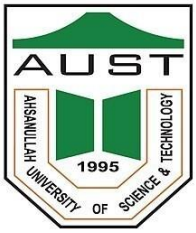


Ahsanullah University of Science & Technology

Department of Computer Science & Engineering

Course No : **CSE4142**

Course Title : **Data Warehousing and Mining Lab**

Assignment No : **04**

Date of Submission : **06.07.2024**

Submitted To : **Mr. Saha Reno & Mr. Raiyan Jahangir**

Submitted By-

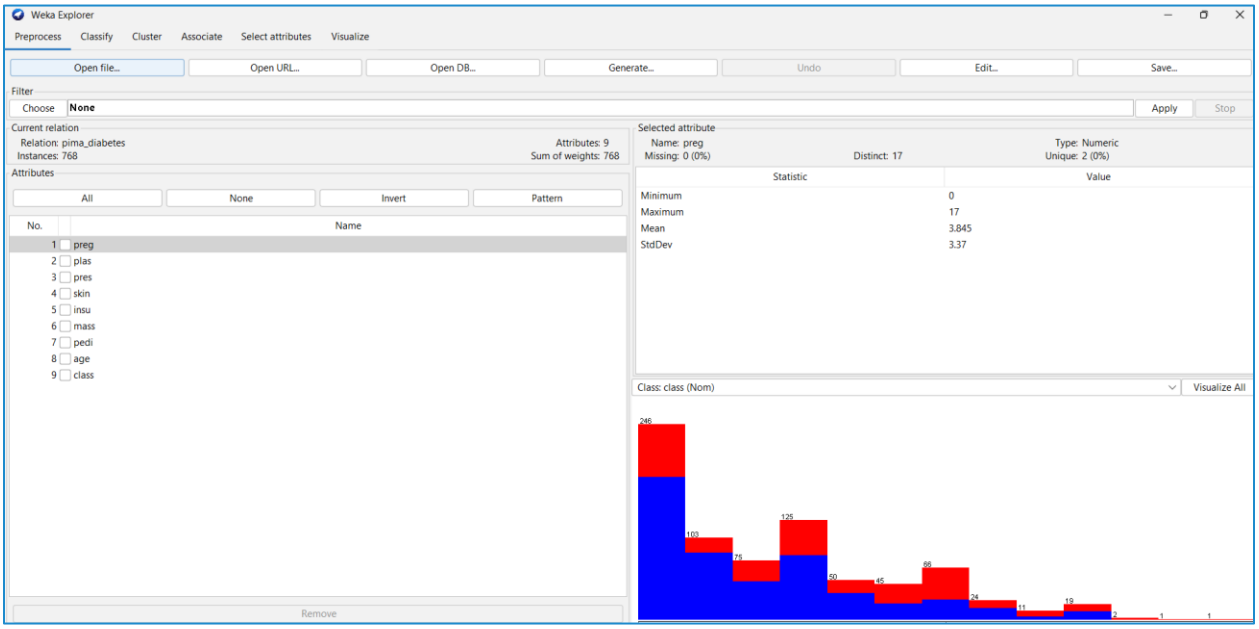
Name: Nabila Rahman

ID: 20200204065

Section: B1

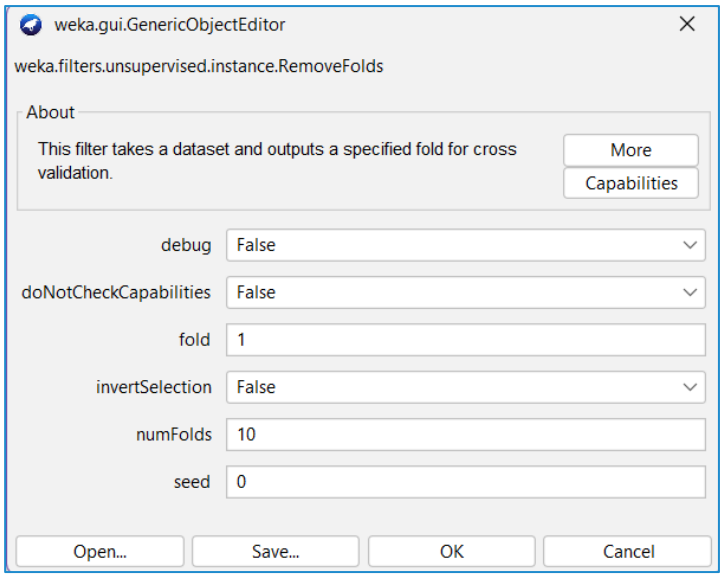
(i) Take any dataset from Weka Repository or Online (Must be in ARFF format)

