Data Mining and Warehousing Practical

Abhishek Roka
10621019

INDEX

<table>
<tr><th>S.No.</th><th>Topic</th><th>Remarks</th></tr>
<tr><td>1.</td><td>Create a relation named “Employee” with the help of data mining tool WEKA, which include attributes given below: -
<table>
<tr><th>Attribute</th><th>DataType</th></tr>
<tr><td>Name</td><td>String</td></tr>
<tr><td>ID</td><td>Numeric</td></tr>
<tr><td>Salary</td><td>{low, medium, high}</td></tr>
<tr><td>Experience</td><td>Numeric</td></tr>
<tr><td>Gender</td><td>{Male, Female}</td></tr>
<tr><td>Phone</td><td>Numeric</td></tr>
</table></td><td></td></tr>
<tr><td>2.</td><td>Create a relation named “Weather” with the help of data mining tool WEKA, which include attributes given below, then apply pre-processing techniques.

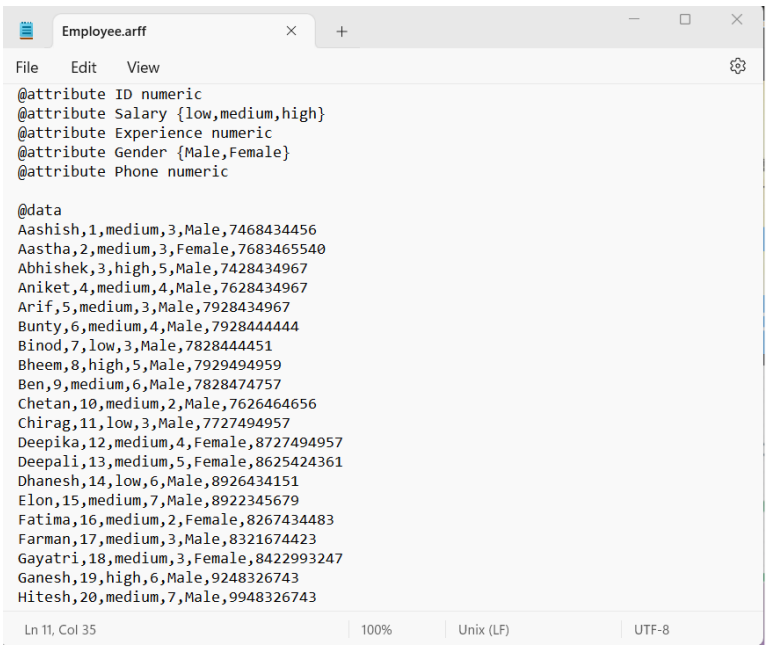
Add attribute name, Climate. Add index number 1, having data type nominal {good,bad}
Remove “windy” attribute using remove filter.
Normalization

<table>
<tr><th>Attribute Name</th><th>Data Type</th></tr>
<tr><td>Outlook</td><td>{Sunny, Rainy, Overcast}</td></tr>
<tr><td>Temperature</td><td>Numeric</td></tr>
<tr><td>Humidity</td><td>Numeric</td></tr>
<tr><td>Windy</td><td>{true,false}</td></tr>
<tr><td>Play</td><td>{yes,no}</td></tr>
</table></td><td></td></tr>
<tr><td>3.</td><td>Implement Association Mining on dataset, “labor.arff” using a priory algorithm using Explorer interface.</td><td></td></tr>
<tr><td>4.</td><td>Remove attribute “Age” and “Skin” from “diabetes” dataset using knowledge flow interface of WEKA</td><td></td></tr>
<tr><td>5.</td><td>Apply “Association Mining” on dataset “Diabetes” using Knowledge Flow. Perform all requisite steps on the given data set.</td><td></td></tr>
</table>

Program 1: Create a relation named "Employee" with the help of data mining tool WEKA, which include attributes given below: -

Attribute	DataType
Name	String
ID	Numeric
Salary	{low, medium, high}
Experience	Numeric
Gender	{Male, Female}
Phone	Numeric

Employee.arff file created on notepad:



