

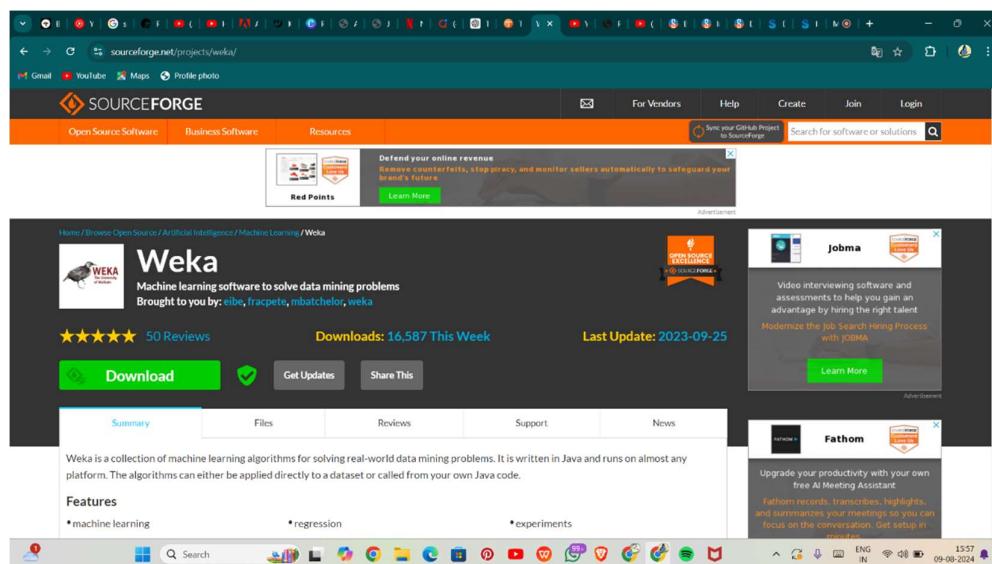
## Experiment 1

TITLE: Introduction & installation of WEKA tool.

THEORY:

Screenshots of WEKA installation:-

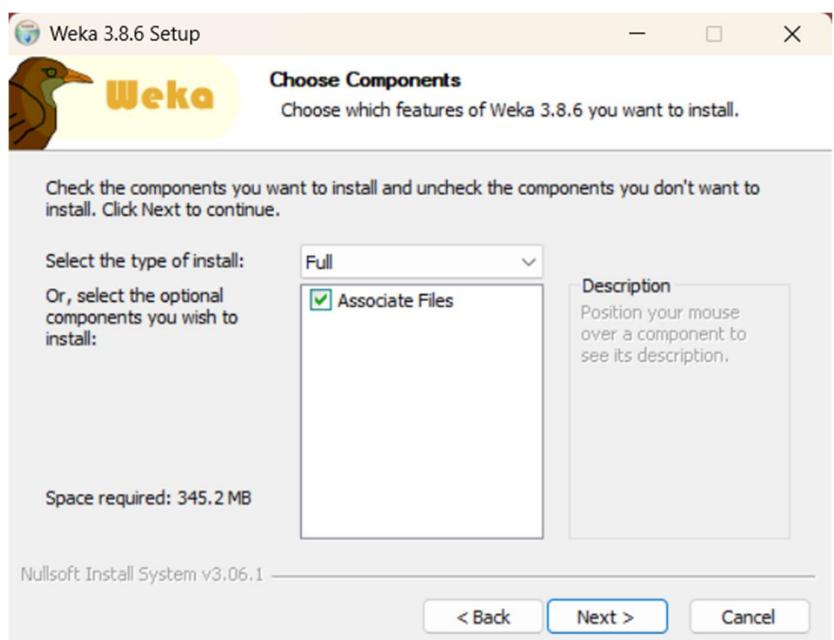
STEP1:



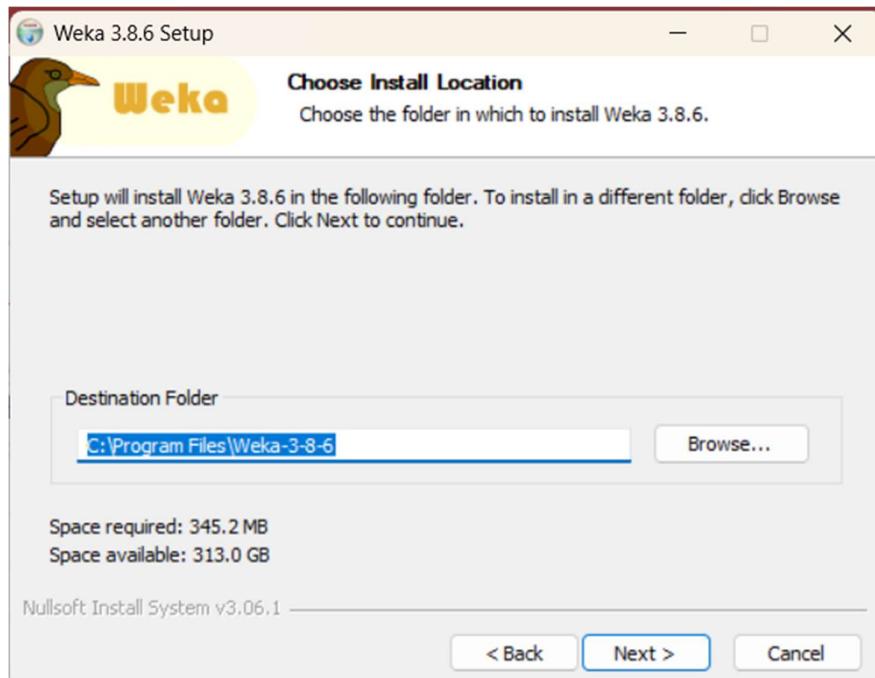
## STEP2:



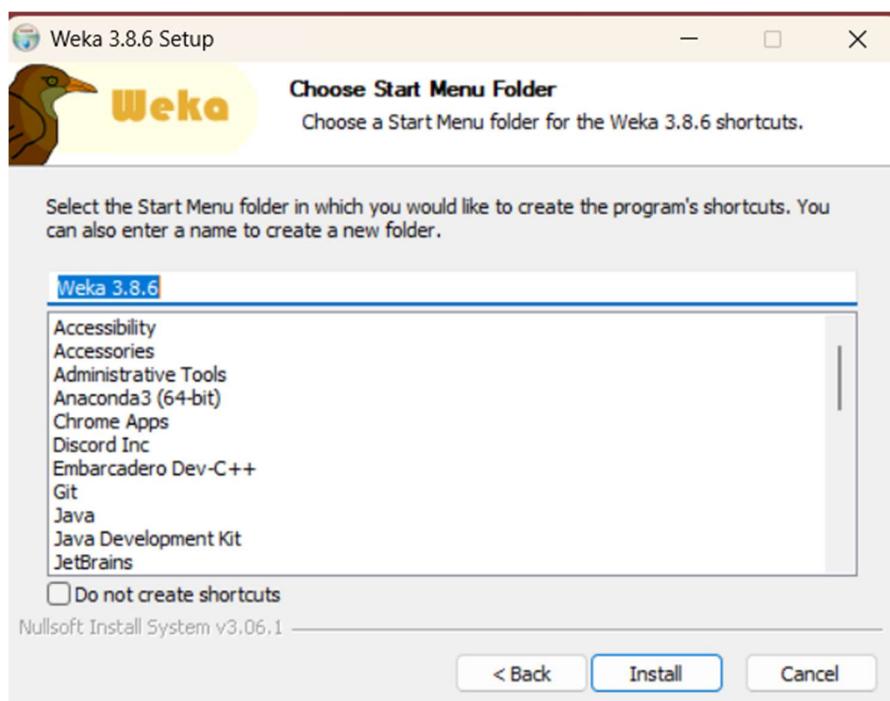
## STEP3:



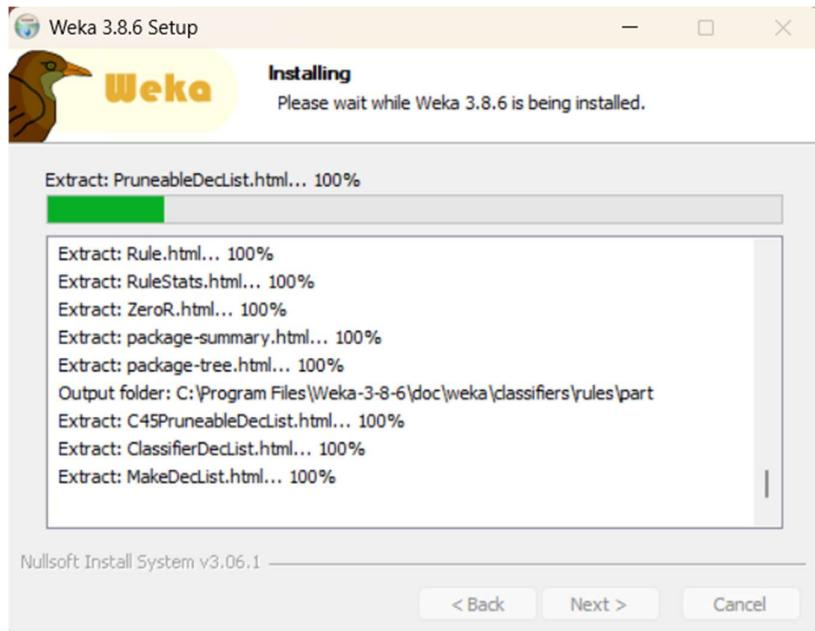
## STEP4:



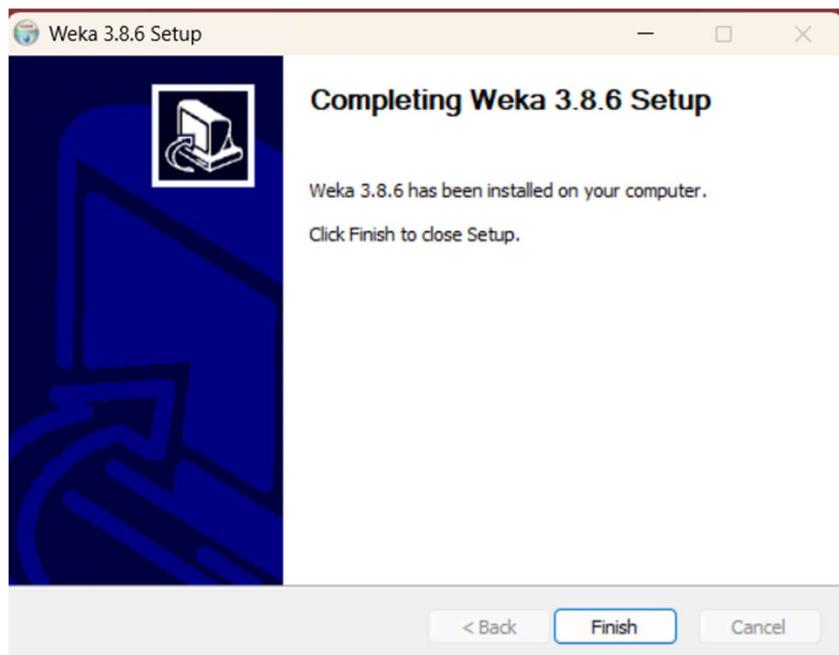
## STEP5:



## STEP6:



## INSTALLATION COMPLETE:



WEKA window:



Learning Outcome:

## **Experiment 2**

**TITLE:** Performing data preprocessing, data cleaning, data transformation, data reduction, data integration on WEKA.

**THEORY:**

Data Preprocessing:

Data Cleaning:

Data Integration:

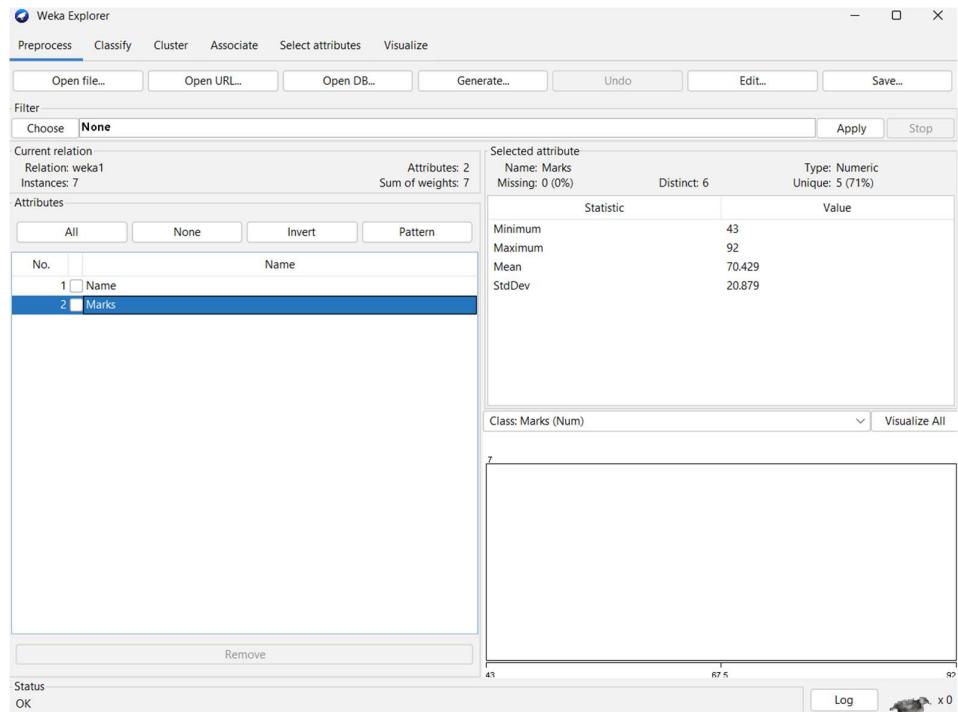
Data Reduction:

Data Transformation:

**.xlxs File**

1	Name	Marks
2	Ram	89
3	shyam	90
4	lakhan	92
5	makan	67
6	lakhshman	45
7	shivam	43
8	rohan	67

**.arff File**



## .text File

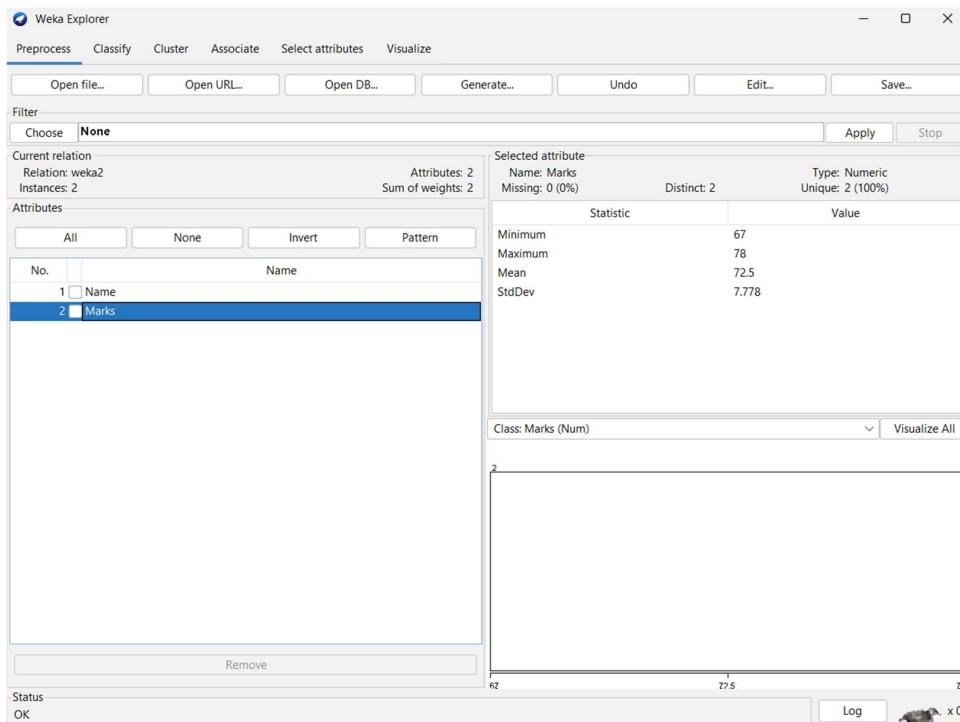
```
File Edit View

@relation weka2

@attribute Name string
@attribute Marks numeric

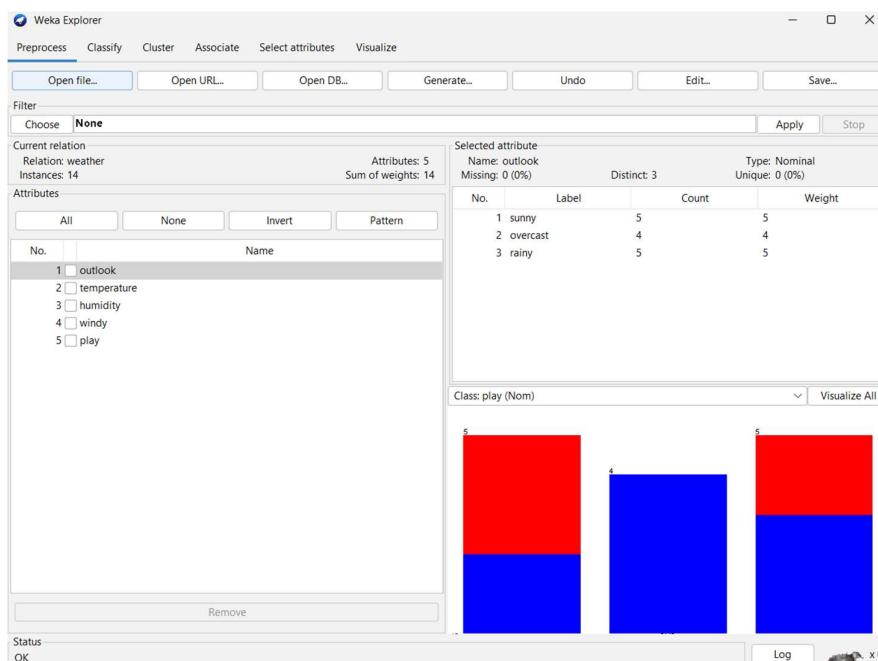
@data
Sita,67
Ram,78
```

## .arff File

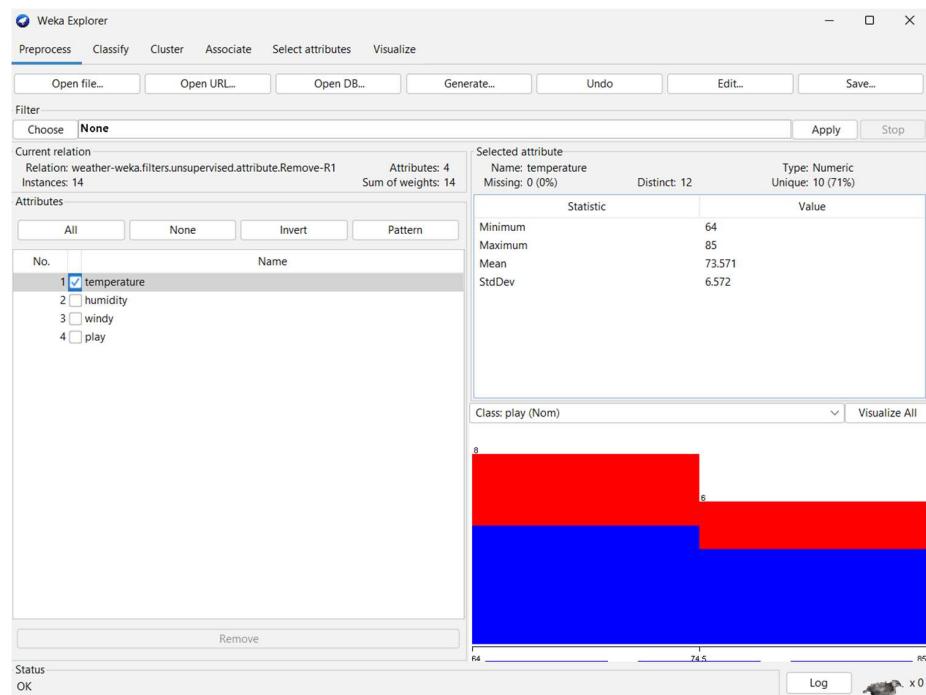


## Data Cleaning

### Before Cleaning

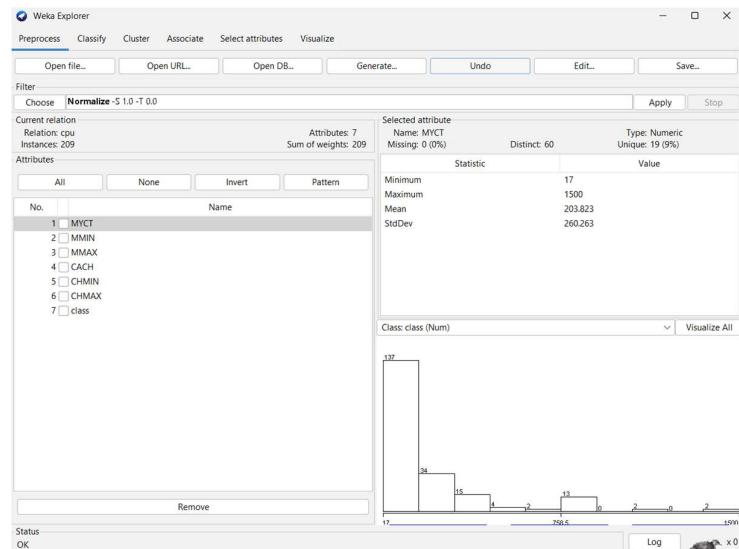


### After Cleaning

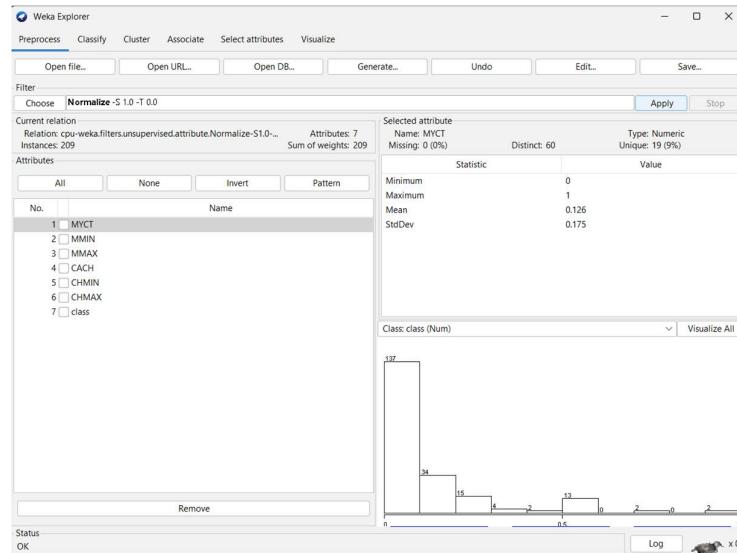


## Data Transformation

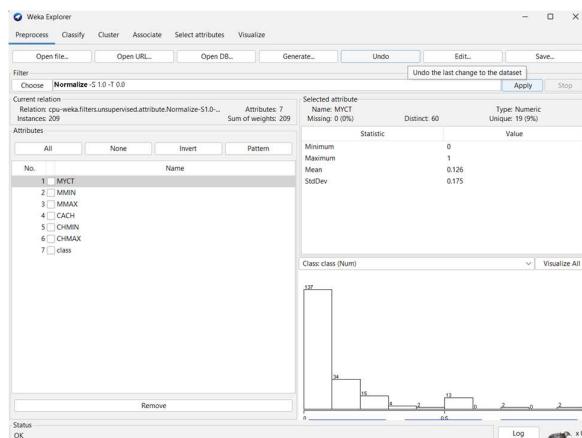
### Before Transformation (Normalisation)



