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: CSE 4142
Course Title: Data Warehousing and Mining Lab

Assignment 3

Date of Submission: 23/6/2024

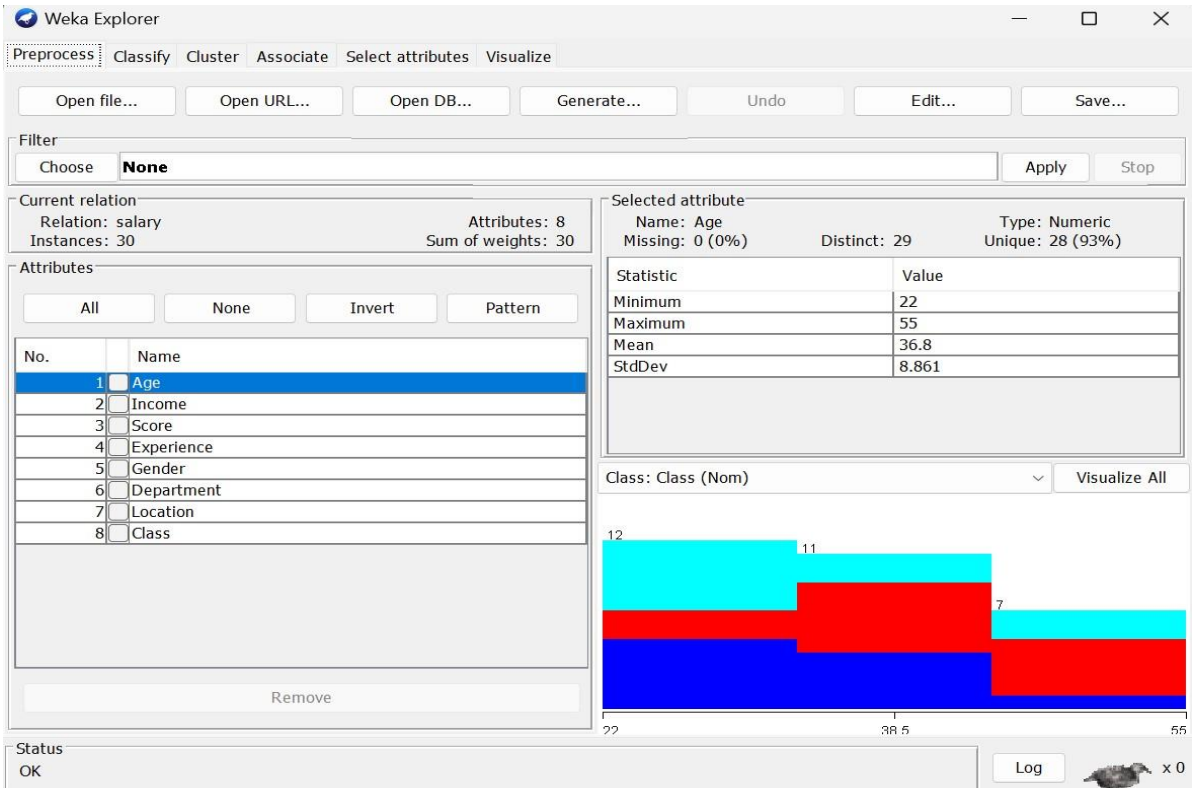
Submitted by,

Name: MD Rafiu Alam Rafi