Ans: Diabetes dataset from Weka Repository-



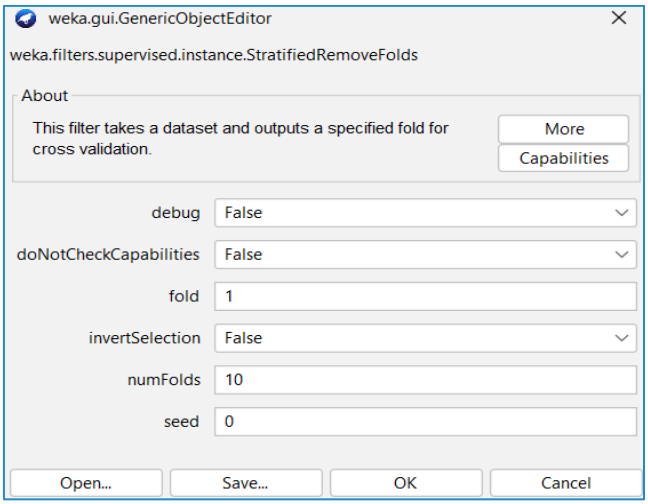
(ii) Extract 1 Non-Stratified Fold from 10-Fold Cross Validation.

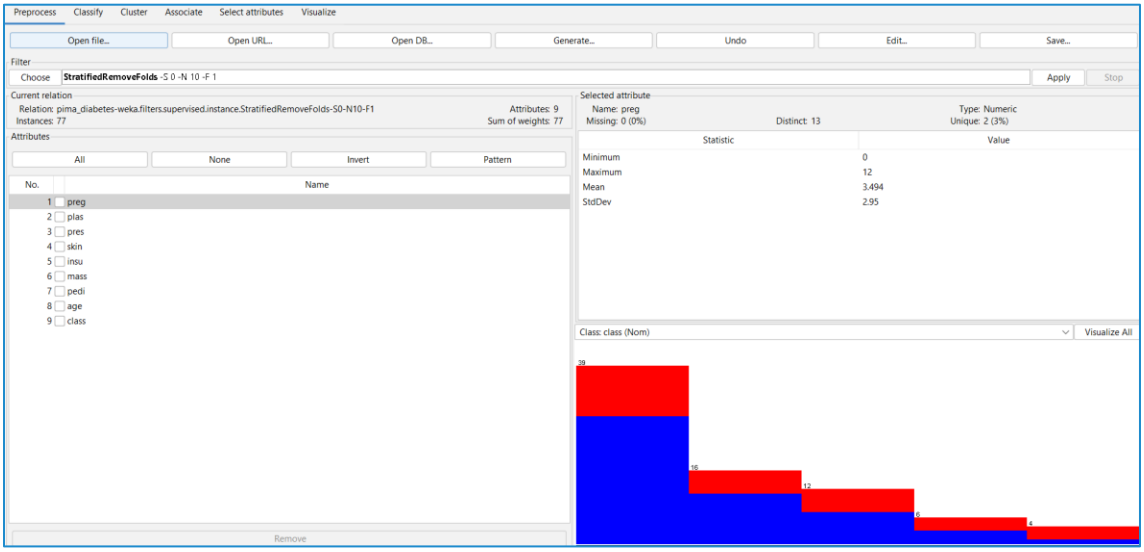
Ans:



(iii) Extract 1 Stratified Fold from 10-Fold Cross Validation.