Employee.arff

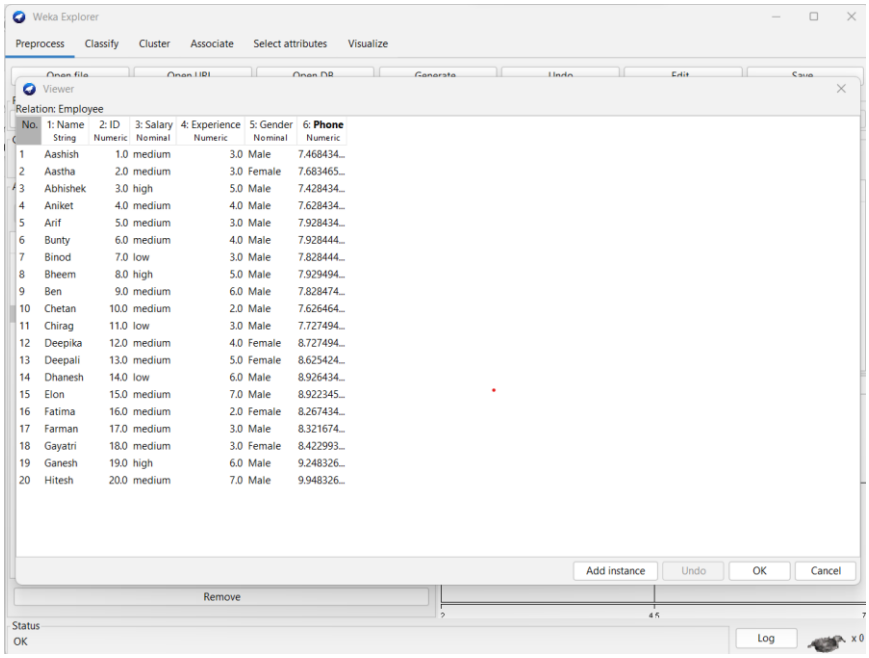
File Edit View

```
@attribute ID numeric
@attribute Salary {low,medium,high}
@attribute Experience numeric
@attribute Gender {Male,Female}
@attribute Phone numeric

@data
Aashish,1,medium,3,Male,7468434456
Aastha,2,medium,3,Female,7683465540
Abhishek,3,high,5,Male,7428434967
Aniket,4,medium,4,Male,7628434967
Arif,5,medium,3,Male,7928434967
Bunty,6,medium,4,Male,7928444444
Binod,7,low,3,Male,7828444451
Bheem,8,high,5,Male,7929494959
Ben,9,medium,6,Male,7828474757
Chetan,10,medium,2,Male,7626464656
Chirag,11,low,3,Male,7727494957
Deepika,12,medium,4,Female,8727494957
Deepali,13,medium,5,Female,8625424361
Dhanesh,14,low,6,Male,8926434151
Elon,15,medium,7,Male,8922345679
Fatima,16,medium,2,Female,8267434483
Farman,17,medium,3,Male,8321674423
Gayatri,18,medium,3,Female,8422993247
Ganesh,19,high,6,Male,9248326743
Hitesh,20,medium,7,Male,9948326743
```

Ln 11, Col 35 100% Unix (LF) UTF-8

Record in WEKA:



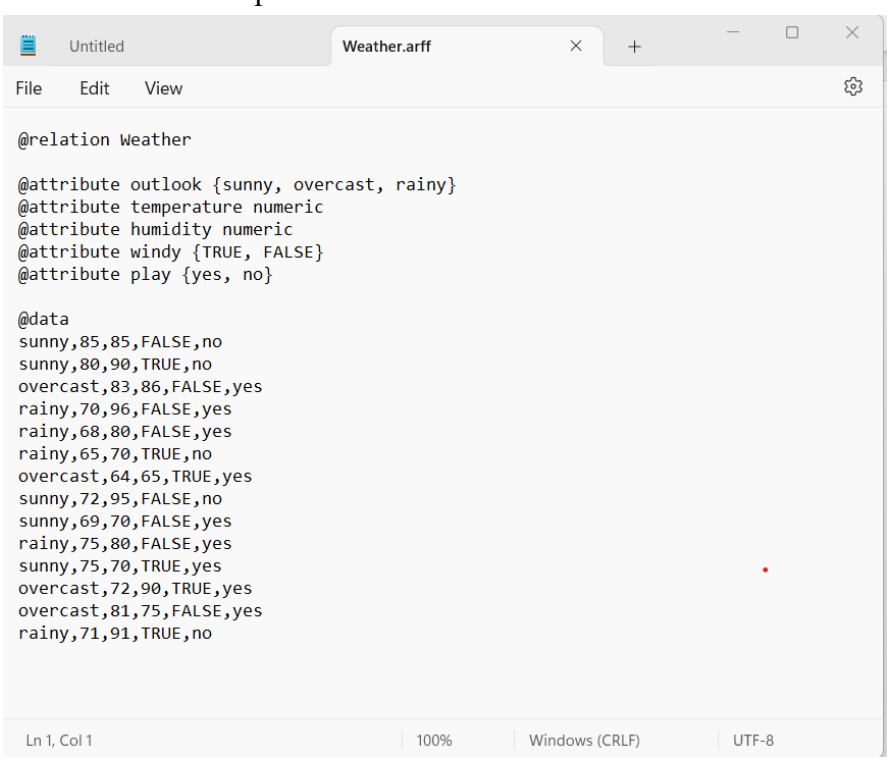
No.	1: Name String	2: ID Numeric	3: Salary Nominal	4: Experience Numeric	5: Gender Nominal	6: Phone Numeric
1	Aashish	1.0	medium	3.0	Male	7.468434...
2	Aastha	2.0	medium	3.0	Female	7.683465...
3	Abhishek	3.0	high	5.0	Male	7.428434...
4	Aniket	4.0	medium	4.0	Male	7.628434...
5	Arif	5.0	medium	3.0	Male	7.928434...
6	Bunty	6.0	medium	4.0	Male	7.928444...
7	Binod	7.0	low	3.0	Male	7.828444...
8	Bheem	8.0	high	5.0	Male	7.929494...
9	Ben	9.0	medium	6.0	Male	7.828474...
10	Chetan	10.0	medium	2.0	Male	7.626464...
11	Chirag	11.0	low	3.0	Male	7.727494...
12	Deepika	12.0	medium	4.0	Female	8.727494...
13	Deepali	13.0	medium	5.0	Female	8.625424...
14	Dhanesh	14.0	low	6.0	Male	8.926434...
15	Elon	15.0	medium	7.0	Male	8.922345...
16	Fatima	16.0	medium	2.0	Female	8.267434...
17	Farman	17.0	medium	3.0	Male	8.321674...
18	Gayatri	18.0	medium	3.0	Female	8.422993...
19	Ganesh	19.0	high	6.0	Male	9.248326...
20	Hitesh	20.0	medium	7.0	Male	9.948326...