Id: 20200204051
Section: A

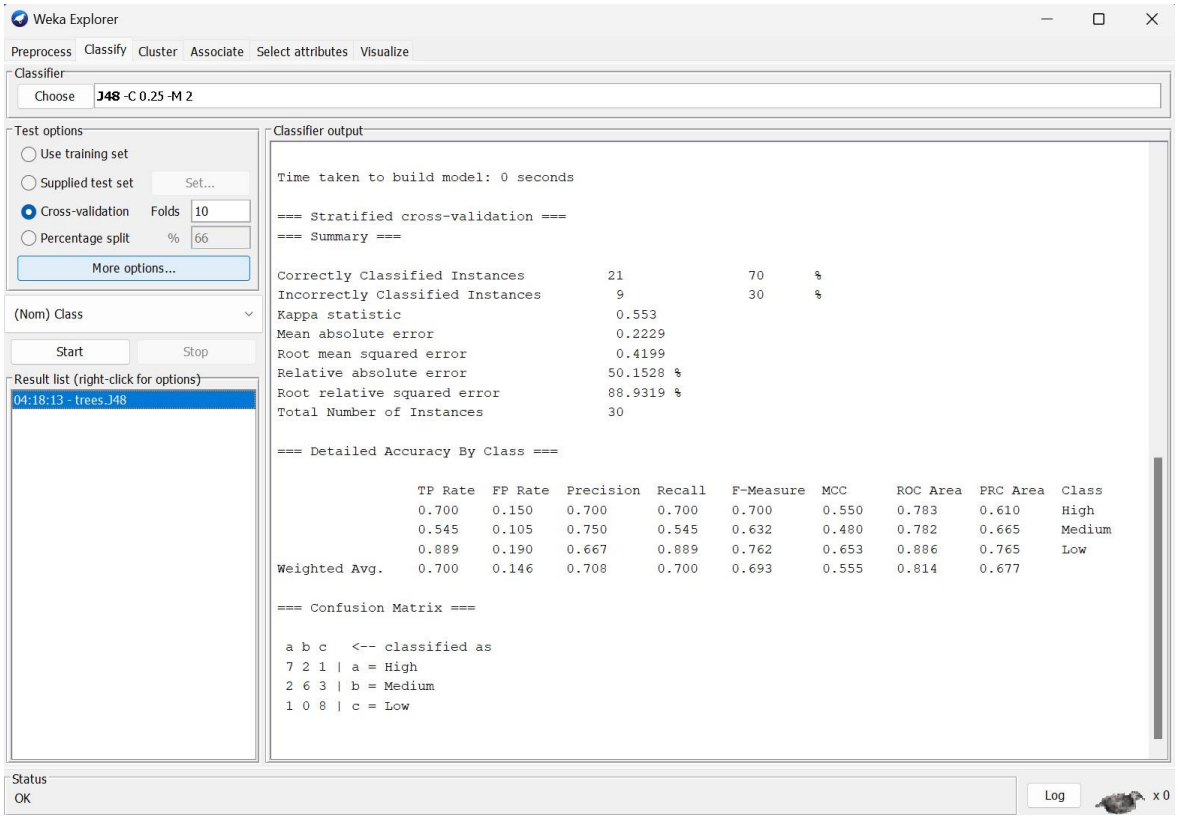
Task Processes:

(i) **Creating a Custom Dataset Which Will Have 8 Attributes: 4 Numeric, 3 Nominal & 1 Class (3 Class Values).** (ii) **Creating 30 Instances of That Dataset.**

