

## **23481A0514**

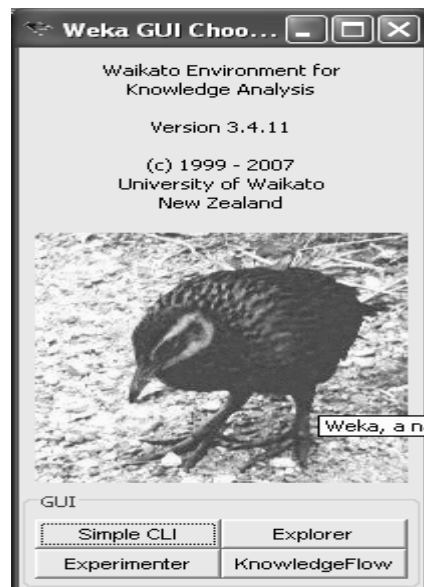
### **Experiment 3:**

**Aim:** Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets

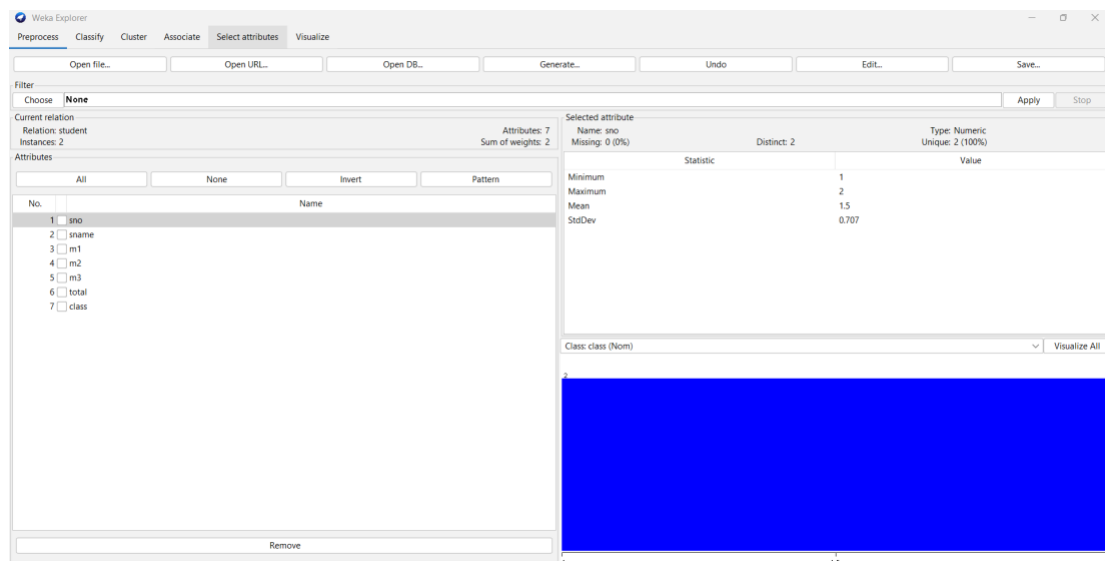
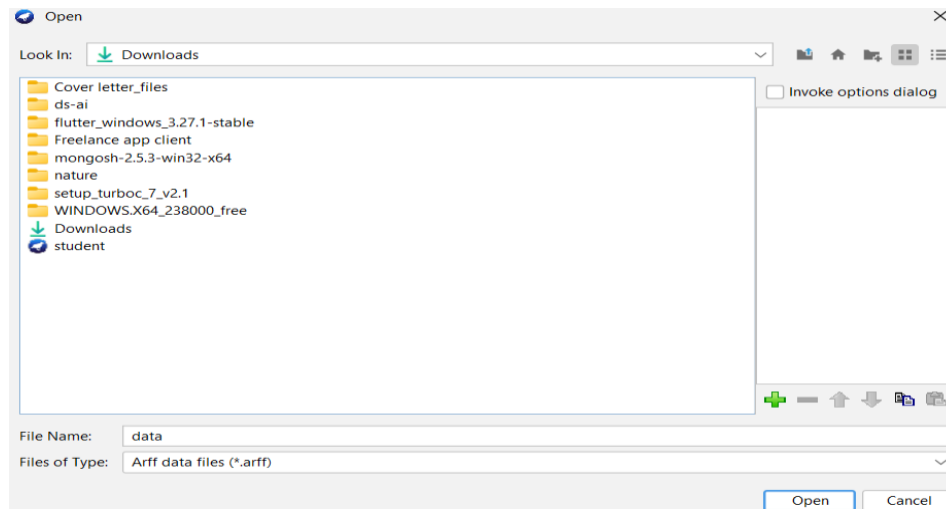
- Explore various options available in Weka for preprocessing data and apply Unsupervised filters like Discretization, Resample filter, etc. on each dataset
  - Load weather. nominal, Iris, Glass datasets into Weka and run Apriori Algorithm with different support and confidence values.
  - Study the rules generated. Apply different discretization filters on numerical attributes and run the Apriori association rule algorithm. Study the rules generated.
  - Derive interesting insights and observe the effect of discretization in the rule generation process.
- i. Add attribute    ii. Add expression    iii. Copy attribute    iv. Remove attribute

### **Filters**

Step-1: - Go to start button then select All Programs and then select weka 3.4.1.1



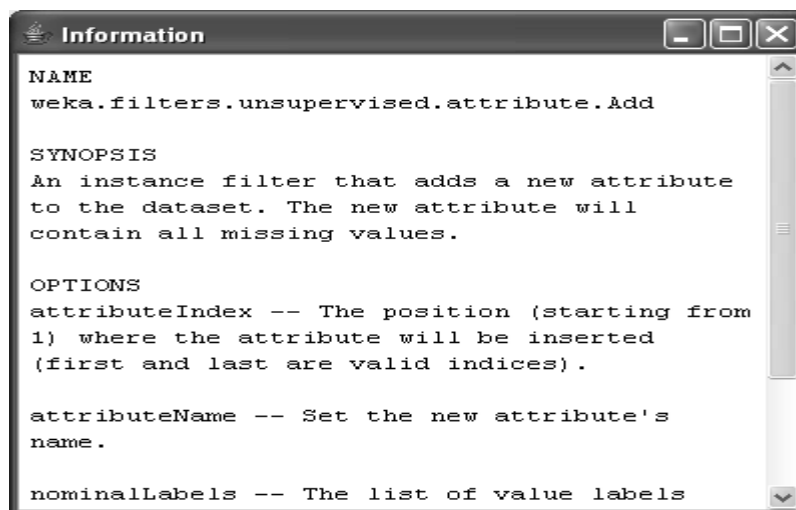
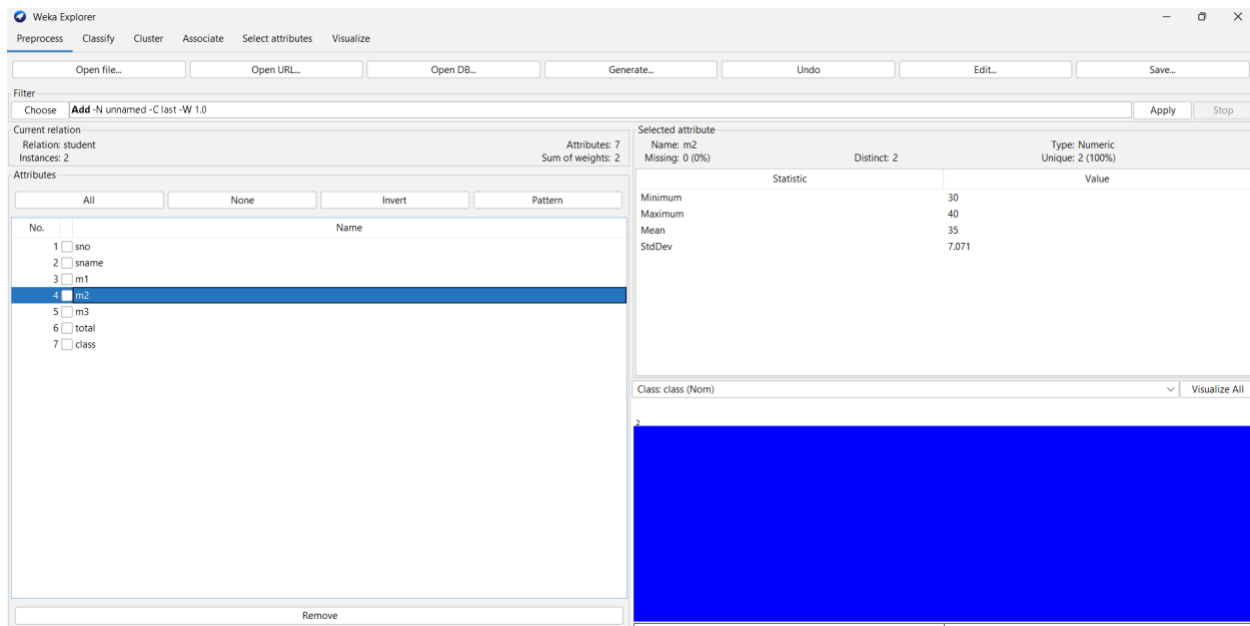
Click on Explorer



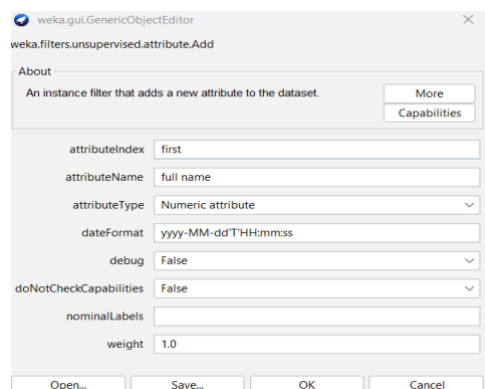
To open the required file

**ADD ATTRIBUTE**

Step-2: - Go to preprocess menu and choose then select attribute of Add option.



Step-3: -To enter the required fields is added to table.




Click on ok button

No.	Name
1	<input type="checkbox"/> full name
2	<input type="checkbox"/> sno
3	<input type="checkbox"/> sname
4	<input type="checkbox"/> m1
5	<input type="checkbox"/> m2
6	<input type="checkbox"/> m3
7	<input type="checkbox"/> total
8	<input type="checkbox"/> class
9	<input type="checkbox"/> avg

Click on Apply button.

Viewer								
relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka								
No.	1: full name	2: sno	3: sname	4: m1	5: m2	6: m3	7: total	8: class
	Numeric	Numeric	String	Numeric	Numeric	Numeric	Numeric	Nominal
1		1.0	a	20.0	30.0	40.0	50.0	A
2		2.0	b	30.0	40.0	10.0	50.0	A

Step-4: - if the index position is last to enter the fields

 weka.gui.GenericObjectEditor

✕

weka.filters.unsupervised.attribute.Add

About

An instance filter that adds a new attribute to the dataset.

More

Capabilities

attributeIndex last

attributeName avg

Click on Ok button.

Weka Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   Visualize

Open file...   Open URL...   Open DB...

Filter  
Choose **Add -N avg -C last -W 1.0**

Current relation  
Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.Ad...   Attribut  
Instances: 2   Sum of weigh

Attributes  
All   None   Invert   Pattern

No.	Name
1	<input checked="" type="checkbox"/> sno
2	<input type="checkbox"/> sname
3	<input type="checkbox"/> m1
4	<input type="checkbox"/> m2
5	<input type="checkbox"/> m3
6	<input type="checkbox"/> total
7	<input type="checkbox"/> class
8	<input type="checkbox"/> full name
9	<input type="checkbox"/> avg

Click on Apply button.

Viewer

Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.A

No.	1: full name Numeric	2: sno Numeric	3: sname String	4: m1 Numeric	5: m2 Numeric	6: m3 Numeric	7: total Numeric	8: class Nominal	9: avg Numeric
1		1.0	a	20.0	30.0	40.0	50.0	A	
2		2.0	b	30.0	40.0	10.0	50.0	A	

Click on ok button.

Step-5: -if the index position is middle

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Add

About

An instance filter that adds a new attribute to the dataset.

More

Capabilities

attributeIndex   5

attributeName   max\_marks


Click on ok button.

Attributes

AllNoneInvert

No.	Name
1	<input type="checkbox"/> full name
2	<input type="checkbox"/> sno
3	<input type="checkbox"/> sname
4	<input type="checkbox"/> m1
5	<input type="checkbox"/> max_marks
6	<input type="checkbox"/> m2
7	<input type="checkbox"/> m3
8	<input type="checkbox"/> total
9	<input type="checkbox"/> class
10	<input type="checkbox"/> avg

Click on Apply button.

 Viewer

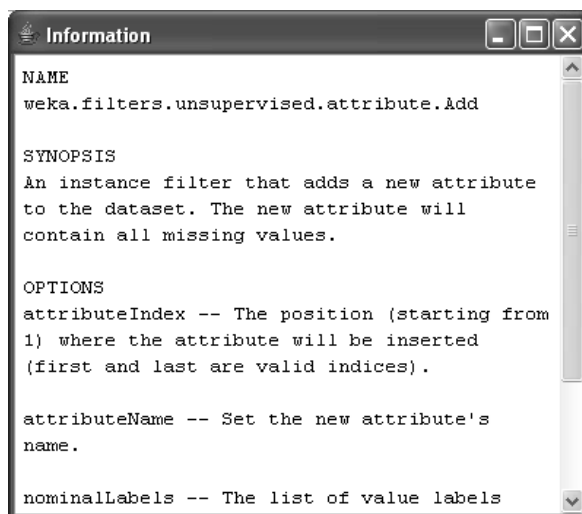
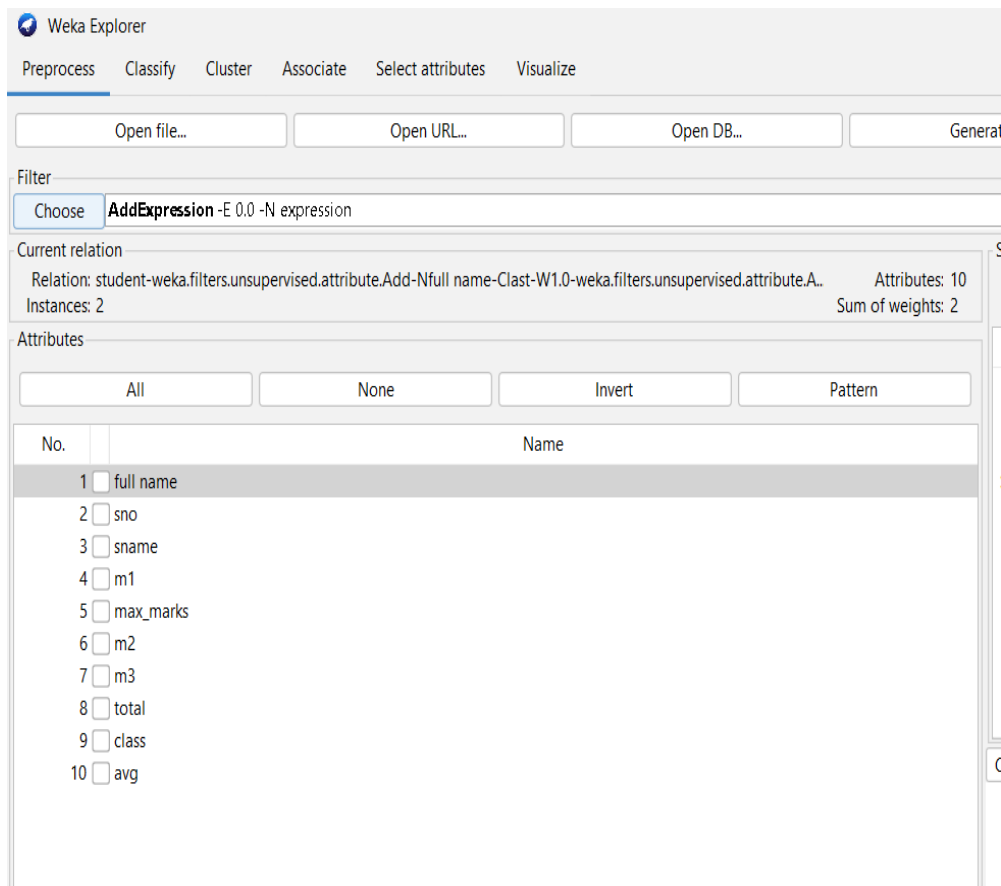
Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.Add-

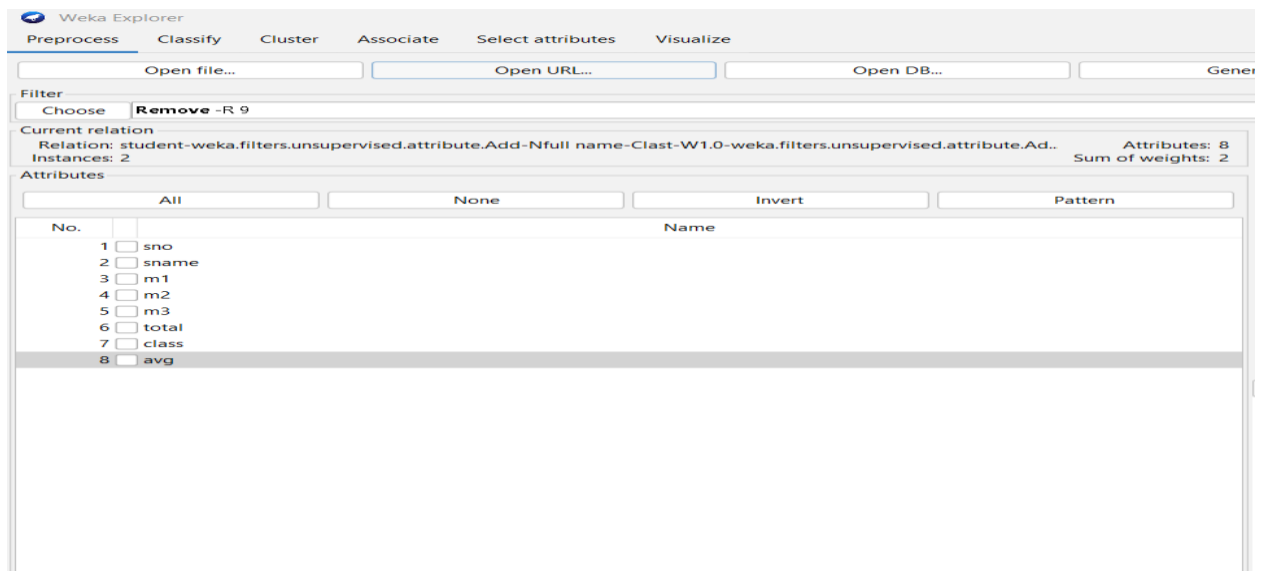
No.	1: full name Numeric	2: sno Numeric	3: sname String	4: m1 Numeric	5: max_marks Numeric	6: m2 Numeric	7: m3 Numeric	8: total Numeric	9: class Nominal	10: avg Numeric
1		1.0	a	20.0		30.0	40.0	50.0	A	
2		2.0	b	30.0		40.0	10.0	50.0	A	

Click on ok button.