Program 2: Create a relation named "Weather" with the help of data mining tool WEKA, which include attributes given below, then apply pre-processing techniques.

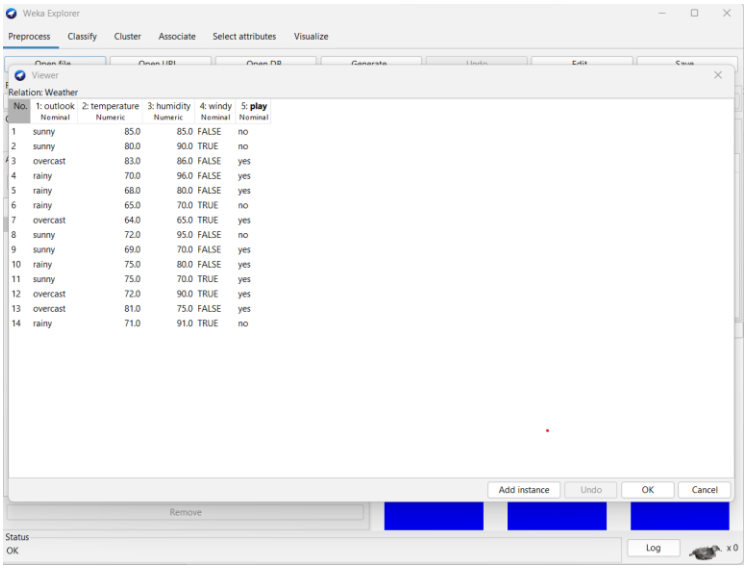
1. **Add attribute name, Climate. Add index number 1, having data type nominal {good,bad}**
2. **Remove "windy" attribute using remove filter.**
3. **Normalization**

Attribute Name	**Data Type**
Outlook	{Sunny, Rainy, Overcast}
Temperature	Numeric
Humidity	Numeric
Windy	{true,false}
Play	{yes,no}

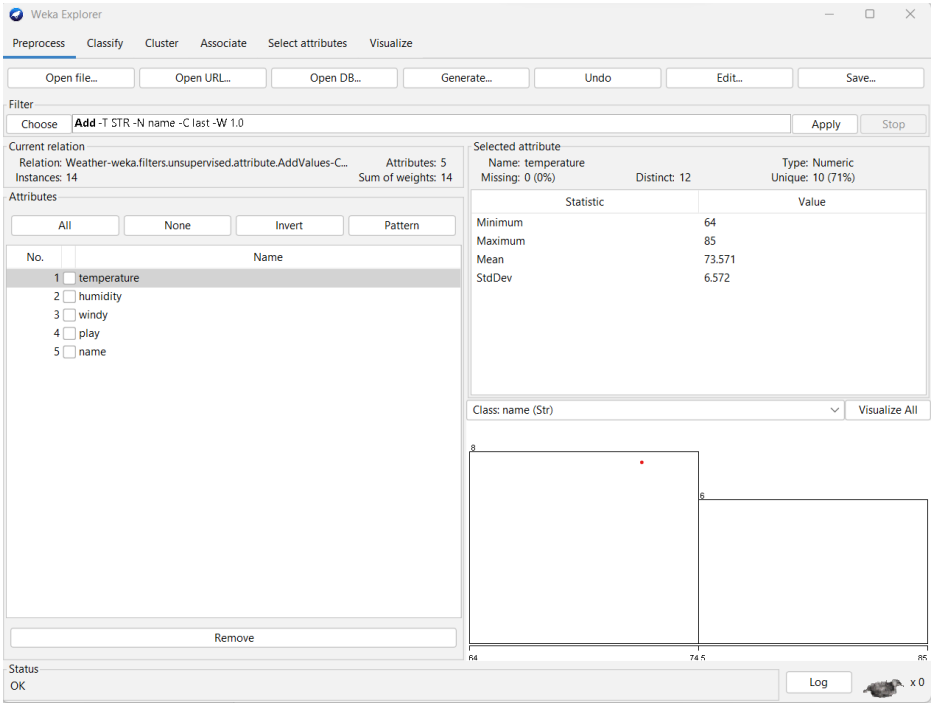
Step 1: Create a .arff file in notepad and enter the weather data.