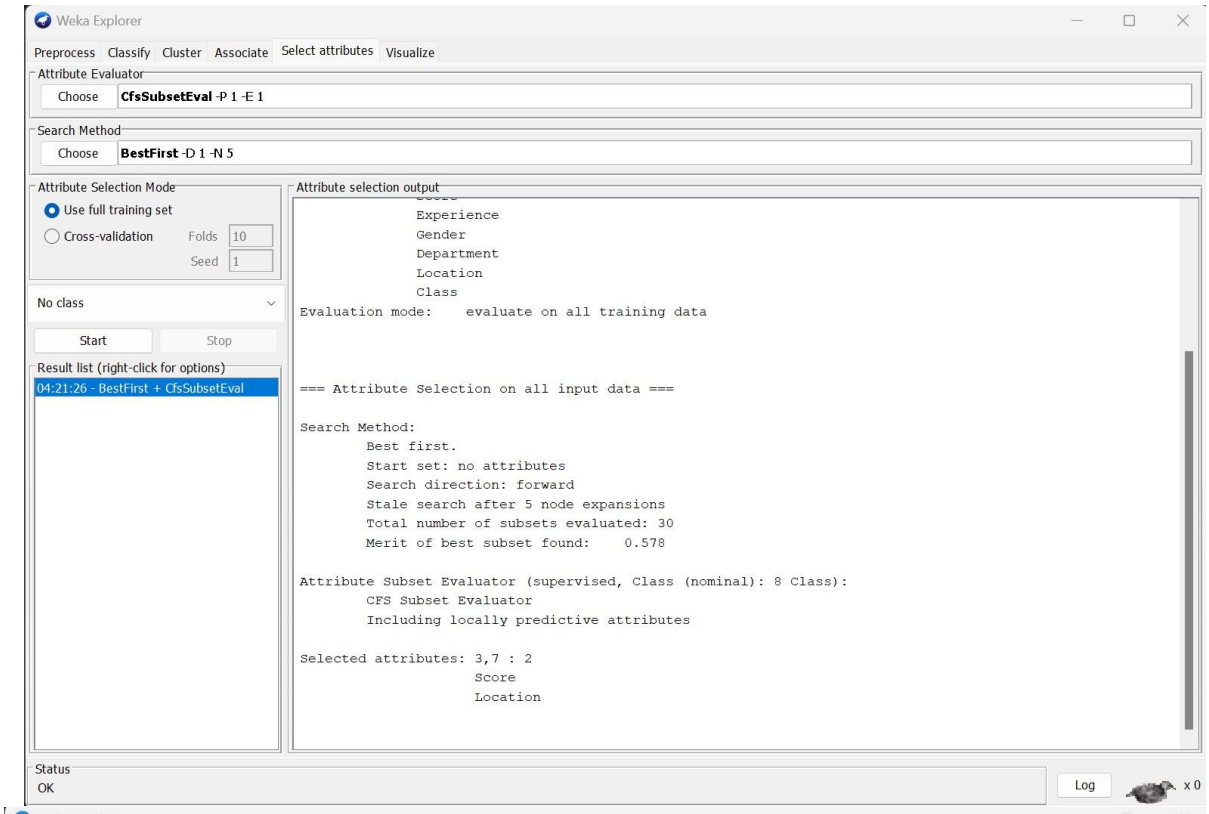
```
@relation salary

@attribute Age numeric
@attribute Income numeric
@attribute Score numeric
@attribute Experience numeric
@attribute Gender {Male, Female, Other}
@attribute Department {Sales, Engineering, Marketing}
@attribute Location {Urban, Suburban, Rural}
@attribute Class {High, Medium, Low}

@data
34, 54000, 75, 10, Male, Sales, Urban, Medium
28, 62000, 88, 5, Female, Engineering, Suburban, High
45, 78000, 52, 20, Male, Marketing, Rural, Low
50, 85000, 60, 25, Female, Sales, Urban, Medium
23, 32000, 70, 2, Other, Engineering, Urban, High
36, 48000, 82, 12, Male, Marketing, Suburban, Medium
29, 74000, 67, 7, Female, Sales, Rural, Low
40, 96000, 90, 18, Other, Engineering, Urban, High
55, 87000, 77, 30, Male, Marketing, Suburban, Medium
32, 45000, 62, 8, Female, Sales, Rural, Low
38, 66000, 85, 15, Male, Engineering, Urban, High
47, 92000, 73, 22, Other, Marketing, Suburban, Medium
22, 31000, 58, 1, Female, Sales, Rural, Low
42, 79000, 80, 17, Male, Engineering, Urban, High
30, 57000, 65, 6, Other, Engineering, Suburban, Medium
48, 83000, 77, 23, Male, Marketing, Rural, Low
35, 71000, 84, 11, Female, Sales, Urban, High
31, 52000, 76, 4, Male, Engineering, Suburban, Medium
27, 40000, 68, 3, Other, Marketing, Rural, Low
49, 88000, 79, 24, Female, Sales, Urban, High
37, 60000, 72, 13, Male, Engineering, Suburban, Medium
26, 55000, 83, 2, Female, Marketing, Rural, High
41, 85000, 66, 19, Other, Sales, Urban, Low
39, 90000, 74, 16, Male, Engineering, Suburban, Medium
33, 68000, 81, 9, Female, Marketing, Rural, High
29, 75000, 64, 7, Other, Sales, Urban, Low
43, 82000, 78, 21, Male, Engineering, Suburban, Medium
25, 47000, 89, 4, Female, Marketing, Rural, High
44, 86000, 61, 22, Other, Sales, Urban, Low
46, 91000, 87, 20, Male, Engineering, Suburban, Medium
```

