AHSANULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

Department of Computer Science and Engineering

Data Warehousing and Mining Lab

Lab Leport

Assignment No: 02

Date of Submission: 30/07/2025

Submited by-

Name: Ali Faruk Shihab

ID: 20210204088

Section: B2

Submitted to-

Mr. Saha Reno

Ms. Tahsina Muthaki

Q1. Create a Custom Dataset Which Will Have 8 Attributes: 3 Numeric, 4 Nominal & 1 Class (3 Class Values)

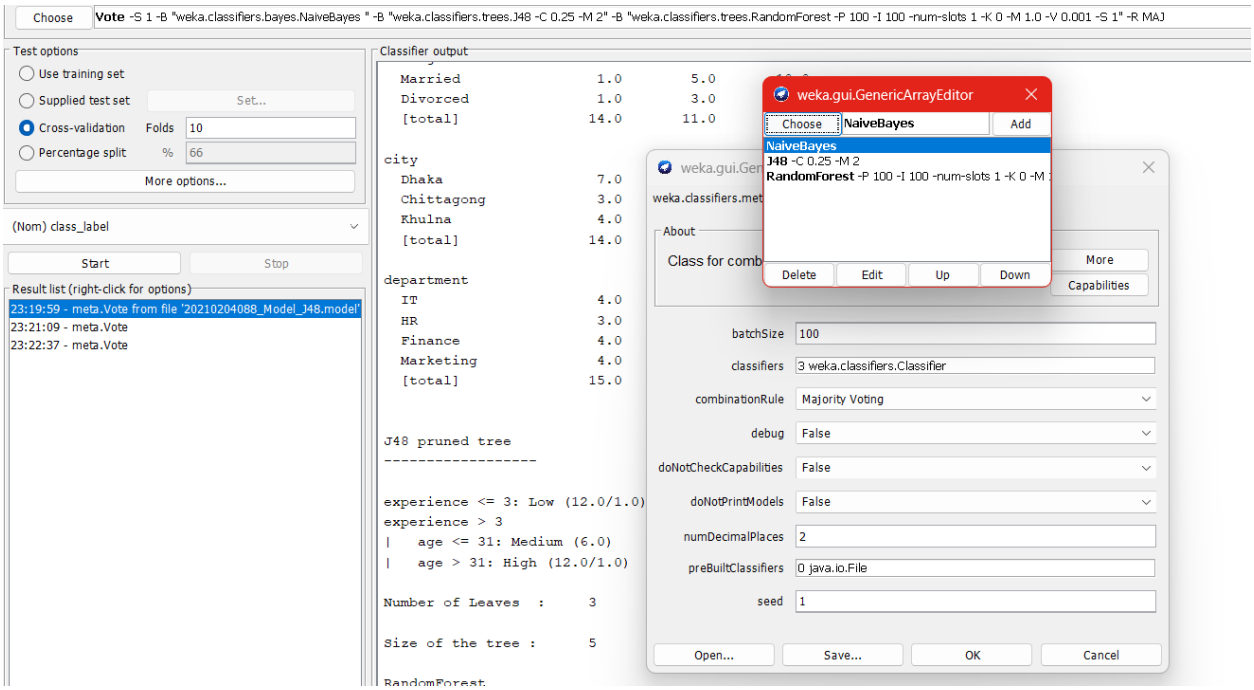
```
@RELATION custom_dataset

@ATTRIBUTE age NUMERIC
@ATTRIBUTE salary NUMERIC
@ATTRIBUTE experience NUMERIC
@ATTRIBUTE education {HighSchool,Bachelor,Master,PhD}
@ATTRIBUTE marital_status {Single,Married,Divorced}
@ATTRIBUTE city {Dhaka,Chittagong,Khulna}
@ATTRIBUTE department {IT,HR,Finance,Marketing}
@ATTRIBUTE class_label {Low,Medium,High}
```