Step 2: Open WEKA, go to explorer tab and open the “Weather.arff” file.

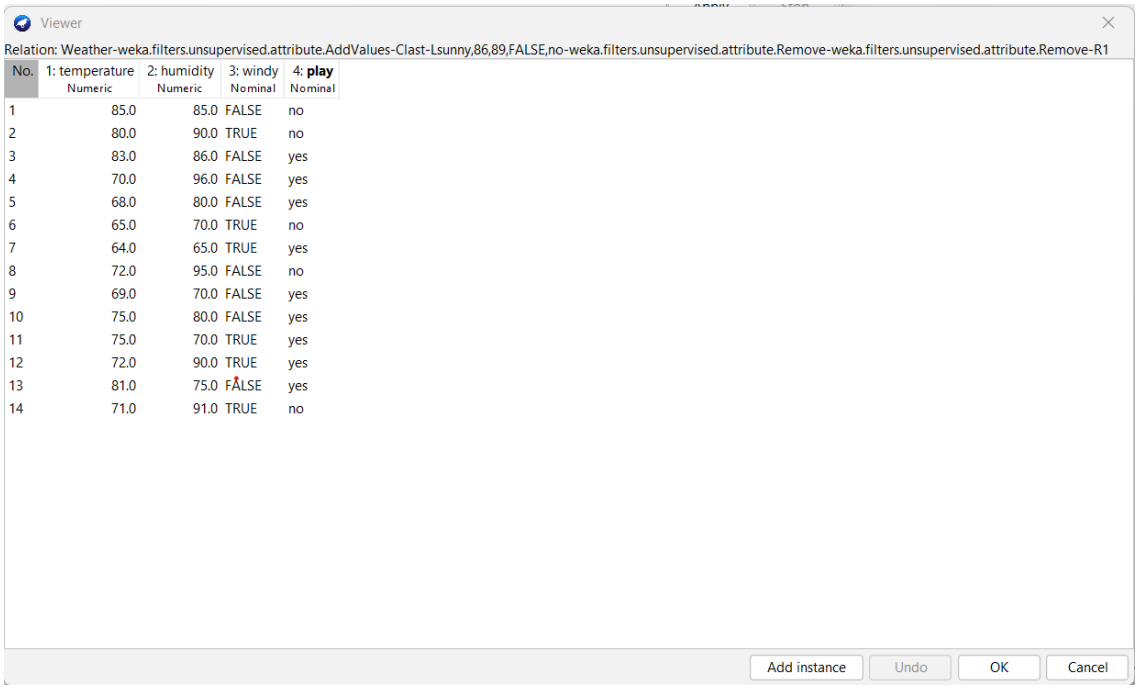


1. In Unsupervised filters, click on “Add filter” and fill the required values and apply it.



Added name attribute.

2. Choose the “remove filter for unsupervised filter, enter the attribute index, we want to remove and apply it.



Viewer

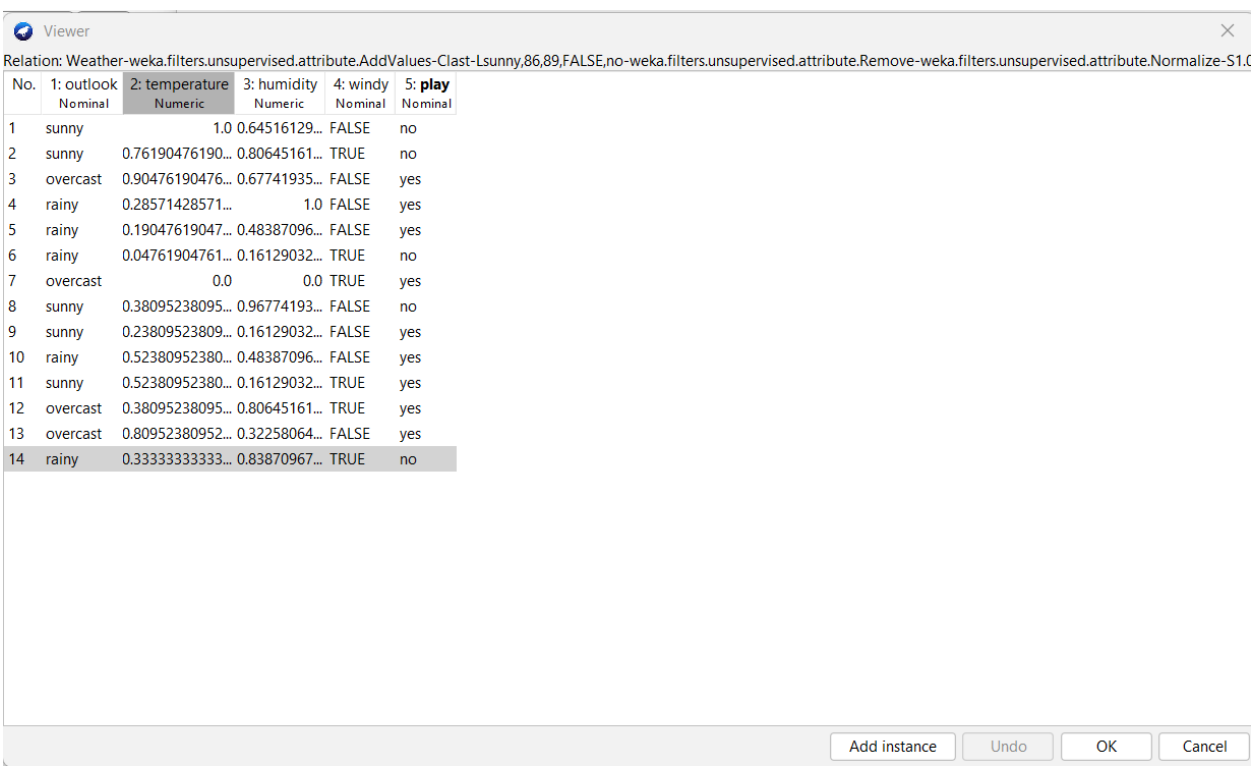
Relation: Weather-weka.filters.unsupervised.attribute.AddValues-Clast-Lsunny,86,89,FALSE,no-weka.filters.unsupervised.attribute.Remove-weka.filters.unsupervised.attribute.Remove-R1