### After Transformation

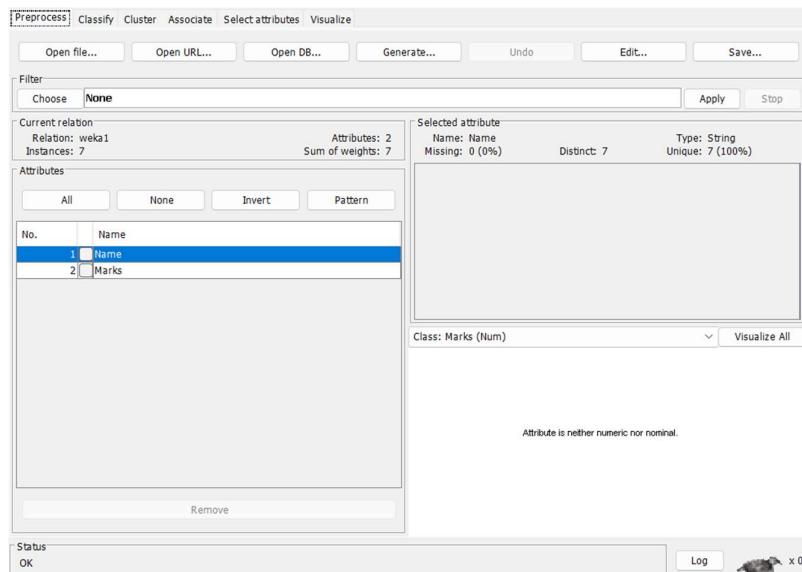


## Undo

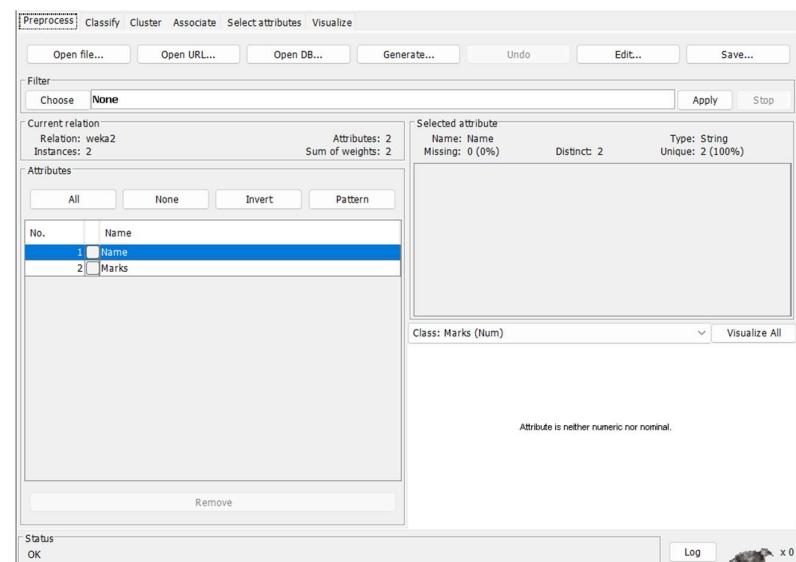


## Data Integration

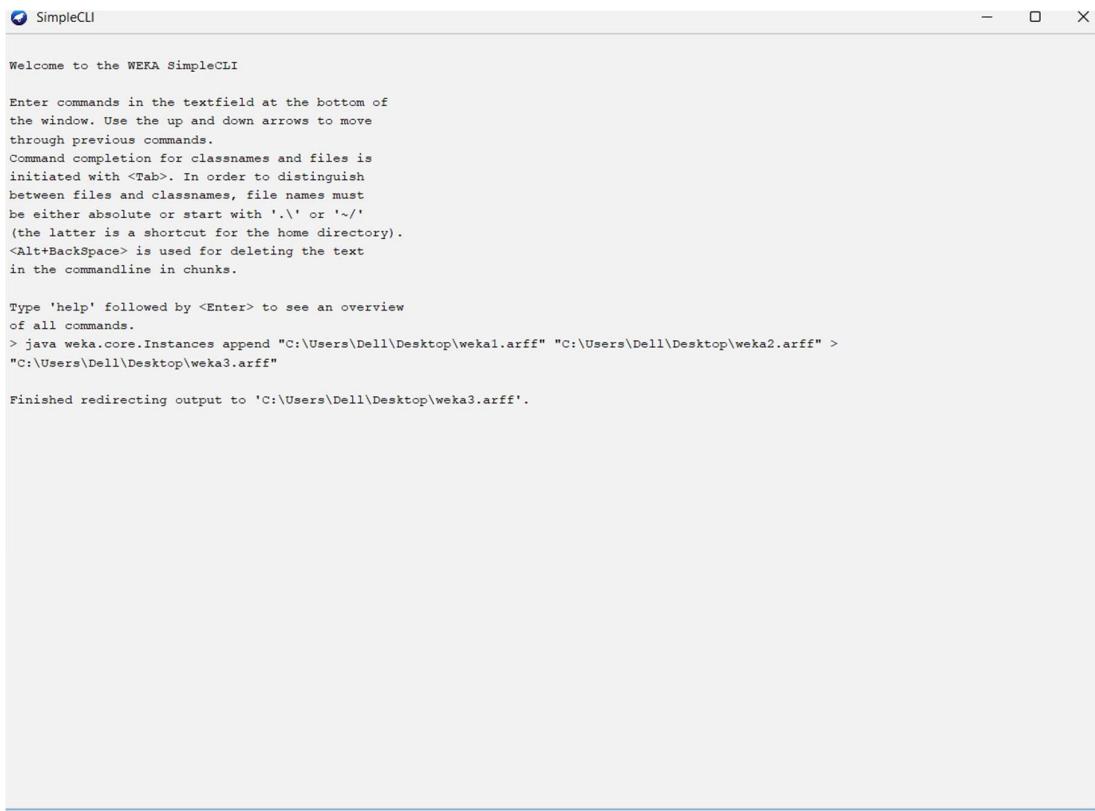
File 1 (Weka1.arff)



File 2 (Weka2.arff)



Simple CLI (weka3.arff)



The screenshot shows a terminal window titled "SimpleCLI". The window contains the following text:

```
Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.
Command completion for classnames and files is
initiated with <Tab>. In order to distinguish
between files and classnames, file names must
be either absolute or start with '.' or '~'
(the latter is a shortcut for the home directory).
<Alt+BackSpace> is used for deleting the text
in the commandline in chunks.

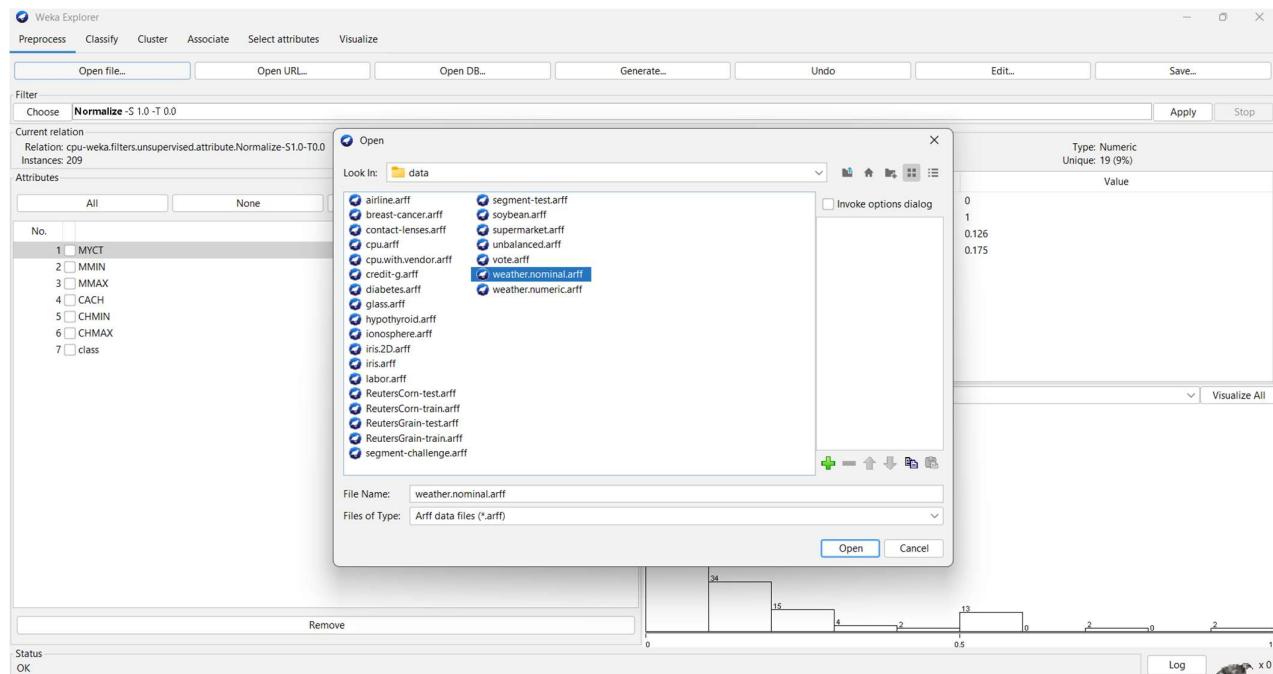
Type 'help' followed by <Enter> to see an overview
of all commands.
> java weka.core Instances append "C:\Users\DELL\Desktop\weka1.arff" "C:\Users\DELL\Desktop\weka2.arff" >
"C:\Users\DELL\Desktop\weka3.arff"
Finished redirecting output to 'C:\Users\DELL\Desktop\weka3.arff'.
```

## Learning Outcome:

Aim: Apply Association Rule Mining on a dataset using Weka.

Theory:

## Opening Data:



Applying Apriori Algorithm

**Weka Explorer**

Preprocess Classify Cluster Associate **Select attributes** Visualize

Associate

Choose **Apriori -N 10 -T 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1**

Start Stop

Result list (right-click for ...)

21/1814 - Apriori  
21/2100 - Apriori

Associate output

```

==== Associate model (full training set) ====

Apriori
=====

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 47
Size of set of large itemsets L(3): 39
Size of set of large itemsets L(4): 6

Best rules found:

1. outlook=overcast 4 ==> play=yes 4   <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
2. temperature<=cool 4 ==> humidity=normal 4   <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. humidity=normal windy=FALSE 4 ==> play=yes 4   <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
4. outlook=sunny play=no 3 ==> humidity=high 3   <conf:(1)> lift:(1.56) lev:(0.11) [1] conv:(1.5)
5. outlook=sunny humidity=high 3 ==> play=no 3   <conf:(1)> lift:(0.8) lev:(0.14) [1] conv:(1.93)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3   <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3   <conf:(1)> lift:(1.56) lev:(0.09) [1] conv:(1.07)
8. temperature<=cool play=yes 3 ==> humidity=normal 3   <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
9. outlook=sunny temperature<=cool humidity=high 2 ==> outlook=sunny 2   <conf:(1)> lift:(0.8) lev:(0.1) [1] conv:(1.1)
10. temperature=hot play=yes 2 ==> outlook=sunny 2   <conf:(1)> lift:(2.8) lev:(0.09) [1] conv:(1.29)

```

Status OK Log x0

## Changing numRules

**Weka Explorer**

Preprocess Classify Cluster Associate **Select attributes** Visualize

Associate

Choose **Apriori -N 35 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1**

Start Stop

Result list (right-click for ...)

21/1814 - Apriori

Associate output

```

==== Associate model (full training set) ====

Apriori
=====

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 47
Size of set of large itemsets L(3): 39
Size of set of large itemsets L(4): 6

Best rules found:

1. outlook=overcast 4 ==> play=yes 4   <conf:(1)>
2. temperature<=cool 4 ==> humidity=normal 4   <conf:
3. humidity=normal windy=FALSE 4 ==> play=yes 4   <conf:
4. outlook=sunny play=no 3 ==> humidity=high 3   <conf:
5. outlook=sunny humidity=high 3 ==> play=no 3   <conf:
6. outlook=rainy play=yes 3 ==> windy=FALSE 3   <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3   <conf:(1)> lift:(1.56) lev:(0.09) [1] conv:(1.07)
8. temperature<=cool play=yes 3 ==> humidity=normal 3   <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
9. outlook=sunny temperature<=cool humidity=high 2 ==> outlook=sunny 2   <conf:(1)> lift:(0.8) lev:(0.1) [1] conv:(1.1)
10. temperature=hot play=yes 2 ==> outlook=sunny 2   <conf:(1)> lift:(2.8) lev:(0.09) [1] conv:(1.29)

```

**weka.gui.GenericObjectEditor**

About

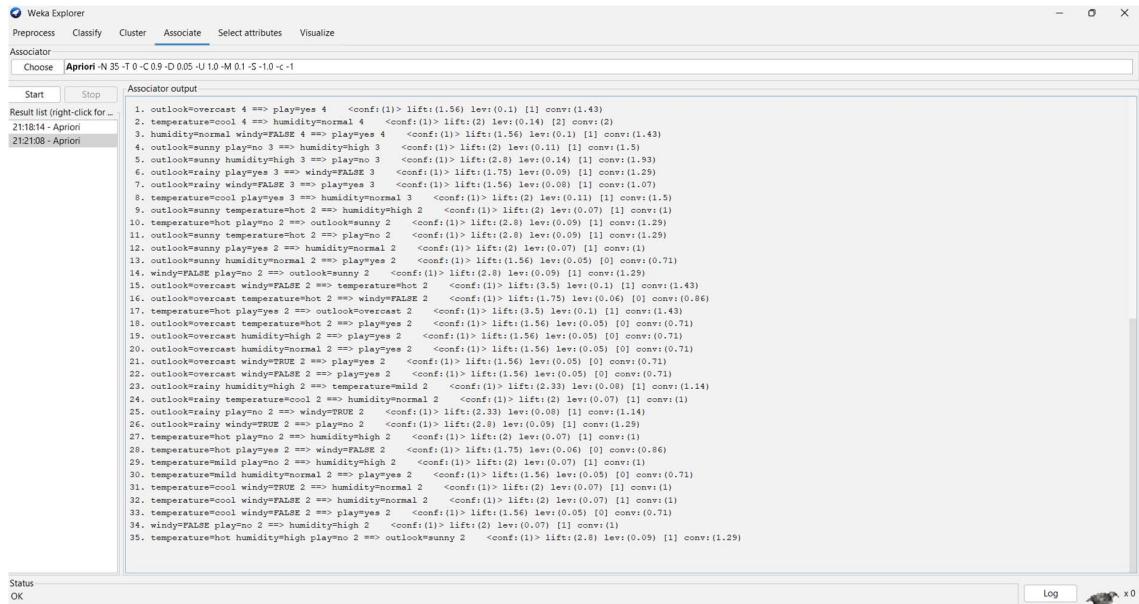
Class implementing an Apriori-type algorithm.

car	False
classIndex	-1
delta	0.05
doNotCheckCapabilities	False
lowerBoundMinSupport	0.1
metricType	Confidence
minMetric	0.9
numRules	35
outputItemSets	False
removeAllMissingCols	False
significanceLevel	-1.0
treatZeroAsMissing	False
upperBoundMinSupport	1.0
verbose	False

Open... Save... OK Cancel

Status OK Log x0

## New Rules



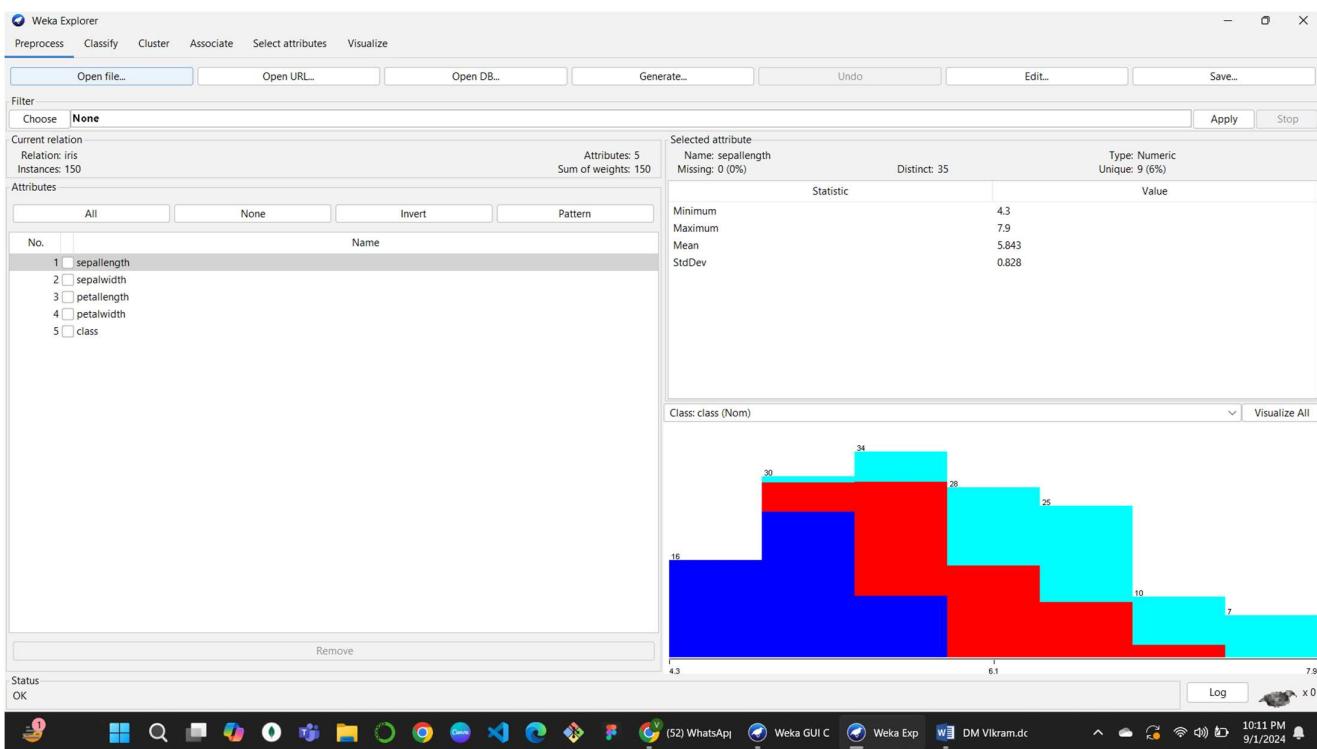
## Learning Outcome:

## Experiment 4

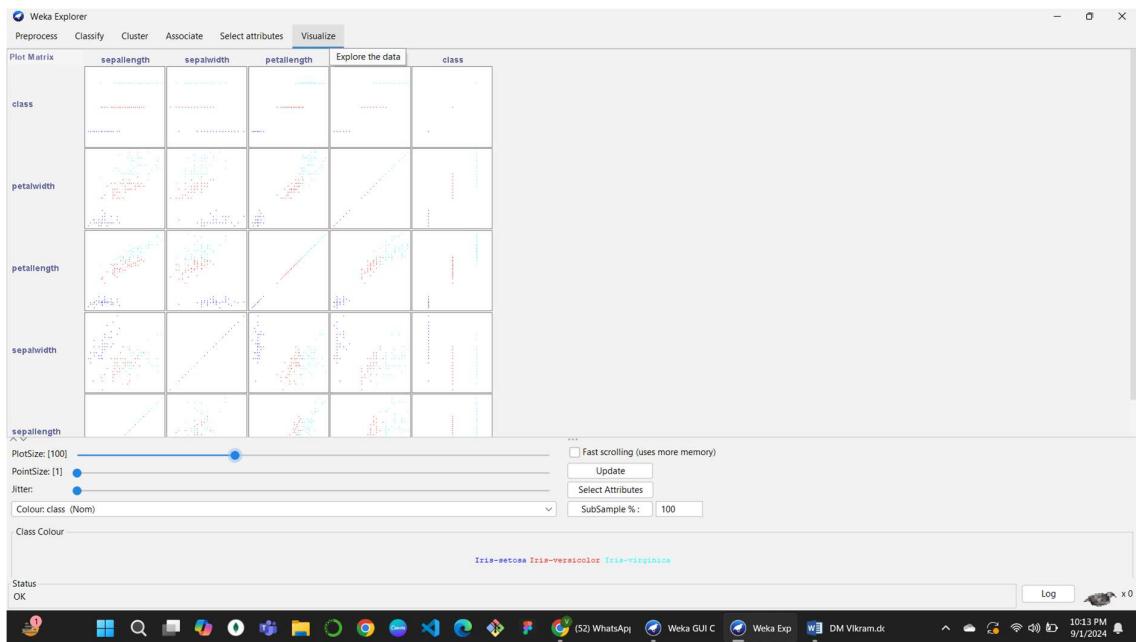
Aim: Apply Visualization on a dataset using Weka.