**Add Expression**

Step-6: - click on choose button and then select AddExpression option.





Click on Apply option.

Viewer

Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-w

No.	1: sno Numeric	2: sname String	3: m1 Numeric	4: m2 Numeric	5: m3 Numeric	6: sum Numeric	7: avg Numeric
1	1.0	a	20.0	30.0	40.0	90.0	30.0
2	2.0	b	30.0	40.0	10.0	80.0	26.666...

Viewer

Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised

No.	1: sno Numeric	2: sname String	3: m1 Numeric	4: m2 Numeric	5: m3 Numeric	6: sum Numeric	7: avg Numeric	8: max_marks Numeric
1	1.0	a	20.0	30.0	40.0	90.0	30.0	40.0
2	2.0	b	30.0	40.0	10.0	80.0	26.666...	40.0

Click on ok button.



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose **AddExpression**  $E \text{ "ifelse}(a3 > a4 \text{ and } a3 > a5, a3, \text{ifelse}(a4 > a5, a4, a5)) \sim -N \text{ max\_marks}$  Apply Stop

Current relation  
 Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.Ad...  
 Instances: 2  
 Attributes: 8  
 Sum of weights: 2

Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> sno
2	<input type="checkbox"/> sname
3	<input type="checkbox"/> m1
4	<input type="checkbox"/> m2
5	<input checked="" type="checkbox"/> m3
6	<input type="checkbox"/> sum
7	<input type="checkbox"/> avg
8	<input type="checkbox"/> max_marks

Remove

Selected attribute  
 Name: m3  
 Missing: 0 (0%)  
 Distinct: 2  
 Type: Numeric  
 Unique: 2 (100%)

Statistic	Value
Minimum	10
Maximum	40
Mean	25
StdDev	21.213

Class: max\_marks (Num) Visualize All

2

10 25 40

**COPY**

Step-7: - Click on choose button and then select copy option.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB...

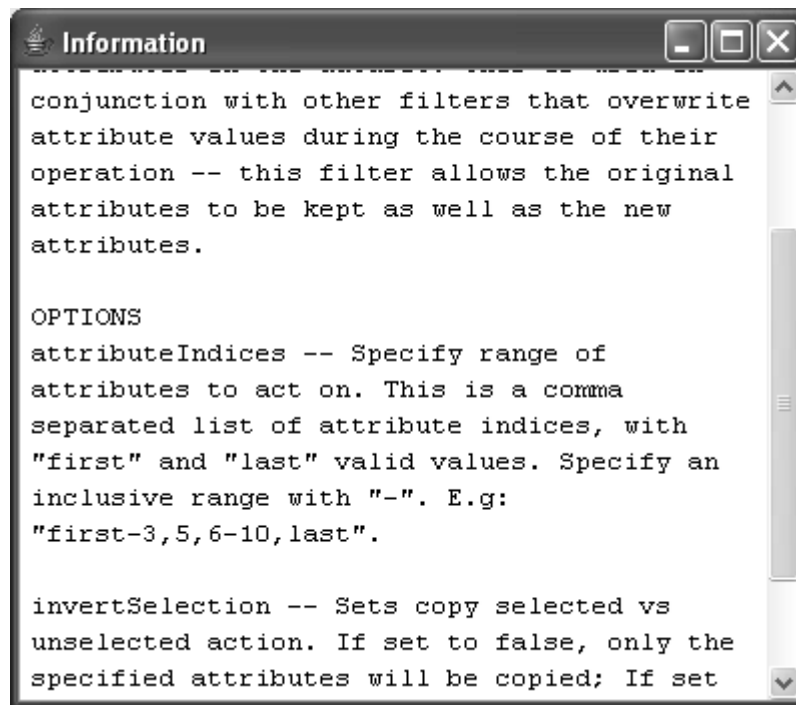
Filter: Choose **Copy**

Current relation  
 Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.Ad...  
 Instances: 2  
 Attributes: 8  
 Sum of weights: 2

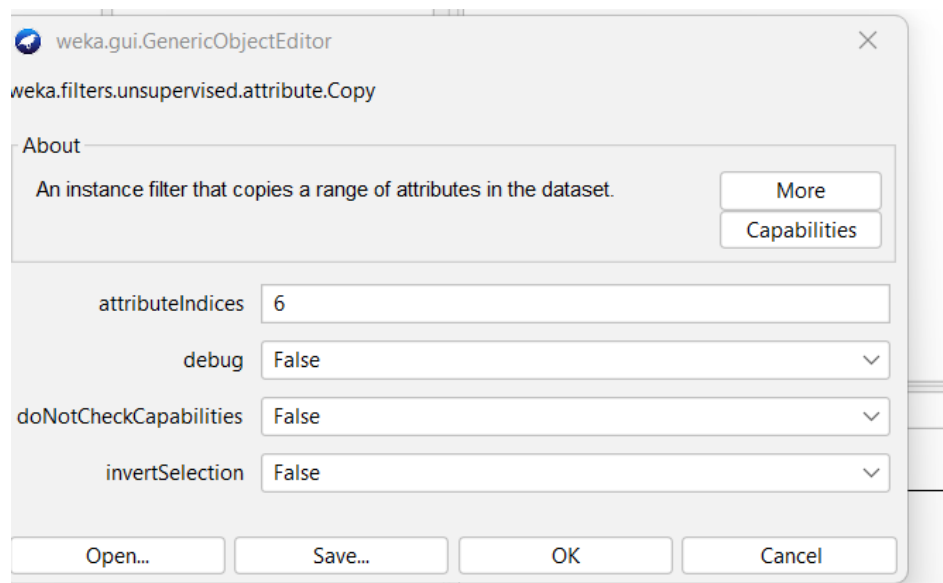
Attributes

All None Invert Pattern

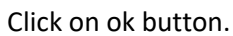
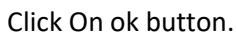
No.	Name
1	<input type="checkbox"/> sno
2	<input type="checkbox"/> sname
3	<input type="checkbox"/> m1
4	<input type="checkbox"/> m2
5	<input checked="" type="checkbox"/> m3
6	<input type="checkbox"/> sum
7	<input type="checkbox"/> avg
8	<input type="checkbox"/> max_marks



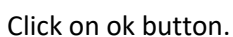
If the index place is any

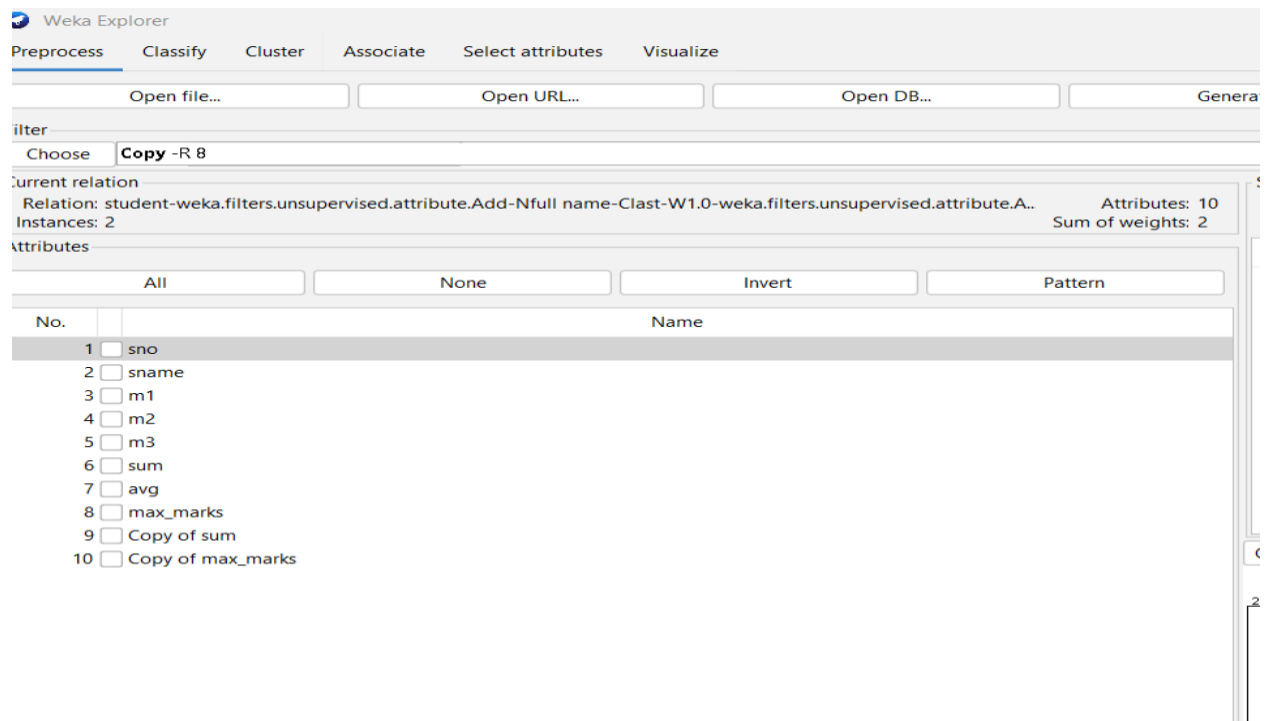


Click on ok button.



If the last position.



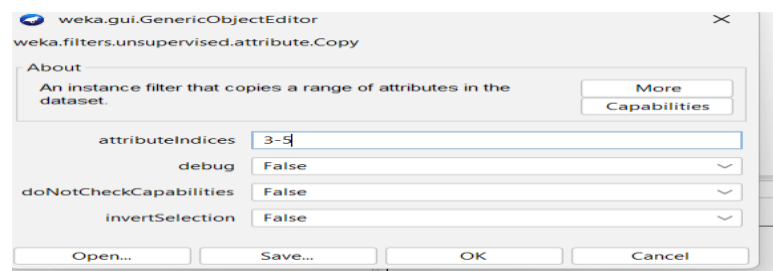


Click on Apply option.

No.	1: sno Numeric	2: sname String	3: m1 Numeric	4: m2 Numeric	5: m3 Numeric	6: sum Numeric	7: avg Numeric	8: max_marks Numeric	9: Copy of sum Numeric	10: Copy of max_marks Numeric
1.0	a		20.0	30.0	40.0	90.0	30.0	40.0	90.0	40.0
2.0	b		30.0	40.0	10.0	80.0	26.666...	40.0	80.0	40.0

Click on ok button.

If the index position range.



Click on ok button.

No.	Name
1	<input checked="" type="checkbox"/> sno
2	<input type="checkbox"/> sname
3	<input type="checkbox"/> m1
4	<input type="checkbox"/> m2
5	<input type="checkbox"/> m3
6	<input type="checkbox"/> sum
7	<input type="checkbox"/> avg
8	<input type="checkbox"/> max_marks
9	<input type="checkbox"/> Copy of sum
10	<input type="checkbox"/> Copy of max_marks
11	<input type="checkbox"/> Copy of m1
12	<input type="checkbox"/> Copy of m2
13	<input type="checkbox"/> Copy of m3

Click on Apply option.

relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.Add-Navg-Clast-W1.0-weka.filters.unsupervised.attribute.Remove-R8-													
No.	1: sno	2: sname	3: m1	4: m2	5: m3	6: sum	7: avg	8: max_marks	9: Copy of sum	10: Copy of max_marks	11: Copy of m1	12: Copy of m2	13: Copy of m3
	Numeric	String	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	1.0	a	20.0	30.0	40.0	90.0	30.0	40.0	90.0	40.0	20.0	30.0	40.0
2	2.0	b	30.0	40.0	10.0	80.0	26.666...	40.0	80.0	40.0	30.0	40.0	10.0

Click on ok button.

**REMOVE**

Step-8:- Click on choose button and then select Remove option

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit

Filter

Choose

Remove

R: 10

Current relation

Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.A-

Attributes: 13

Sum of weights: 2

Selected attribute

Name: sno

Missing: 0 (0%)

Instances: 2

Attributes

All

None

Invert

Pattern

Minimum

Statistic

Distinct: 2

No.	Name
1	<input checked="" type="checkbox"/> sno
2	<input type="checkbox"/> sname
3	<input type="checkbox"/> m1
4	<input type="checkbox"/> m2
5	<input type="checkbox"/> m3
6	<input type="checkbox"/> sum
7	<input type="checkbox"/> avg
8	<input type="checkbox"/> max_marks
9	<input type="checkbox"/> Copy of sum
10	<input type="checkbox"/> Copy of max_marks
11	<input type="checkbox"/> Copy of m1
12	<input type="checkbox"/> Copy of m2
13	<input type="checkbox"/> Copy of m3

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Remove

About

A filter that removes a range of attributes from the dataset.

More

Capabilities

attributeIndices

1-13

debug

False

doNotCheckCapabilities

False

invertSelection

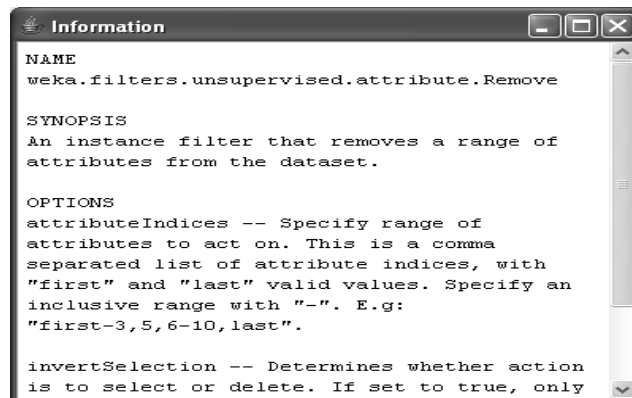
False

Open...

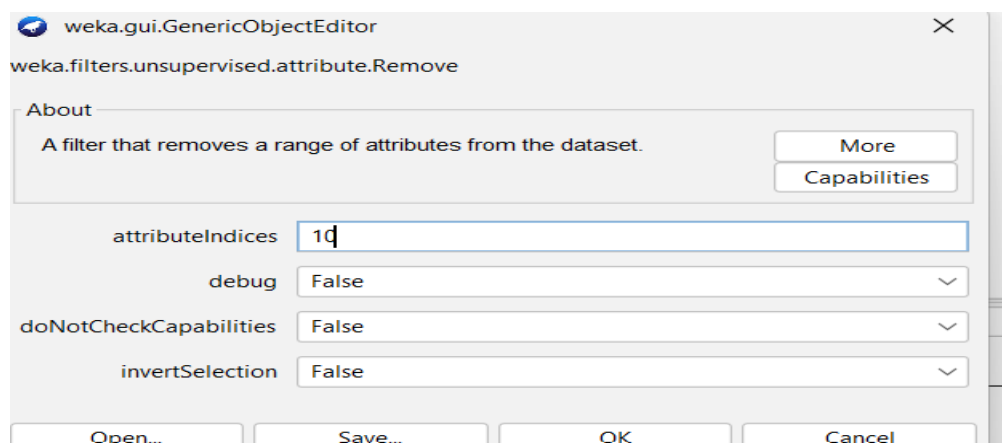
Save...