No.	1: temperature Numeric	2: humidity Numeric	3: windy Nominal	4: **play** Nominal
1	85.0	85.0	FALSE	no
2	80.0	90.0	TRUE	no
3	83.0	86.0	FALSE	yes
4	70.0	96.0	FALSE	yes
5	68.0	80.0	FALSE	yes
6	65.0	70.0	TRUE	no
7	64.0	65.0	TRUE	yes
8	72.0	95.0	FALSE	no
9	69.0	70.0	FALSE	yes
10	75.0	80.0	FALSE	yes
11	75.0	70.0	TRUE	yes
12	72.0	90.0	TRUE	yes
13	81.0	75.0	FALSE	yes
14	71.0	91.0	TRUE	no

Add instance | Undo | OK | Cancel

Step 3: Normalization

Click the choose button to select the filter and select unsupervised attribute, normalize and apply it.



Viewer

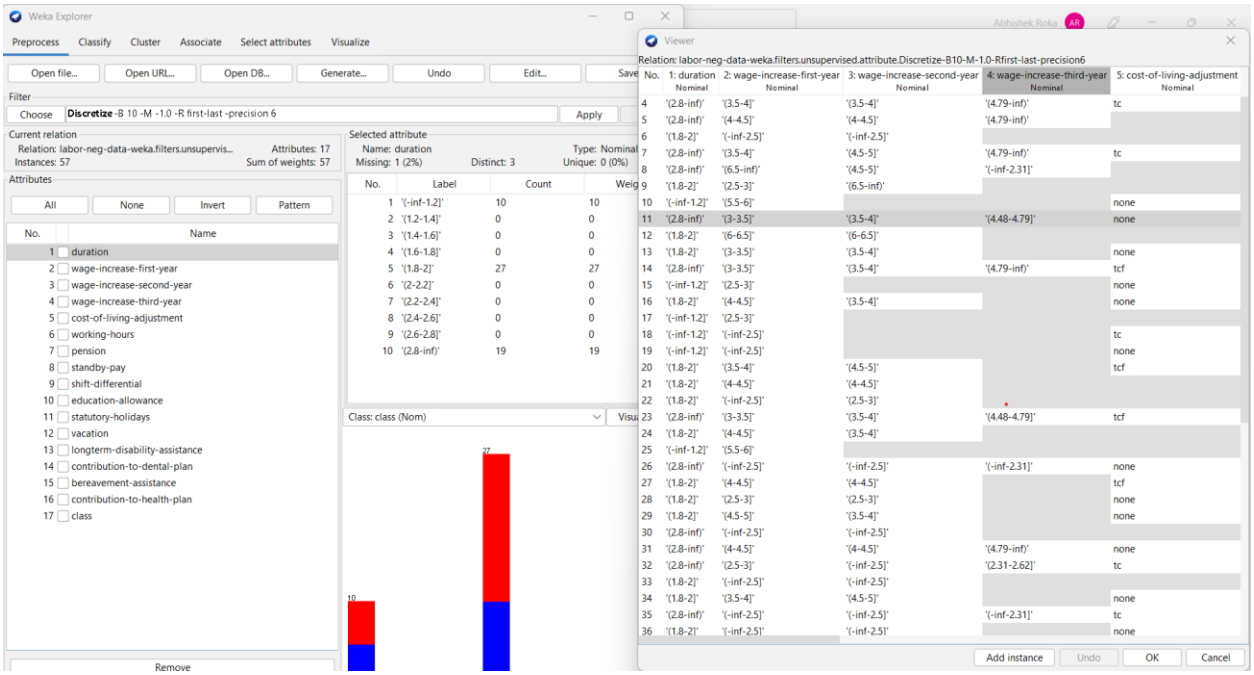
Relation: Weather-weka.filters.unsupervised.attribute.AddValues-Clast-Lsunny,86,89,FALSE,no-weka.filters.unsupervised.attribute.Remove-weka.filters.unsupervised.attribute.Normalize-S1.0