Theory:

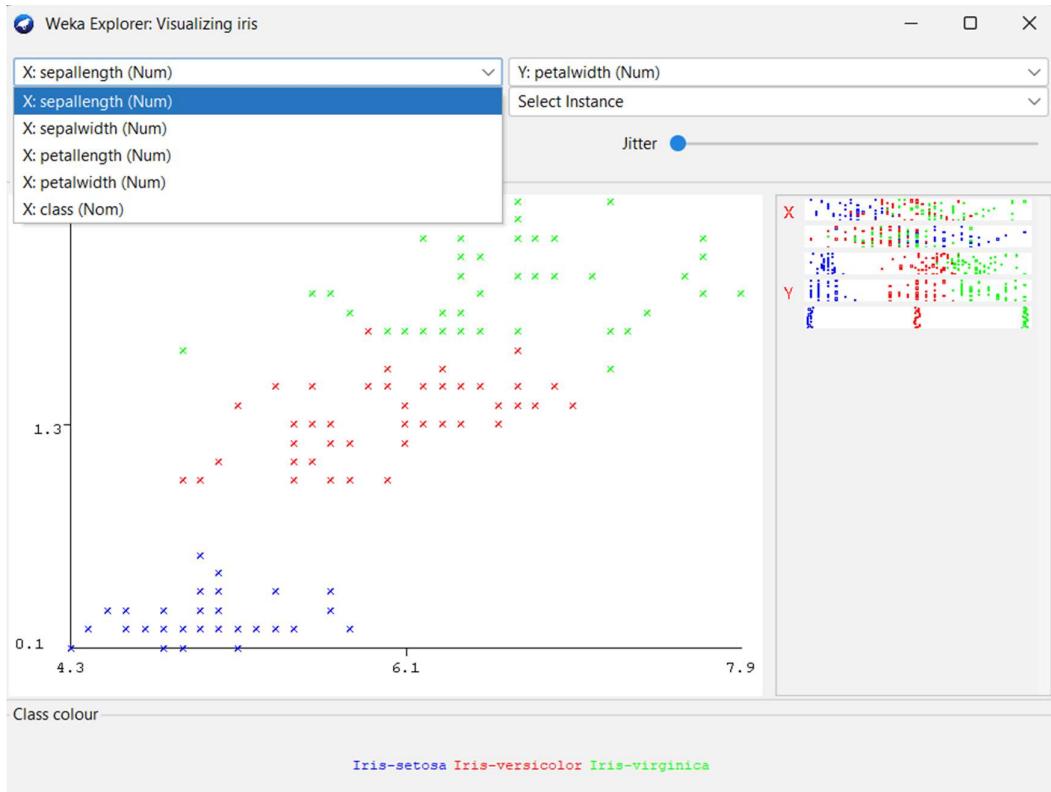
Open Iris Dataset:



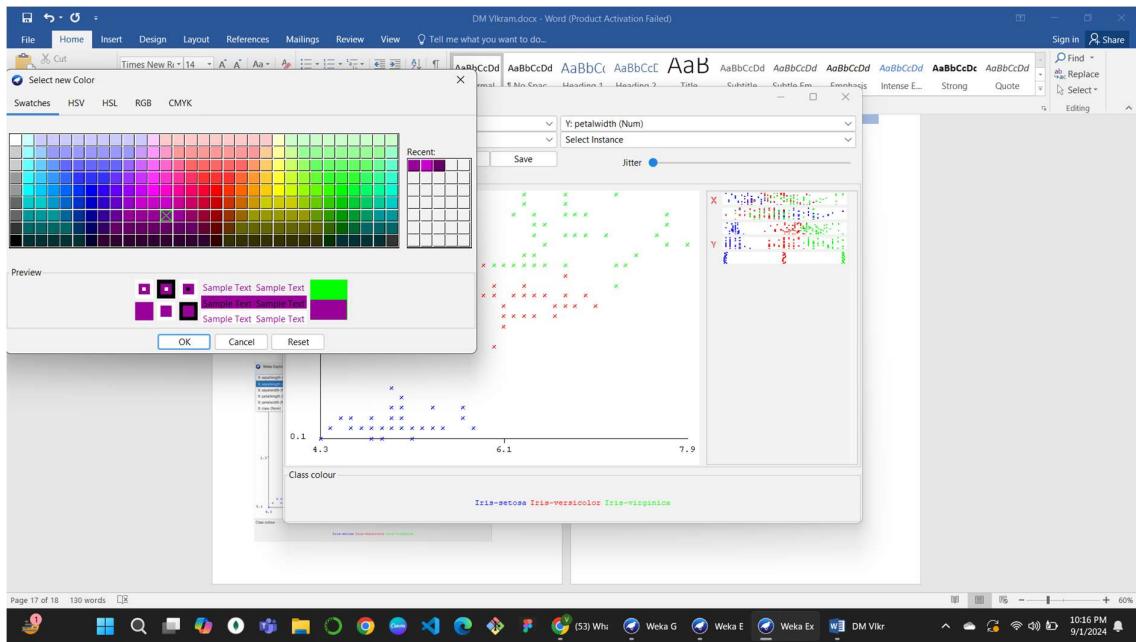
Open visualize section:



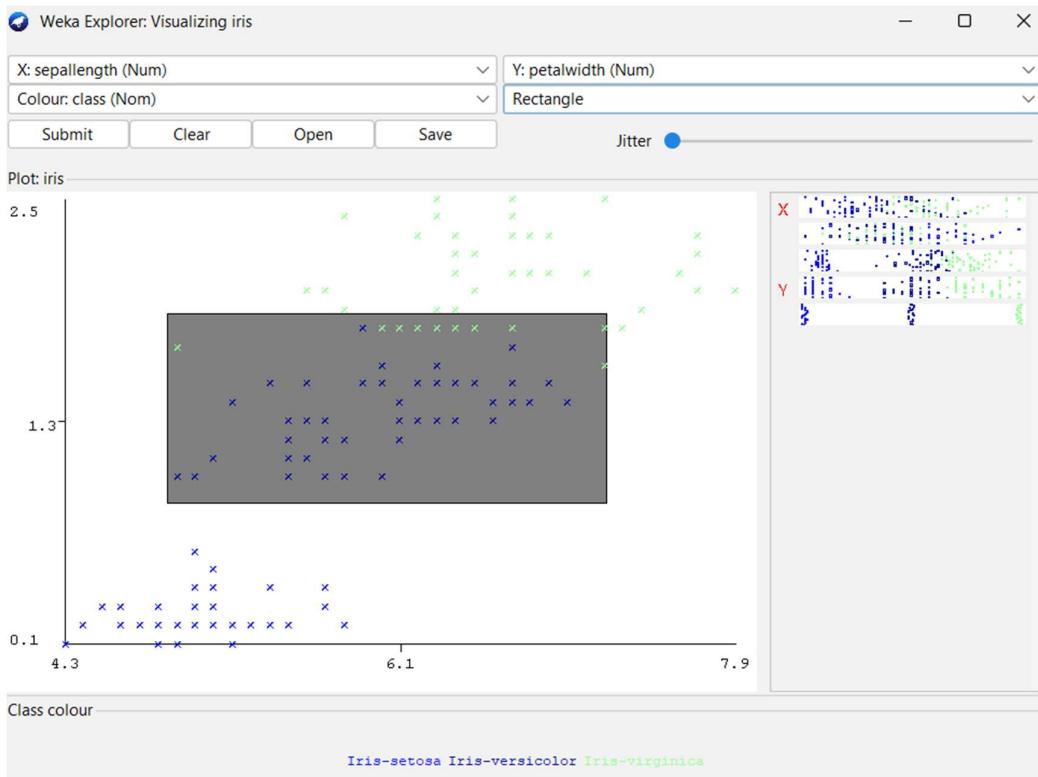
## Visualizing dataset in different scenarios:



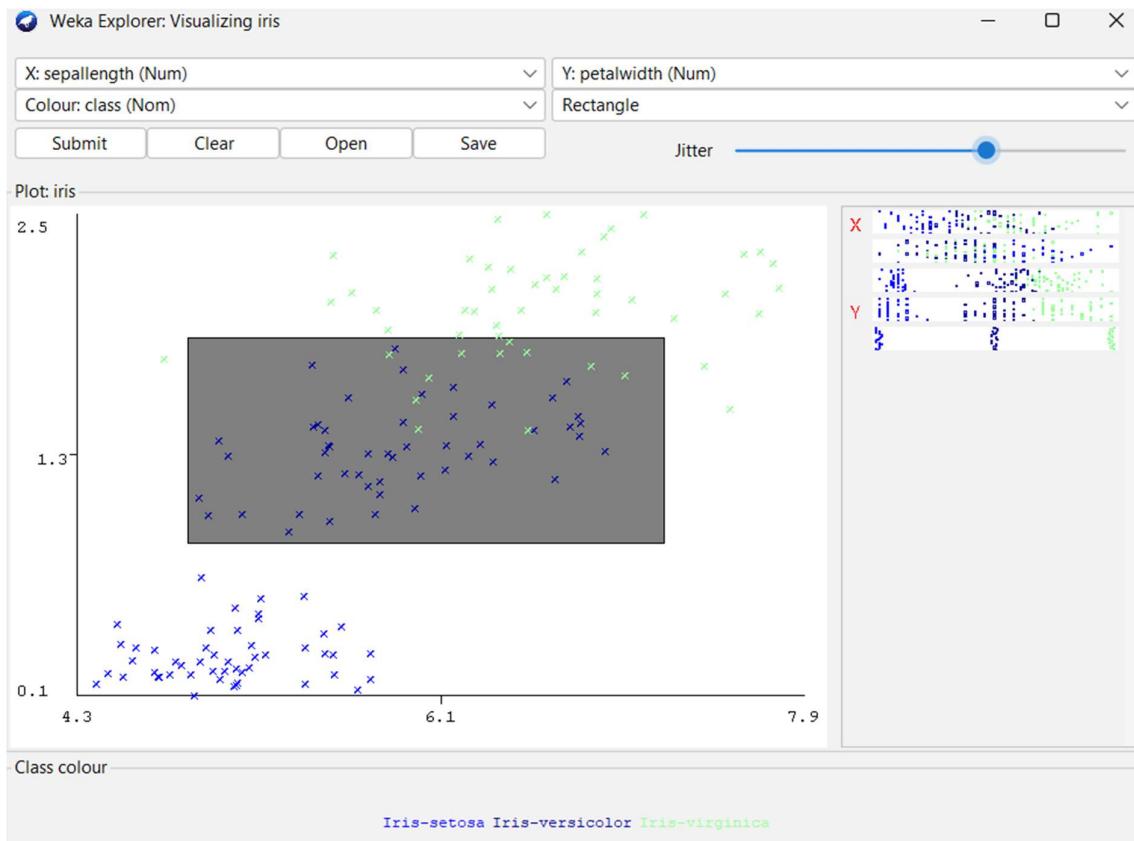
## Exploration of colour tool:



## Selecting Instances:



Changing the jitter:



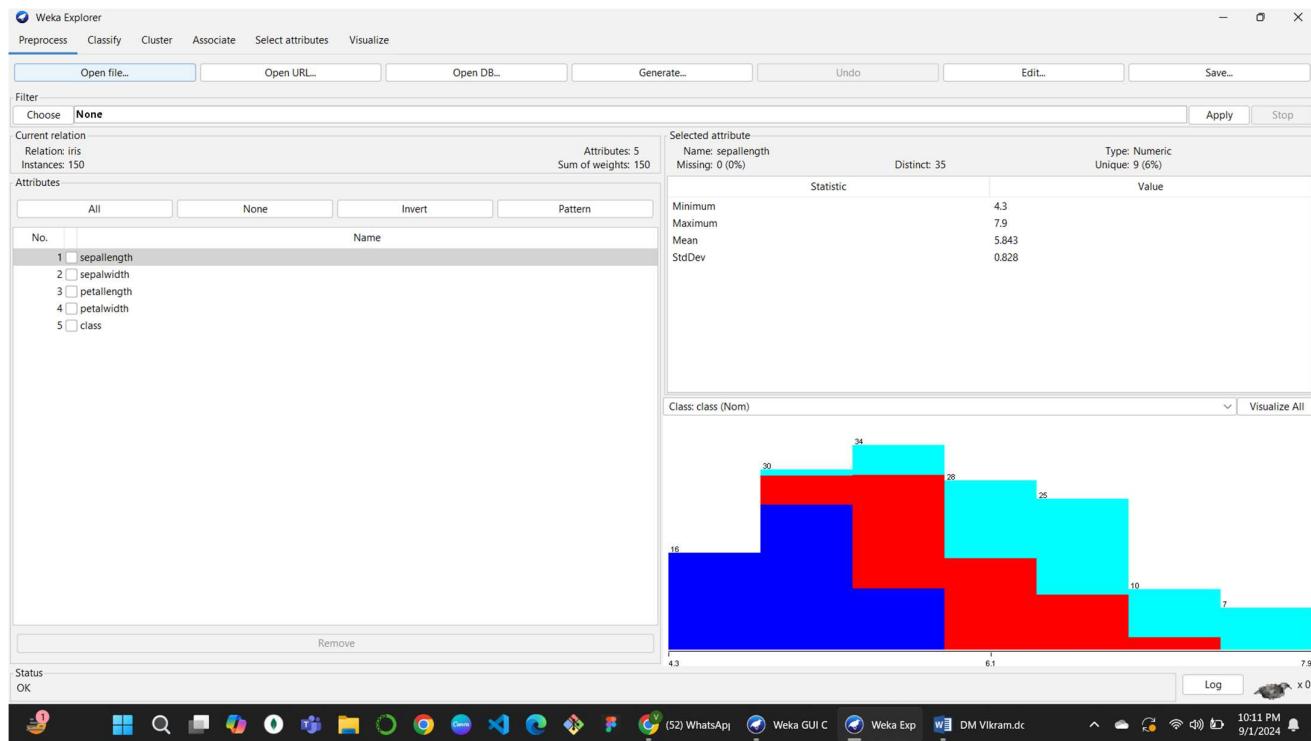
## Learning Outcome:

## Experiment 5

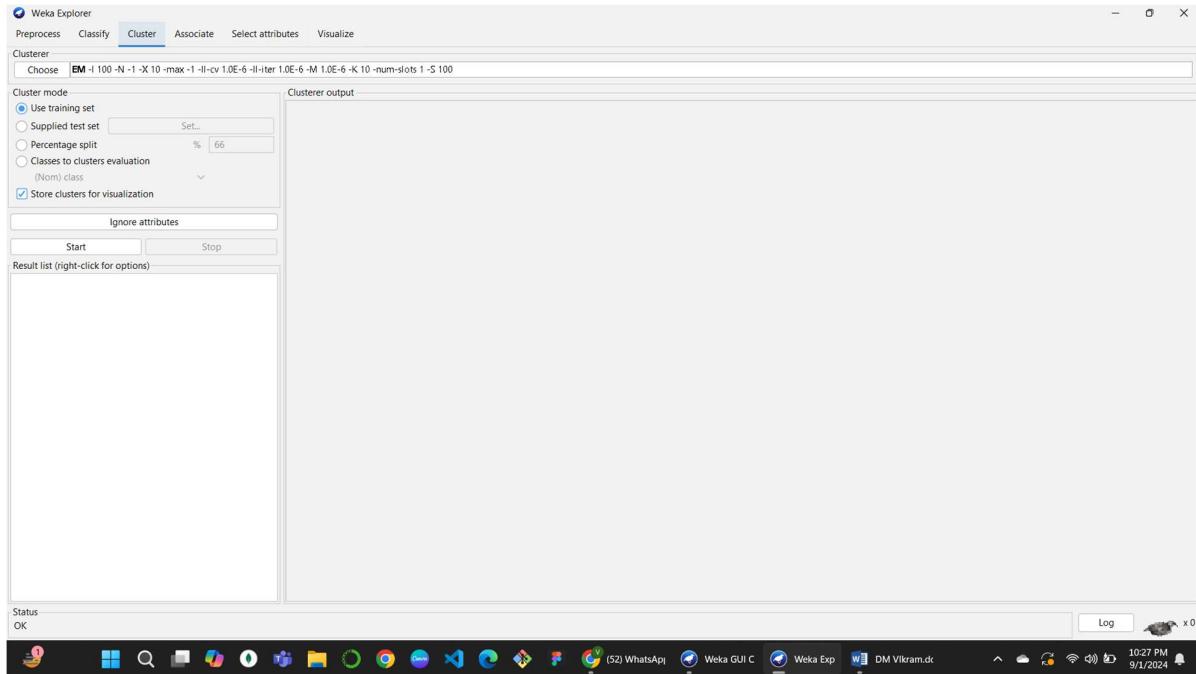
Aim: Apply Clustering on a dataset using Weka.

Theory:

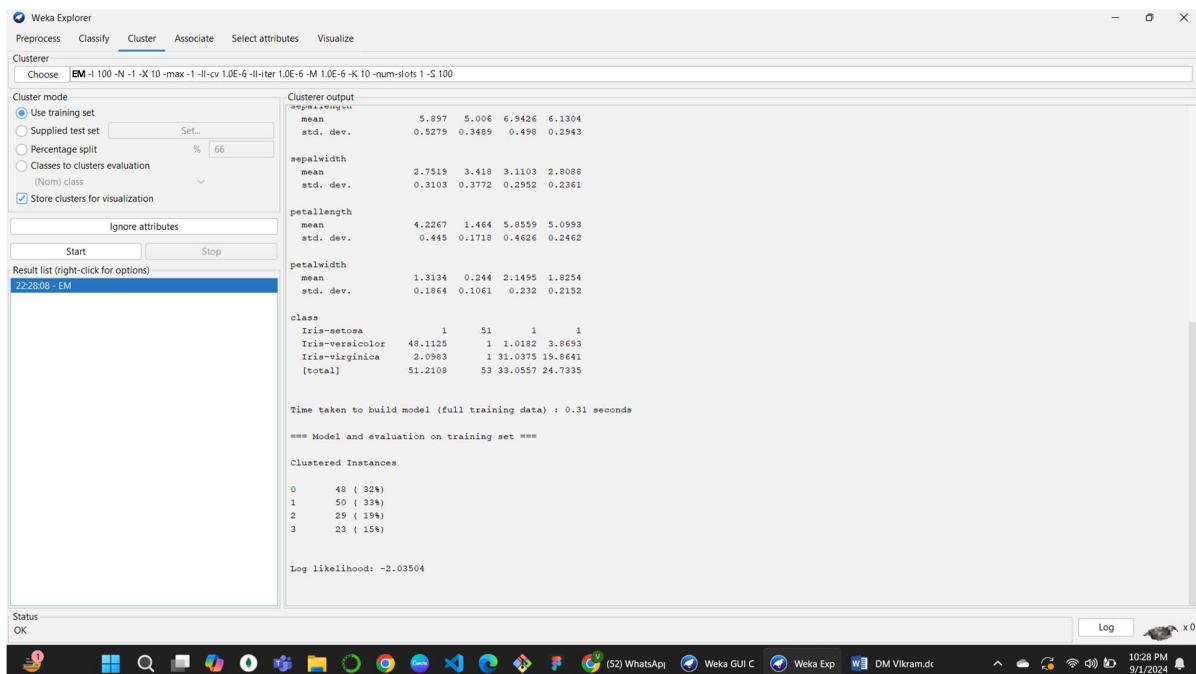
Open Iris Dataset:



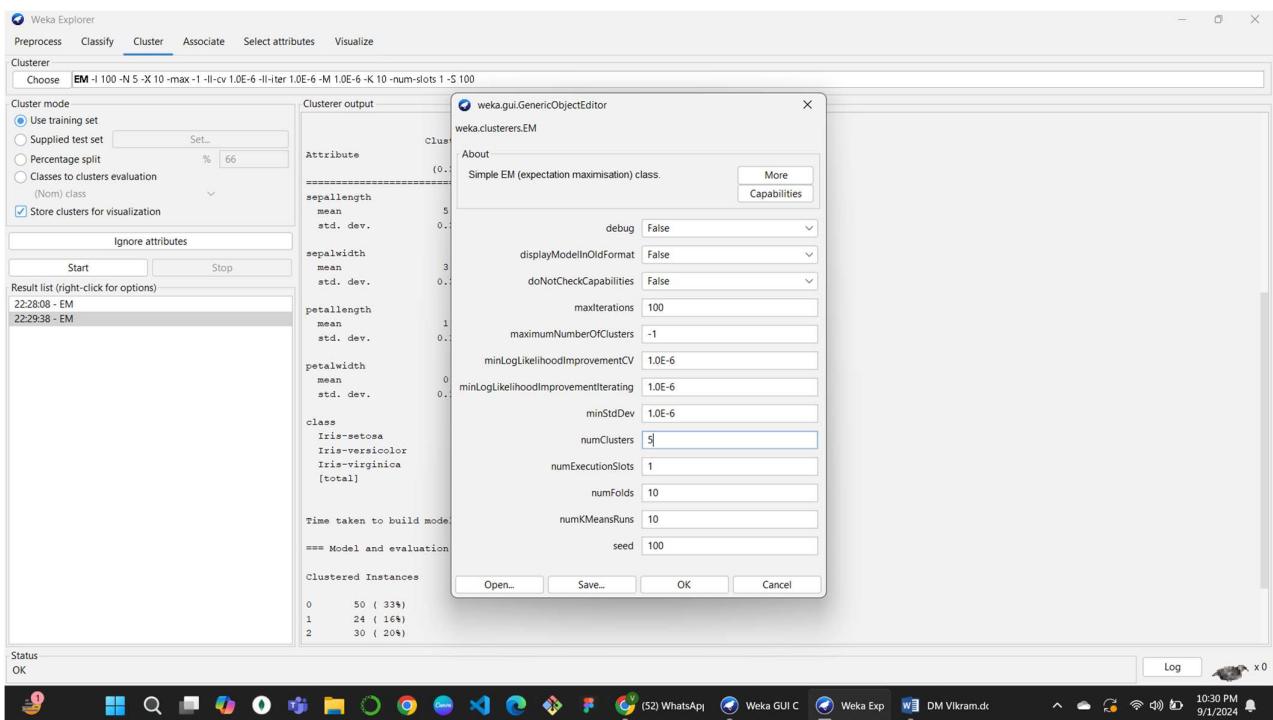
## Selecting cluster option:



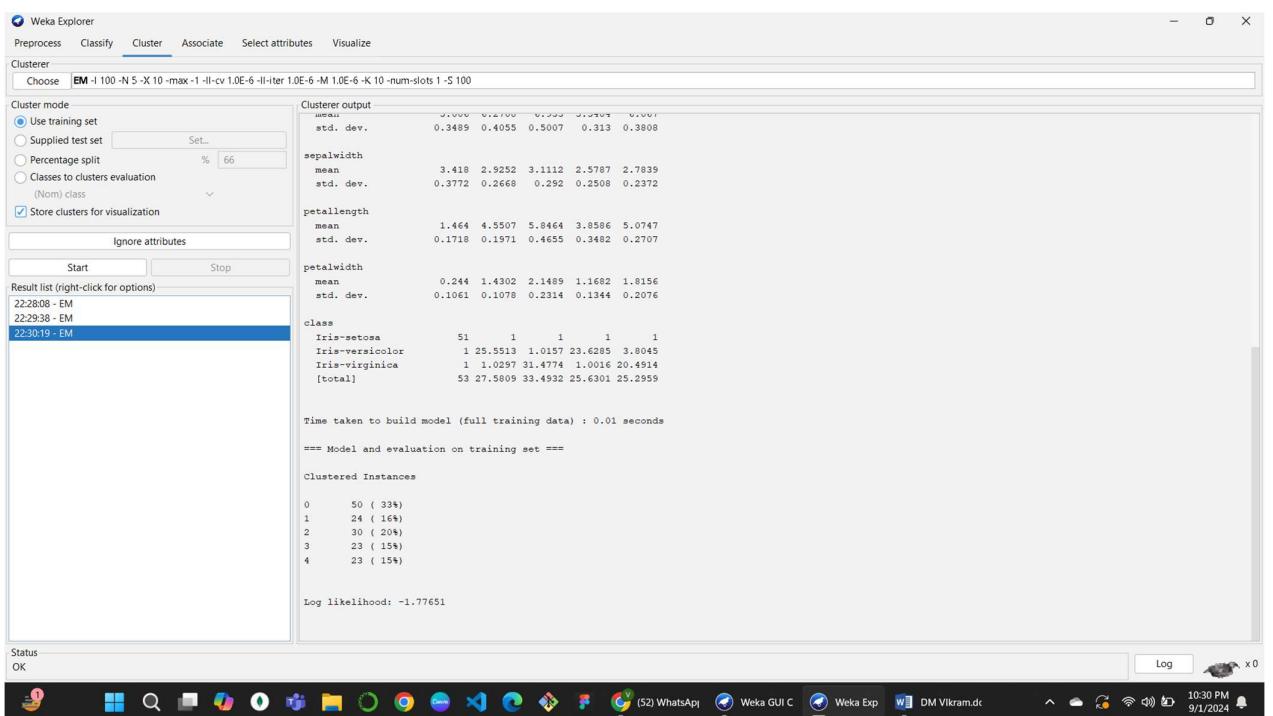
## Applied clustering:



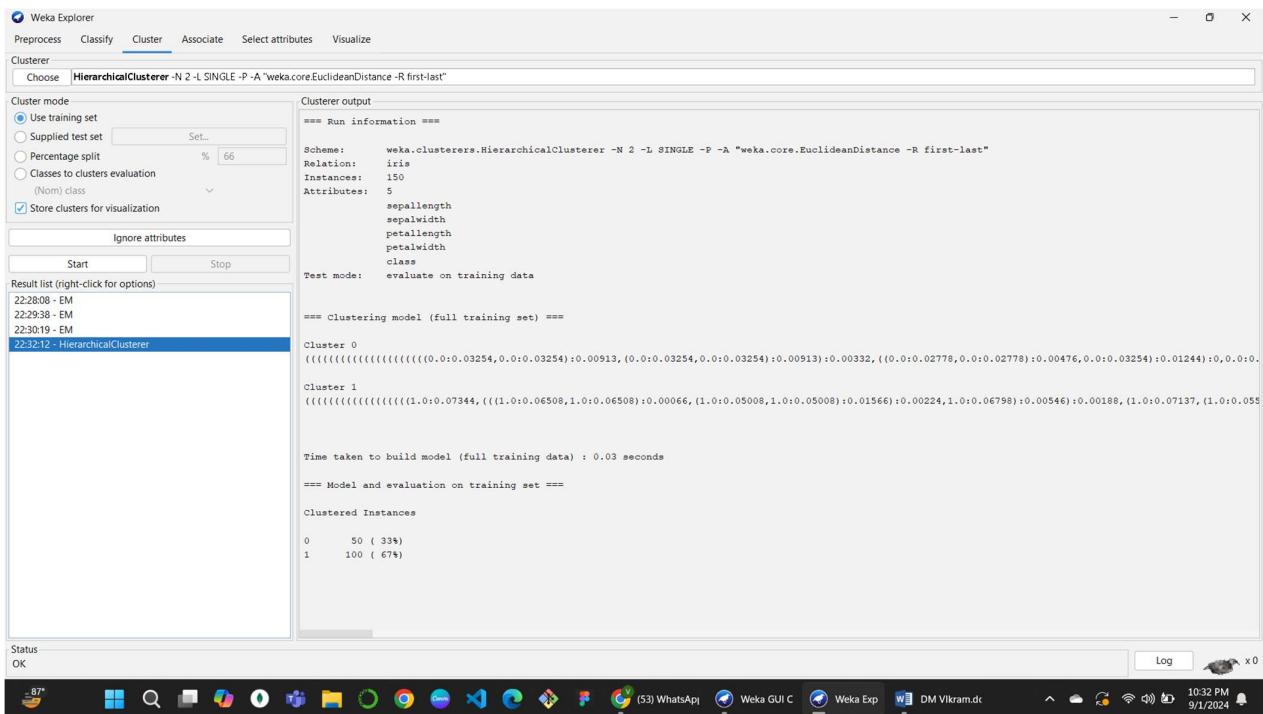
## Changed numCluster:



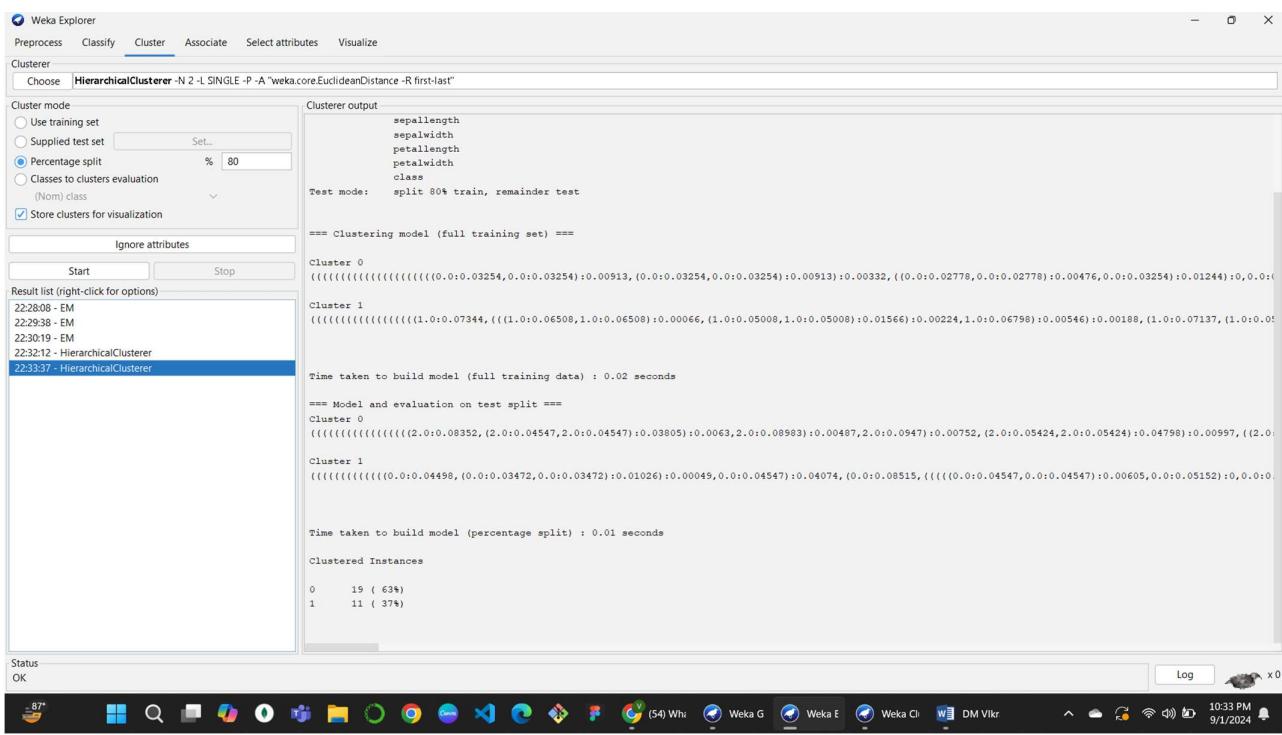
## Output after changing numcluster:



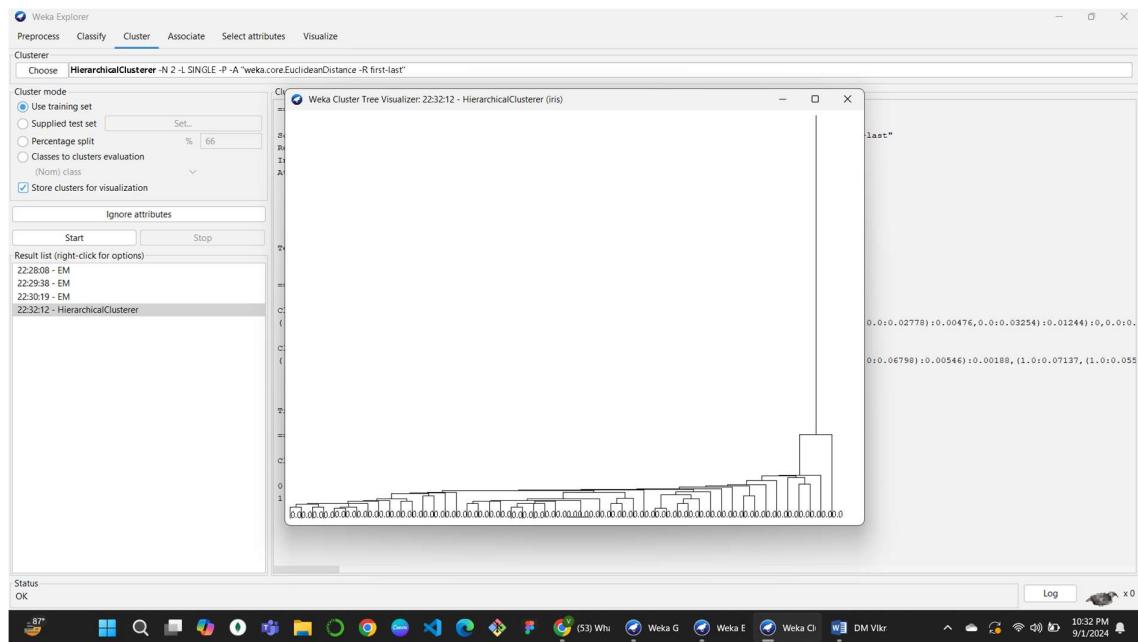
## Applied Hierarchical Cluster:



## Output of Hierarchical Cluster:



## Visualization of Hierarchical Cluster Tree:



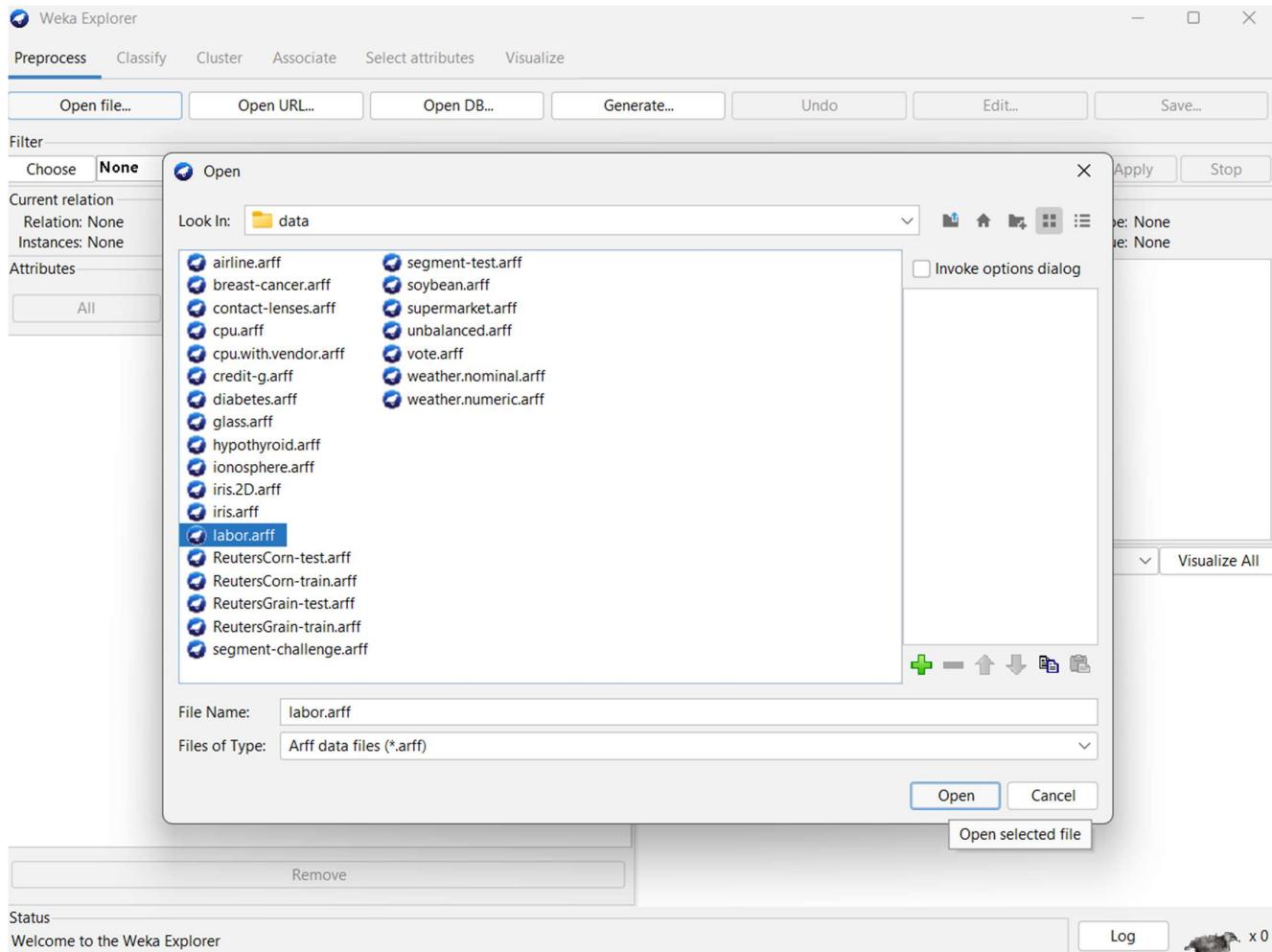
## Learning Outcome:

## **Experiment 6**

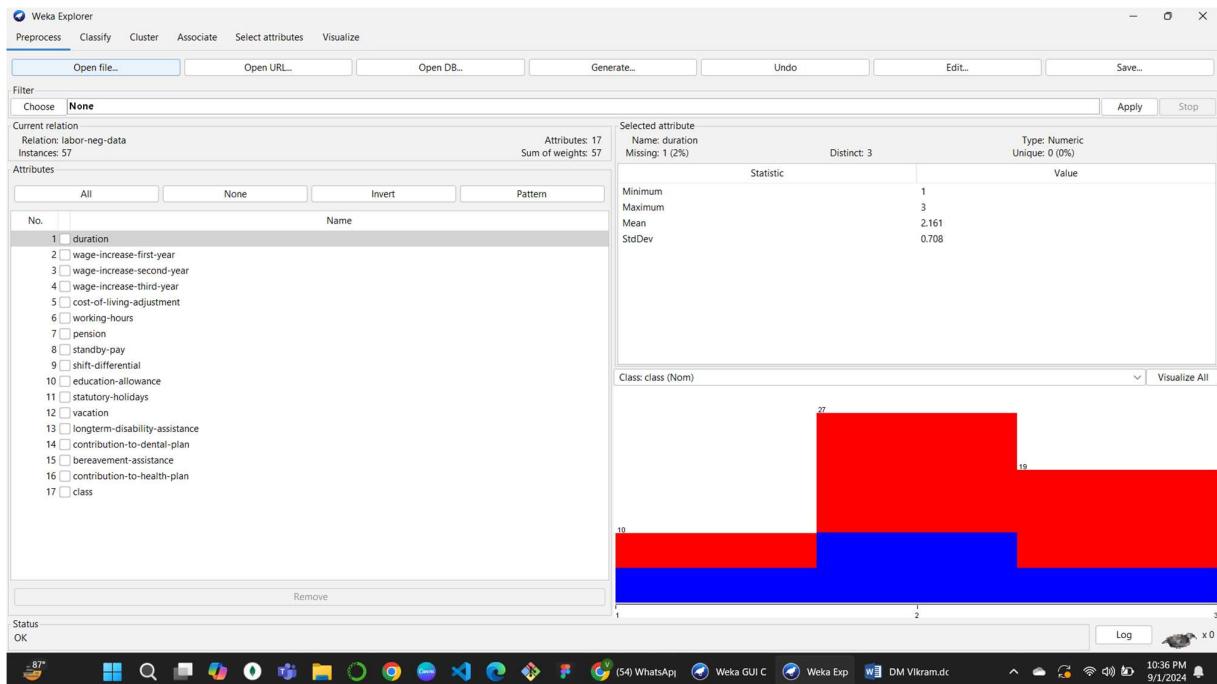
Aim: Apply Classification on a dataset using Weka.

Theory:

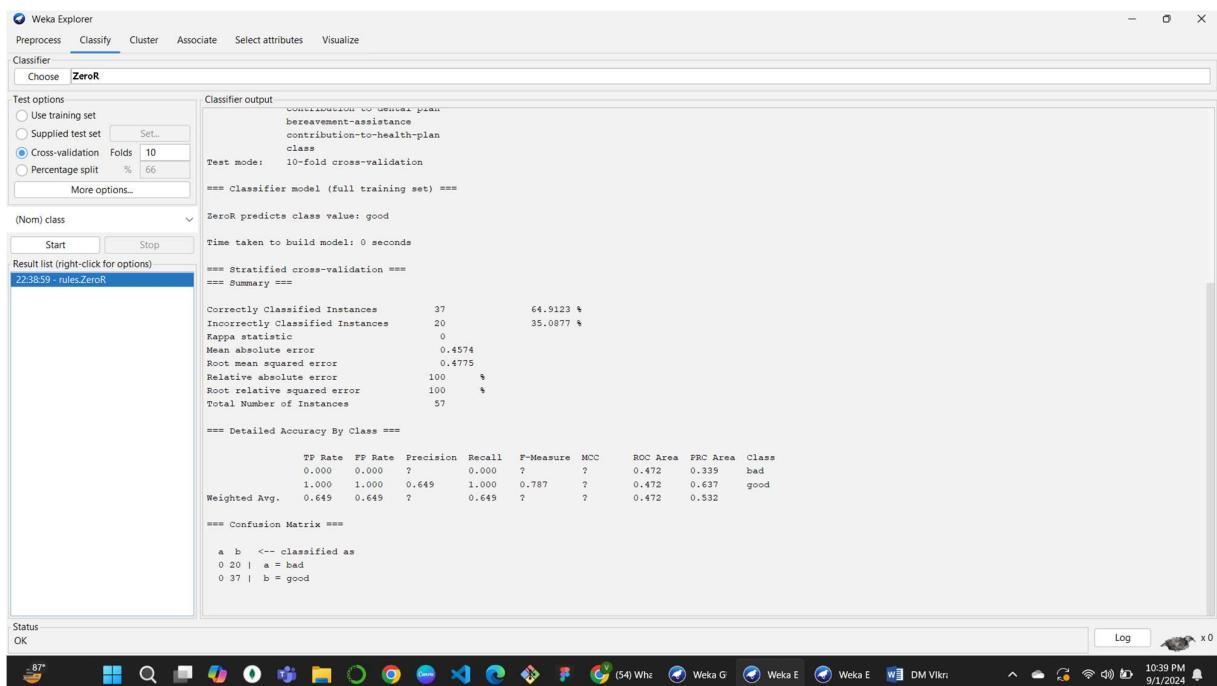
Selection of Labor dataset:



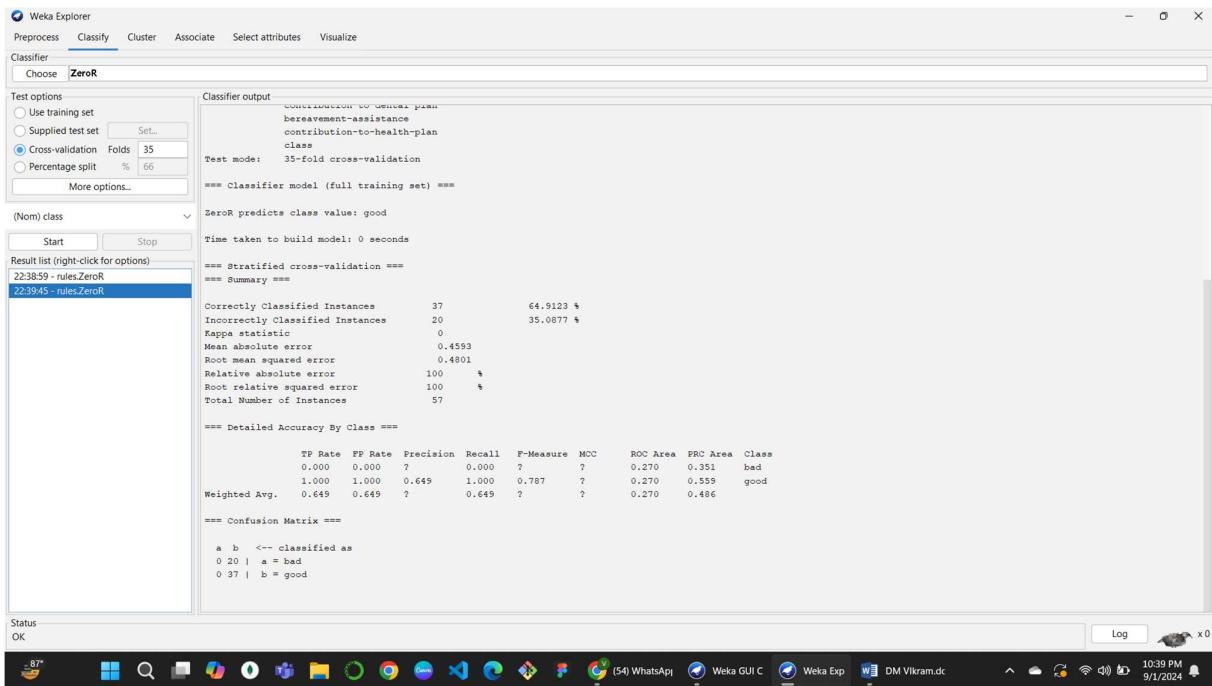
## Analyzation of Labor Dataset:



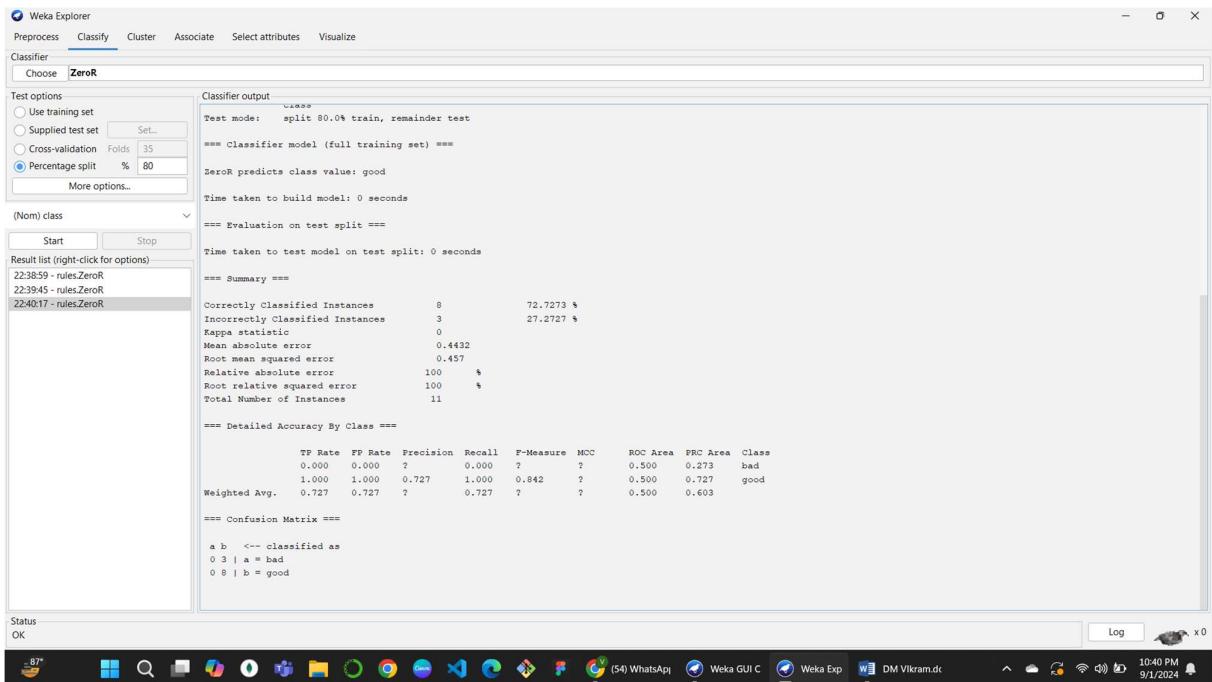
## Applying ZeroR Classification on labor dataset:



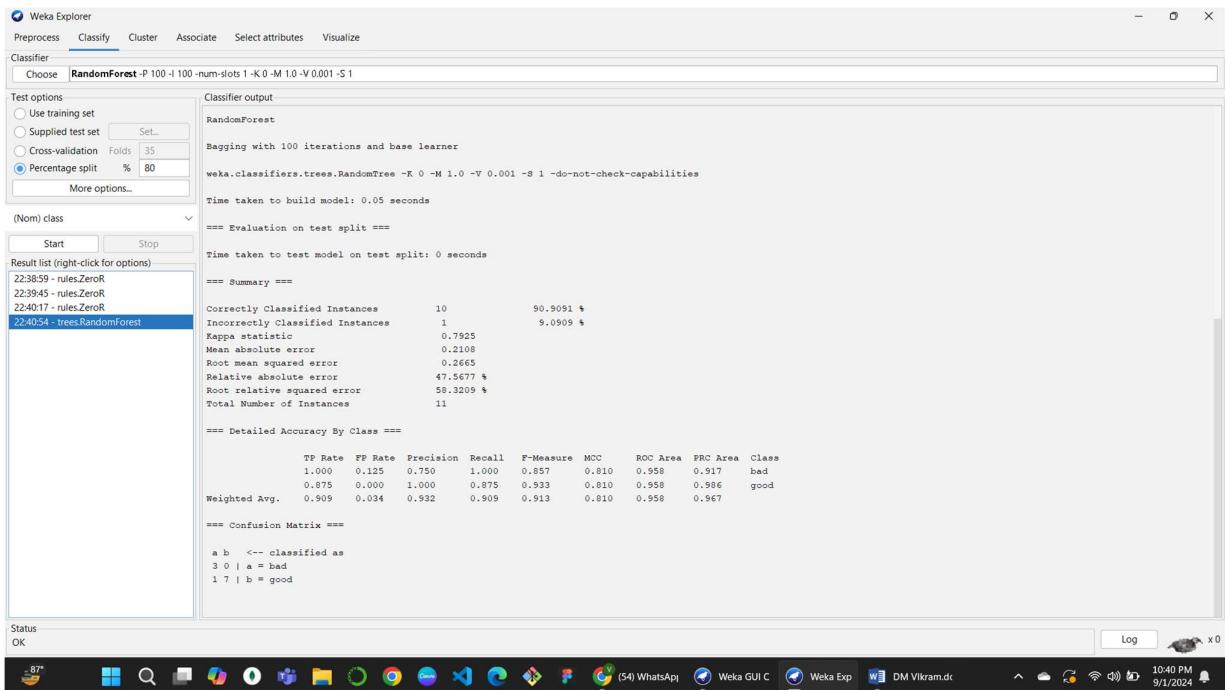
## Changing number of Cross-Validation Folds:



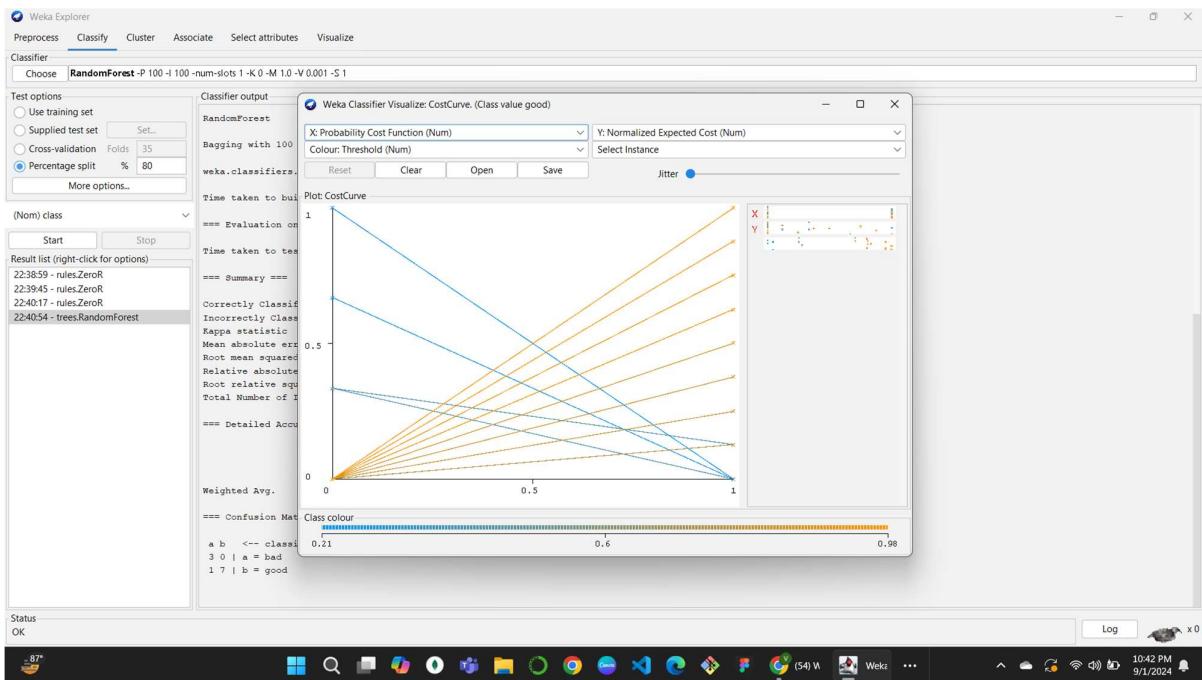
## Splitting the dataset into training and testing sets:



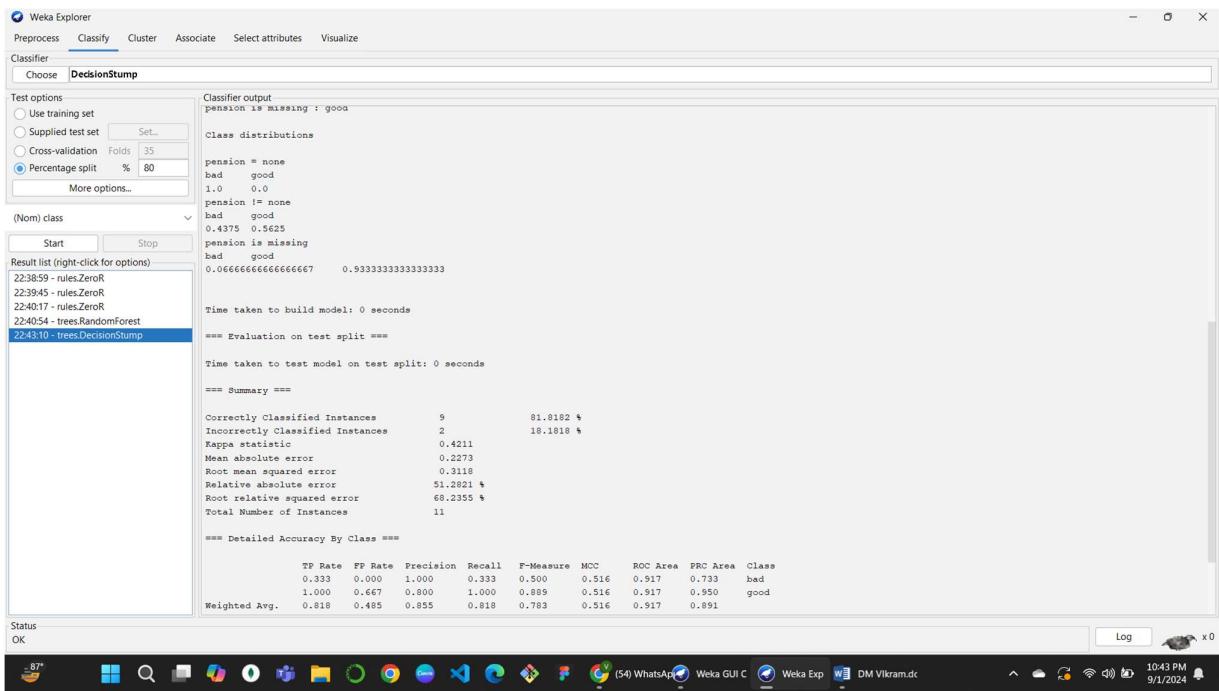
## Applying Random Forest Classification on labor dataset:



## Visualization of cost curve:



## Applying Decision Stump Classification on labor dataset:



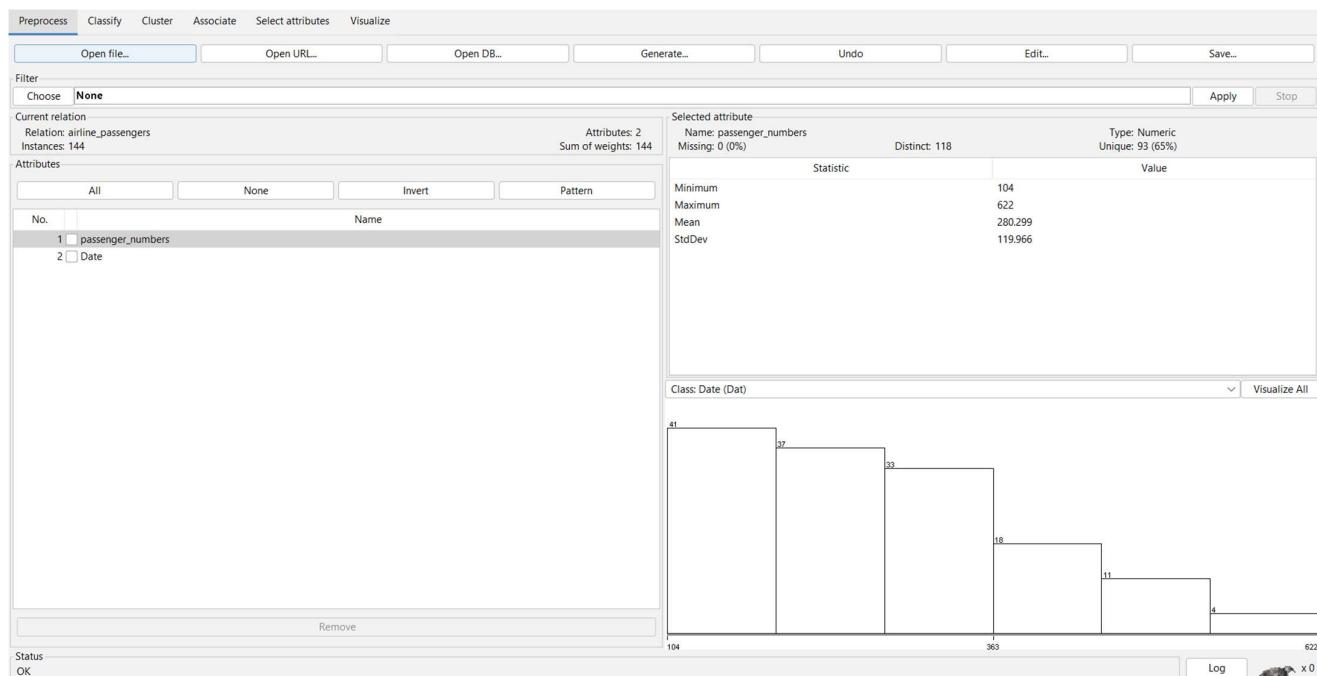
## Learning Outcome:

## Experiment 6 B

Aim: Apply Regression on a dataset using Weka.

Theory:

Applying Passengers Number dataset on weka.



## Applying Linear Regression:

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **LinearRegression** -S 0 -R 1.0E-8 -num-decimal-places 4

Test options Left-click to edit properties for this object, right-click/Alt+Shift+left-click for menu  
== Run information ==

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 15  
 Percentage split % 80  
More options...

(Dat) Date Start Stop

Result list (right-click for options)  
21:11:33 - functions.LinearRegression

```

Linear Regression Model

Date =
844774835.7902 * passenger_numbers +
-711559013173.6233

Time taken to build model: 0.05 seconds

== Cross-validation ==
== Summary ==

Correlation coefficient 0.9198
Mean absolute error 35244048789.1212
Root mean squared error 42516973182.1379
Relative absolute error 37.0674 %
Root relative squared error 39.0974 %
Total Number of Instances 144

```

Status OK Log x 0

## Applying Additive Regression

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **AdditiveRegression** -S 1.0 -I 10 -W weka.classifiers.trees.DecisionStump

Test options Left-click to edit properties for this object, right-click/Alt+Shift+left-click for menu  
== Run information ==

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 15  
 Percentage split % 80  
More options...

(Dat) Date Start Stop

Result list (right-click for options)  
21:11:33 - functions.LinearRegression  
21:20:05 - functions.SimpleLinearRegression  
21:20:55 - meta.AdditiveRegression

```

Classifier output

Model number 8
Decision Stump

Classifications

passenger_numbers <= 506.5 : -1.827021712938578E9
passenger_numbers > 506.5 : 4.202149393758715E10
passenger_numbers is missing : -5.828009711371527E-6

Model number 9
Decision Stump

Classifications

passenger_numbers <= 138.0 : -1.944700776074563E10
passenger_numbers > 138.0 : 2.430875870093209E5
passenger_numbers is missing : 4.344516330295139E-6

Time taken to build model: 0.02 seconds

== Cross-validation ==
== Summary ==

Correlation coefficient 0.9325
Mean absolute error 33222706113.6629
Root mean squared error 39540717721.6844
Relative absolute error 34.9415 %
Root relative squared error 36.0216 %
Total Number of Instances 144

```

Status OK Log x 0

## Simple Linear regression

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose AdditiveRegression -S 1.0 -I 10 -W weka.classifiers.trees.DecisionStump

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 15
- Percentage split % 80

(Data) Date

Start Stop

Result list (right-click for options)

21:11:33 - functions.LinearRegression  
21:20:05 - functions.SimpleLinearRegression  
21:20:55 - meta.AdditiveRegression

```

Classifier output
==== Run information ====
Scheme:      weka.classifiers.functions.SimpleLinearRegression
Relation:    airline_passengers
Instances:   144
Attributes:  2
              passenger_numbers
              Date
Test mode:   15-fold cross-validation

==== Classifier model (full training set) ====
Linear regression on passenger_numbers
844774835.85 * passenger_numbers - 711559013190.18

Predicting 0 if attribute value is missing.

Time taken to build model: 0 seconds

==== Cross-validation ====
==== Summary ====
Correlation coefficient          0.9198
Mean absolute error            352440408707.9874
Root mean squared error        42916973182.3167
Relative absolute error         37.0674 %
Root relative squared error   39.0974 %
Total Number of Instances       144

```

## Visualizing Linear Regression



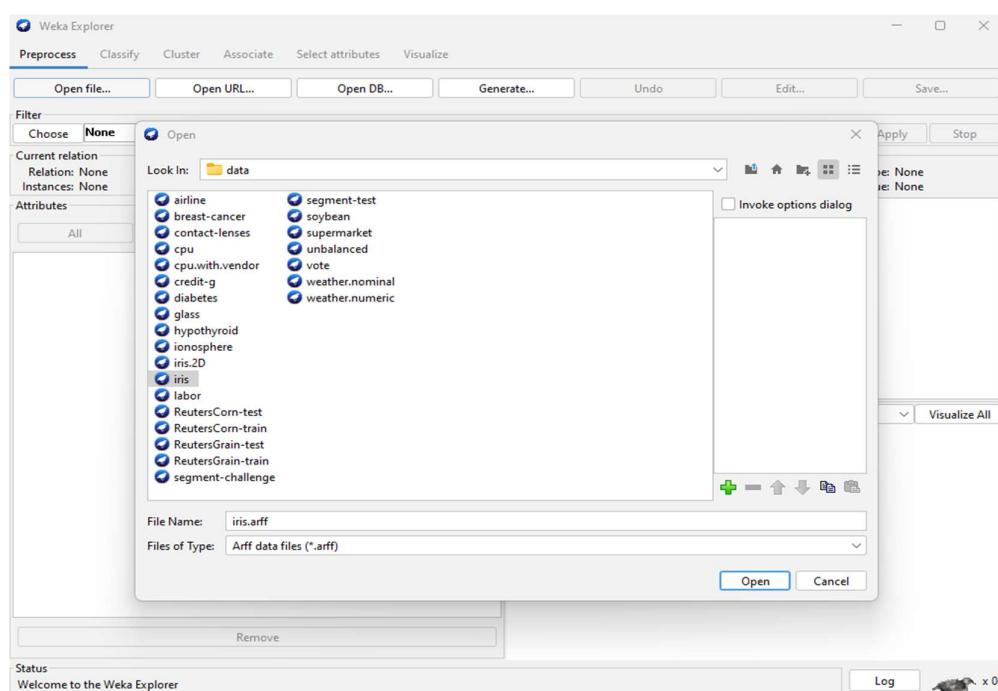
Learning Outcome:

## Experiment 7

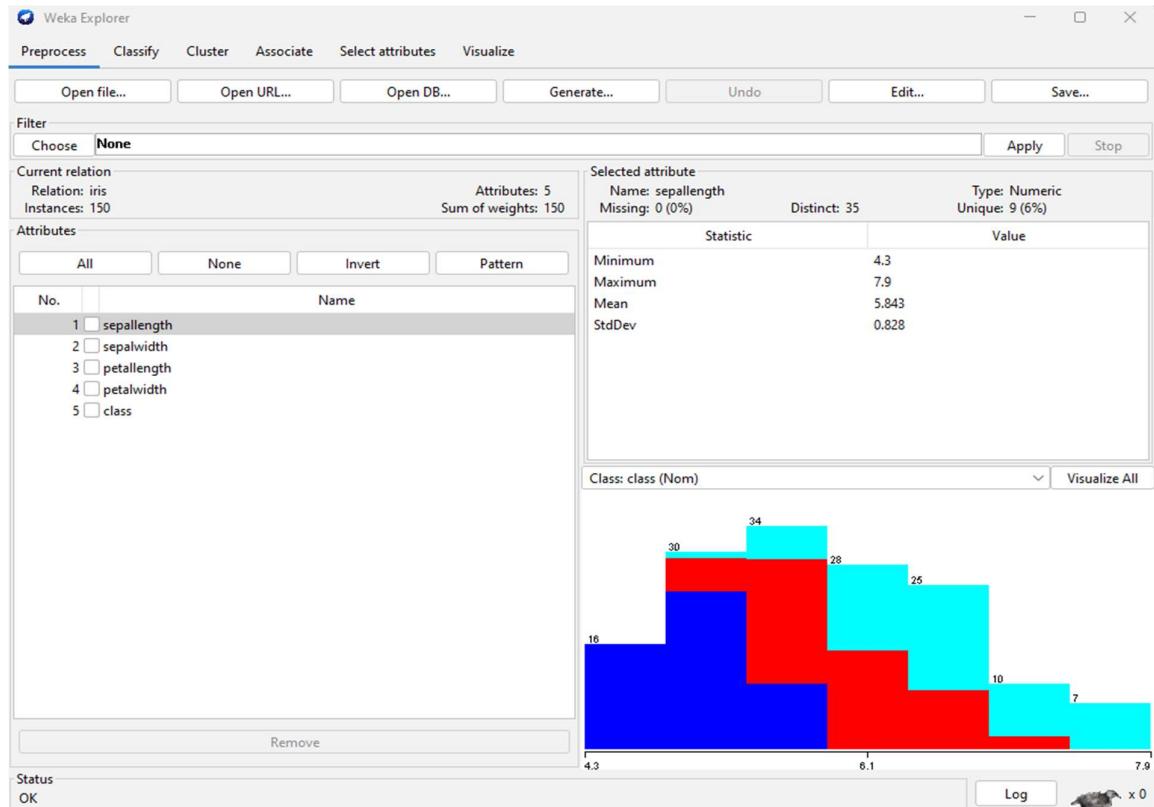
Aim: Evaluate performance using various metrics: Precision, Recall, F1-score, MCC, and ROC.