OK

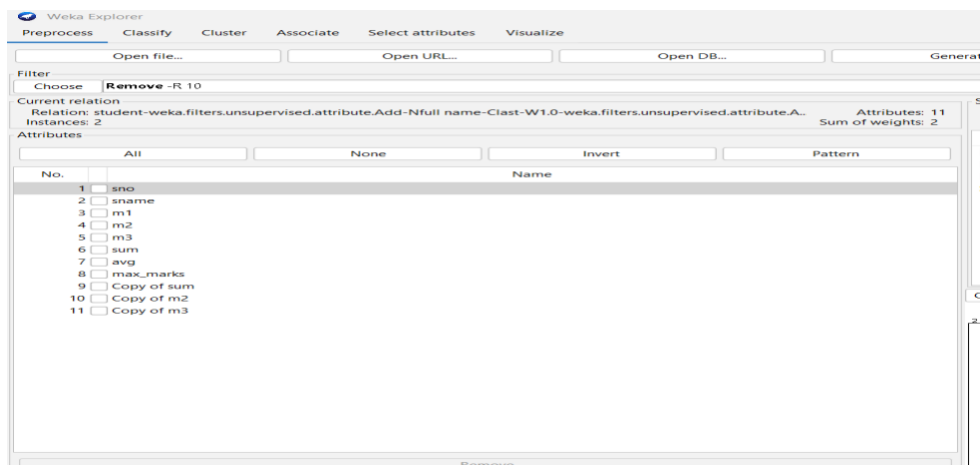
Cancel



If the index position is first.

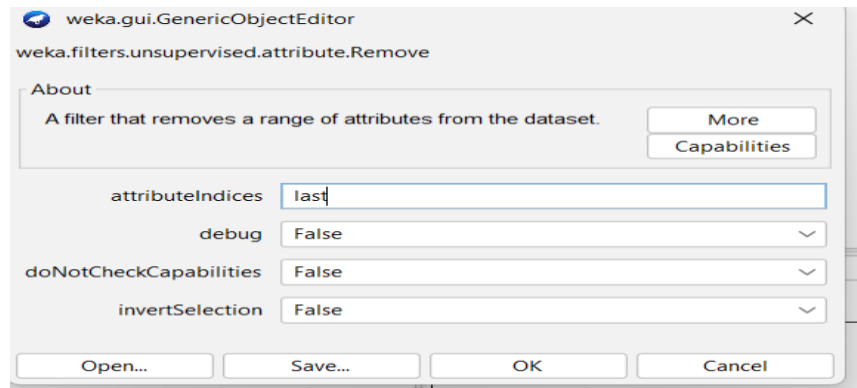


Click on ok button.

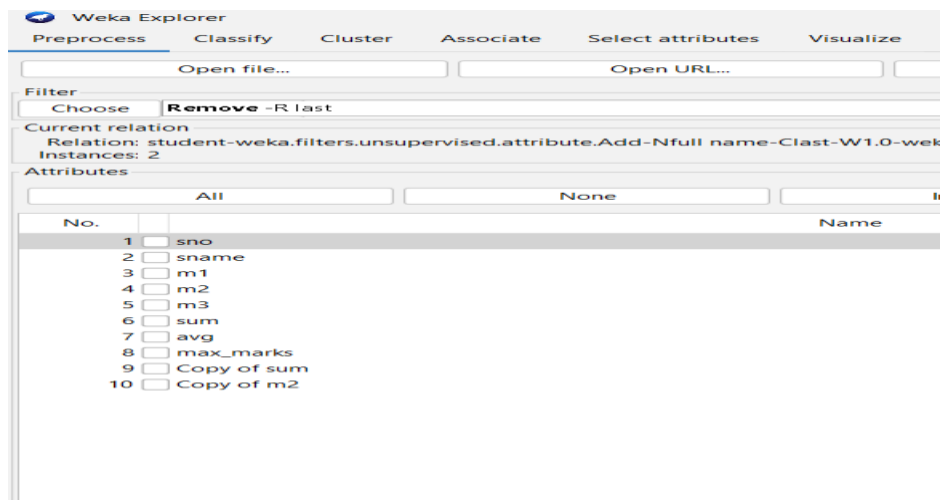


Click on Apply option.

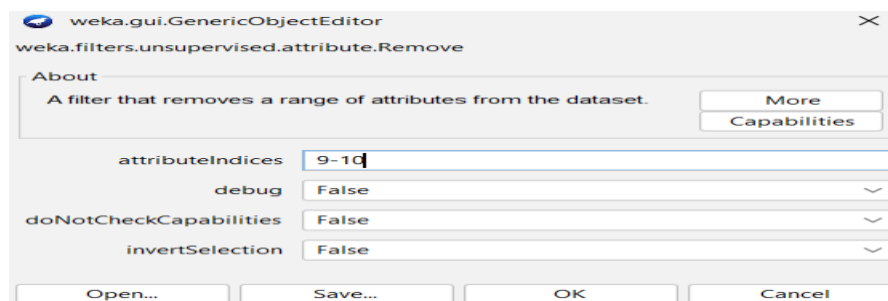
If the index position is last.



Click on ok button.



If the range .



Click on ok button.

Weka Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   Visualize

Open file...   Open URL...   Open DB...

Filter

Choose   Remove -R 9-10

Current relation

Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.Ad...

Instances: 2

Attributes

All   None   Invert   Pa

No.		Name
1	<input type="checkbox"/>	sno
2	<input type="checkbox"/>	sname
3	<input type="checkbox"/>	m1
4	<input type="checkbox"/>	m2
5	<input type="checkbox"/>	m3
6	<input type="checkbox"/>	sum
7	<input type="checkbox"/>	avg
8	<input type="checkbox"/>	max_marks

Click on Apply option.

Viewer

Relation: student-weka.filters.unsupervised.attribute.Add-Nfull name-Clast-W1.0-weka.filters.unsupervised.attribute.Add-Navg-Clast-V

No.	1: sno	2: sname	3: m1	4: m2	5: m3	6: sum	7: avg	8: max_marks
	Numeric	String	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	1.0	a	20.0	30.0	40.0	90.0	30.0	40.0
2	2.0	b	30.0	40.0	10.0	80.0	26.666...	40.0

Click on ok button.



## FILE FORMATES FOR WEKA

### 1. Create CSV(Comma Separated Values) file.

**Step1:** Create an excel file and save with specified format as CSV(Comma Delimited).

**Step2:** Now open with notepad and check the values. Here, the fields of data in each row are delimited with a comma and individual rows are separated by new line.

### 2. Create arff(Attribute Relation File Format) file.

**Step1:** Open a notepad and type the data as instructed below:

ARFF files have two distinct sections. The first section is the **Header** information, which is followed the **Data** information.

The **Header** of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types. An example header on the standard IRIS dataset looks like this:

```
% 1. Title: Iris Plants Database
%
% 2. Sources:
%   (a) Creator: Balu
%   (b) Donor: Shiva
%   (c) Date: Feb 2023
%
@RELATION iris

@ATTRIBUTE   sepallength      NUMERIC
@ATTRIBUTE   sepalwidth      NUMERIC
@ATTRIBUTE   petallength     NUMERIC
@ATTRIBUTE   petalwidth      NUMERIC
@ATTRIBUTE   class           {Iris-setosa,Iris-versicolor,Iris-virginica}
```

The **Data** of the ARFF file looks like the following:

```
@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
```

5.4,3.9,1.7,0.4,Iris-setosa  
4.6,3.4,1.4,0.3,Iris-setosa  
5.0,3.4,1.5,0.2,Iris-setosa  
4.4,2.9,1.4,0.2,Iris-setosa  
4.9,3.1,1.5,0.1,Iris-setosa

Lines that begin with a % are comments. The **@RELATION**, **@ATTRIBUTE** and **@DATA** declarations are case insensitive.

**Step 2:** Save the file as .arff.

**Step 3:** Open with Weka Explorer and check the file values.

### 3. Convert CSV to ARFF file format.

Step1: Open CSV file with notepad

Step 2: To fill header and data section in CSV file.

Step 3: save the file type as arff.

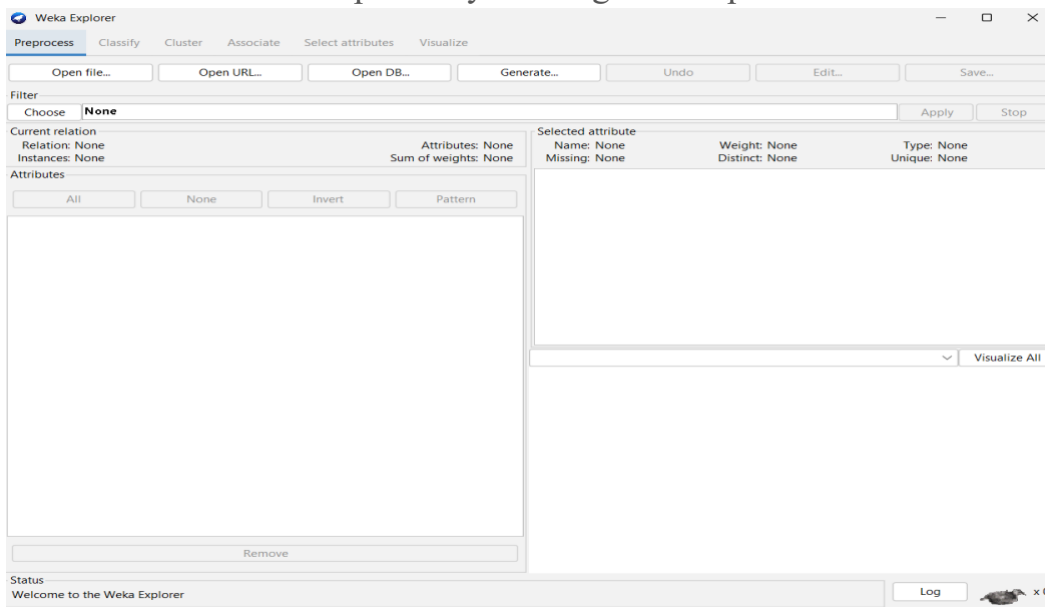
Step 4: Open with Weka Explorer.

Load CSV Files in the Weka Explorer

1.Start the weka chooser



2. Launch the Weka Explorer by clicking the “Explorer” button.



3. Click the “Open File.....” button

4. Navigate to your current working directory. Change the “Files of Type” to “CSV data files (\*.csv)”. Select your file and click the “Open” button.

## Data PreProcessing

### Data Cleaning

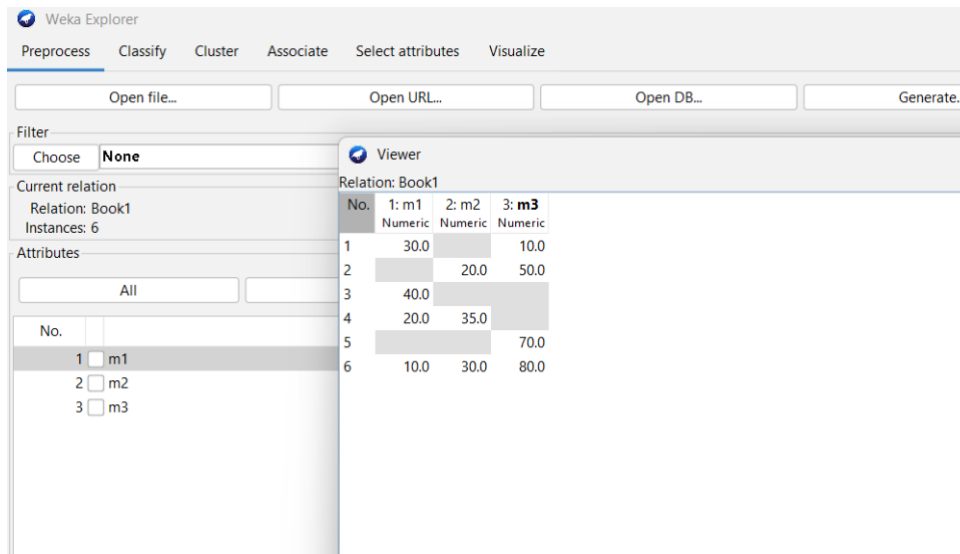
#### 1. Missing-values

##### 1. Remove with

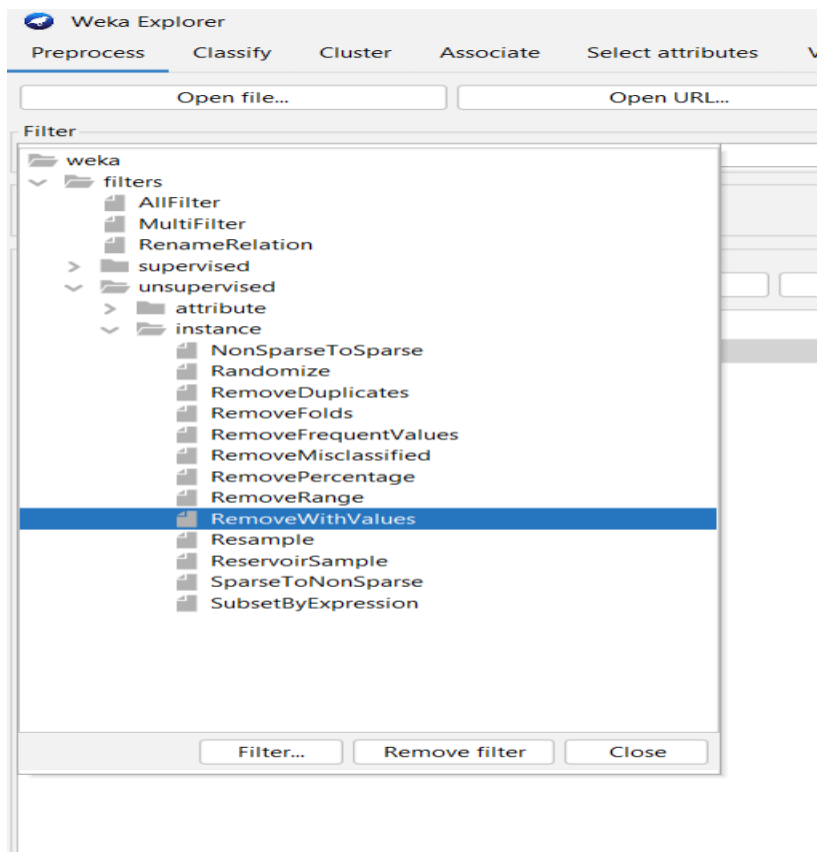
Value 1. Start Weka and choose Explorer



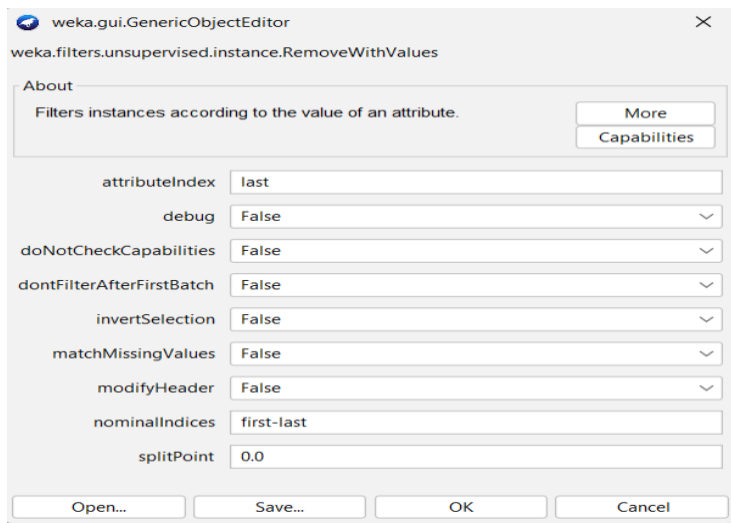
## 2. Load missing.csv data set.