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	sunny	1.0	0.64516129...	FALSE	no
2	sunny	0.76190476190...	0.80645161...	TRUE	no
3	overcast	0.90476190476...	0.67741935...	FALSE	yes
4	rainy	0.28571428571...	1.0	FALSE	yes
5	rainy	0.19047619047...	0.48387096...	FALSE	yes
6	rainy	0.04761904761...	0.16129032...	TRUE	no
7	overcast	0.0	0.0	TRUE	yes
8	sunny	0.38095238095...	0.96774193...	FALSE	no
9	sunny	0.23809523809...	0.16129032...	FALSE	yes
10	rainy	0.52380952380...	0.48387096...	FALSE	yes
11	sunny	0.52380952380...	0.16129032...	TRUE	yes
12	overcast	0.38095238095...	0.80645161...	TRUE	yes
13	overcast	0.80952380952...	0.32258064...	FALSE	yes
14	rainy	0.33333333333...	0.83870967...	TRUE	no

Add instance | Undo | OK | Cancel

Program 3: Implement Association Mining on dataset, “labor.arff” using a priory algorithm using Explorer interface.

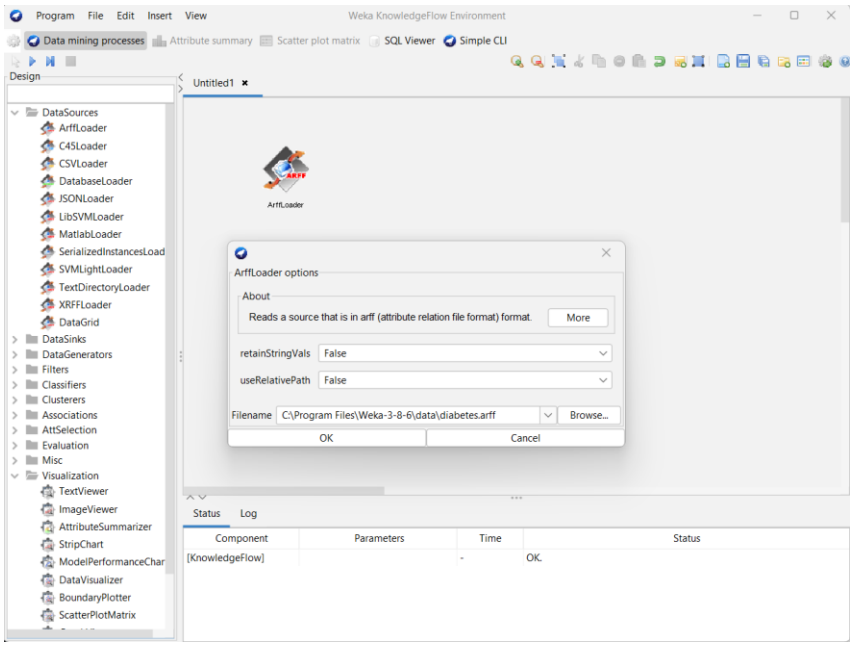
Solution: Click the “choose” button to select a filter and select unsupervised -> attribute -> discretize -> Apply it.



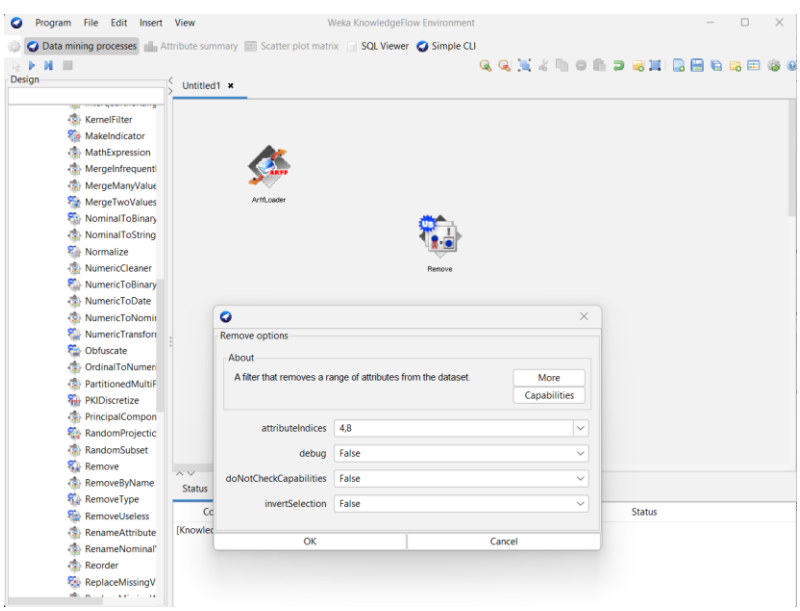
Program 4: Remove attribute “Age” and “Skin” from “diabetes” dataset using knowledge flow interface of WEKA

