

AHSANULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY (AUST)
141 & 142, Love Road, Tejgaon Industrial Area, Dhaka-1208.



Department of Computer Science and Engineering
Program: Bachelor of Science in Computer Science and Engineering

Course No: 4142 Course Title: Data Warehousing and Mining Lab

Assignment 1

Date of Submission: 25/5/2024

Submitted by,

Name: **Md Rafiu Alam Rafi**

Id: 20200204051

Section: A

Task-1:

First I created a custom dataset and save it as **20200204051_originalTraining.arff**,

```
@relation Employee_Bonus

@attribute performance {poor, satisfactory, good, excellent}
@attribute experience numeric
@attribute satisfaction {low, medium, high}
@attribute education {high_school, bachelors, masters, phd}
@attribute age numeric
@attribute department {sales, marketing, finance, engineering, HR}
@attribute salary numeric
@attribute bonus {yes, no}

@data
poor,2,low,high_school,25,sales,30000,yes
satisfactory,5,medium,bachelors,30,marketing,40000,no
good,8,high,masters,35,finance,50000,yes
excellent,10,high,phd,40,engineering,60000,yes
satisfactory,4,low,bachelors,28,HR,35000,no
excellent,12,high,masters,45,engineering,70000,yes
poor,1,low,high_school,22,sales,28000,no
good,7,medium,bachelors,32,marketing,45000,yes
satisfactory,6,high,masters,38,finance,55000,no
good,9,medium,bachelors,33,engineering,65000,yes
poor,3,low,high_school,27,HR,32000,no
excellent,11,high,phd,42,engineering,68000,yes
satisfactory,5,medium,masters,31,finance,42000,no
good,8,high,bachelors,34,marketing,47000,yes
poor,2,low,high_school,26,sales,31000,no
excellent,12,high,masters,48,engineering,72000,yes
satisfactory,4,medium,bachelors,29,HR,38000,no
good,8,high,masters,36,finance,51000,yes
satisfactory,6,medium,masters,39,engineering,59000,yes
excellent,10,high,phd,41,marketing,48000,yes
satisfactory,3,low,bachelors,26,HR,33000,no
poor,1,low,high_school,23,sales,29000,no
good,7,medium,masters,37,finance,52000,yes
satisfactory,5,medium,bachelors,30,engineering,56000,yes
excellent,9,high,phd,44,engineering,73000,yes
good,8,high,masters,35,marketing,49000,yes
satisfactory,6,medium,masters,38,HR,37000,no
poor,2,low,high_school,24,sales,30000,no
excellent,11,high,masters,46,engineering,71000,yes
good,7,medium,bachelors,31,finance,53000,yes
satisfactory,4,low,masters,27,HR,34000,no
poor,3,low,high_school,28,sales,32000,no
good,9,high,masters,36,marketing,50000,yes
excellent,12,high,phd,43,engineering,74000,yes
satisfactory,5,medium,bachelors,32,HR,39000,no
good,8,high,masters,37,finance,54000,yes
excellent,10,high,masters,47,engineering,68000,yes
satisfactory,6,medium,bachelors,33,sales,31000,no
```

And a test dataset and saved it as 20200204051_originalTest.arff

```
@relation Employee_Bonus

@attribute performance {poor, satisfactory, good, excellent}
@attribute experience numeric
@attribute satisfaction {low, medium, high}
@attribute education {high_school, bachelors, masters, phd}
@attribute age numeric
@attribute department {sales, marketing, finance, engineering, HR}
@attribute salary numeric
@attribute bonus {yes, no}

@data
poor,1,low,high_school,23,HR,32000,no
good,6,medium,masters,35,finance,50000,yes
satisfactory,4,medium,bachelors,29,HR,36000,no
excellent,10,high,masters,42,engineering,68000,yes
poor,2,low,high_school,26,sales,29000,no
good,8,medium,bachelors,30,marketing,48000,yes
satisfactory,5,high,masters,37,finance,53000,no
excellent,11,high,phd,44,engineering,70000,yes
```

Task-2:

Then I selected the dataset from ... preprocess > Open file > Desktop >

20200204051_originalTraining.arff > Open

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose Discretize -B 10 -M -1.0 -R first-last -precision 6 Apply Stop

Current relation: Relation: Employee_Bonus, Instances: 38, Attributes: 8, Sum of weights: 38

Attributes: All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> performance
2	<input type="checkbox"/> experience
3	<input type="checkbox"/> satisfaction
4	<input type="checkbox"/> education
5	<input type="checkbox"/> age
6	<input type="checkbox"/> department
7	<input type="checkbox"/> salary
8	<input type="checkbox"/> bonus

Remove

Selected attribute: Name: performance, Missing: 0 (0%), Distinct: 4, Type: Nominal, Unique: 0 (0%)