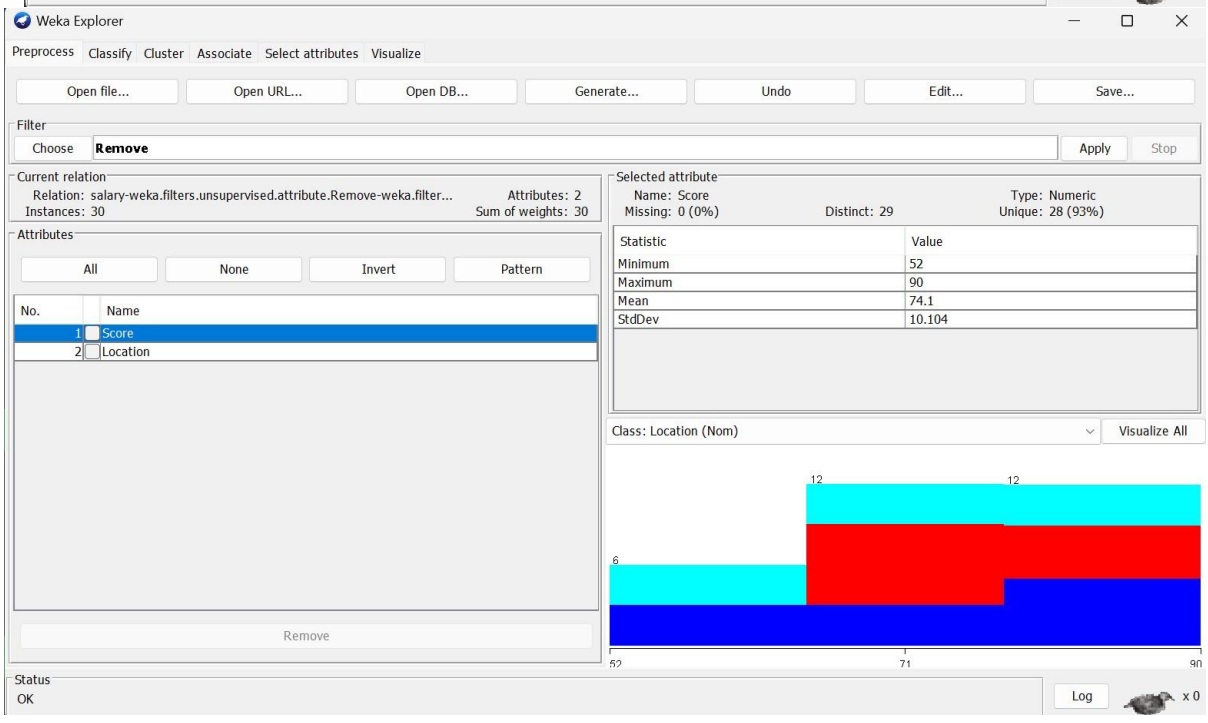


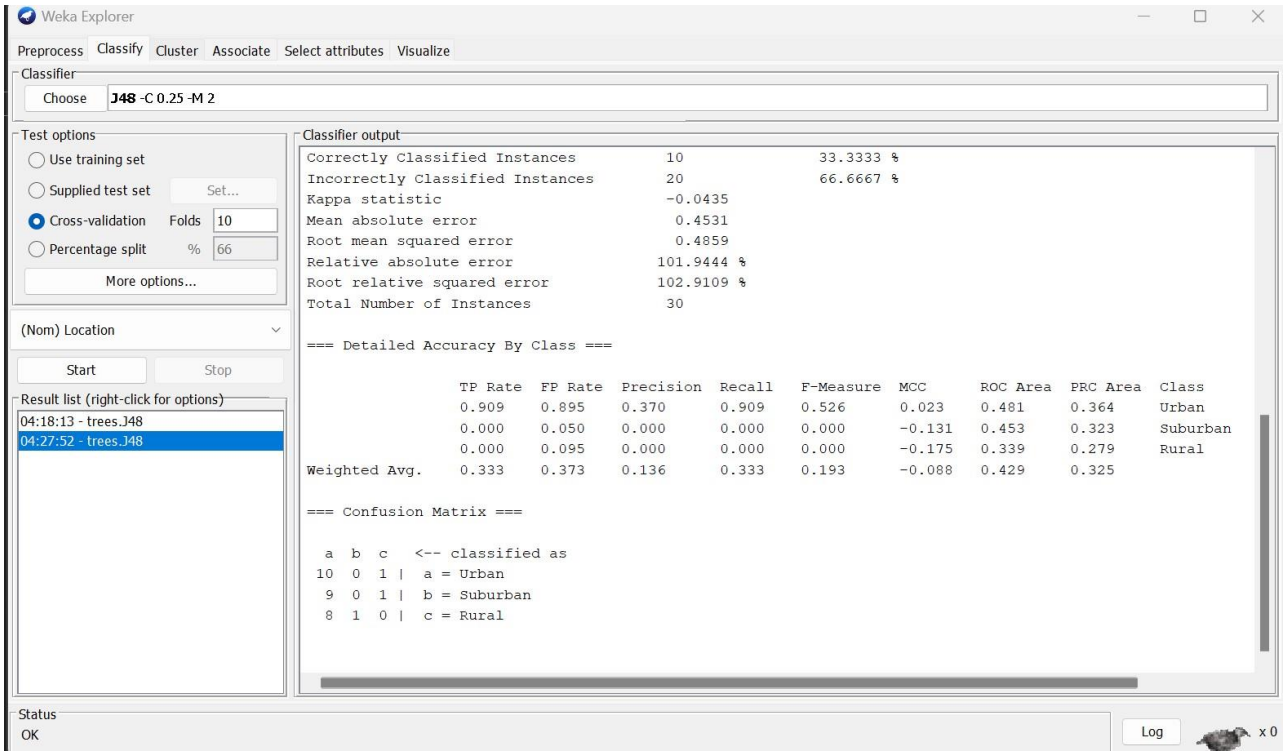
(iii) **Constructing a Classification Model using J48 Decision Tree Algorithm using 10-Fold Cross Validation.**