Step 1:

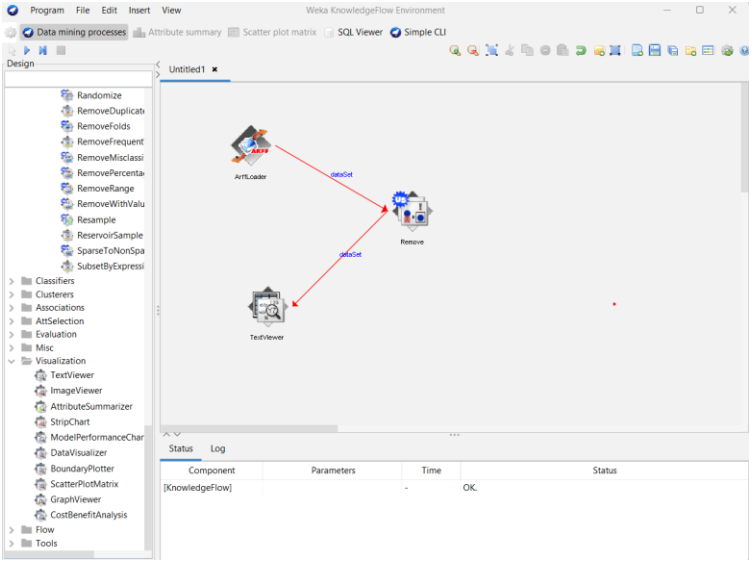
1. Open “Knowledge Flow” tab from WEKA main window.
2. Add arff loader in the “Knowledge Flow Environment”.
3. Select the “diabetes” data and Click “Ok”.



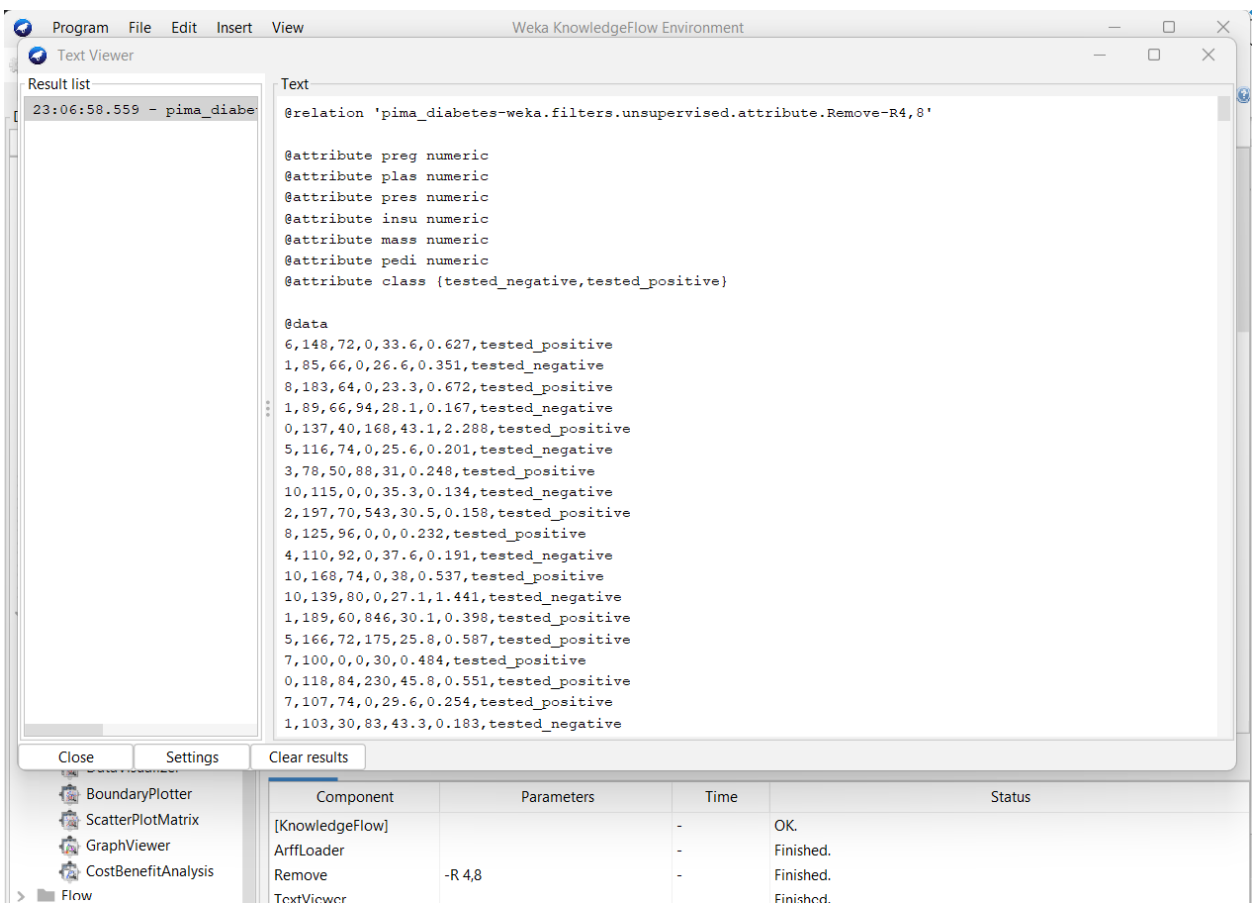
Step 2: Now, add “remove” filter from “Unsupervised” filters, connect it to “ArffLoader” and enter values (4,8) to remove age and skin attribute. Click “OK”