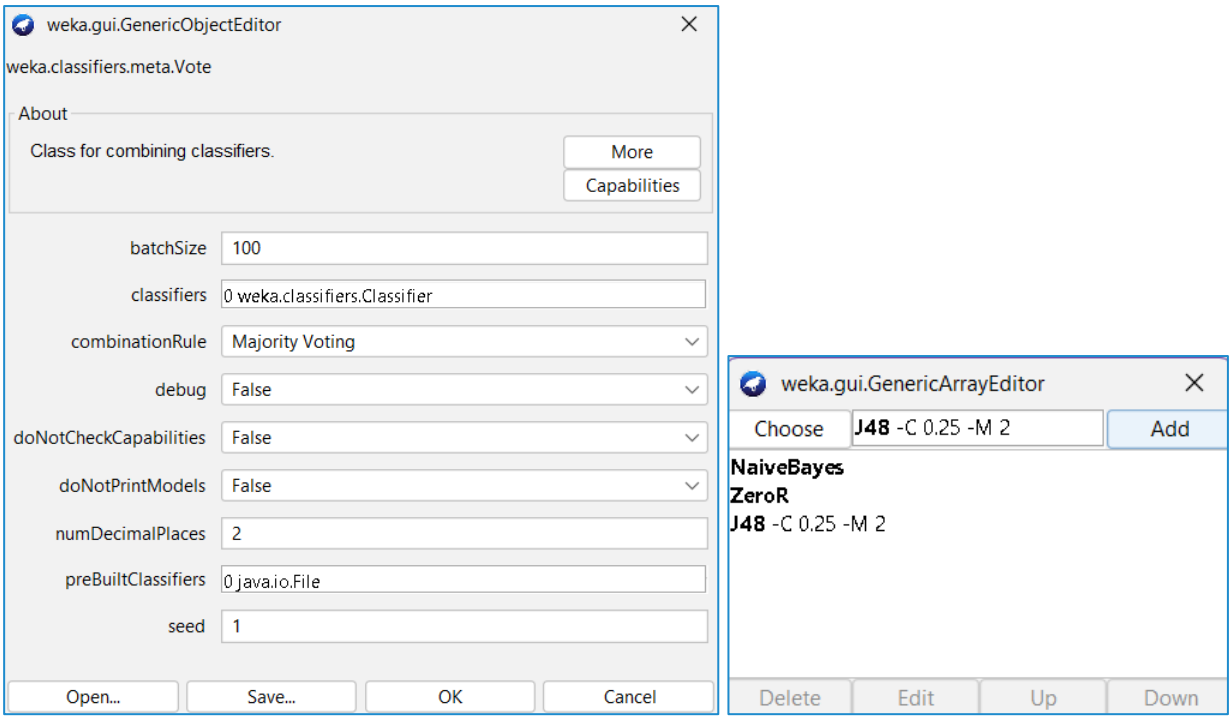
Ans:





(iv) 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:



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

Ans: First Non-Stratified Fold is taken as Test Data then the Stratified Fold us used as the Test Data.

(vi) Show the outputs for both classifications from (v).

Ans: From (v), for Non-Stratified fold the output is-

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

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

```
=== Summary ===

Correctly Classified Instances          58               75.3247 %
Incorrectly Classified Instances        19               24.6753 %
Kappa statistic                          0.4705
Mean absolute error                      0.2468
Root mean squared error                  0.4967
Relative absolute error                 50.7609 %
Root relative squared error            100.7945 %
Total Number of Instances               77

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.889    0.438    0.741      0.889    0.808      0.486    0.726     0.723     tested_negative
                 0.563    0.111    0.783      0.563    0.655      0.486    0.726     0.622     tested_positive
Weighted Avg.    0.753    0.302    0.758      0.753    0.744      0.486    0.726     0.681

=== Confusion Matrix ===

  a  b   <-- classified as
 40  5 |  a = tested_negative
 14 18 |  b = tested_positive
```

For Stratified fold the output is-

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

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

```
=== Summary ===

Correctly Classified Instances          63               81.8182 %
Incorrectly Classified Instances        14               18.1818 %
Kappa statistic                          0.5553
Mean absolute error                      0.1818
Root mean squared error                  0.4264
Relative absolute error                 39.8272 %
Root relative squared error             89.3571 %
Total Number of Instances               77

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.980    0.481    0.790      0.980    0.875      0.601    0.749     0.788     tested_negative
                 0.519    0.020    0.933      0.519    0.667      0.601    0.749     0.653     tested_positive
Weighted Avg.    0.818    0.320    0.840      0.818    0.802      0.601    0.749     0.740

=== Confusion Matrix ===

  a  b   <-- classified as
 49  1 |  a = tested_negative
 13 14 |  b = tested_positive
```