## 3. Click “Choose” in tab Preproces -> Filter-> unsupervised->instance->RemovewithValues which you will find in.

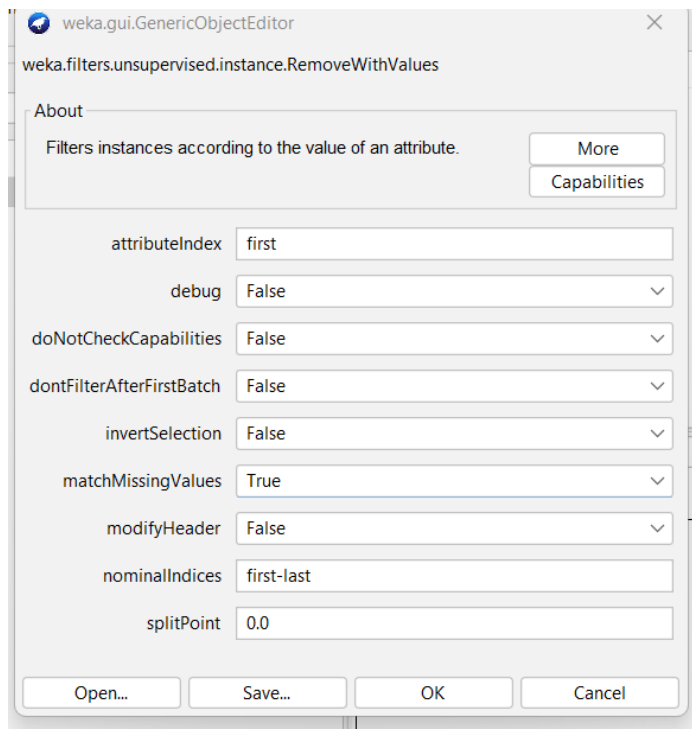


## 4.click on filter tag to configure:



5.Set attributeIndex to first


6.Set matchMissingValues to True



7.Press "OK", in order to use filter configuration.

8.Press "Apply", in order to use selected filter.


Before applying result

 Viewer

Relation: Book1

No.	1: m1 Numeric	2: m2 Numeric	3: m3 Numeric
1	30.0		10.0
2		20.0	50.0
3	40.0		
4	20.0	35.0	
5			70.0
6	10.0	30.0	80.0

After Applying result

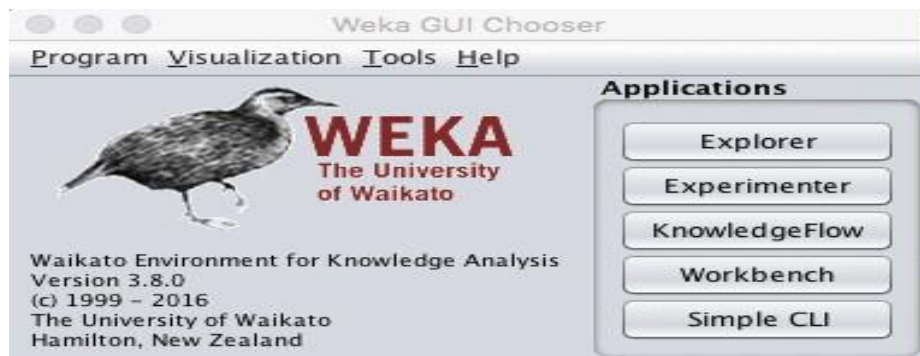
 Viewer

Relation: Book1-weka.filters.unsupervised.instance.Re

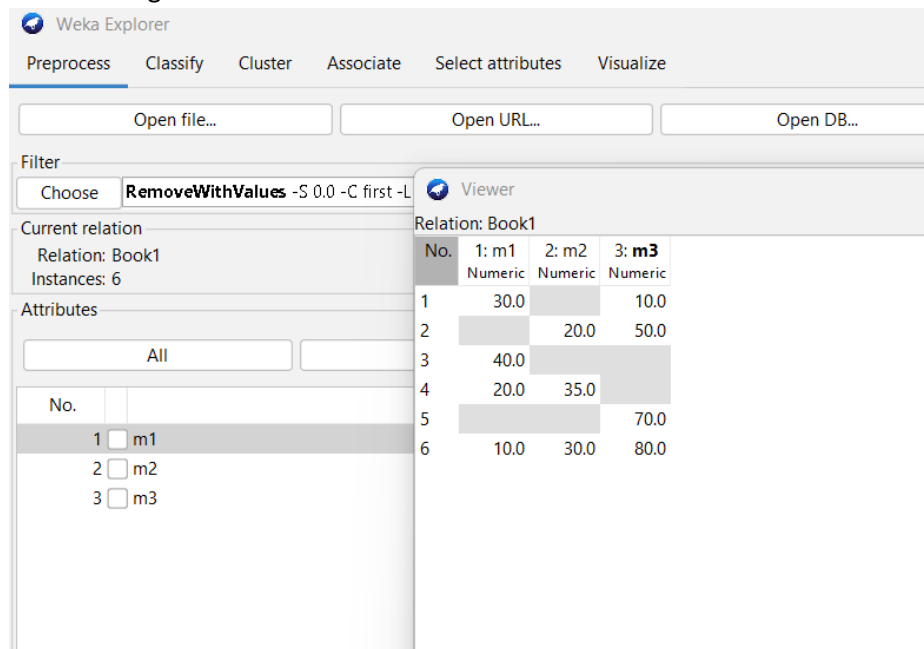
No.	1: m1 Numeric	2: m2 Numeric	3: m3 Numeric
1	30.0		10.0
2	40.0		
3	20.0	35.0	
4	10.0	30.0	80.0

## 2.Replacing Missing with userConstant

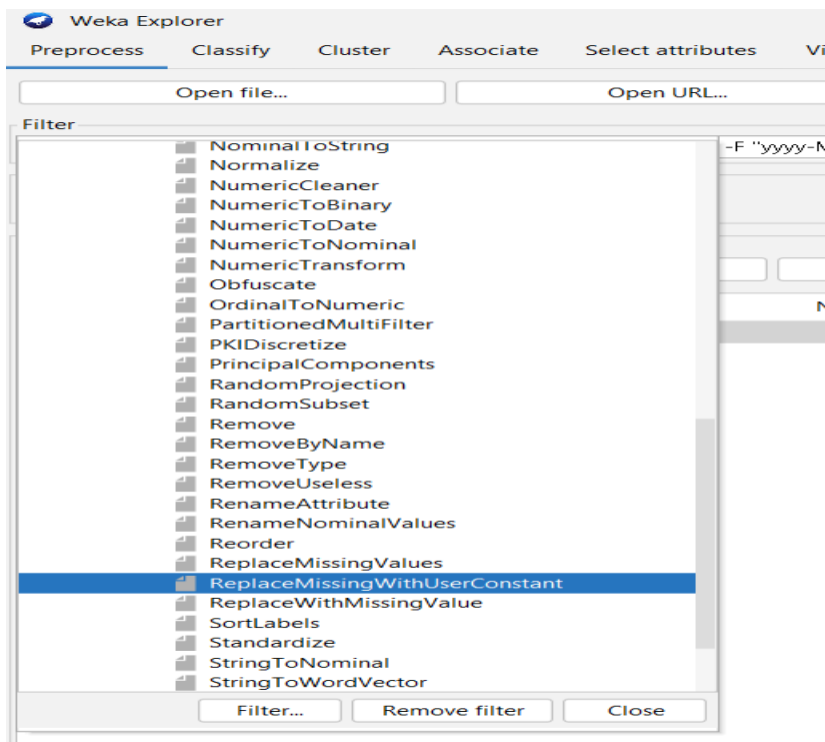
1. Start Weka and choose Explorer



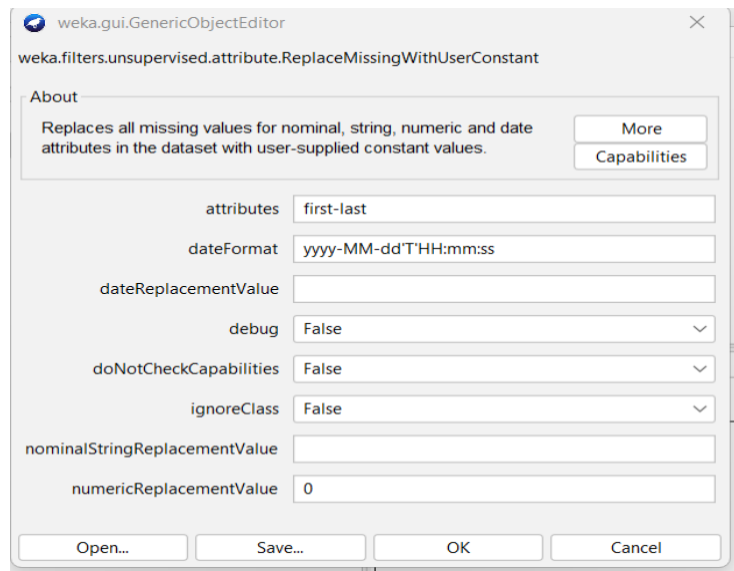
2. Load missing.csv data set.



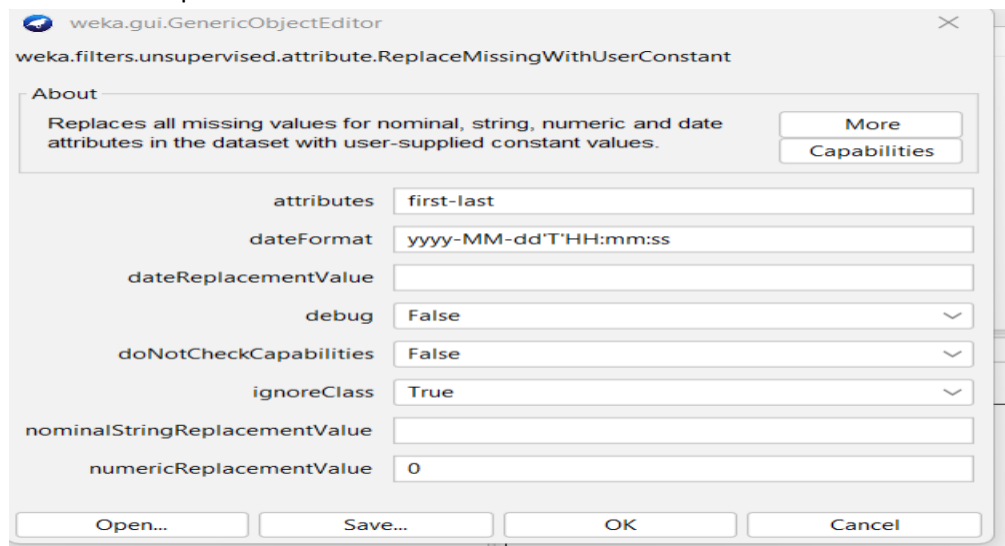
3. Click "Choose" in tab Preprocess -> Filter-> unsupervised->attribute->ReplaceMissingWithUserConstant which you will find in.



- Click on filter tab to configure it



- Set attributeIndex to first
- Set ignoreClass to True
- set numericReplacementValue to 0



- Press "OK", in order to use filter configuration.



9. Press "Apply", in order to use selected filter.

Before applying result

Viewer

Relation: Book1

No.	1: m1 Numeric	2: m2 Numeric	3: m3 Numeric
1	30.0		10.0
2		20.0	50.0
3	40.0		
4	20.0	35.0	
5			70.0
6	10.0	30.0	80.0

After Applying result

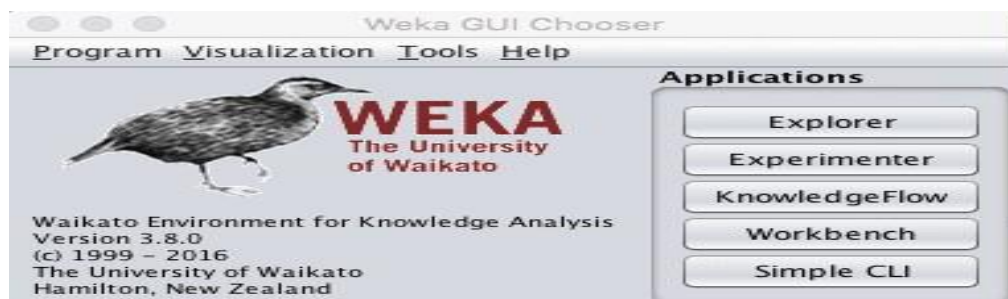
Viewer

Relation: Book1-weka.filters.unsupervised.attribute.Replace

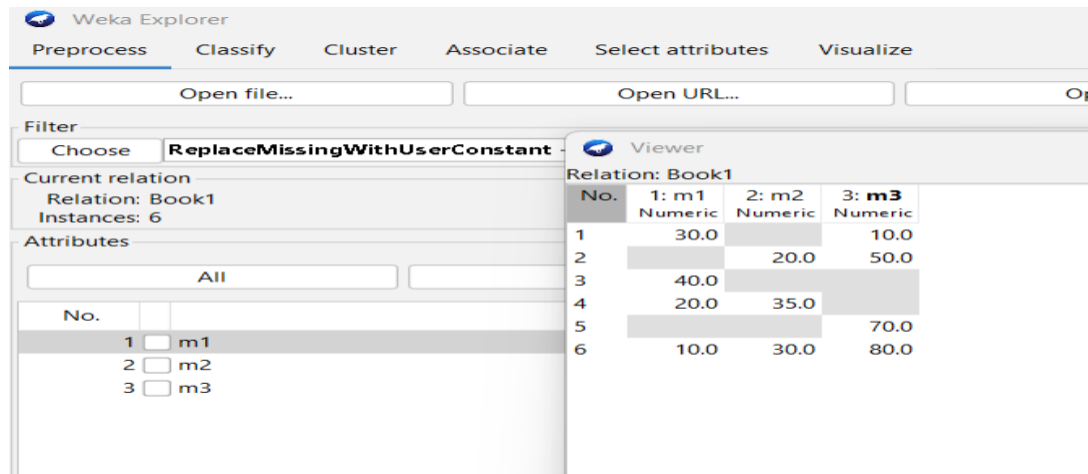
No.	1: m1 Numeric	2: m2 Numeric	3: m3 Numeric
1	30.0	0.0	10.0
2	0.0	20.0	50.0
3	40.0	0.0	
4	20.0	35.0	
5	0.0	0.0	70.0
6	10.0	30.0	80.0

### 3. ReplaceMissingValues

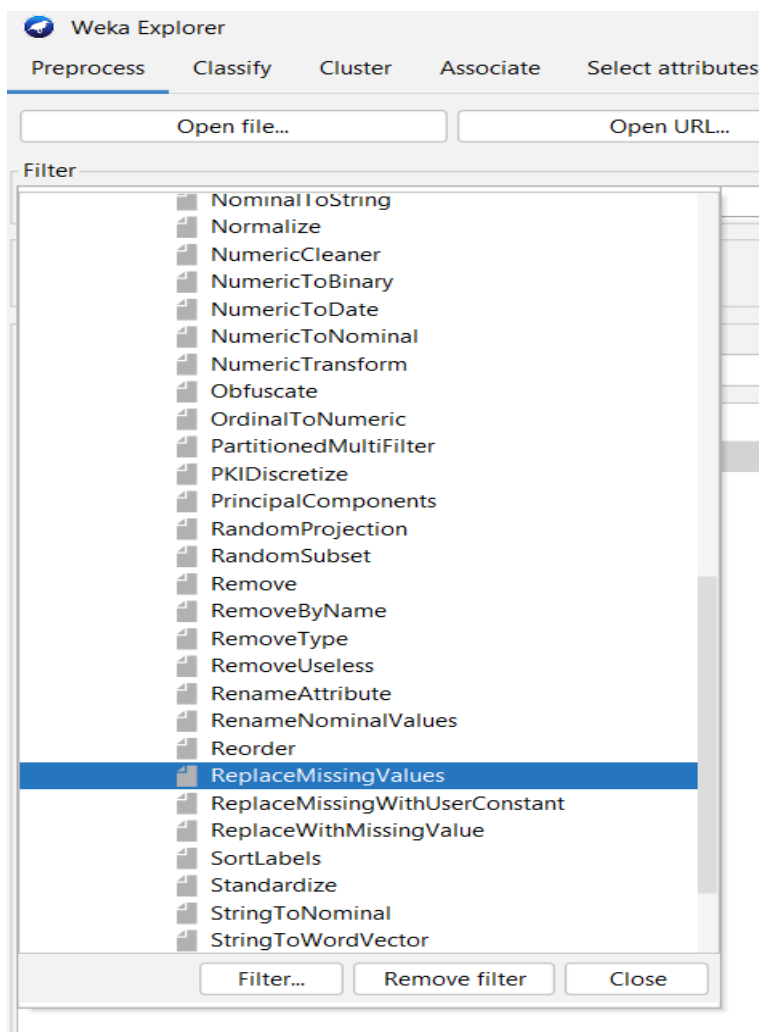
Start Weka and choose Explorer



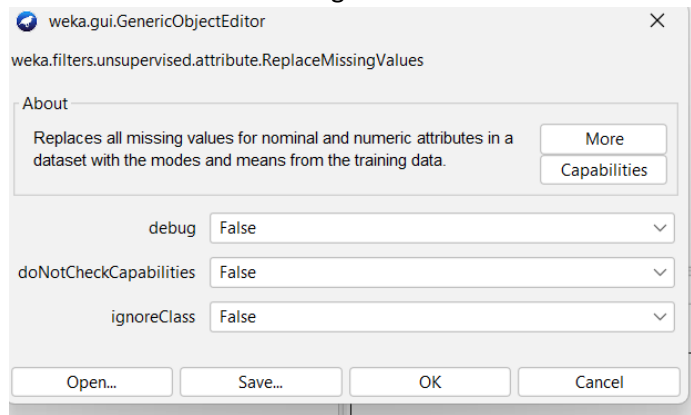
1. Load missing.csv data set.