Step 3: Add “TextViewer” for visualization and connect it to remove component.



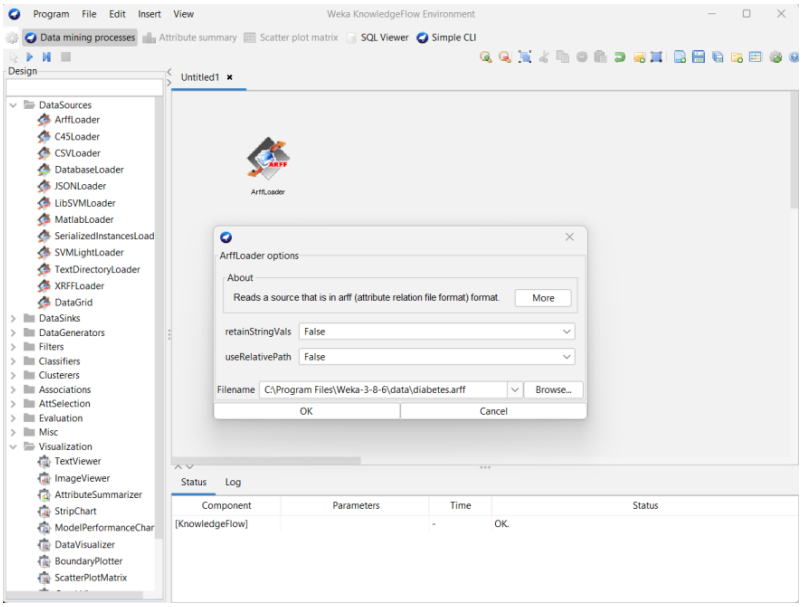
Step 4: Finally, run the “Knowledge Flow” and see the result in “TextViewer” by clicking on “Show Results” option.



Program 5: Apply "Association Mining" on dataset "Diabetes" using Knowledge Flow. Perform all requisite steps on the given data set.

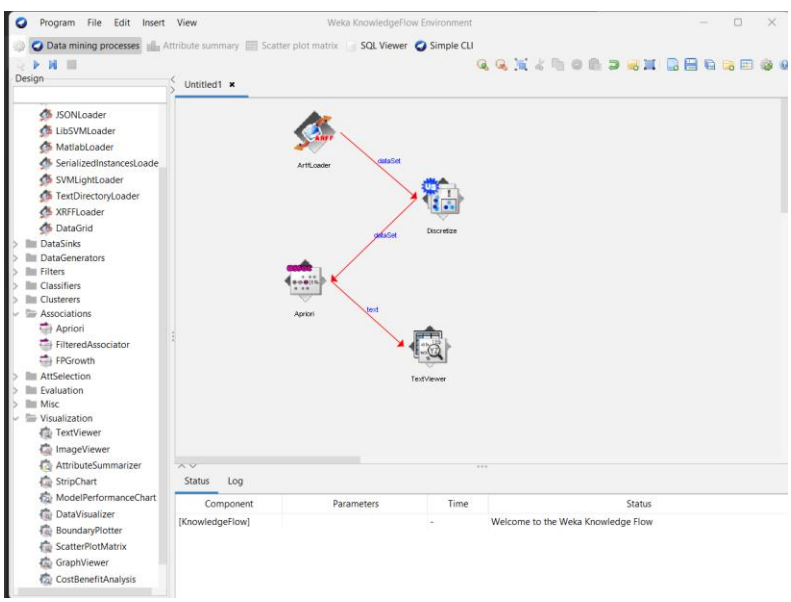
Step 1:

1. Open "Knowledge Flow" tab from WEKA main window.
2. Add "ArffLoader" in the Knowledge Flow Environment.
3. Select the Diabetes data and click OK button.



Step 2:

1. Add "Discretize" from "Unsupervised", "Apriori" from "Associations" and "TextViewer" from "Visualization".
2. Connect them with each other.



Step 3: Finally, run the “Knowledge Flow” and see the result in “TextViewer” by clicking on “Show Results” option.