Theory:

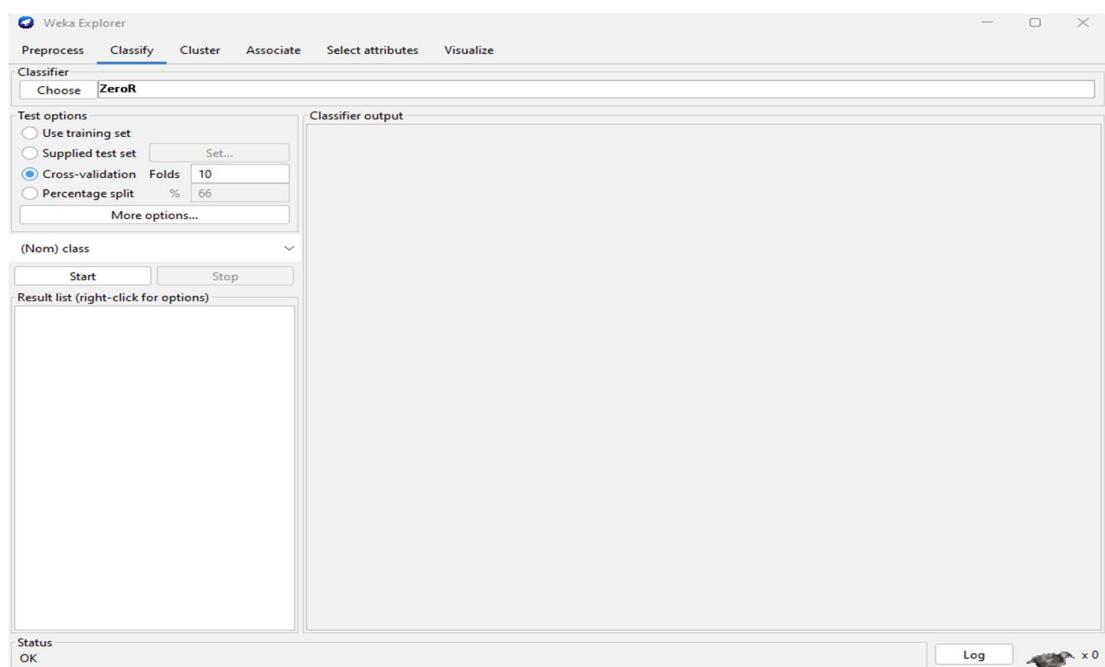
### Step 1:



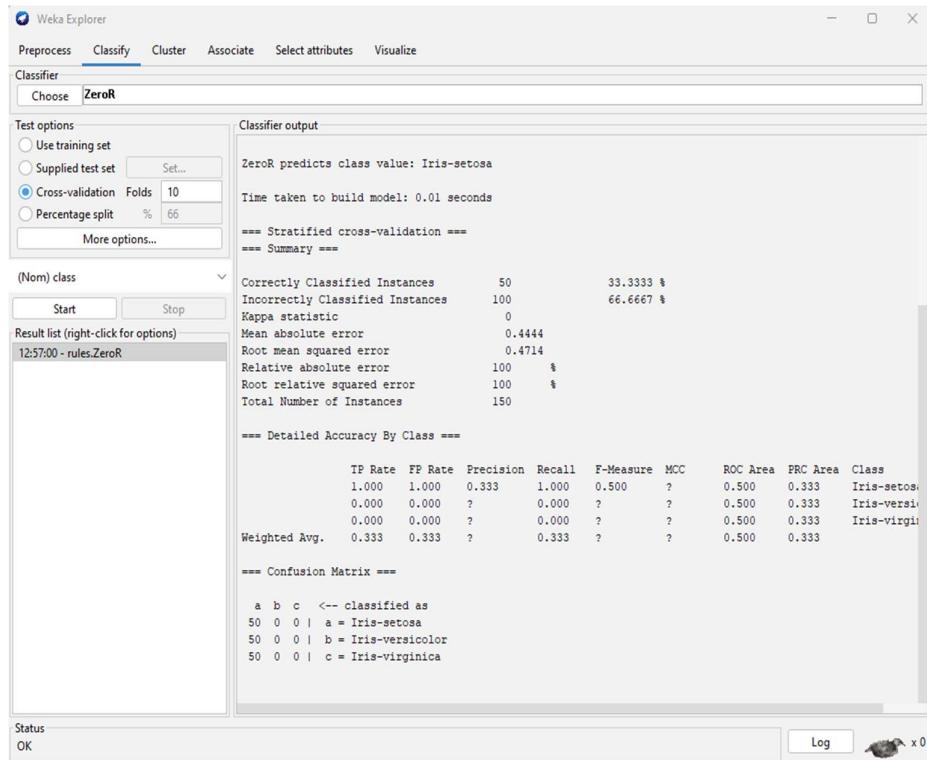
## Step 2:



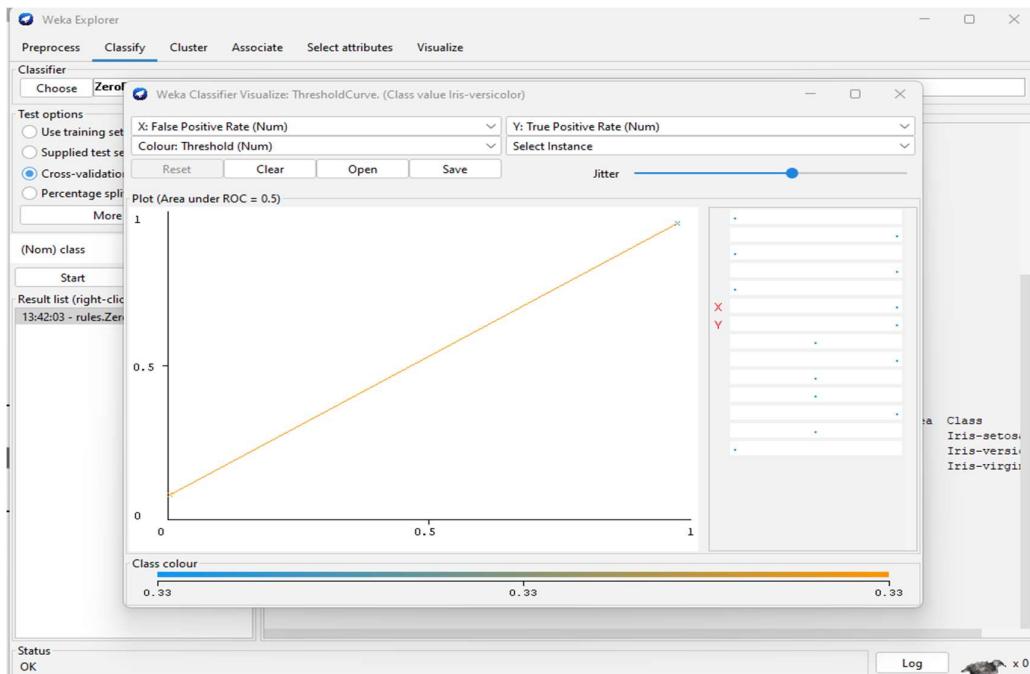
## Step 3:



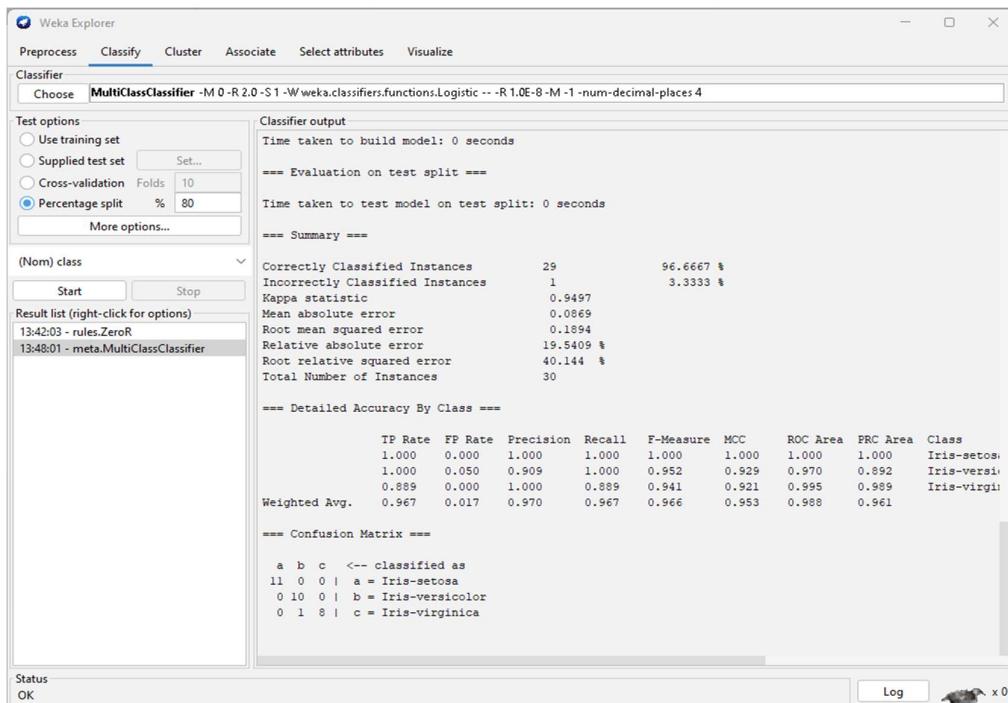
## Step 4:



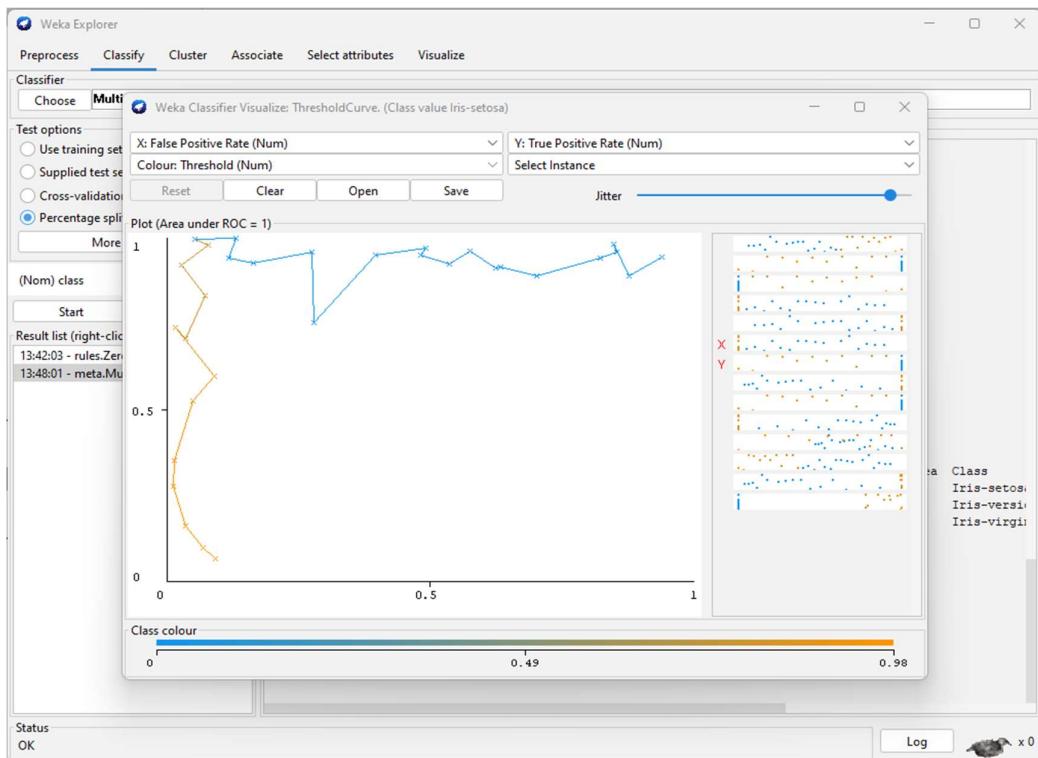
## Step 5:



## Step 6:



## Step 7:



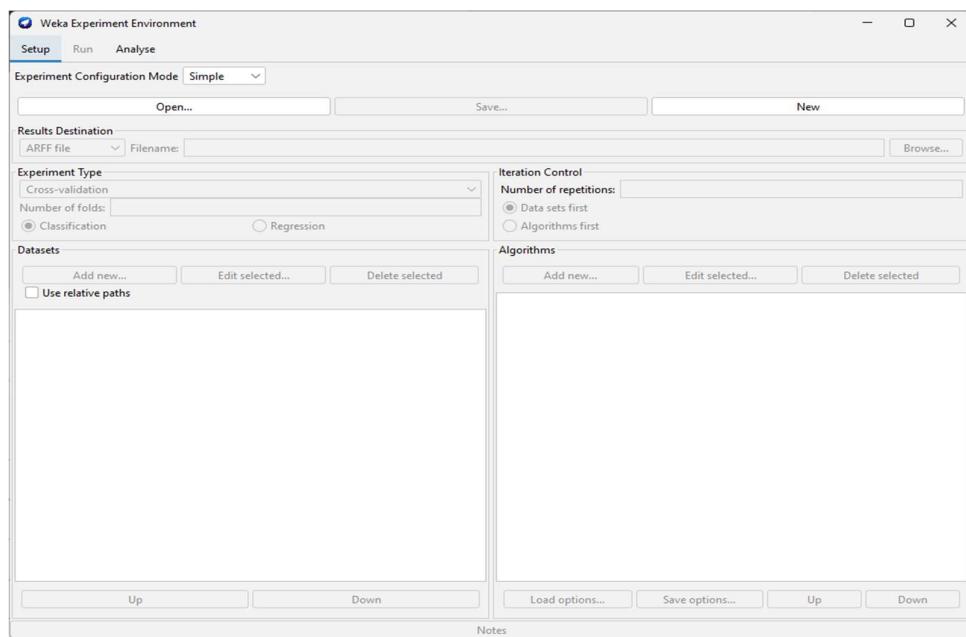
### Learning Outcomes:

## Experiment 8

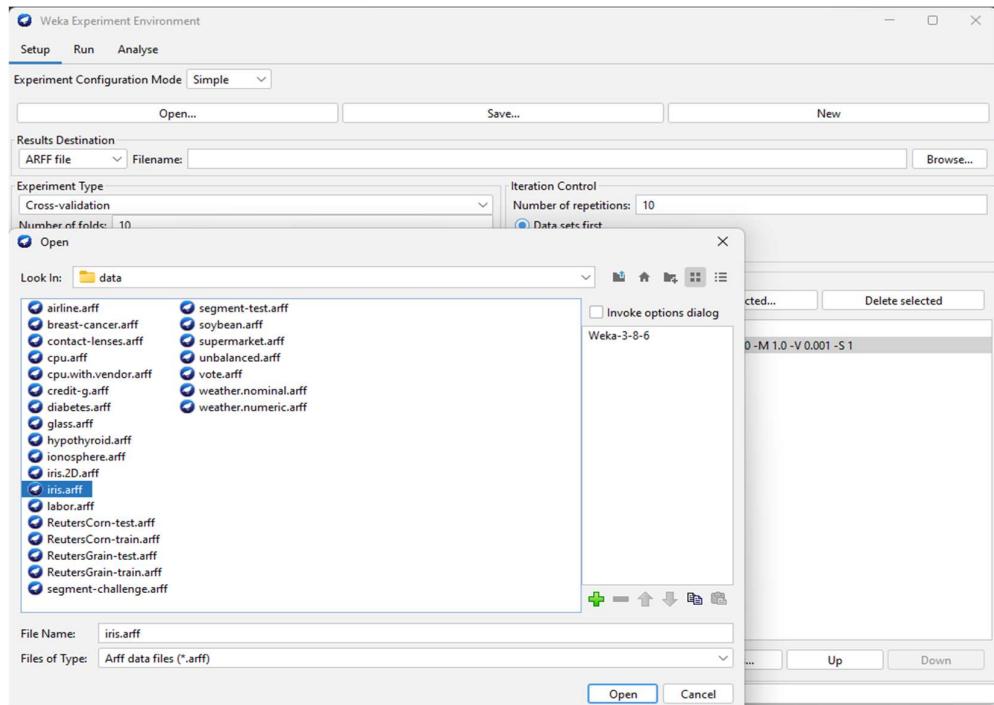
### Experiment 8: Implementation of Bagging and Boosting techniques on ARFF files using WEKA.

#### Theory:

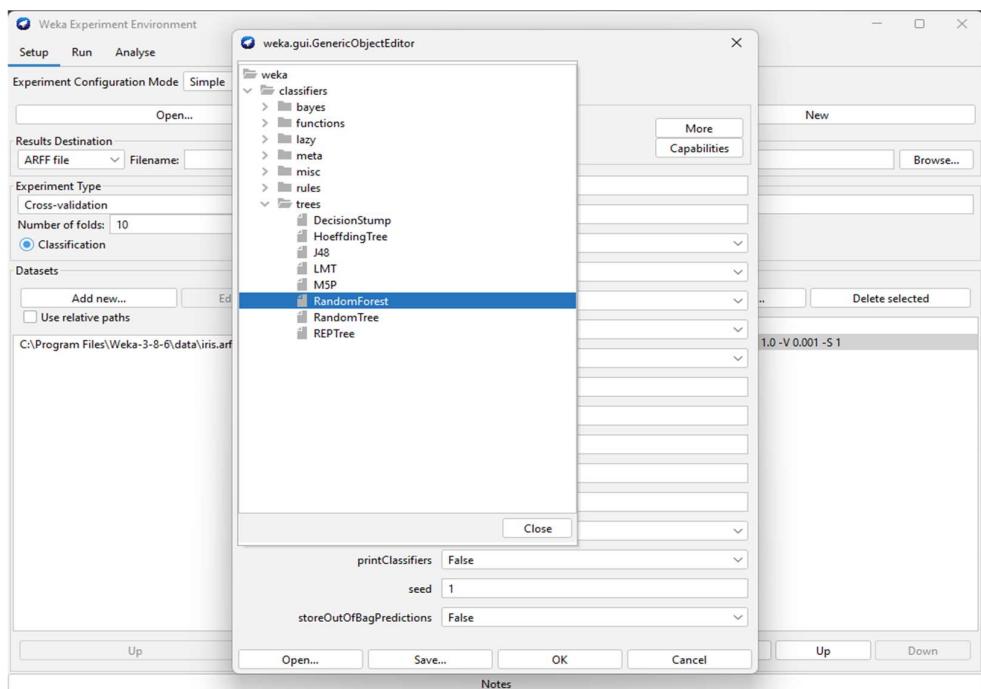
#### Step 1:



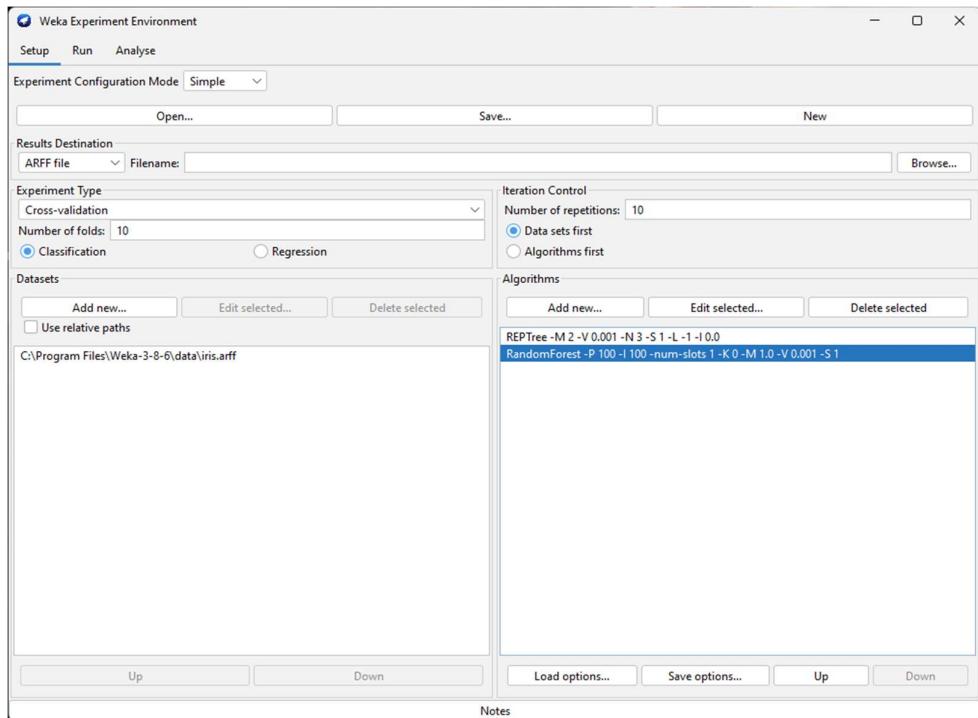
## Step 2:



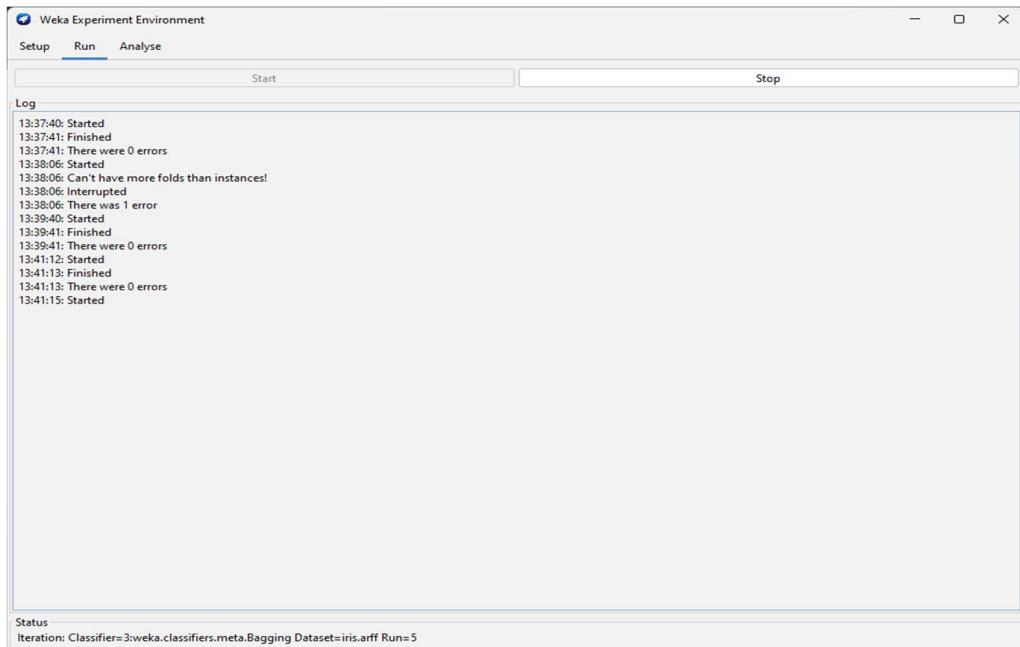
## Step 3:



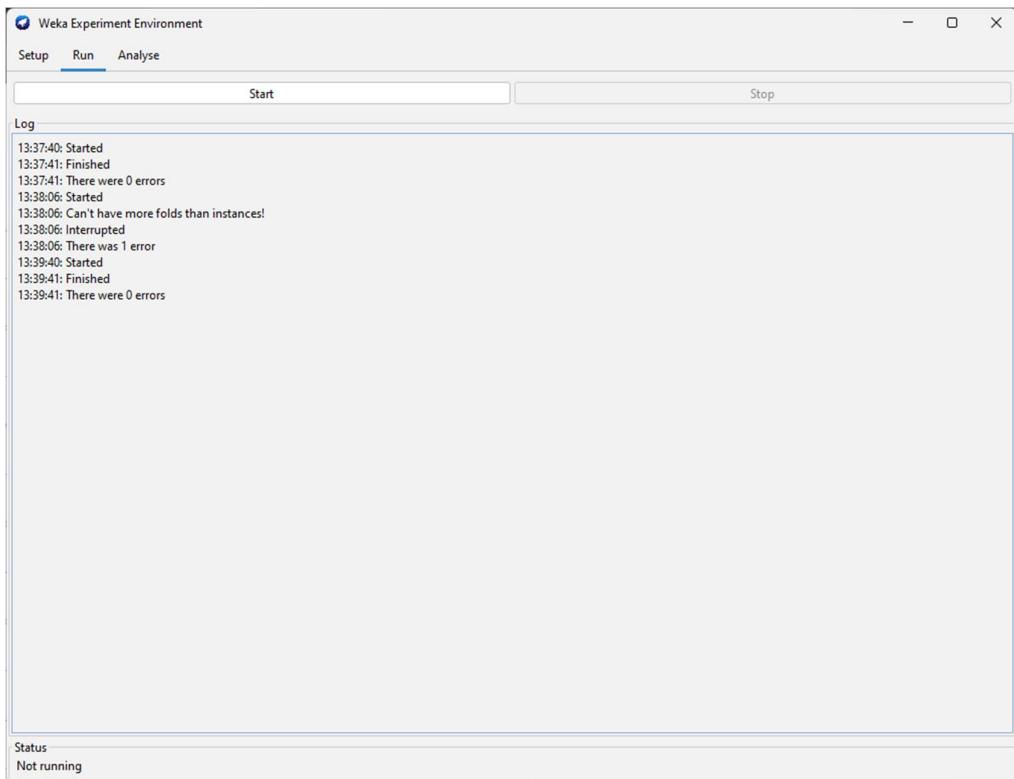
## Step 4:



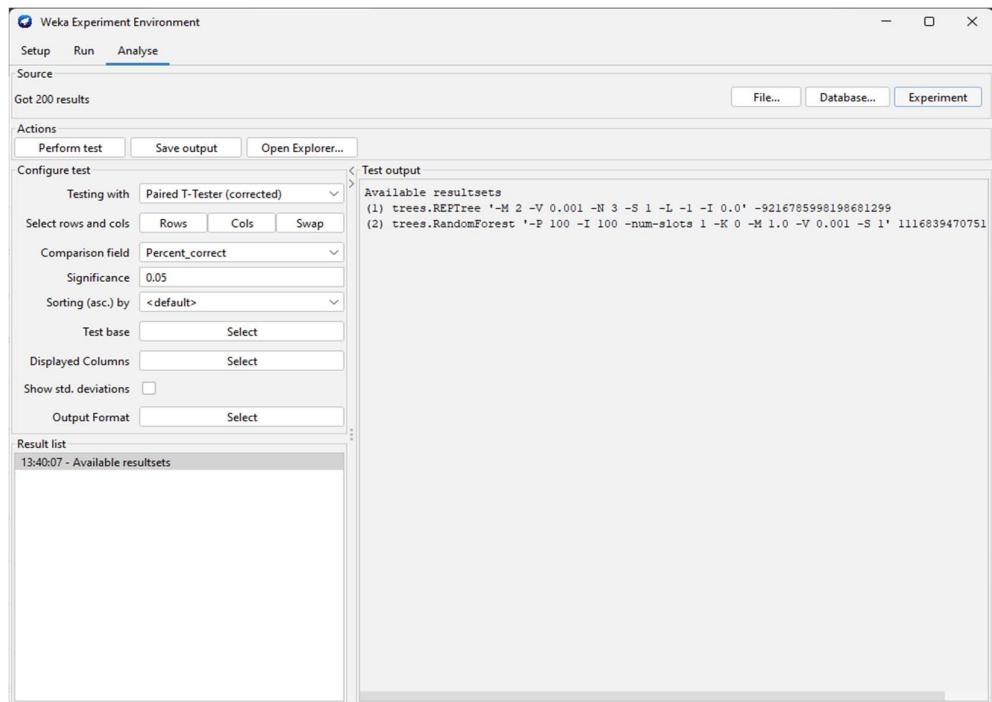
## Step 5:



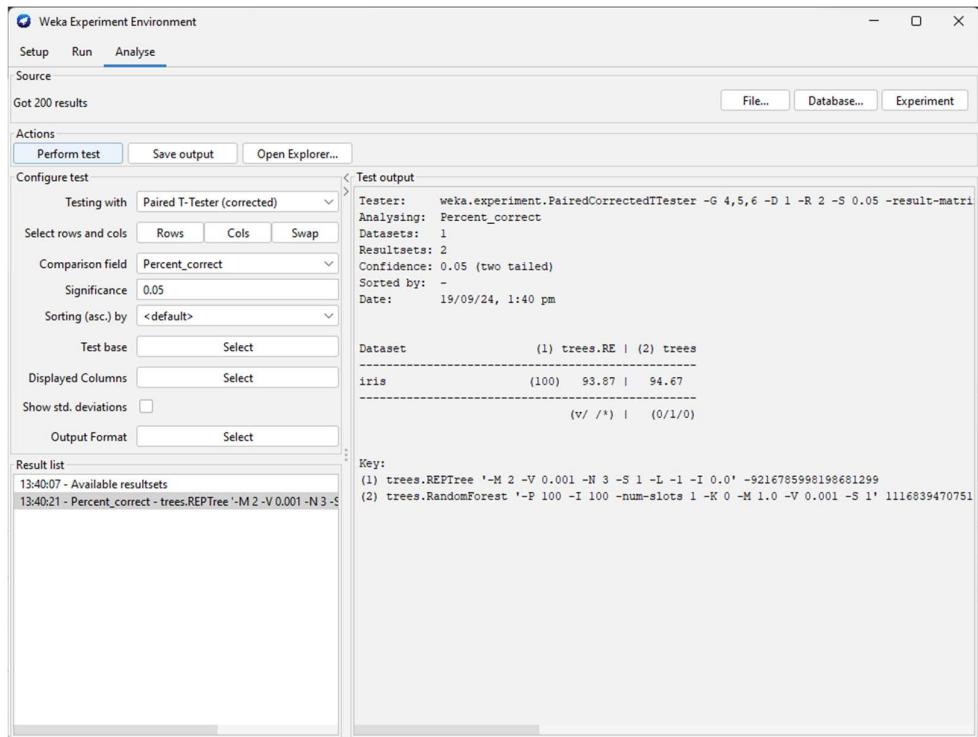
## Step 6:



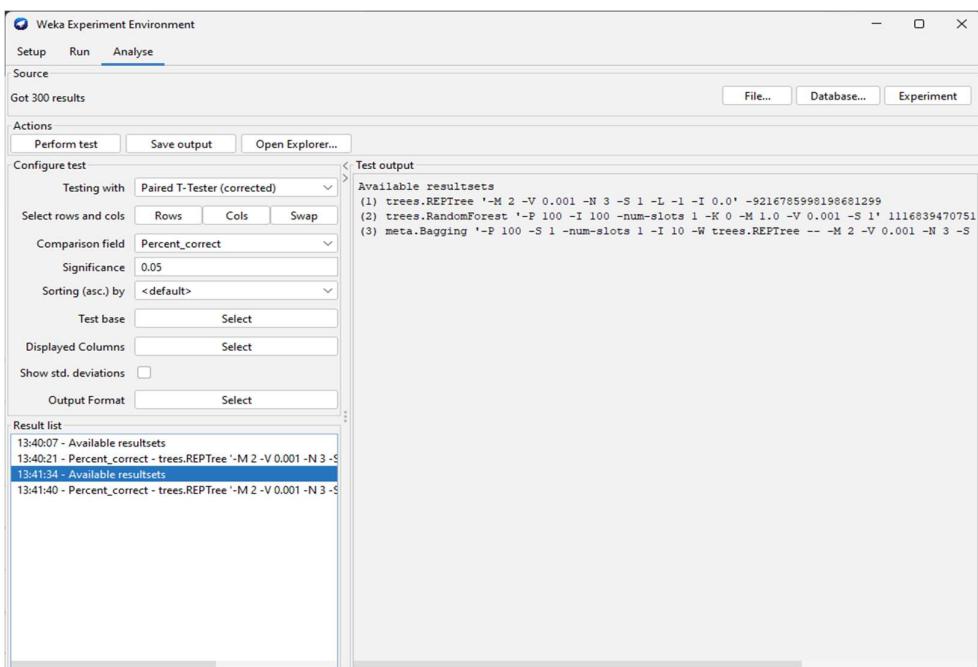
## Step 7:



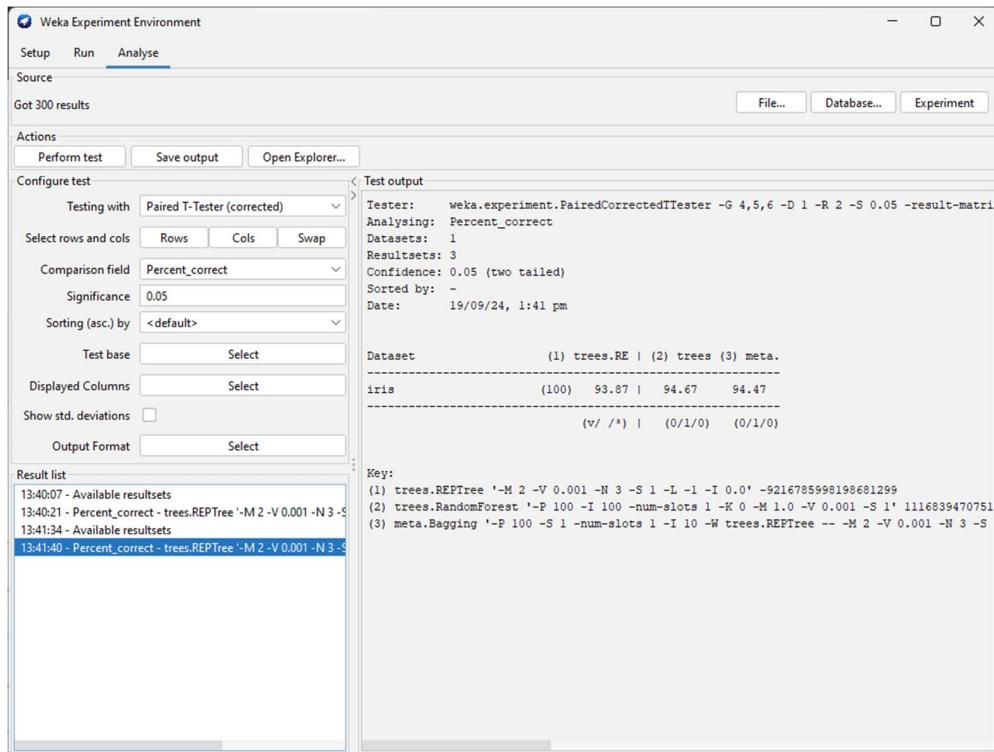
## Step 8:



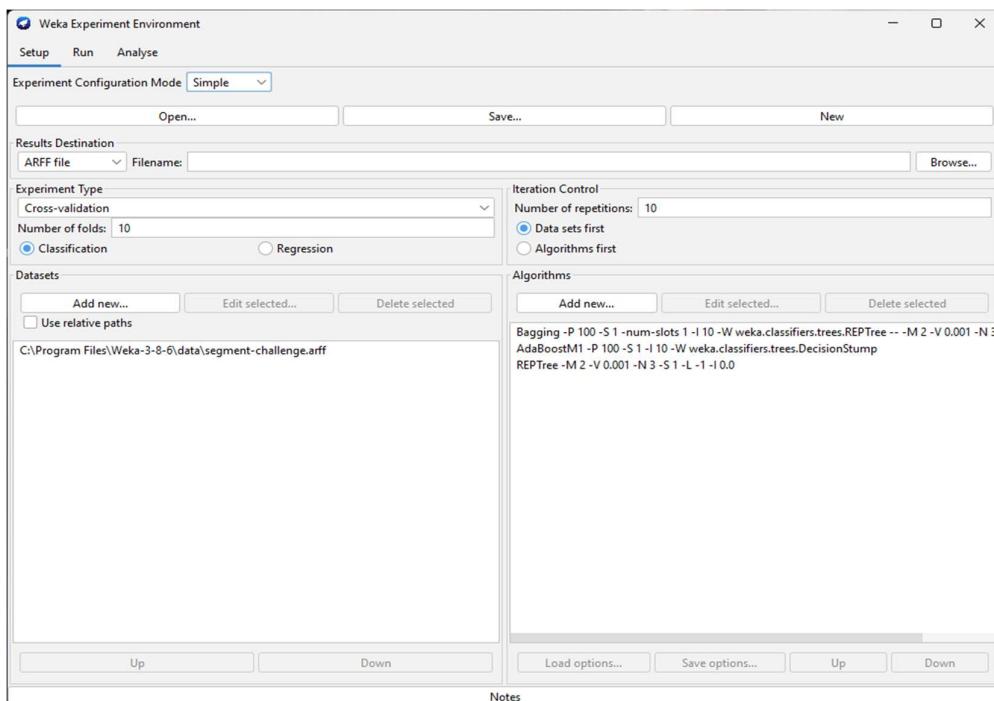
## Step 9:



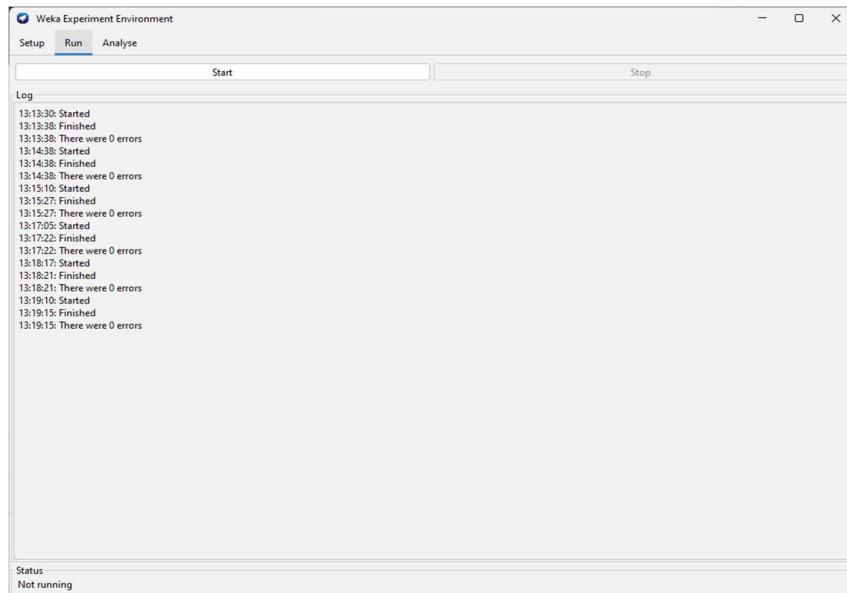
## Step 10:



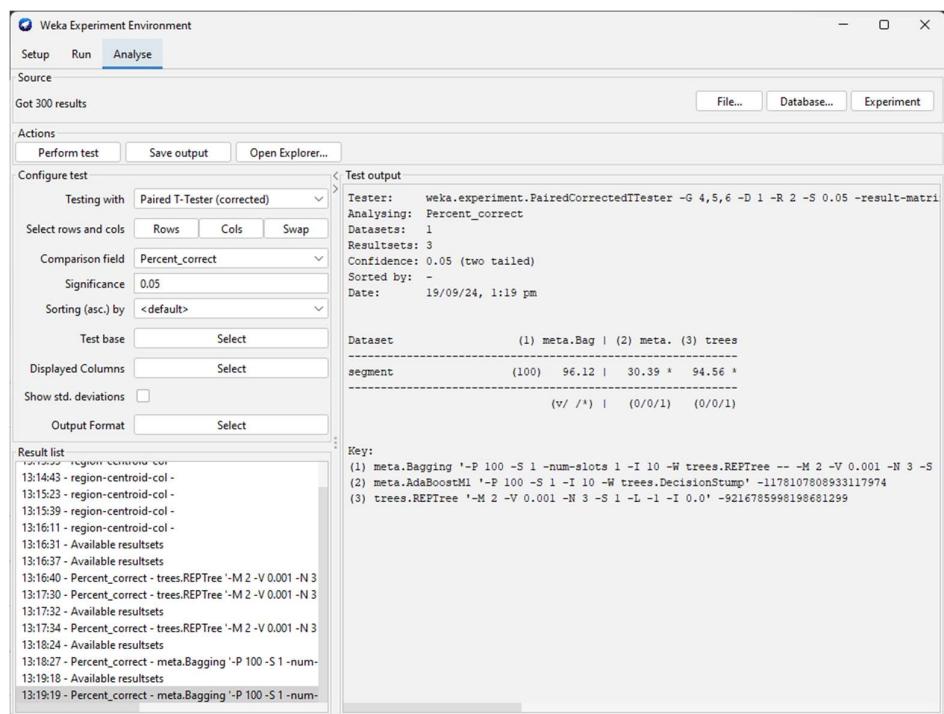
## Step 11:



## Step 12:



## Step 13:



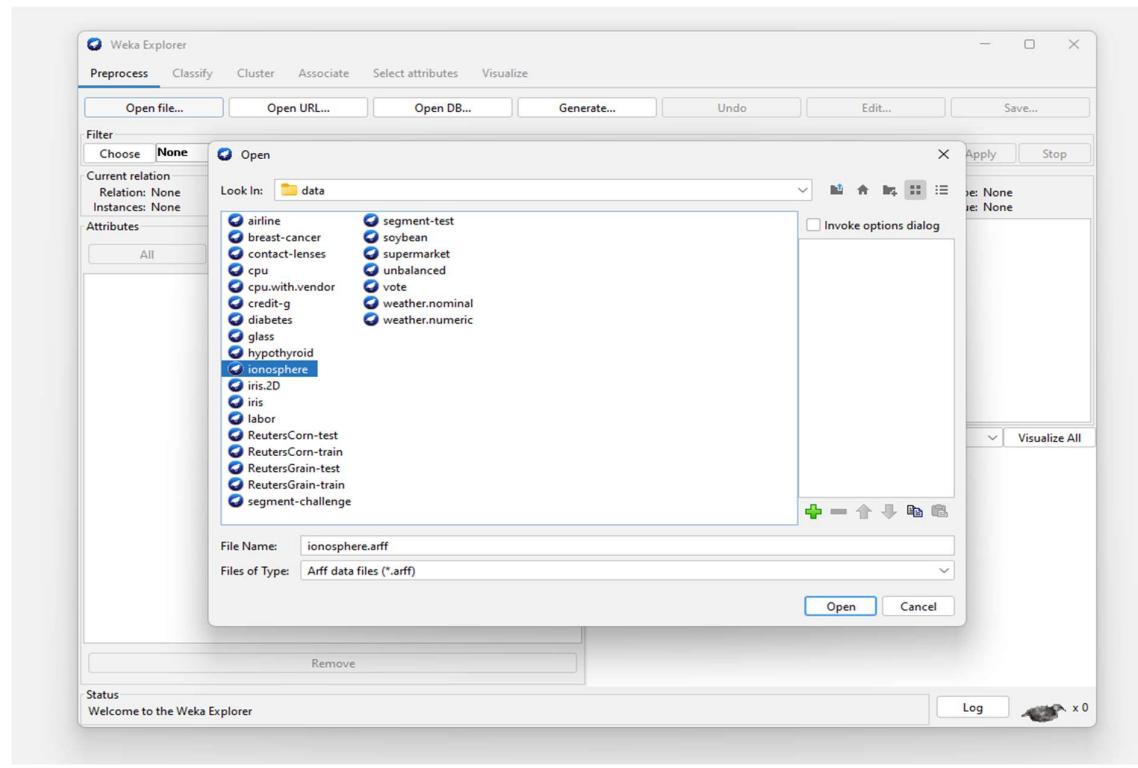
## Learning Outcomes:

## Experiment 9

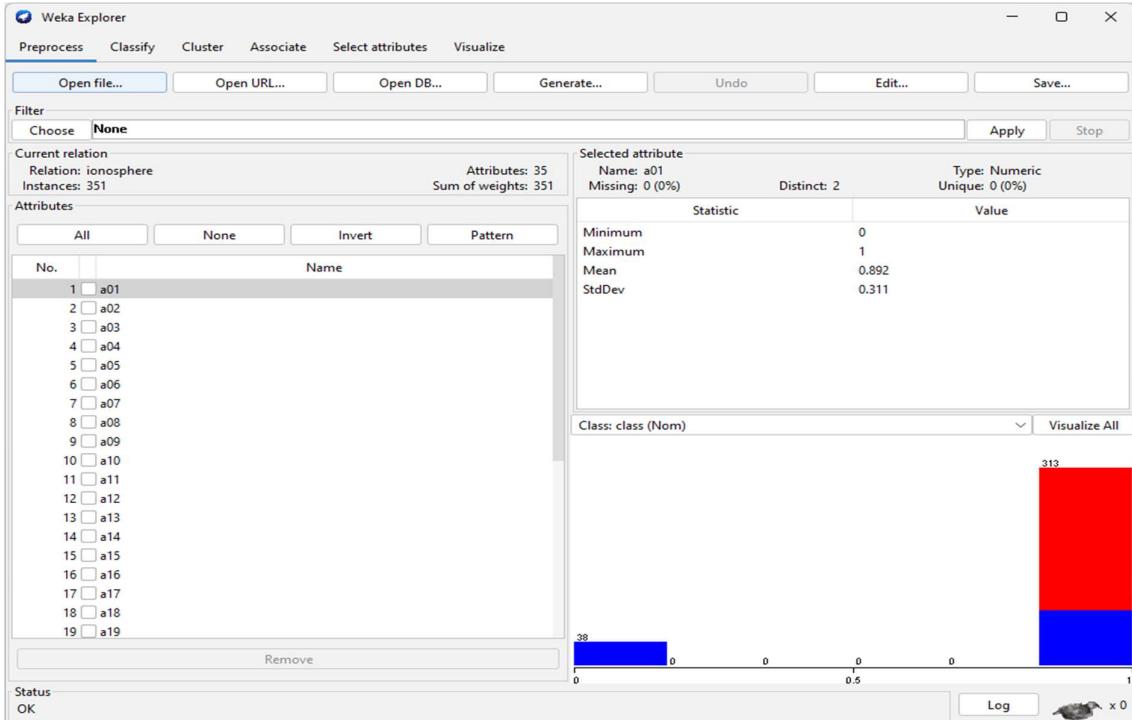
**Experiment 9: Apply the concept of Voting ensemble method to ARFF files and compare the results with single classifiers.**