- Click "Choose" in tab Preprocess -> Filter-> unsupervised->attribute->ReplaceMissingWithUserConstant which you will find in.



### 3. Click on filter tab to configure it



3. Set ignoreClass to True
4. Press "OK", in order to use filter configuration.
5. Press "Apply", in order to use selected filter.

### Before applying result

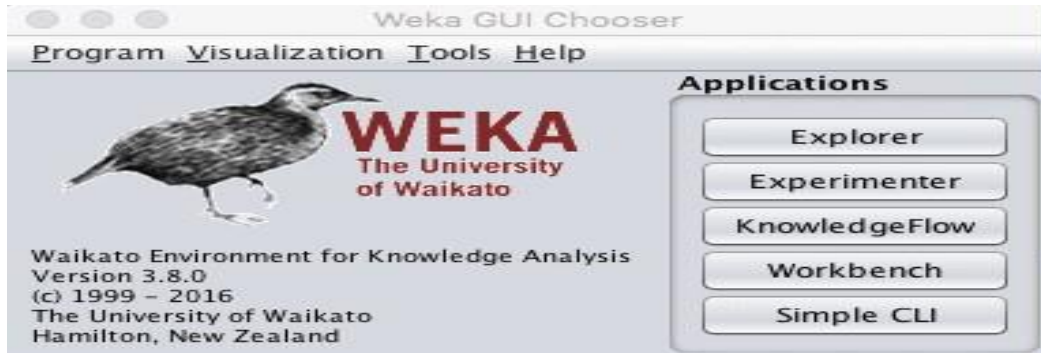
No.	1: m1 Numeric	2: m2 Numeric	3: m3 Numeric
1	30.0		10.0
2		20.0	50.0
3	40.0		
4	20.0	35.0	
5			70.0
6	10.0	30.0	80.0

### After Applying result

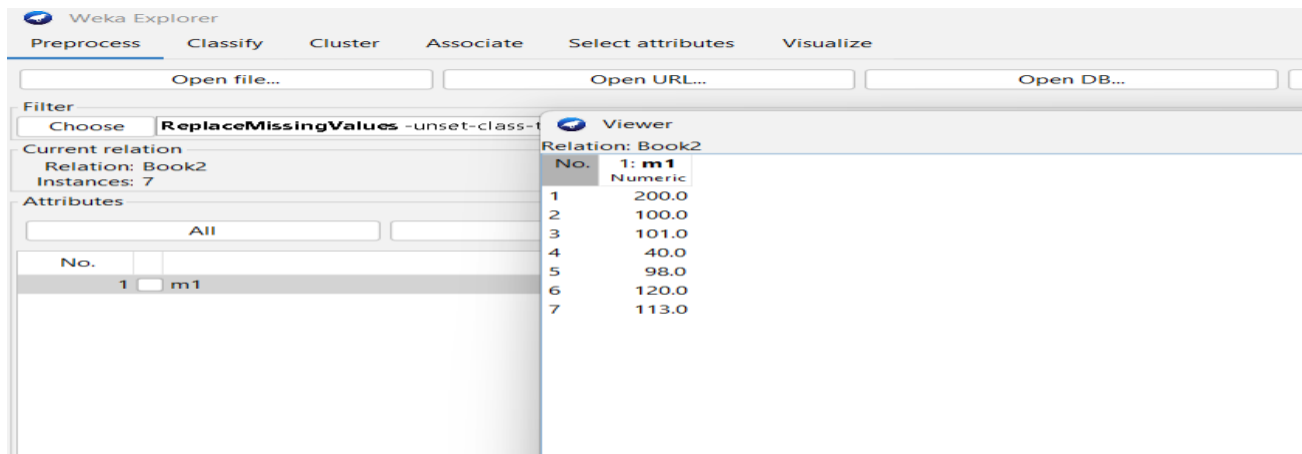
No.	1: m1 Numeric	2: m2 Numeric	3: m3 Numeric
1	30.0	28.333...	10.0
2	25.0	20.0	50.0
3	40.0	28.333...	52.5
4	20.0	35.0	52.5
5	25.0	28.333...	70.0
6	10.0	30.0	80.0

## Noisy Data- outlier / Extreme Values

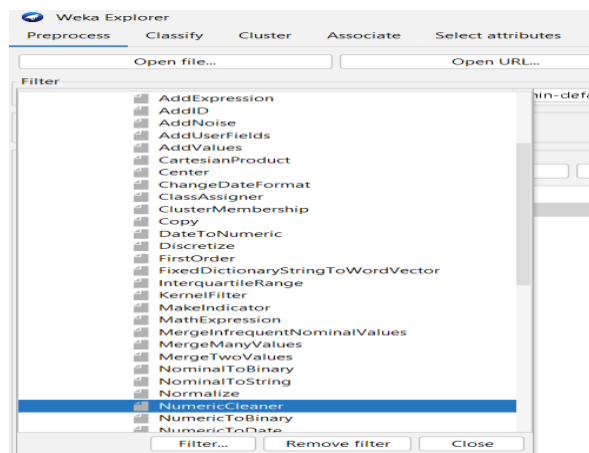
1. Start Weka and choose Explorer



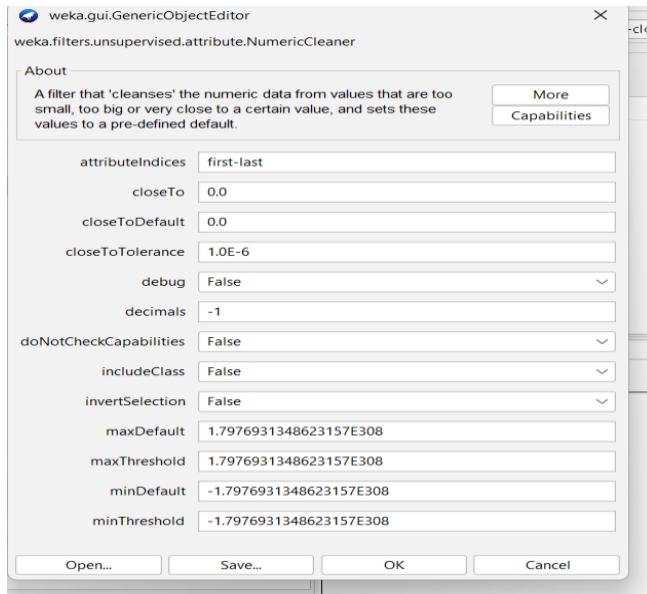
2. Load noisy.csv data set.



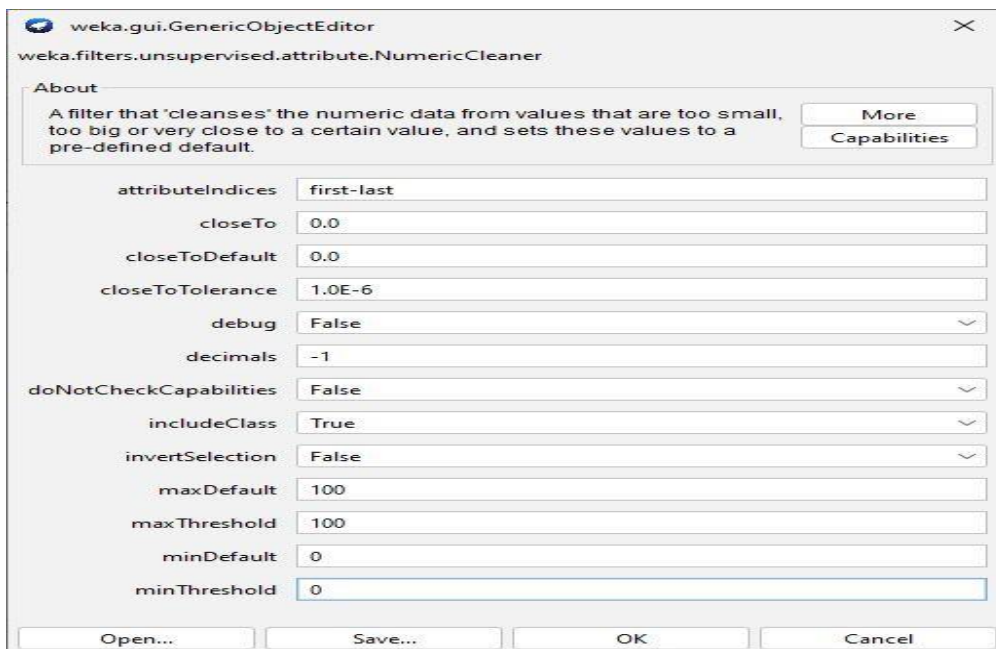
3. Click "Choose" in tab Preproces -> Filter-> unsupervised->attribute->NumericCleaner which you will find in.



- Click on filter tab to configure it



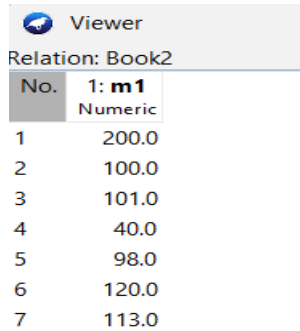
- Set includeClass to True
- Set maxDefault to 100
- Set maxThreshold to 100
- Set minDefault to 0
- Set minThreshold to 0



9. Press "OK", in order to use filter configuration.

10. Press "Apply", in order to use selected filter.

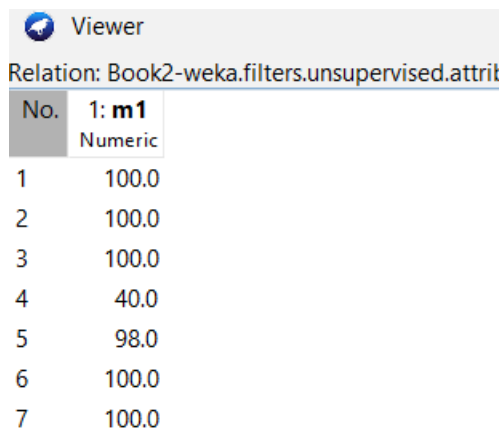
11. Before applying result



Viewer  
Relation: Book2

No.	1: m1 Numeric
1	200.0
2	100.0
3	101.0
4	40.0
5	98.0
6	120.0
7	113.0

After Applying result

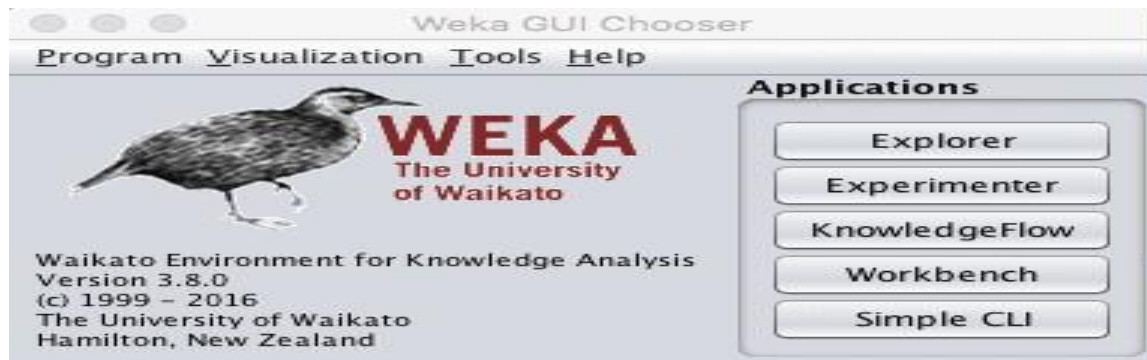


Viewer  
Relation: Book2-weka.filters.unsupervised.attrit

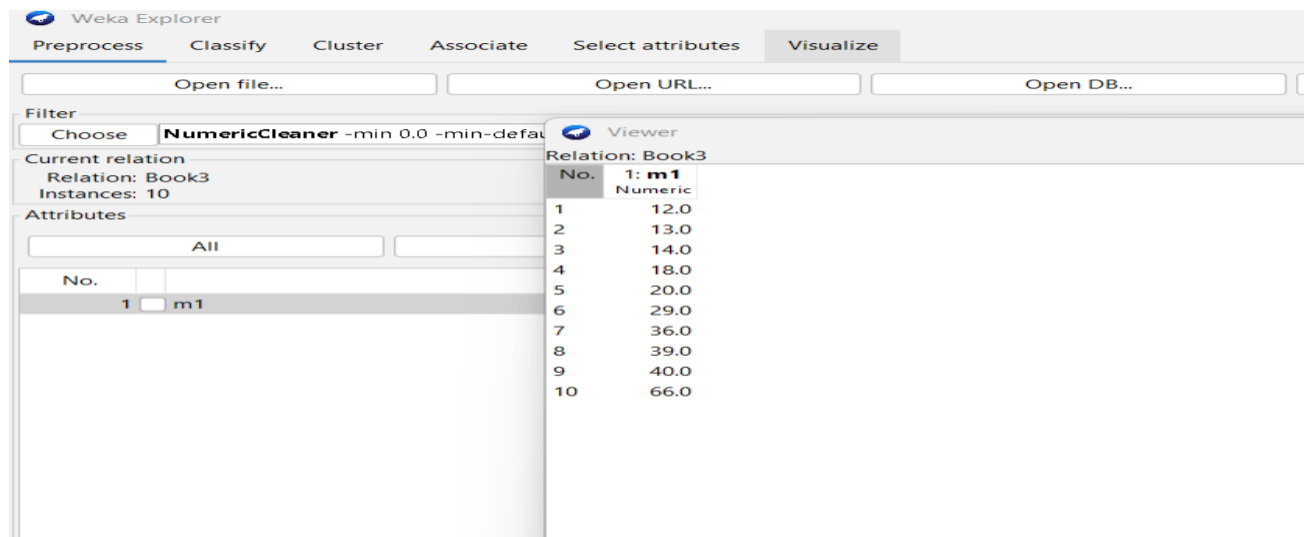
No.	1: m1 Numeric
1	100.0
2	100.0
3	100.0
4	40.0
5	98.0
6	100.0
7	100.0

## Discretization

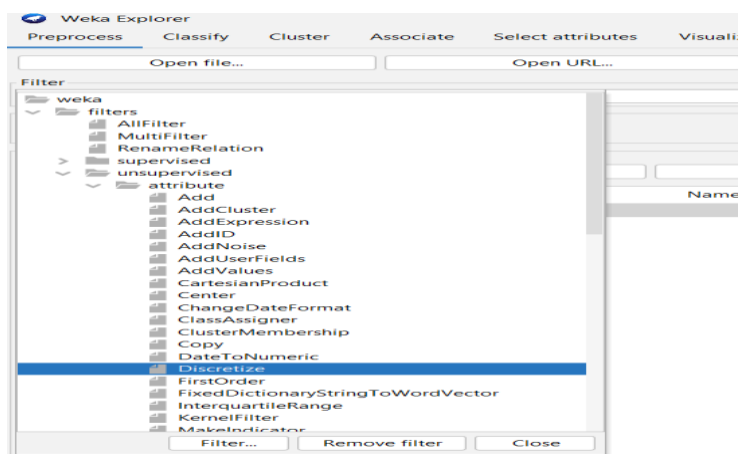
1. Start Weka and choose Explorer



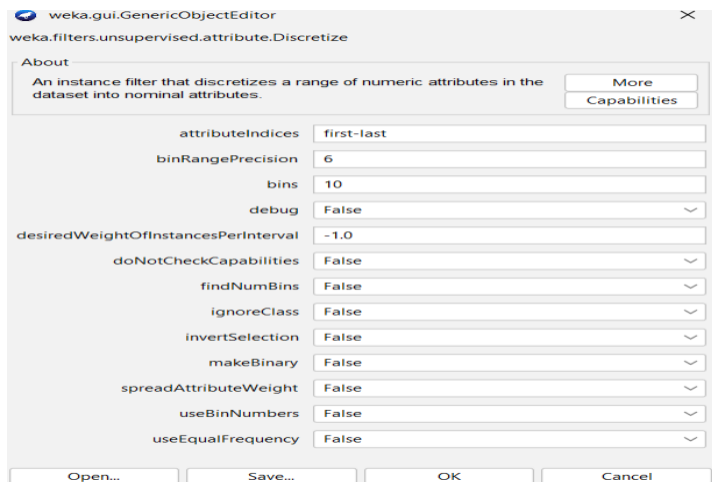
2. Load bin.csv data set.



3. Click "Choose" in tab Preproces -> Filter-> unsupervised->attribute->Discretize which you will find in.



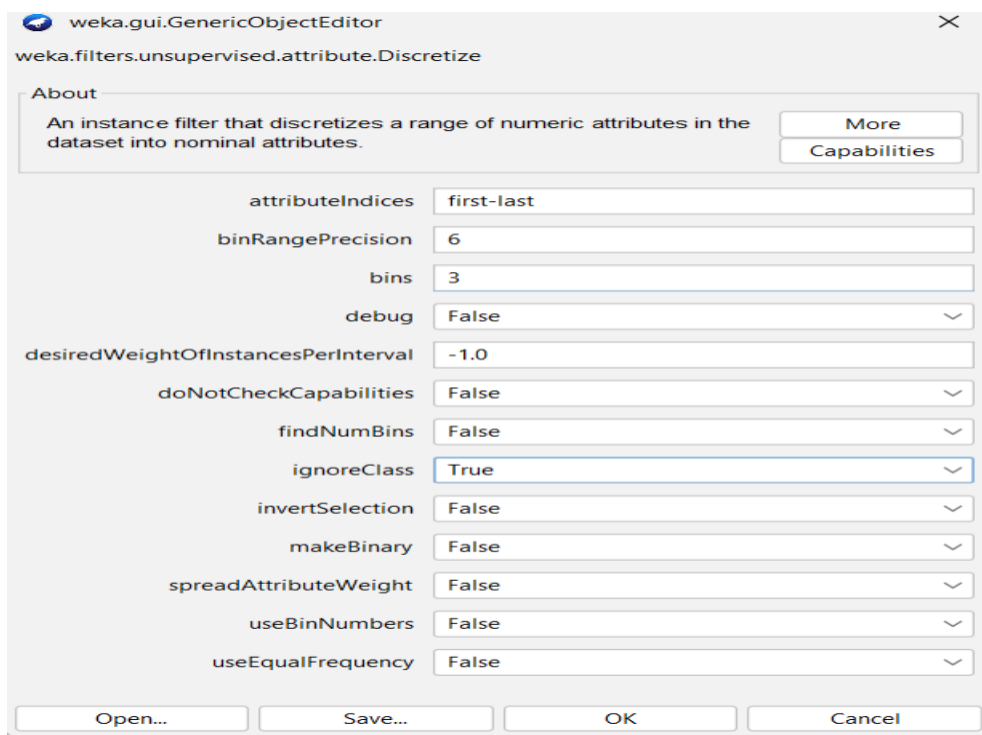
4. Click on filter tab to configure it



5. Set attributeIndices to first-last

4. set bins to 3

5. Set ignoreClass to True

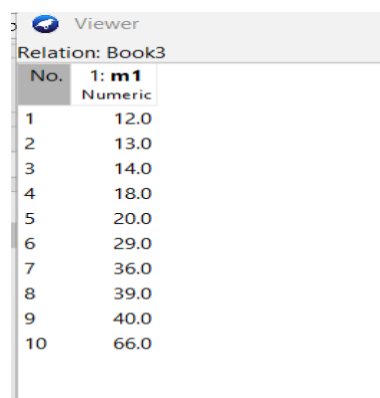




6. Press "OK", in order to use filter configuration.

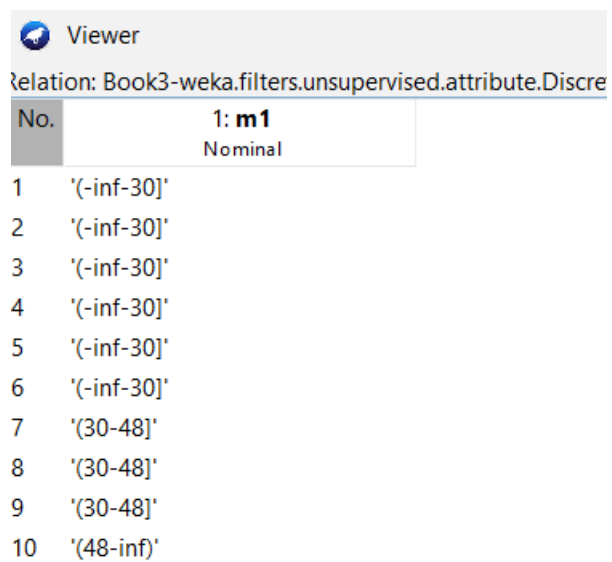
7. Press "Apply", in order to use selected filter.

Before applying result



No.	1: m1 Numeric
1	12.0
2	13.0
3	14.0
4	18.0
5	20.0
6	29.0
7	36.0
8	39.0
9	40.0
10	66.0

After applying result

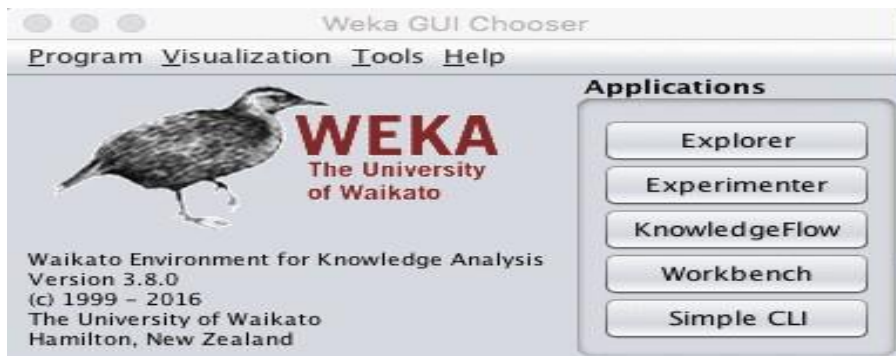


No.	1: m1 Nominal
1	'(-inf-30]'
2	'(-inf-30]'
3	'(-inf-30]'
4	'(-inf-30]'
5	'(-inf-30]'
6	'(-inf-30]'
7	'(30-48]'
8	'(30-48]'
9	'(30-48]'
10	'(48-inf)'

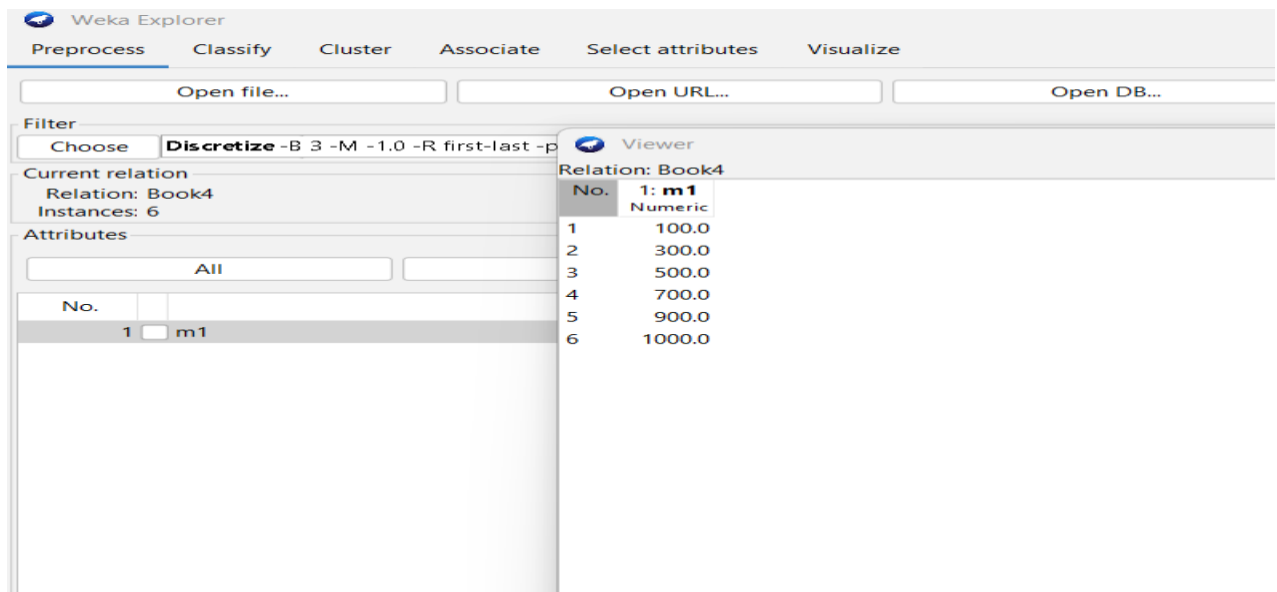
**Normalization**

## Min-Max-Normalization

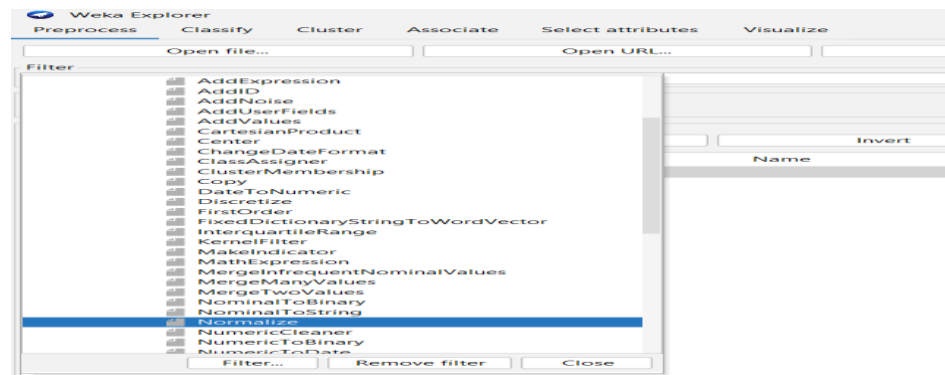
1. Start Weka and choose Explorer



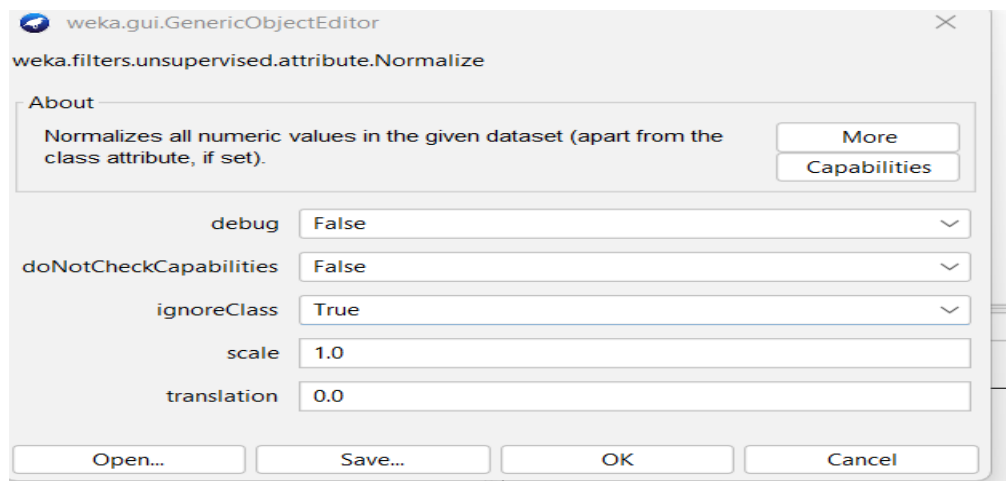
2. Load bin.csv data set.



3. Click “Choose” in tab Preproces -> Filter-> unsupervised->attribute->Normalize which you will find in.



4. Click on filter tab to configure it



5. Set ignoreClass to True

6. Press "OK", in order to use filter configuration.


7. Press "Apply", in order to use selected filter.

Before applying result

The screenshot shows the 'Viewer' window with the title 'relation: Book4'. It displays a table with 6 rows and 2 columns. The first column is labeled 'No.' and the second column is labeled '1: m1 Numeric'. The data values are: 1, 100.0; 2, 300.0; 3, 500.0; 4, 700.0; 5, 900.0; 6, 1000.0.

No.	1: m1 Numeric
1	100.0
2	300.0
3	500.0
4	700.0
5	900.0
6	1000.0

After applying result

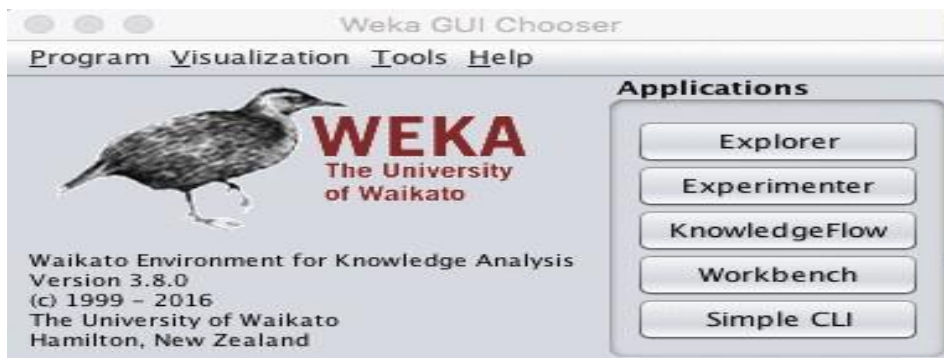
 Viewer

Relation: Book4-weka.filters.unsupervised.a

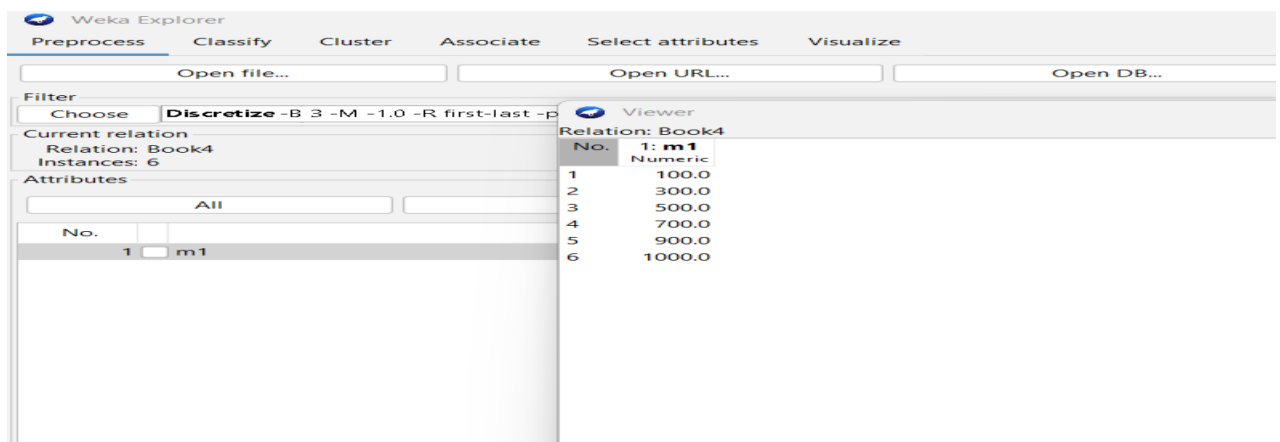
No.	1: m1
	Numeric
1	0.0
2	0.2222...
3	0.4444...
4	0.6666...
5	0.8888...
6	1.0

## Z-Score-Normalization

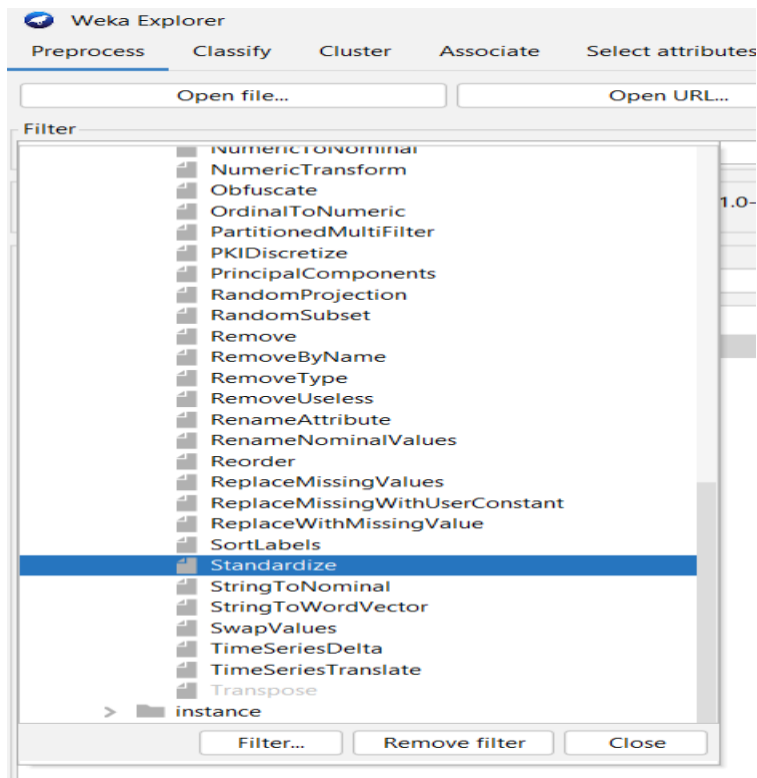
1. Start Weka and choose Explorer



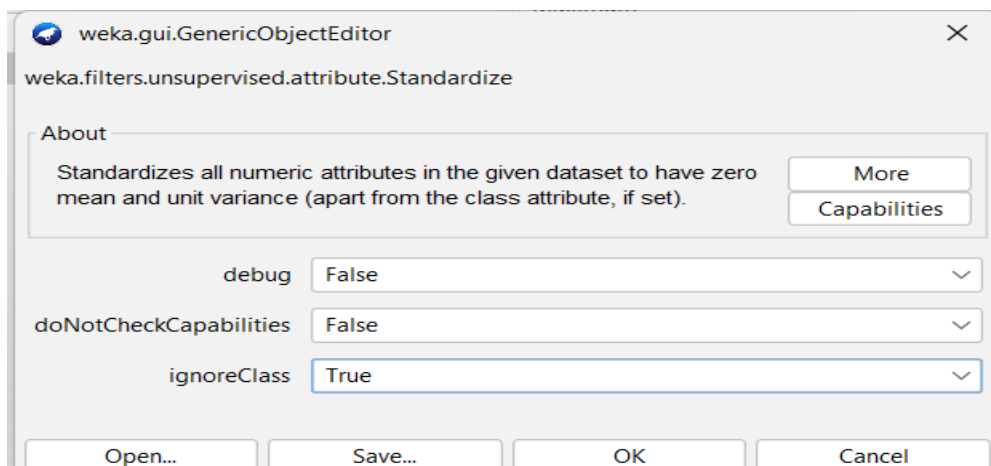
2. Load bin.csv data set.



Click "Choose" in tab Preproces -> Filter-> unsupervised->attribute->Standardize which you will find in.



3. Click on filter tab to configure it



4. Set ignoreClass to True

5. Press "OK", in order to use filter configuration.

6. Press "Apply", in order to use selected filter.

Before applying result

Viewer	
Relation: Book4	
No.	1: m1 Numeric
1	100.0
2	300.0
3	500.0
4	700.0
5	900.0
6	1000.0

After applying result

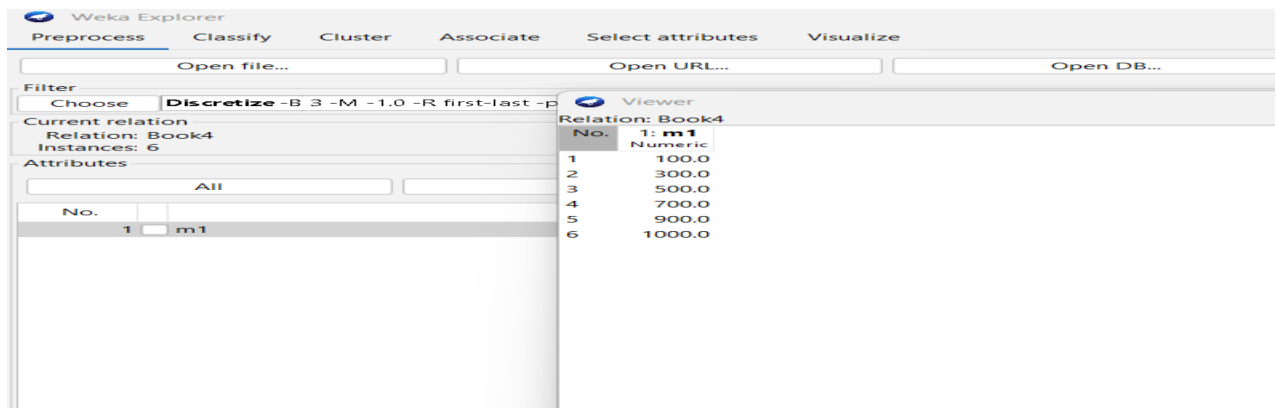
Viewer	
Relation: Book4-weka.filters.unsupervised.attribu	
No.	1: m1 Numeric
1	-1.385...
2	-0.812...
3	-0.238...
4	0.3344...
5	0.9078...
6	1.1945...

## Decimal-Scale-Normalization

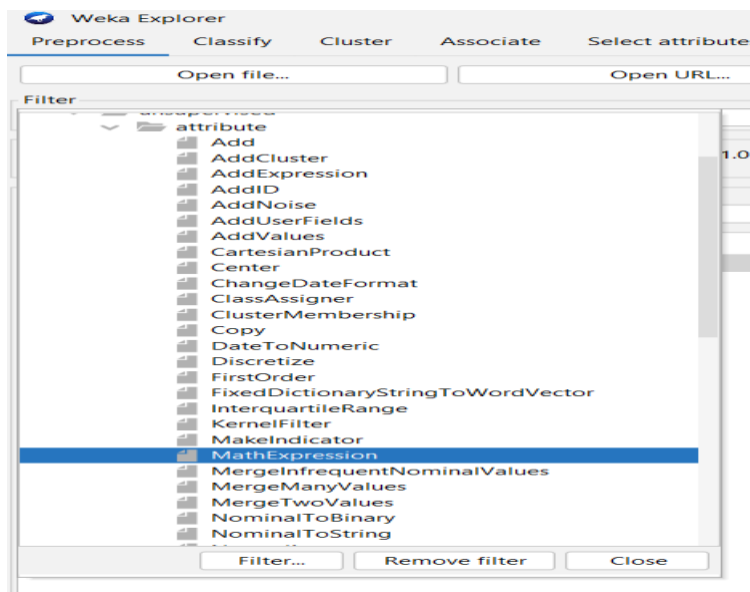
1. Start Weka and choose Explorer



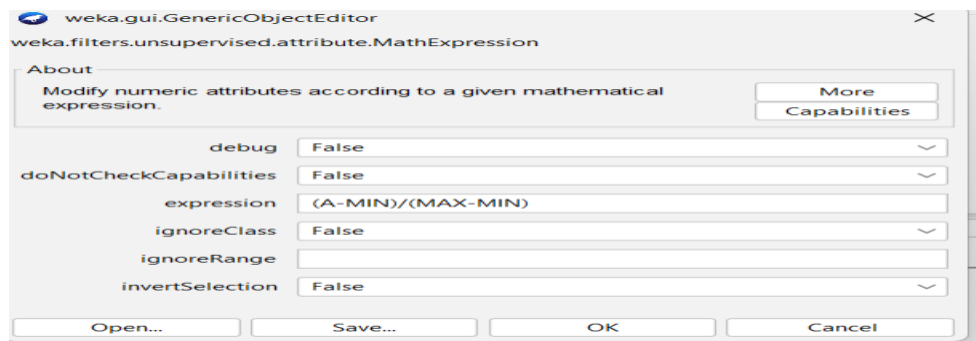
2. Load bin.csv data set.



3. Click "Choose" in tab Preproces -> Filter-> unsupervised->attribute->MathExpression which you will find in.

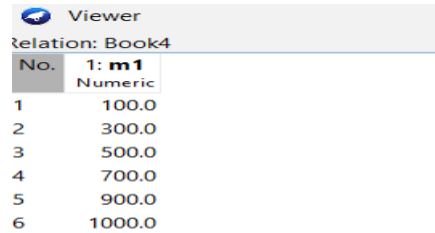


4. Click on filter tab to configure it to specify expression



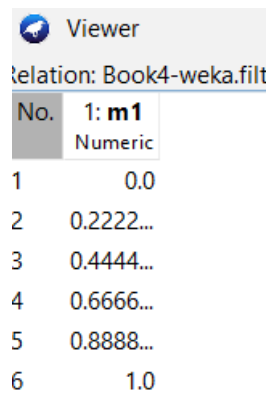
5. Set ignoreClass to True
6. Press "OK", in order to use filter configuration.
7. Press "Apply", in order to use selected filter.

8. Before applying result



No.	1: m1 Numeric
1	100.0
2	300.0
3	500.0
4	700.0
5	900.0
6	1000.0

After applying result



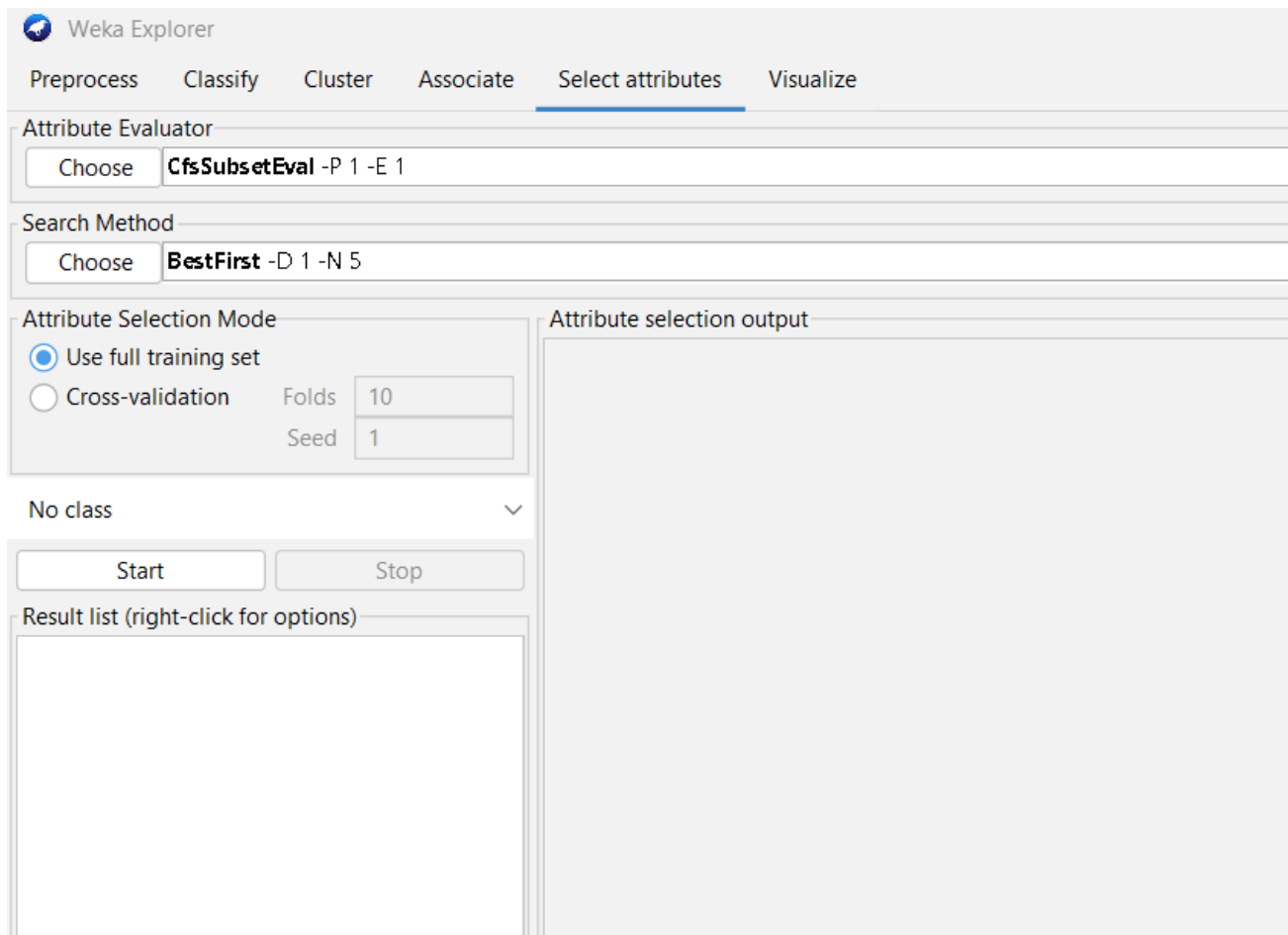
No.	1: m1 Numeric
1	0.0
2	0.2222...
3	0.4444...
4	0.6666...
5	0.8888...
6	1.0

## Feature Selection-Measure

### Correlation Based Feature Selection

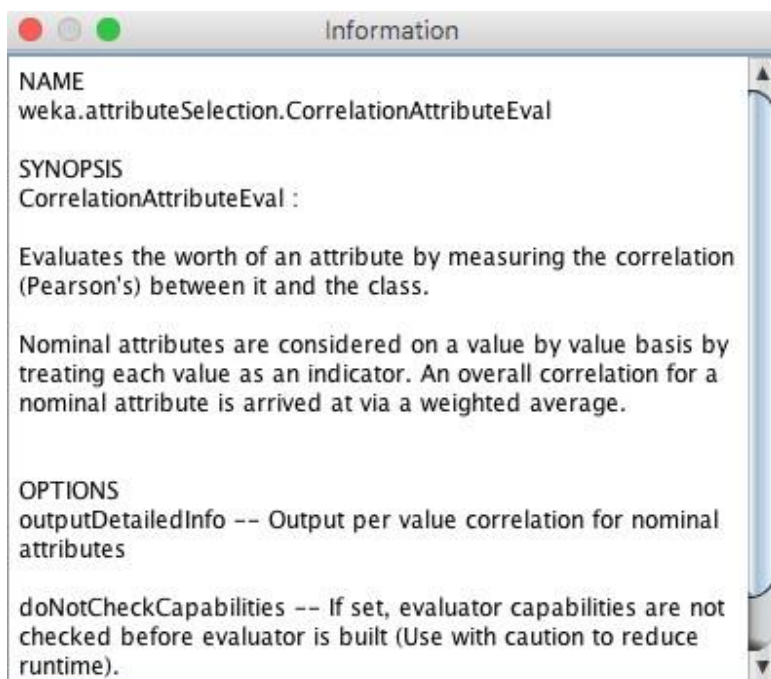
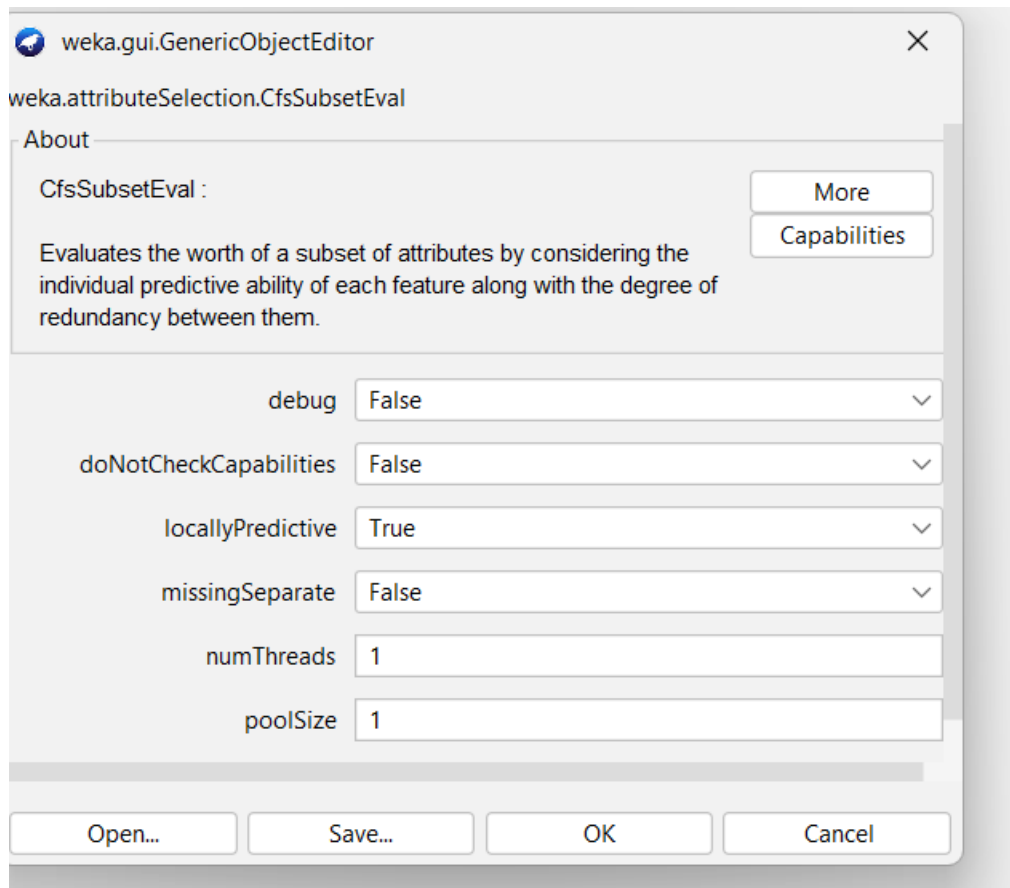
1. Open the Weka GUI Chooser.
2. Click the "Explorer" button to launch the Explorer.
3. Open the Pima Indians dataset.
4. Click the "Select attributes" tab to access the feature selection methods.

