

# 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 : Data Mining Lab

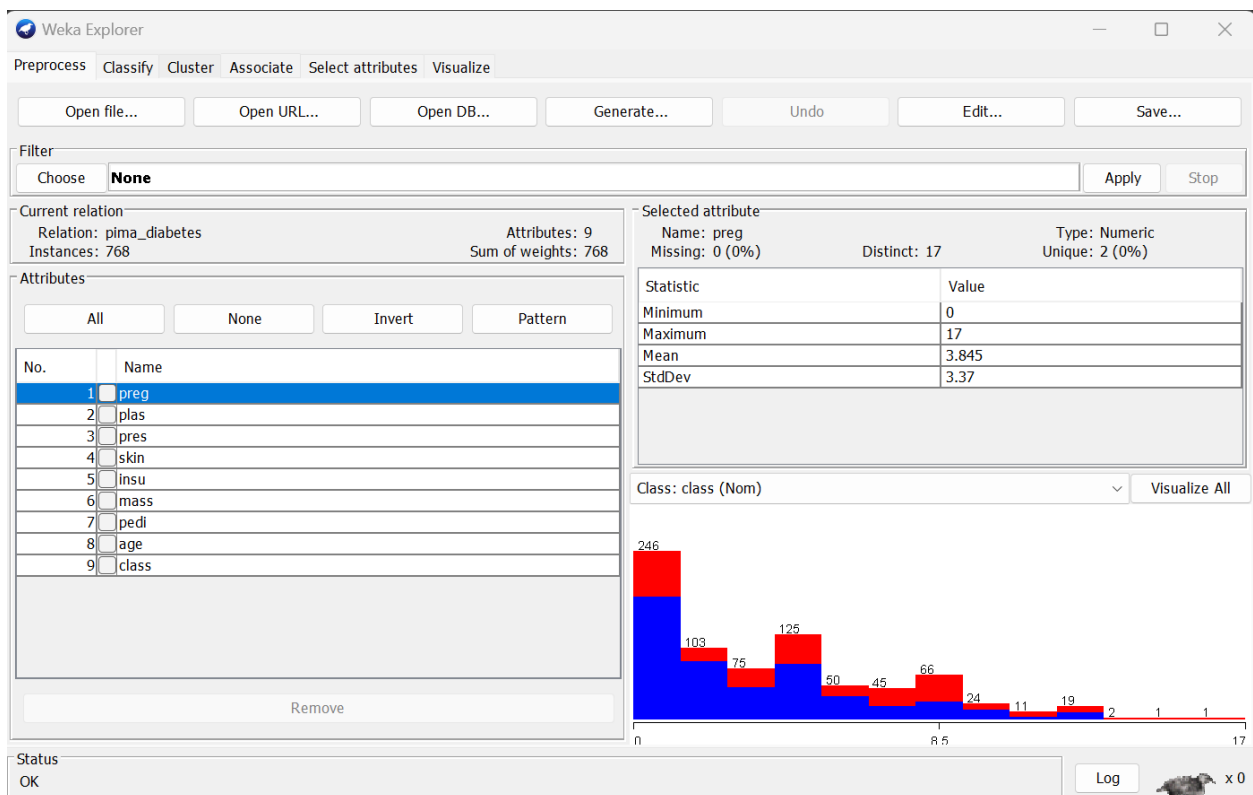
Assignment: 4

## 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

**Weka Explorer**

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

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter  
Choose **RemoveFolds -S 0 -N 10 -F 1** Apply Stop

Current relation  
Relation: plima\_diabetes-weka.filters.unsupervised.instance.... Attributes: 9 Sum of weights: 77  
Instances: 77

Attributes  
All None Invert Pattern

No.	Name
1	preg
2	plas
3	pres
4	skin
5	insu
6	mass
7	pedi
8	age
9	class

Remove

Selected attribute  
Name: preg  
Missing: 0 (0%) Distinct: 13 Type: Numeric  
Unique: 0 (0%)

Statistic	Value
Minimum	0
Maximum	13
Mean	4.805
StdDev	3.517

Class: class (Nom) Visualize All

Status OK Log x 0

**Weka Explorer**

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

Classifier  
Choose **ZeroR**

Test options  
☐ Use training set  
☐ Supplied test set Set...  
☒ Cross-validation Folds **10**  
☐ Percentage split % **66**  
 More options...

(Nom) class  
Start Stop

Result list (right-click for options)  
23:13:09 - rules.ZeroR

Classifier output

Correctly Classified Instances	45	58.4416 %
Incorrectly Classified Instances	32	41.5584 %
Kappa statistic	0	
Mean absolute error	0.4867	
Root mean squared error	0.4934	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	77	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC A
0.000	1.000	0.000	0.584	1.000	0.738	?	0.448	0.558
Weighted Avg.	0.584	0.584	?	0.584	?	?	0.448	0.488

=== Confusion Matrix ===

```

a b <-- classified as
45 0 | a = tested_negative
32 0 | b = tested_positive

```

Status OK Log x 0

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

The screenshot shows the Weka Explorer interface with the 'Filter' tab selected. The 'StratifiedRemoveFolds' filter is applied with parameters: -S 0 -N 10 -F 1. The 'Current relation' is 'pima\_diabetes-weka.filters.unsupervised.instance....' with 9 attributes and 8 instances. The 'Attributes' list shows 'preg' selected. The 'Selected attribute' section displays statistics for 'preg': Name: preg, Missing: 0 (0%), Distinct: 6, Type: Numeric, Unique: 5 (63%). A table of statistics for 'preg' is shown:

Statistic	Value
Minimum	0
Maximum	6
Mean	2.625
StdDev	1.996

The 'Class: class (Nom)' dropdown is set to 'Visualize All'. A bar chart shows the distribution of the 'class' attribute, with 6 instances in the 'tested\_negative' class (red bar) and 2 instances in the 'tested\_positive' class (blue bar).

Using 8 folds validation

The screenshot shows the Weka Explorer interface with the 'Classifier' tab selected. The 'ZeroR' classifier is chosen. The 'Test options' section shows 'Cross-validation' selected with 'Folds' set to 8. The 'Classifier output' section displays the following results:

Metric	Value	Unit
Correctly Classified Instances	0	%
Incorrectly Classified Instances	8	%
Kappa statistic	-1	
Mean absolute error	0.5556	
Root mean squared error	0.5556	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	8	

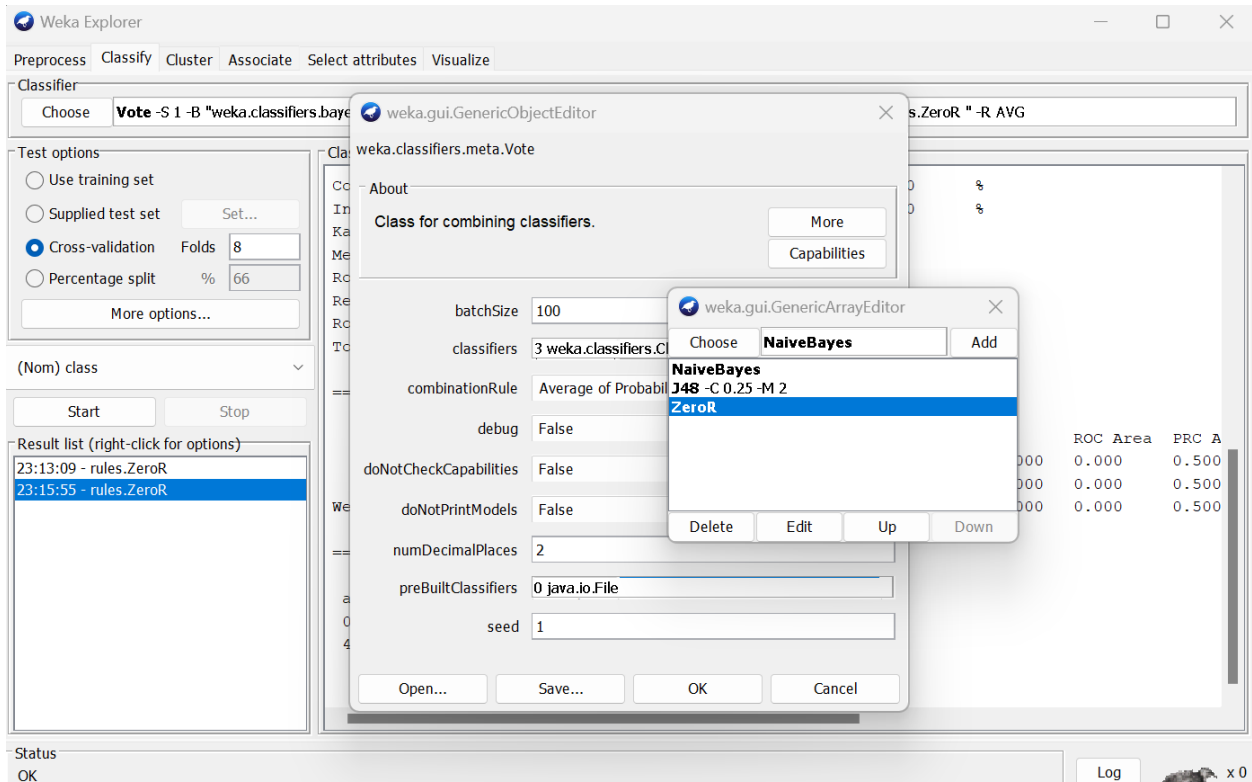
The 'Detailed Accuracy By Class' section shows the following results:

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC A
tested_negative	0.000	1.000	0.000	0.000	0.000	-1.000	0.000	0.500
tested_positive	0.000	1.000	0.000	0.000	0.000	-1.000	0.000	0.500
Weighted Avg.	0.000	1.000	0.000	0.000	0.000	-1.000	0.000	0.500

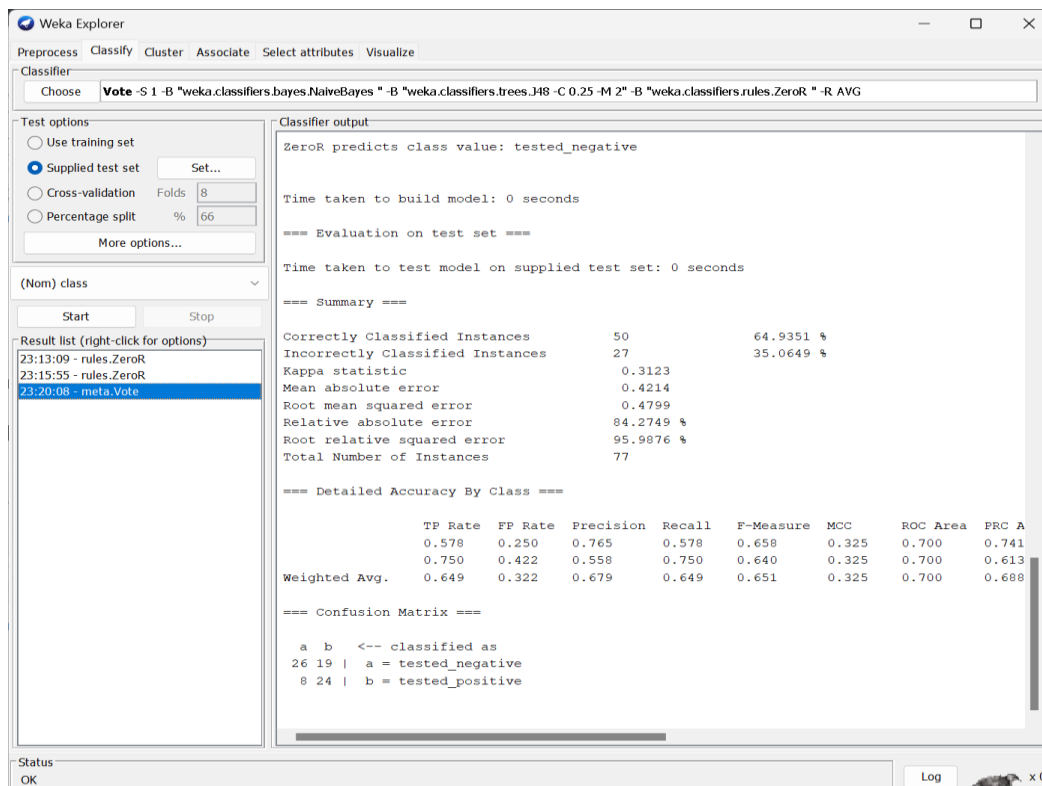
The 'Confusion Matrix' section shows the following results:

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

(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

☐ Cross-validation Folds

☐ Percentage split %

(Nom) class

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

 x 0