Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

Program: Bachelor of Science in Computer Science and Engineering

Name : MD Rafiu Alam Rafi

Student ID : 20200204051

Section : A

Semester : 4.1

Course :Datamining Lab

Assignment:4

Questions:

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

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

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

    (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.

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

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

Answer:

    (i)   Dataset
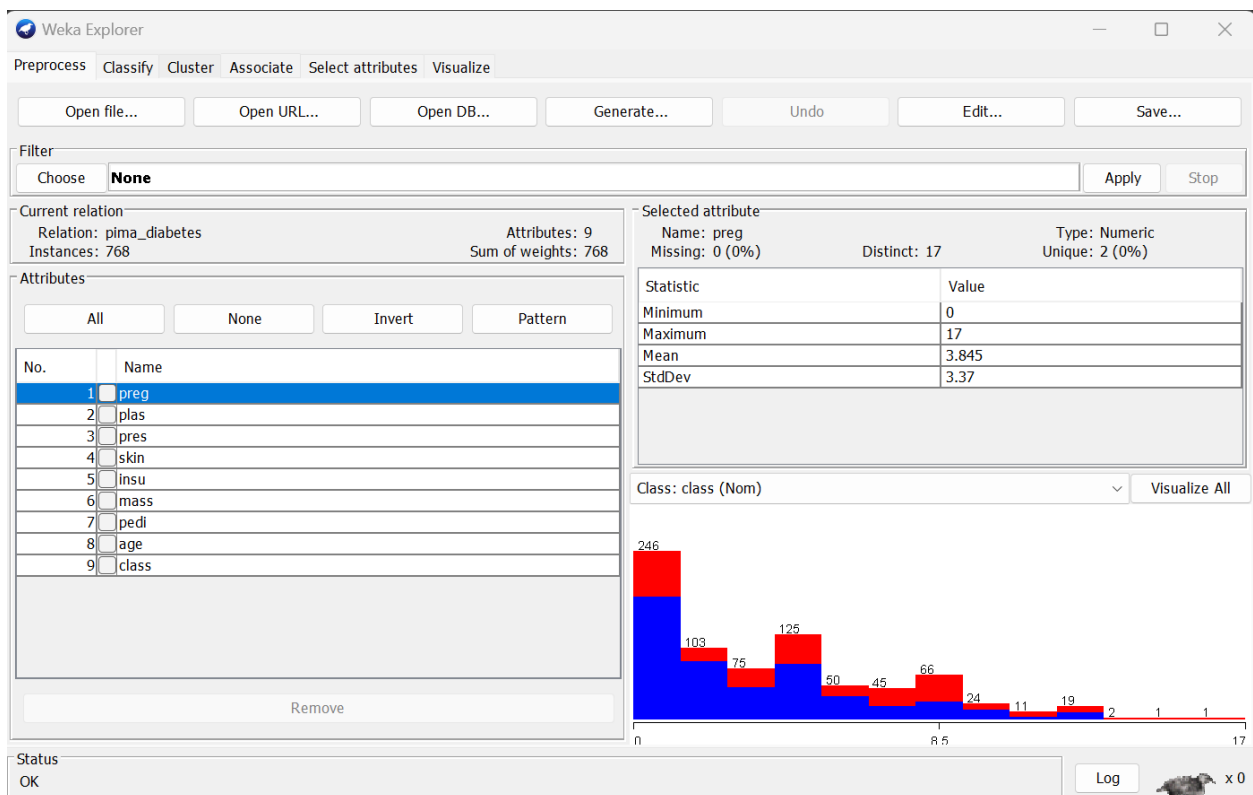
(ii)   for extracting :
       Preprocess > choose > filter > unsupervised > instance > removefolds >apply

(iii)   for stratified :choose>filter>supervised>instances>stratifiedremovefolds>apply

Using 8 folds validation

(iv)   for voting or majority method: Choose>meta>vote>classifier>add.

(v)+(vi) then 1 st with non stratified data result:
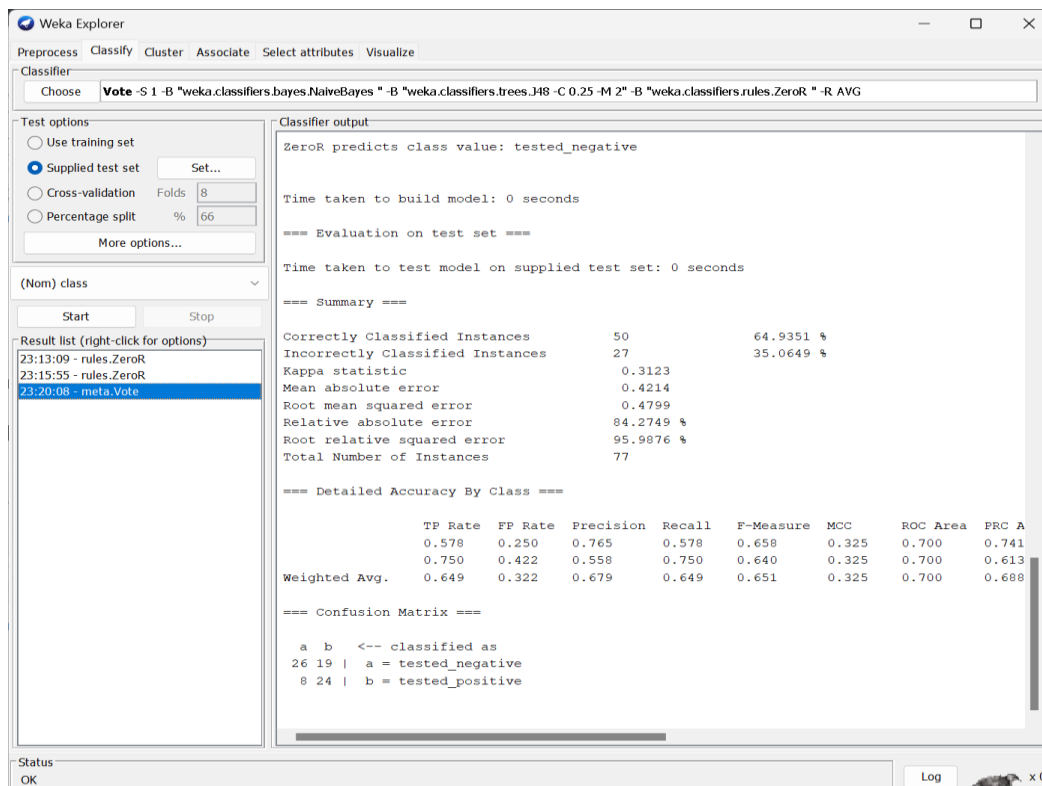
for stratified data:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose | Vote -S 1 -B "weka.classifiers.bayes.NaiveBayes " -B "weka.classifiers.trees.J48 -C 0.25 -M 2" -B "weka.classifiers.rules.ZeroR " -R AVG