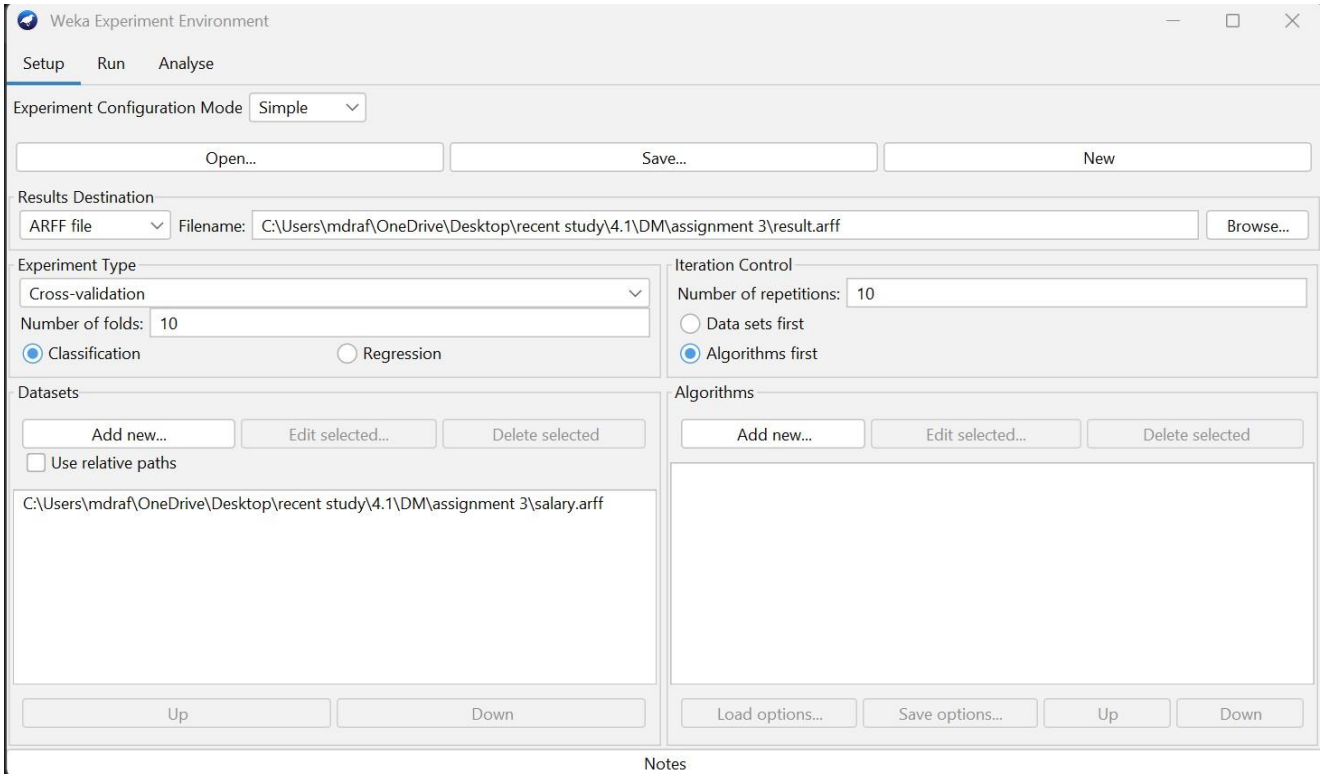
(iv) **Using Filter Method to find out those Attributes for Which the J48 Model Performs the Best (Least Number of "Inaccurately Classified Instances"). All Other Attributes are Removed.**