**Theory:**

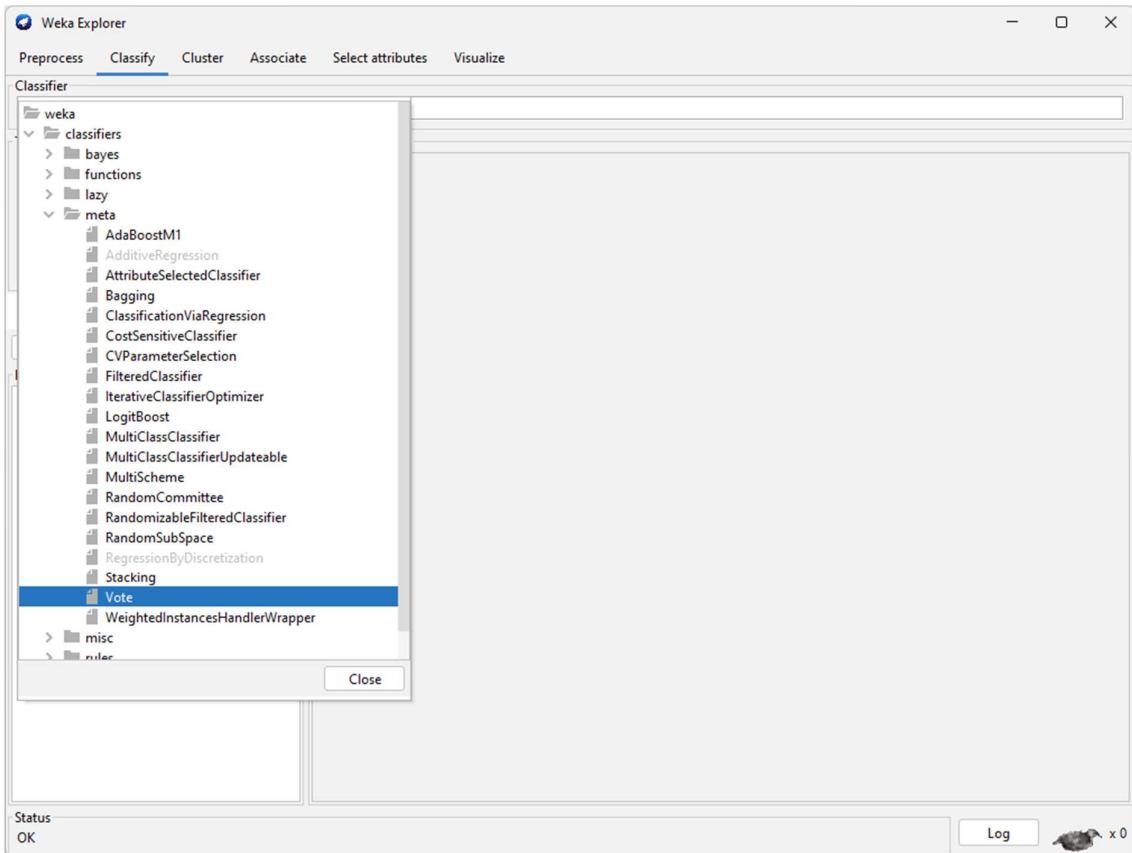
**Step 1:**



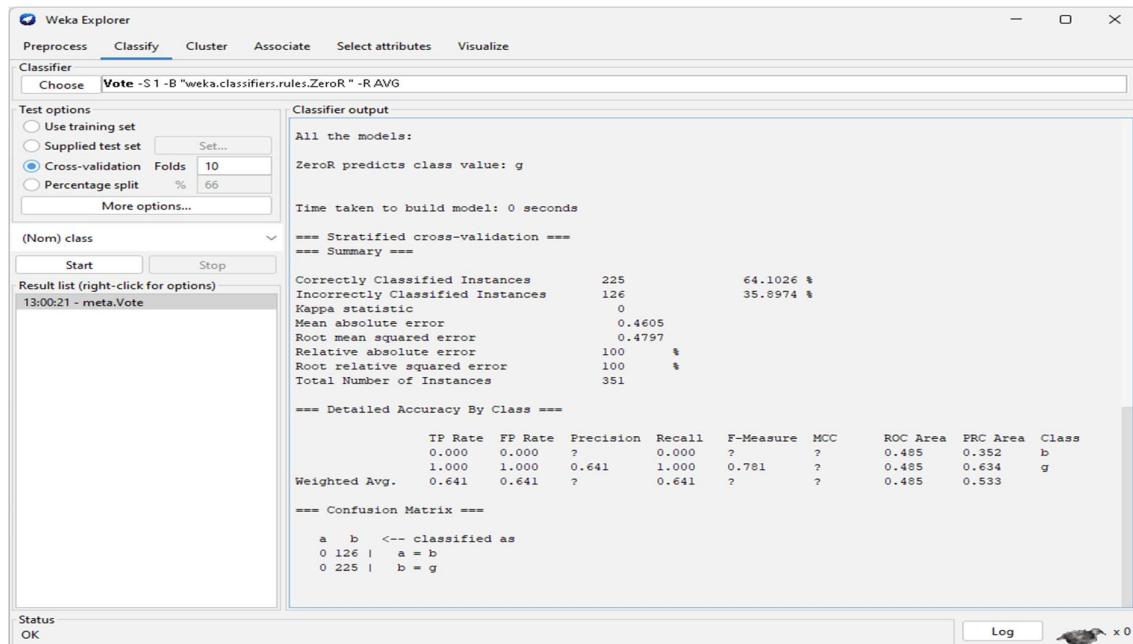
## Step 2:



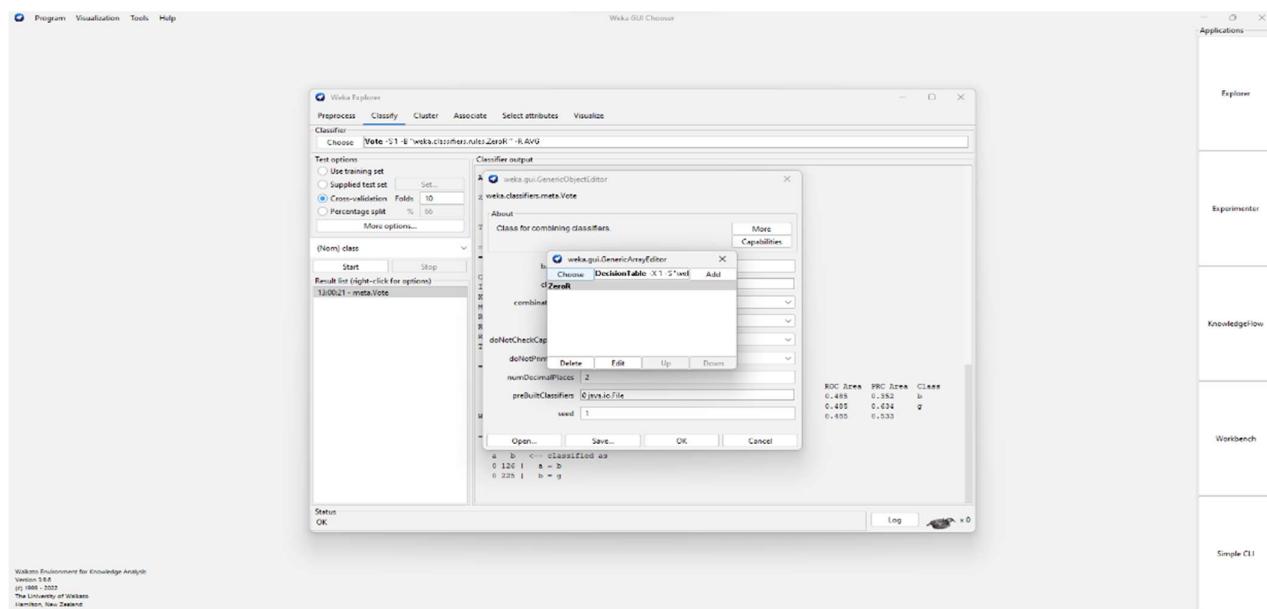
## Step 3:



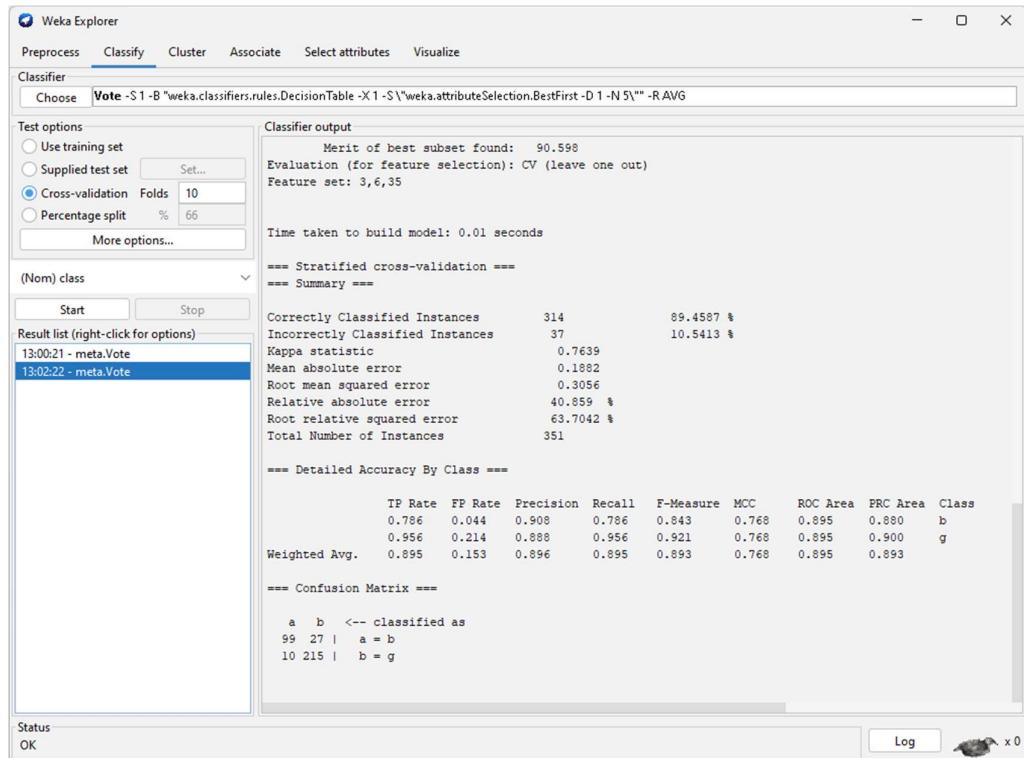
## Step 4:



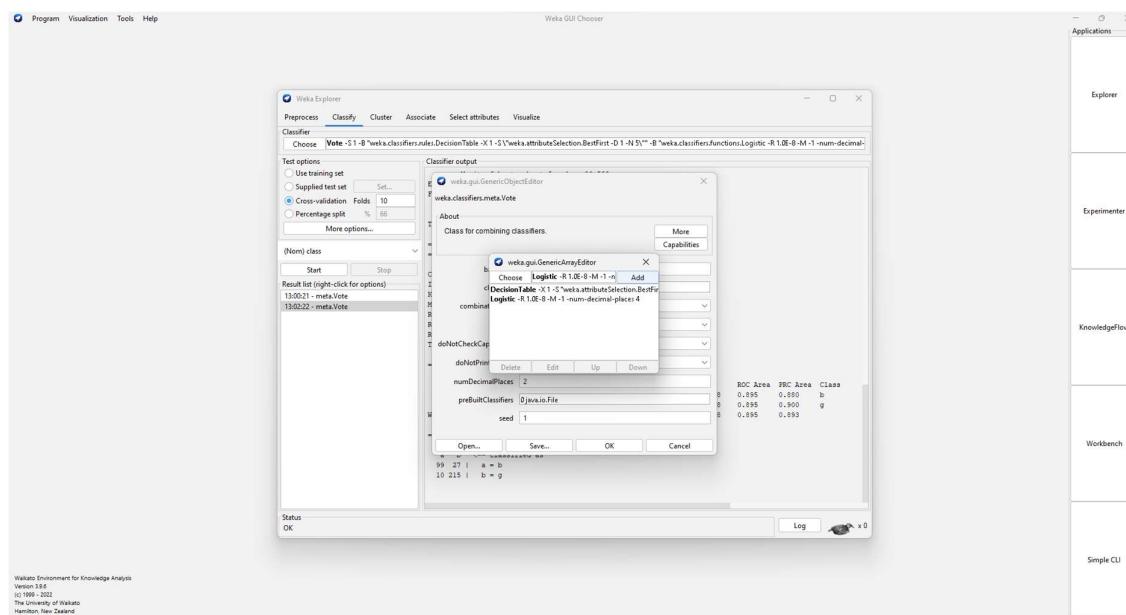
## Step 5:



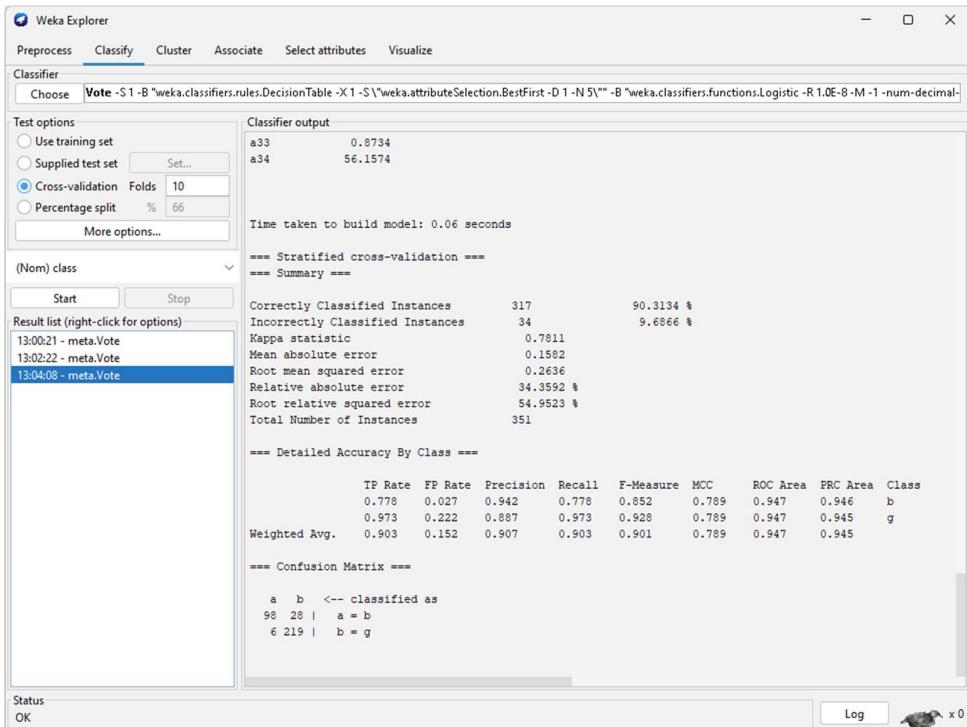
## Step 6:



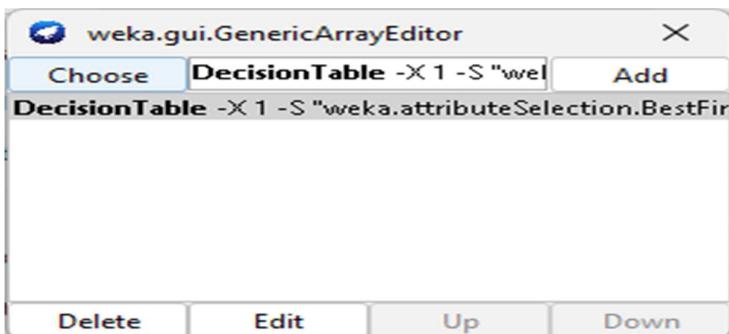
## Step 7:



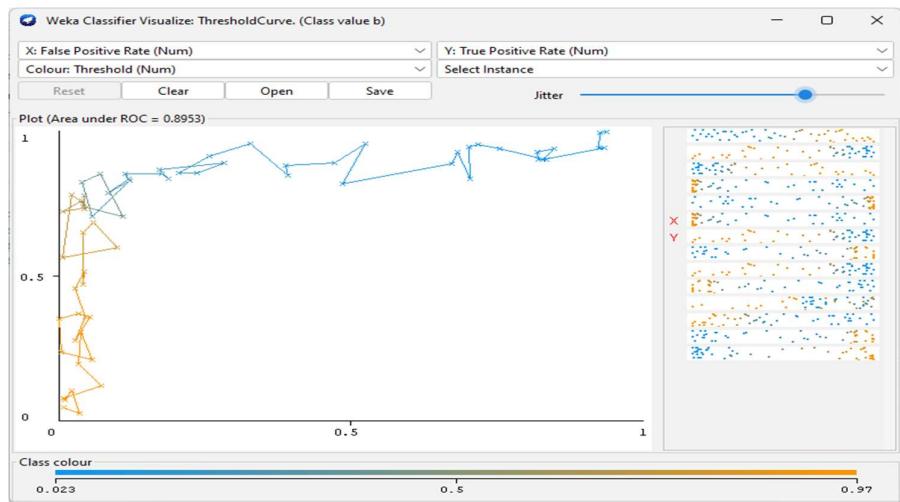
## Step 8:



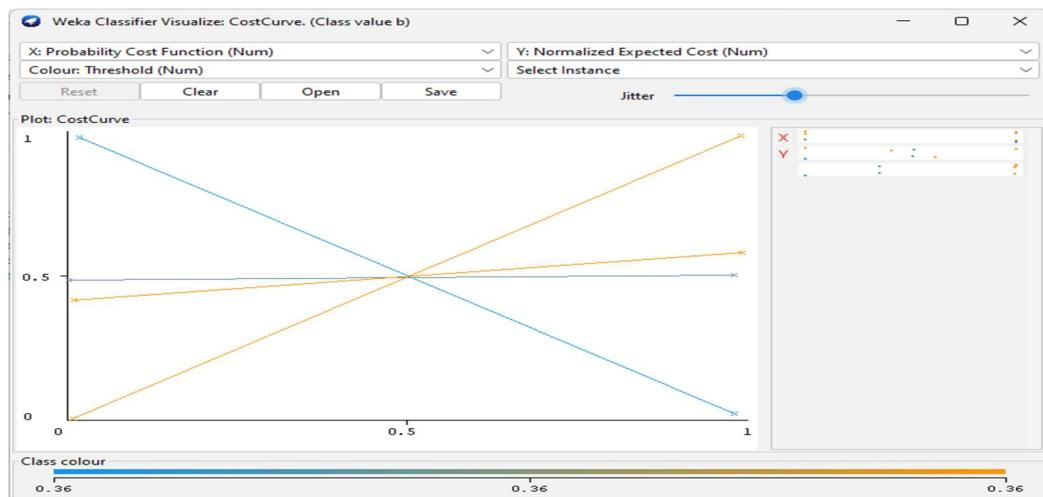
## Step 9:



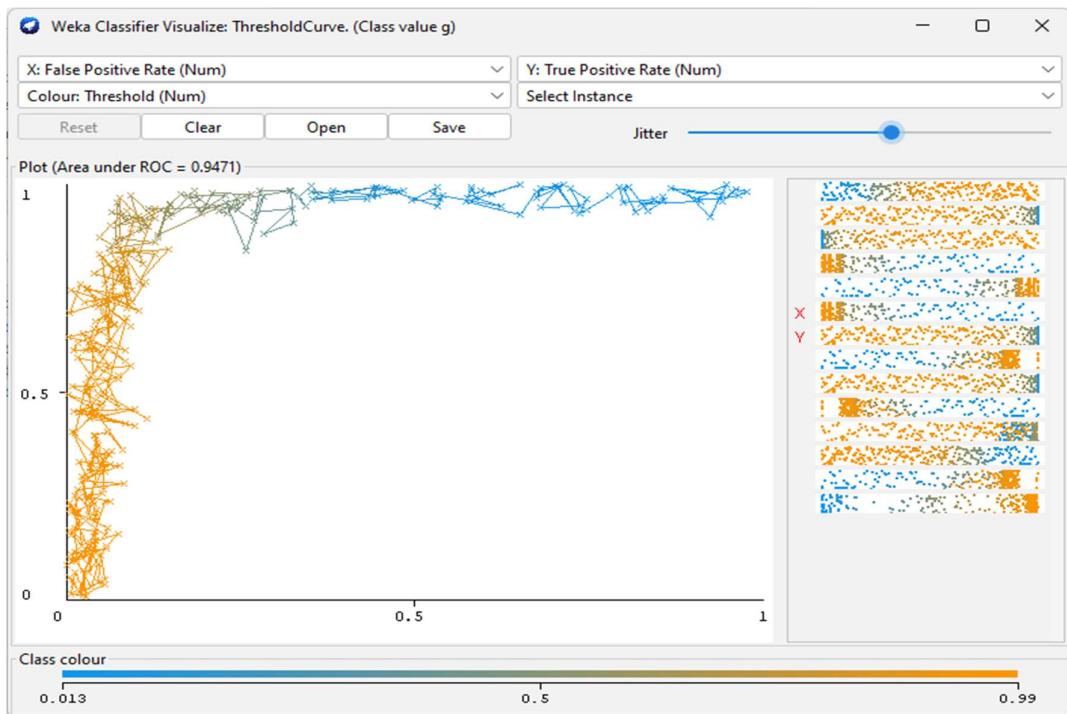
## Step 10:



### Step 11:



### Step 12:



### Learning Outcomes: