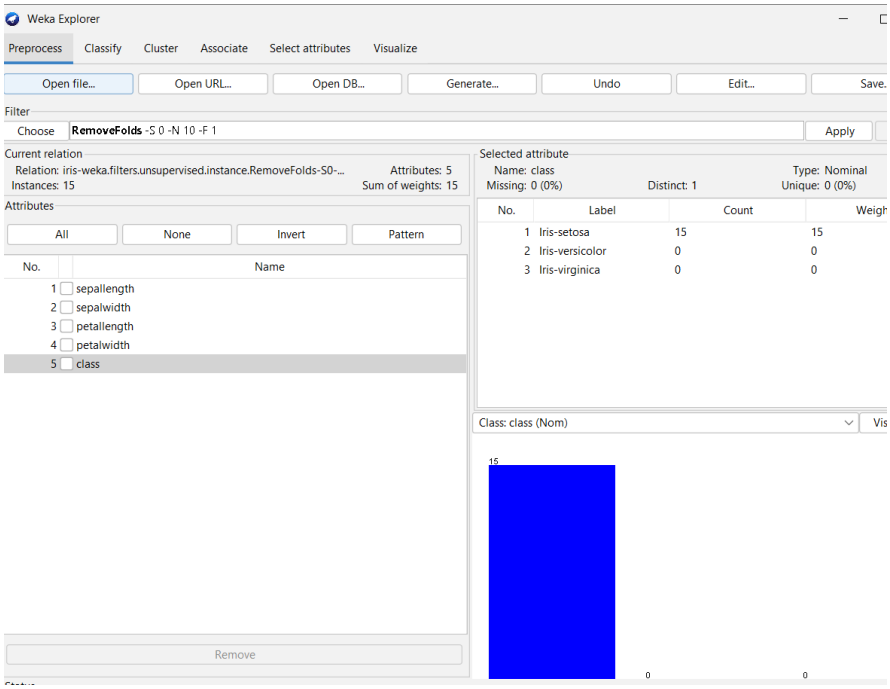
Ans:



2) Extract 1 Non-Stratified Fold from 10-Fold Cross Validation.

Ans: For extracting in non stratified method, I have followed this :

Pre process->choose->filter->unsupervised->instance->remove folds. Then I select classify and and apply 10 fold cross validation. Then I save it as non stratified.arff.

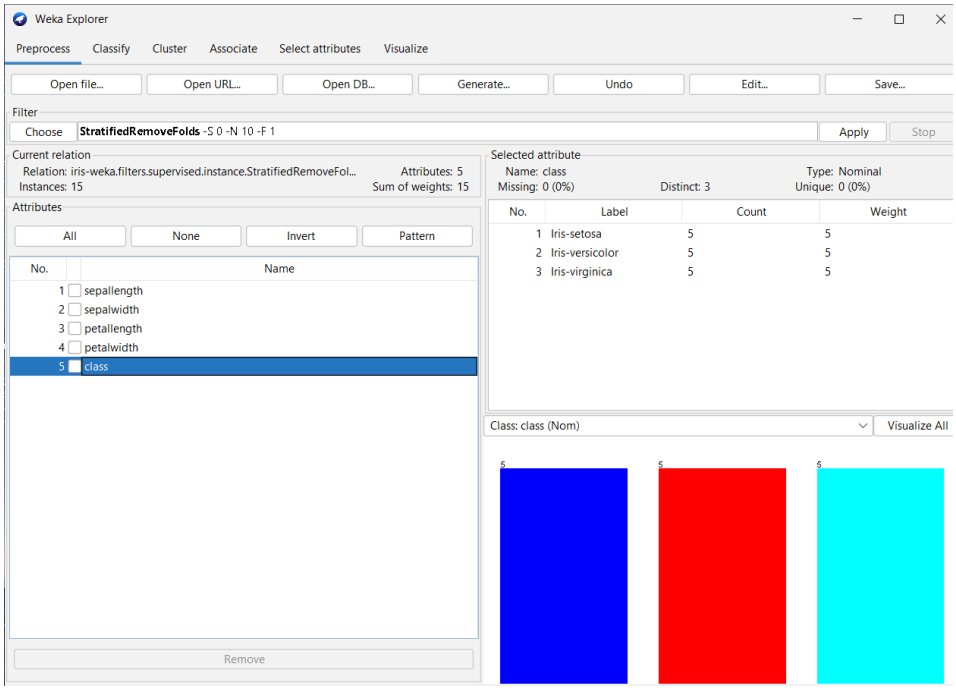


```
Correctly Classified Instances          15               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.0817
Root mean squared error                  0.0867
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               15

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    ?        1.000      1.000    1.000      ?        ?         1.000     Iris-setosa
                 ?        0.000    ?          ?        ?          ?        ?         ?         Iris-versicolor
                 ?        0.000    ?          ?        ?          ?        ?         ?         Iris-virginica
Weighted Avg.    1.000    ?        1.000      1.000    1.000      ?        ?         1.000

=== Confusion Matrix ===

  a  b  c   <-- classified as
 15  0  0 |  a = Iris-setosa
  0  0  0 |  b = Iris-versicolor
  0  0  0 |  c = Iris-virginica
```

3) Extract 1 Stratified Fold from 10-Fold Cross Validation.

Ans: for stratified method I have followed;

Pre process->choose->filter->supervised->instances->StratifiedRemoveFold. Then save it as Stratified.arff dataset and apply 10 fold cross validation.

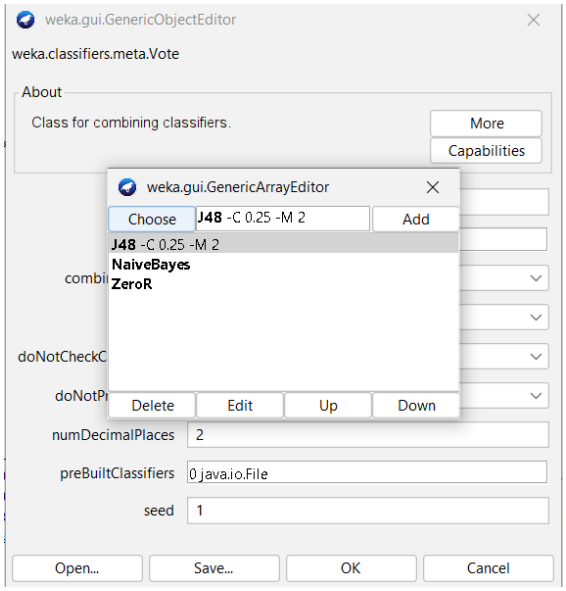



```
Correctly Classified Instances           0                0      %
Incorrectly Classified Instances        15              100      %
Kappa statistic                         -0.5
Mean absolute error                      0.4624
Root mean squared error                  0.4909
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               15

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    0.500    0.000      0.000    0.000      -0.500   0.250     0.333     Iris-setosa
                 0.000    0.000    ?          0.000    ?          ?        0.250     0.333     Iris-versicolor
                 0.000    1.000    0.000      0.000    0.000      -1.000   0.000     0.333     Iris-virginica
Weighted Avg.    0.000    0.500    ?          0.000    ?          ?        0.167     0.333

=== Confusion Matrix ===

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

4) Using Ensemble technique, create any 3 models/classifiers and follow Voting approach to classify test instances. You must select "Majority Voting" option as the Rule of Combination.

Ans: for ensemble technique I have reloaded irish.arff data in preprocess tab. Then I select classify and follow these steps:

Choose->meta->vote->classifier->choose-> 3 classifier->add. I have used zeroR, naïve bayes and j48 tree .



5) At first, use the Non-Stratified Fold as the Test Data. After that, use the Stratified Fold as the Test Data.

6) Show the outputs for both classifications from(5)

Ans: Result for non stratified test data.

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

Vote combines the probability distributions of these base learners:
        weka.classifiers.trees.J48 -C 0.25 -M 2
        weka.classifiers.bayes.NaiveBayes
        weka.classifiers.rules.ZeroR
using the 'Majority Voting' combination rule
```

```
Correctly Classified Instances          15               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               15

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    ?        1.000      1.000    1.000      ?        ?         1.000     Iris-setosa
                 ?        0.000    ?          ?        ?          ?        ?         ?         Iris-versicolor
                 ?        0.000    ?          ?        ?          ?        ?         ?         Iris-virginica
Weighted Avg.    1.000    ?        1.000      1.000    1.000      ?        ?         1.000

=== Confusion Matrix ===

  a  b  c   <-- classified as
 15  0  0 |  a = Iris-setosa
  0  0  0 |  b = Iris-versicolor
  0  0  0 |  c = Iris-virginica
```

Result for Statified test data:

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

Vote combines the probability distributions of these base learners:
        weka.classifiers.trees.J48 -C 0.25 -M 2
        weka.classifiers.bayes.NaiveBayes
        weka.classifiers.rules.ZeroR
using the 'Majority Voting' combination rule
```

```
Correctly Classified Instances          14               93.3333 %
Incorrectly Classified Instances         1                6.6667 %
Kappa statistic                          0.9
Mean absolute error                      0.0444
Root mean squared error                  0.2108
Relative absolute error                 10      %
Root relative squared error             44.7214 %
Total Number of Instances               15

=== 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
                 0.800    0.000    1.000      0.800    0.889      0.853    0.900     0.867     Iris-versicolor
                 1.000    0.100    0.833      1.000    0.909      0.866    0.950     0.833     Iris-virginica
Weighted Avg.    0.933    0.033    0.944      0.933    0.933      0.906    0.950     0.900

=== Confusion Matrix ===

 a b c   <-- classified as
 5 0 0 | a = Iris-setosa
 0 4 1 | b = Iris-versicolor
 0 0 5 | c = Iris-virginica
```