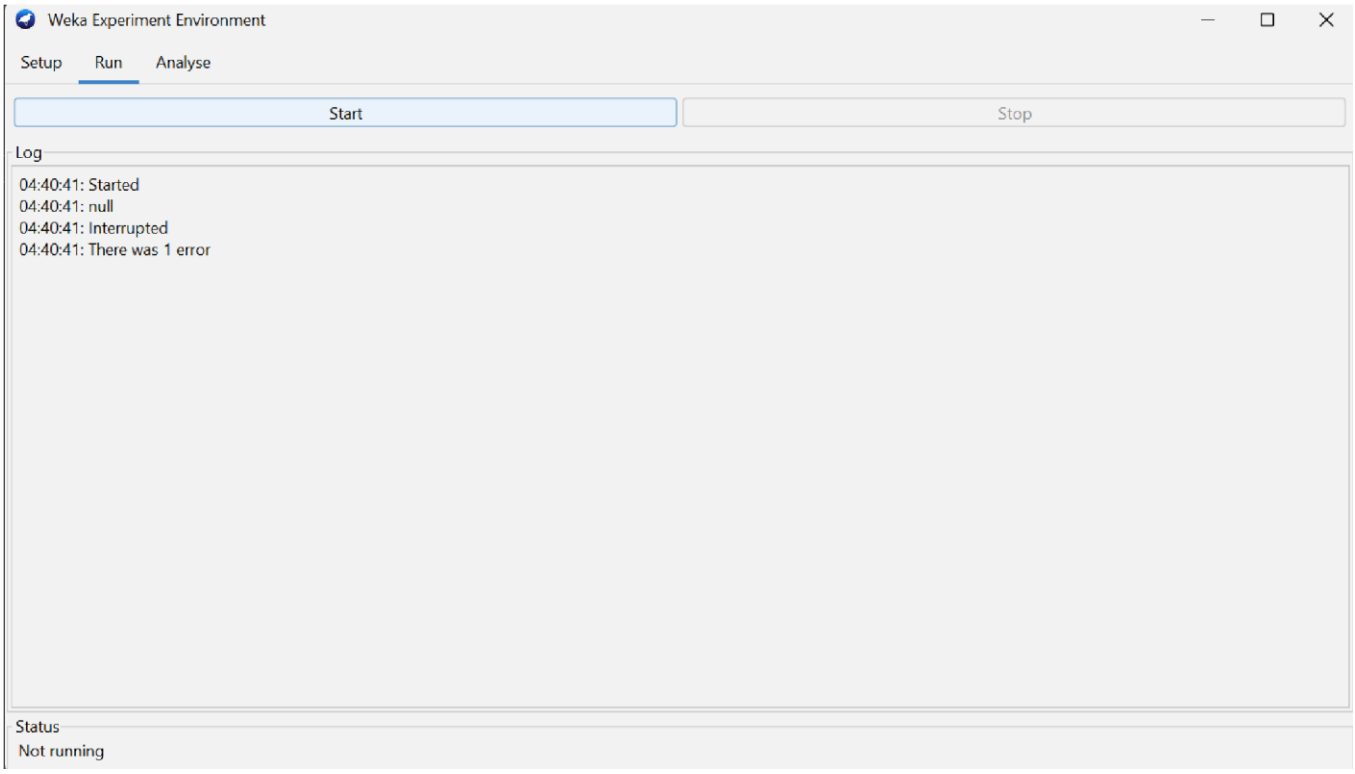

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Attribute Evaluator
Choose CfsSubsetEval -P 1 -E 1
Search Method
Choose BestFirst -D 1 -N 5
Attribute Selection Mode
Use full training set
Cross-validation Folds 10
Seed 1
No class
Start Stop
Result list (right-click for options)
04:21:26 - BestFirst + CfsSubsetEval
Attribute selection output
Experience
Gender
Department
Location
Class
Evaluation mode: evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 30
Merit of best subset found: 0.578
Attribute Subset Evaluator (supervised, Class (nominal): 8 Class):
CFS Subset Evaluator
Including locally predictive attributes
Selected attributes: 3,7 : 2
Score
Location
Status
OK
Log
x 0

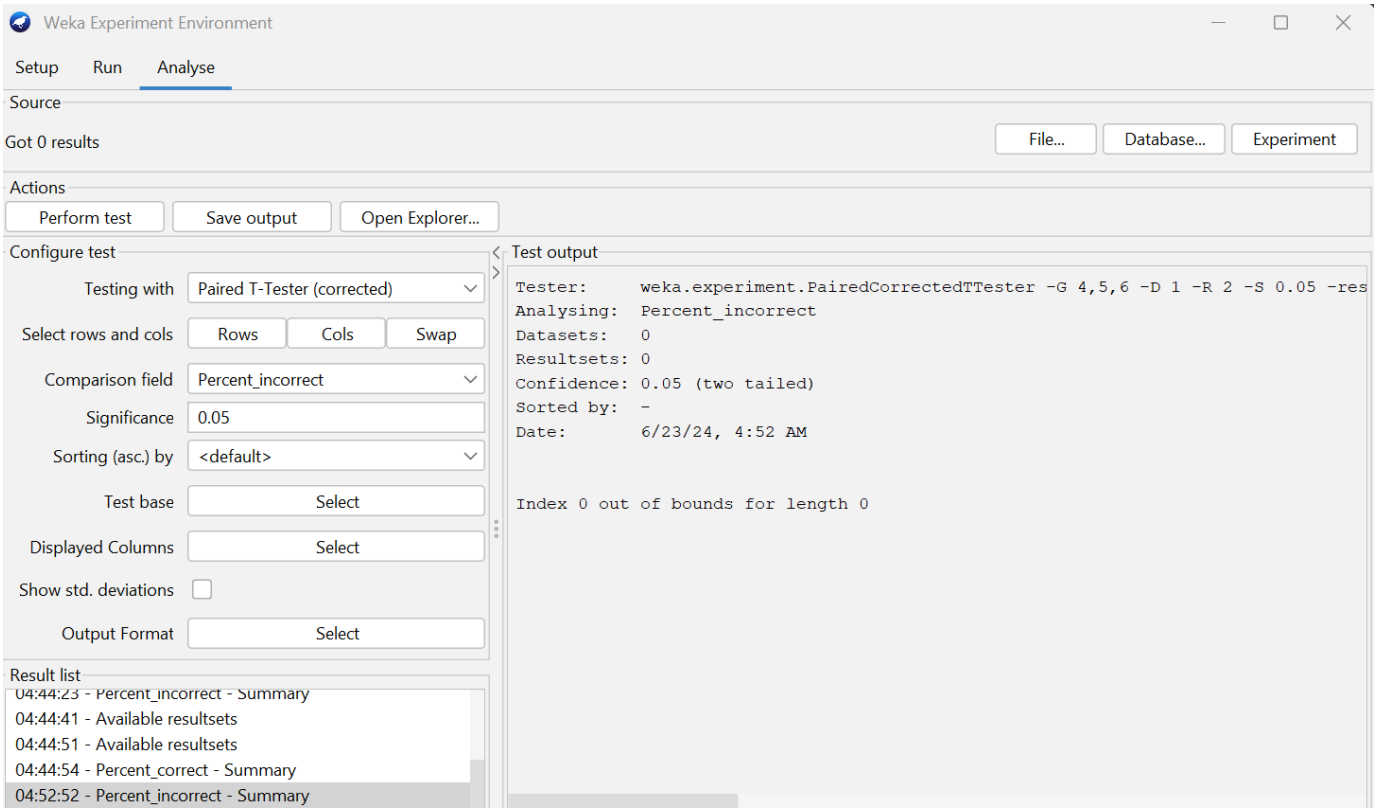

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Open file... Open URL... Open DB... Generate... Undo Edit... Save...
Filter
Choose Remove
Apply Stop
Current relation
Relation: salary-weka.filters.unsupervised.attribute.Remove-weka.filter...
Instances: 30
Attributes: 2
Sum of weights: 30
Attributes
All None Invert Pattern
No. Name
1 Score
2 Location
Remove
Selected attribute
Name: Score
Missing: 0 (0%)
Distinct: 29
Type: Numeric
Unique: 28 (93%)
Statistic Value
Minimum 52
Maximum 90
Mean 74.1
StdDev 10.104
Class: Location (Nom)
Visualize All
12
12
6
52
71
90
Status
OK
Log
x 0



(v) **Lastly, finding out The Percentage of Incorrect Classification for Each Folds using Weka Experiment Environment.**




Weka Experiment Environment
Setup
Run
Analyse
Start
Stop
Log
04:40:41: Started
04:40:41: null
04:40:41: Interrupted
04:40:41: There was 1 error
Status
Not running


Weka Experiment Environment
Setup
Run
Analyse
Source
Got 0 results
File...
Database...
Experiment
Actions
Perform test
Save output
Open Explorer...
Configure test
Testing with
Paired T-Tester (corrected)
Select rows and cols
Rows
Cols
Swap
Comparison field
Percent_incorrect
Significance
0.05
Sorting (asc.) by
<default>
Test base
Select
Displayed Columns
Select
Show std. deviations
Output Format
Select
Result list
04:44:41 - Available resultsets
04:44:51 - Available resultsets
04:44:54 - Percent_correct - Summary
04:52:52 - Percent_incorrect - Summary
Test output
Tester: weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -res
Analysing: Percent_incorrect
Datasets: 0
Resultsets: 0
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 6/23/24, 4:52 AM
Index 0 out of bounds for length 0