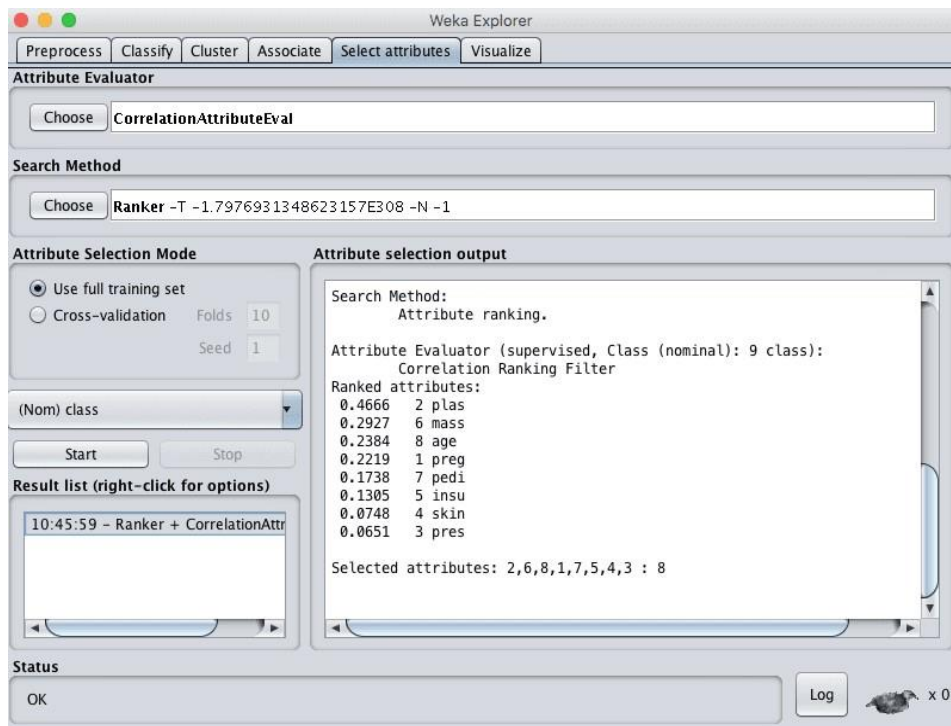




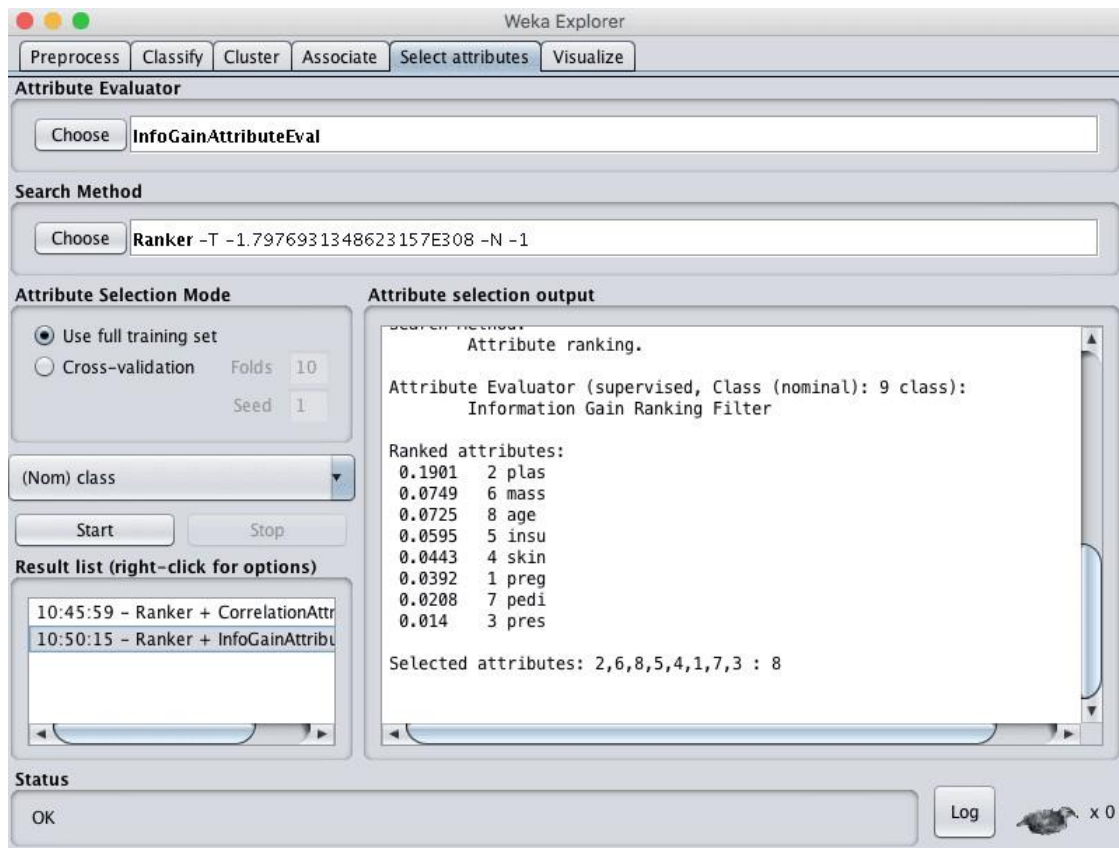
Feature selection is divided into two parts:

- Attribute
- Evaluator □
- Search Method.





## Information Gain Based Feature Selection



## Association:

1)open the weka tool

2)create a csv file

	A	B	C	D	E	F	G
1	Transacti onID	Milk	Bread	Eggs	Butter	Chips	
2	T1	yes	yes			yes	
3	T2		yes	yes	yes		
4	T3	yes			yes	yes	
5	T4	yes	yes	yes			
6	T5		no	yes		yes	
7							
8							
9							

3)open the csv file in weka tool

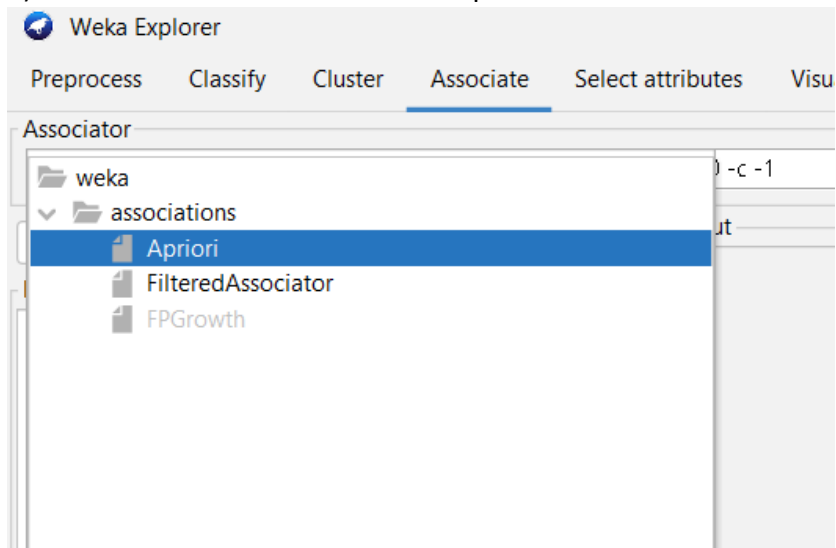
The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Current relation' is 'super market' with 5 instances. The 'Attributes' list includes TransactionID, Milk, Bread, Eggs, Butter, and Chips. A 'Viewer' window is open, displaying the data for the 'super market' relation. The data is as follows:

No.	1: TransactionID	2: Milk	3: Bread	4: Eggs	5: Butter	6: Chips
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	T1	yes	yes			yes
2	T2		yes	yes	yes	
3	T3	yes			yes	yes
4	T4	yes	yes	yes		
5	T5		no	yes		yes

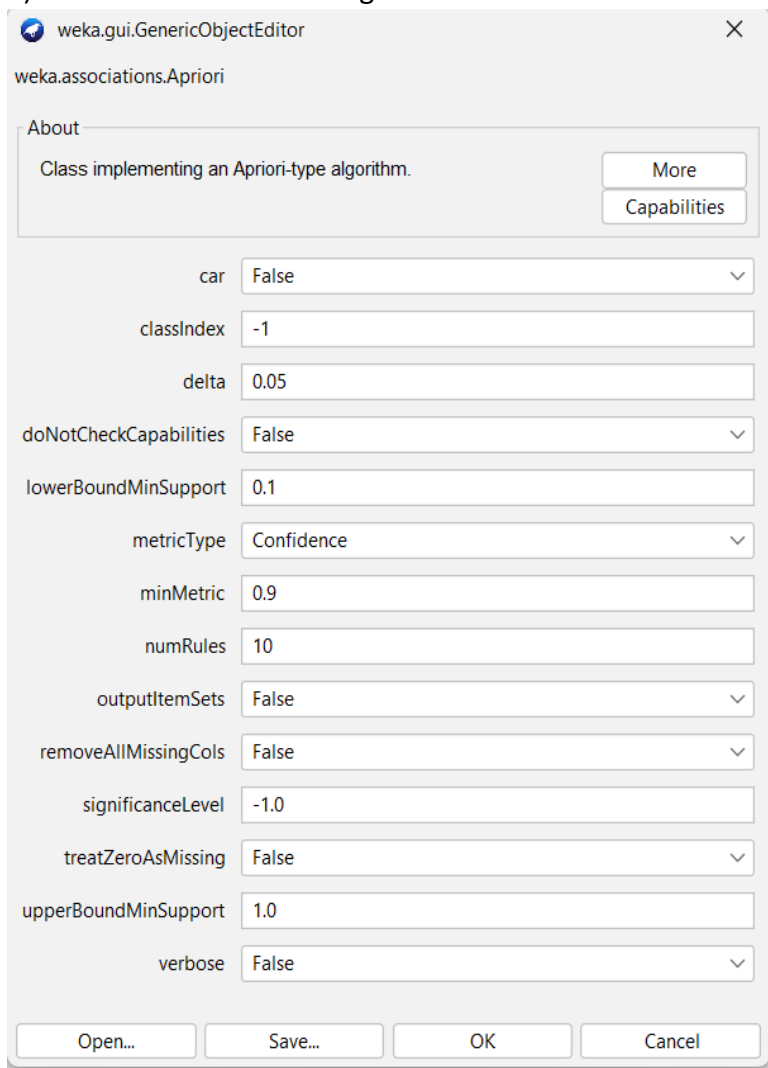
4)select the Associate option

The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Associate' window is open, showing the 'Associate output' tab. The output is empty, indicating that the association analysis has not yet been performed.

5) Click "choose" in tab associate->Apriori



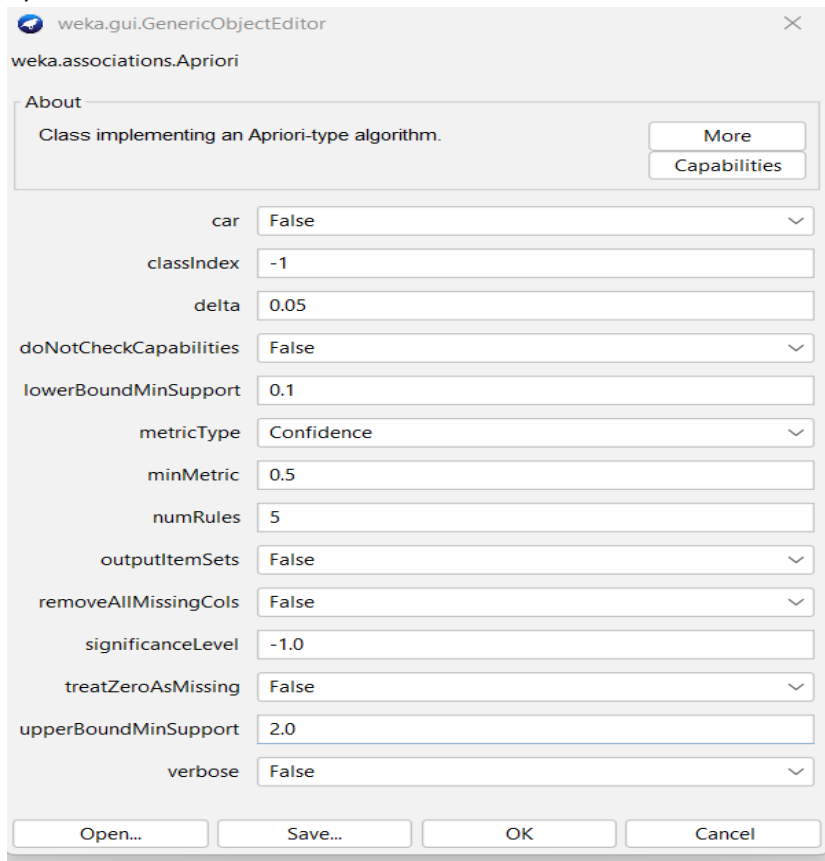
6) Click on filter tab to configure it



7)set upper bound min support 2.0

8)set num rules 5

9)set min metric 0.5



10)Press "ok",in order to use filter configuration.

11)press "start",In order to use selected filter

Associate output:

#### Associator output

=== Run information ===

Scheme: weka.associations.Apriori -N 5 -T 0 -C 0.5 -D 0.05 -U 2.0 -M 0.1 -S -1.0 -c -1  
Relation: super market  
Instances: 5  
Attributes: 6  
TransactionID  
Milk  
Bread  
Eggs  
Butter  
Chips

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.5 (2 instances)  
Minimum metric <confidence>: 0.5  
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 3

Best rules found:

1. Bread=yes 3 ==> Milk=yes 2 <conf:(0.67)> lift:(1.11) lev:(0.04) [0] conv:(0.6)
2. Milk=yes 3 ==> Bread=yes 2 <conf:(0.67)> lift:(1.11) lev:(0.04) [0] conv:(0.6)
3. Chips=yes 3 ==> Milk=yes 2 <conf:(0.67)> lift:(1.11) lev:(0.04) [0] conv:(0.6)
4. Milk=yes 3 ==> Chips=yes 2 <conf:(0.67)> lift:(1.11) lev:(0.04) [0] conv:(0.6)
5. Eggs=yes 3 ==> Bread=yes 2 <conf:(0.67)> lift:(1.11) lev:(0.04) [0] conv:(0.6)

Set numrules 10

Output:

#### Associator output

```
=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.5 -D 0.05 -U 2.0 -M 0.1 -S -1.0 -c -1
Relation:     super market
Instances:    5
Attributes:   6
              TransactionID
              Milk
              Bread
              Eggs
              Butter
              Chips

=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.3 (1 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 27

Size of set of large itemsets L(3): 20

Size of set of large itemsets L(4): 5

Best rules found:

1. TransactionID=T1 1 ==> Milk=yes 1    <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
2. TransactionID=T1 1 ==> Bread=yes 1    <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
3. TransactionID=T1 1 ==> Chips=yes 1    <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
4. TransactionID=T2 1 ==> Bread=yes 1    <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
5. TransactionID=T2 1 ==> Eggs=yes 1     <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
6. TransactionID=T2 1 ==> Butter=yes 1   <conf:(1)> lift:(2.5) lev:(0.12) [0] conv:(0.6)
7. TransactionID=T3 1 ==> Milk=yes 1     <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
8. TransactionID=T3 1 ==> Butter=yes 1   <conf:(1)> lift:(2.5) lev:(0.12) [0] conv:(0.6)
9. TransactionID=T3 1 ==> Chips=yes 1    <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
10. TransactionID=T4 1 ==> Milk=yes 1    <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
```

#### Output:

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.5 -D 0.05 -U 2.0 -M 0.1 -S -1.0 -c  
-1  
Relation: super market  
Instances: 5  
Attributes: 6  
TransactionID  
Milk  
Bread



Eggs  
Butter  
Chips

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.3 (1 instances)

Minimum metric <confidence>: 0.5

Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 27

Size of set of large itemsets L(3): 20

Size of set of large itemsets L(4): 5

Best rules found:

1. TransactionID=T1 1 ==> Milk=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
2. TransactionID=T1 1 ==> Bread=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
3. TransactionID=T1 1 ==> Chips=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
4. TransactionID=T2 1 ==> Bread=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
5. TransactionID=T2 1 ==> Eggs=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
6. TransactionID=T2 1 ==> Butter=yes 1 <conf:(1)> lift:(2.5) lev:(0.12) [0] conv:(0.6)
7. TransactionID=T3 1 ==> Milk=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
8. TransactionID=T3 1 ==> Butter=yes 1 <conf:(1)> lift:(2.5) lev:(0.12) [0] conv:(0.6)
9. TransactionID=T3 1 ==> Chips=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] conv:(0.4)
10. TransactionID=T4 1 ==> Milk=yes 1 <conf:(1)> lift:(1.67) lev:(0.08) [0] con