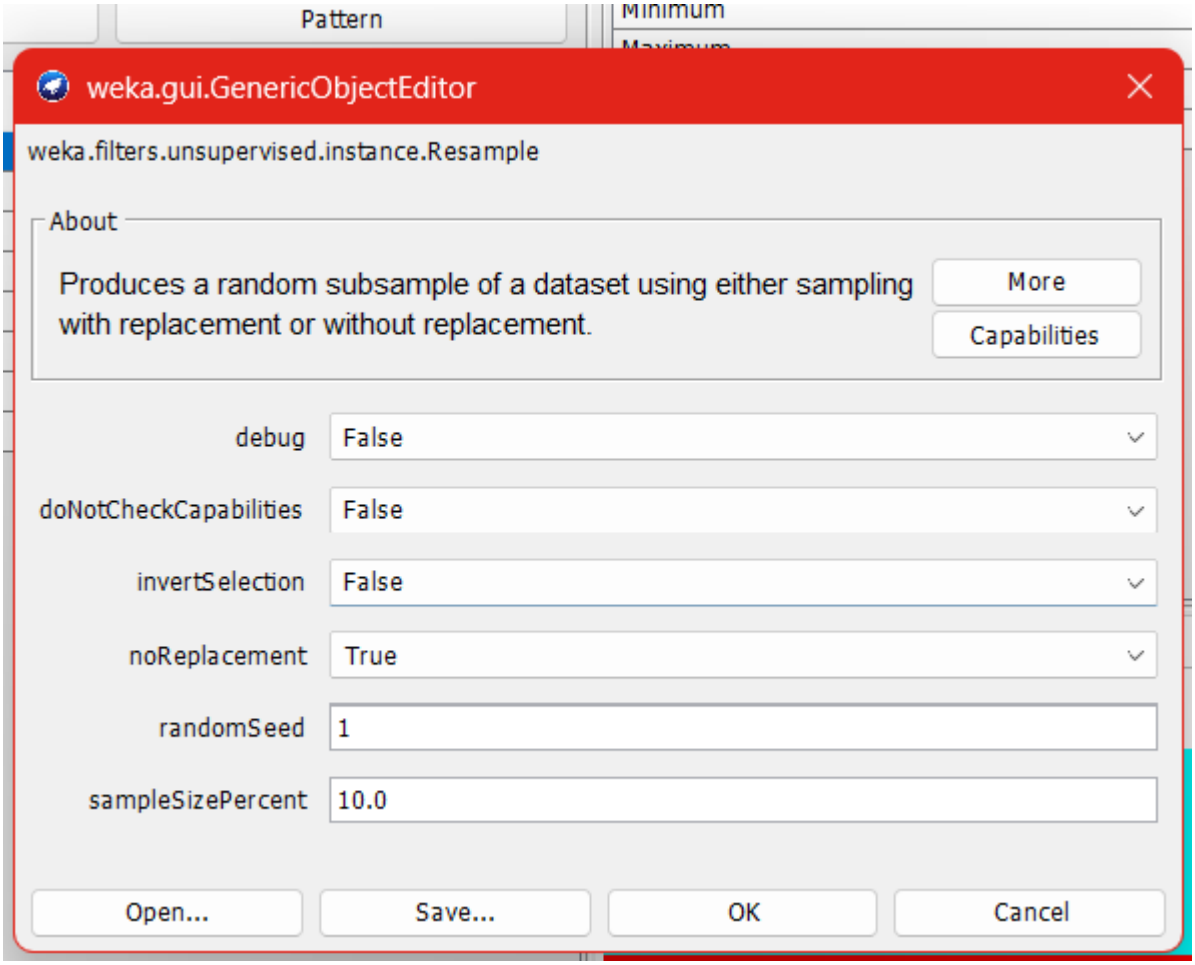
Q2. Create 30 Instances of That Dataset

```
@DATA
22,25000,1,Bachelor,Single,Dhaka,IT,Low
28,40000,3,Master,Married,Chittagong,Finance,Medium
35,60000,10,PhD,Married,Dhaka,HR,High
30,45000,5,Master,Single,Khulna,IT,Medium
26,30000,2,Bachelor,Single,Dhaka,Marketing,Low
40,75000,15,PhD,Married,Chittagong,Finance,High
29,38000,4,Master,Divorced,Khulna,HR,Medium
33,50000,6,Bachelor,Married,Dhaka,Marketing,High
24,28000,2,HighSchool,Single,Khulna,IT,Low
37,68000,12,Master,Married,Dhaka,Finance,High
27,32000,3,Bachelor,Single,Chittagong,Marketing,Low
32,55000,7,PhD,Divorced,Khulna,HR,High
25,31000,2,HighSchool,Single,Dhaka,Finance,Low
31,42000,5,Bachelor,Married,Chittagong,IT,Medium
36,60000,10,Master,Married,Dhaka,Marketing,High
23,26000,1,Bachelor,Single,Khulna,Finance,Low
38,72000,14,PhD,Married,Dhaka,HR,High
34,58000,8,Master,Divorced,Chittagong,Finance,High
21,24000,0,HighSchool,Single,Dhaka,Marketing,Low
29,37000,4,Bachelor,Single,Khulna,IT,Medium
35,62000,9,Master,Married,Chittagong,HR,High
28,39000,3,HighSchool,Single,Dhaka,Finance,Low
33,51000,7,Master,Married,Khulna,Marketing,Medium
30,46000,5,Bachelor,Divorced,Chittagong,IT,Medium
26,34000,2,HighSchool,Single,Dhaka,HR,Low
39,70000,13,PhD,Married,Khulna,Finance,High
24,27000,1,Bachelor,Single,Chittagong,IT,Low
31,43000,6,Master,Married,Dhaka,Marketing,Medium
27,35000,3,HighSchool,Single,Khulna,HR,Low
36,64000,11,PhD,Married,Dhaka,Finance,High
```

Q3. Construct a Classification Model using Voting Algorithm [Classifiers: Naive Bayes, J48, Random Forest; Apply "Majority Voting" Technique], Use 10-Fold Cross Validation



Q4. Then, Extract 1 Fold from the Training Set, Which Will Not Guarantee That All Class Values Are Present in That Fold and the Fold Remains Balanced



Q5. Finally, Supply That Fold as Test Set and Show the Model's (Voting) Performance on That Test Set

Classifier output

```
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities


Time taken to build model: 0.02 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Correctly Classified Instances           3               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                3

=== 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     Low
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     Medium
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     High
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000

=== Confusion Matrix ===

 a b c   <-- classified as
 1 0 0 | a = Low
 0 1 0 | b = Medium
 0 0 1 | c = High
```