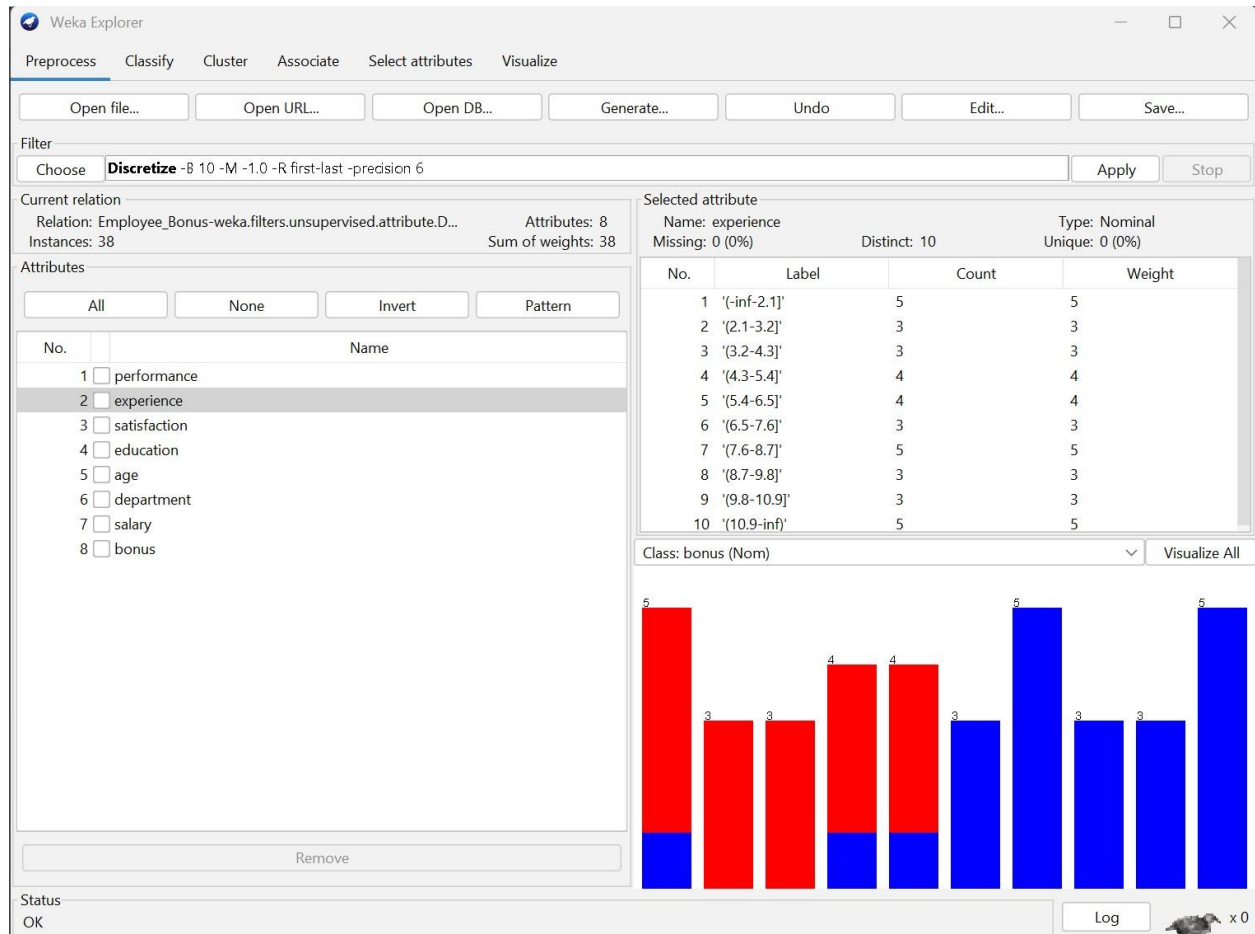
No.	Label	Count	Weight
1	poor	7	7
2	satisfactory	12	12
3	good	10	10
4	excellent	9	9

Class: bonus (Nom) Visualize All

Status: OK Log x 0

Task-3:

Then I normalized the **original and test** dataset , Preprocess>Choose > filters> unsupervised >Discretize >Apply



Then saved as ,

20200204051_modTraining.arff

&

20200204051_modTest.arff

Task-4-5:

Then I classified the dataset , (selecting “20200204051_modTraining.arff”)

Classify > Choose > NaiveBayes > Supplied test data (set) > open file > 20200204051_modTest.arff > Start

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'NaiveBayes'. Under 'Test options', 'Supplied test set' is selected with a 'Set...' button. The 'Percentage split' is set to 80%. The 'Start' button is visible. The 'Result list' on the left shows several entries, with '21:48:04 - misc.InputMappedClassifier' selected. The 'Classifier output' pane displays the following results:

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	7	87.5	%
Incorrectly Classified Instances	1	12.5	%
Kappa statistic	0.75		
Mean absolute error	0.1198		
Root mean squared error	0.2956		
Relative absolute error	23.9547	%	
Root relative squared error	58.46	%	
Total Number of Instances	8		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.250	0.800	1.000	0.889	0.775	1.000	1.000	yes
	0.750	0.000	1.000	0.750	0.857	0.775	1.000	1.000	no
Weighted Avg.	0.875	0.125	0.900	0.875	0.873	0.775	1.000	1.000	

=== Confusion Matrix ===

```
a b  <-- classified as
4 0 | a = yes
1 3 | b = no
```

The status bar at the bottom shows 'Status OK' and a 'Log' button.

Then I right clicked on my model and saved the model as "20200204051_model"