Test options

○ Use training set
● Supplied test set | Set...
○ Cross-validation | Folds | 8
○ Percentage split | % | 66

More options...

(Nom) class

Start | Stop

Result list (right-click for options)

23:13:09 - rules.ZeroR
23:15:55 - rules.ZeroR
23:20:08 - meta.Vote
23:21:34 - meta.Vote

Classifier output

```
ZeroR predicts class value: tested_negative


Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances           8              100       %
Incorrectly Classified Instances         0                0       %
Kappa statistic                          1
Mean absolute error                      0.2637
Root mean squared error                  0.2862
Relative absolute error                 52.7363 %
Root relative squared error             57.2386 %
Total Number of Instances                8

=== Detailed Accuracy By Class ===

                  TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC A
                  1.000    0.000    1.000      1.000    1.000      1.000   1.000     1.000
                  1.000    0.000    1.000      1.000    1.000      1.000   1.000     1.000
Weighted Avg.     1.000    0.000    1.000      1.000    1.000      1.000   1.000     1.000

=== Confusion Matrix ===

 a b   <-- classified as
 4 0 | a = tested_negative
 0 4 | b = tested_positive
```
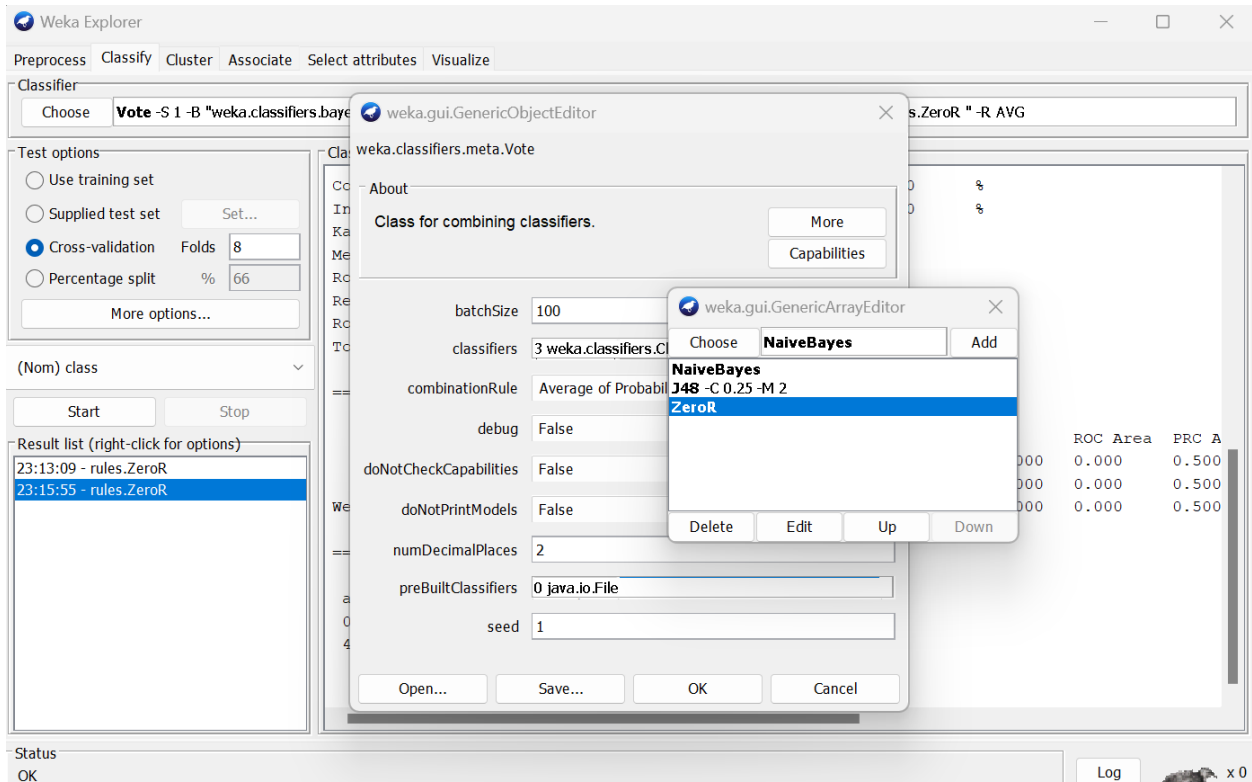
Status

OK

Log | x 0