Abhishek Roka

10621019

Data Mining and Warehousing Practical File

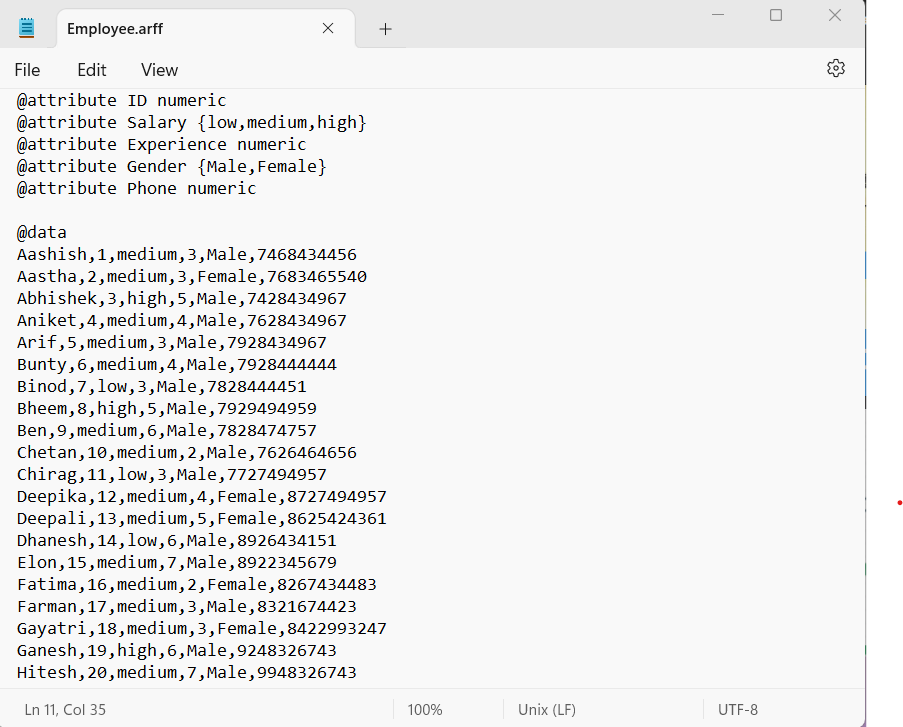
[Document subtitle]

**INDEX**

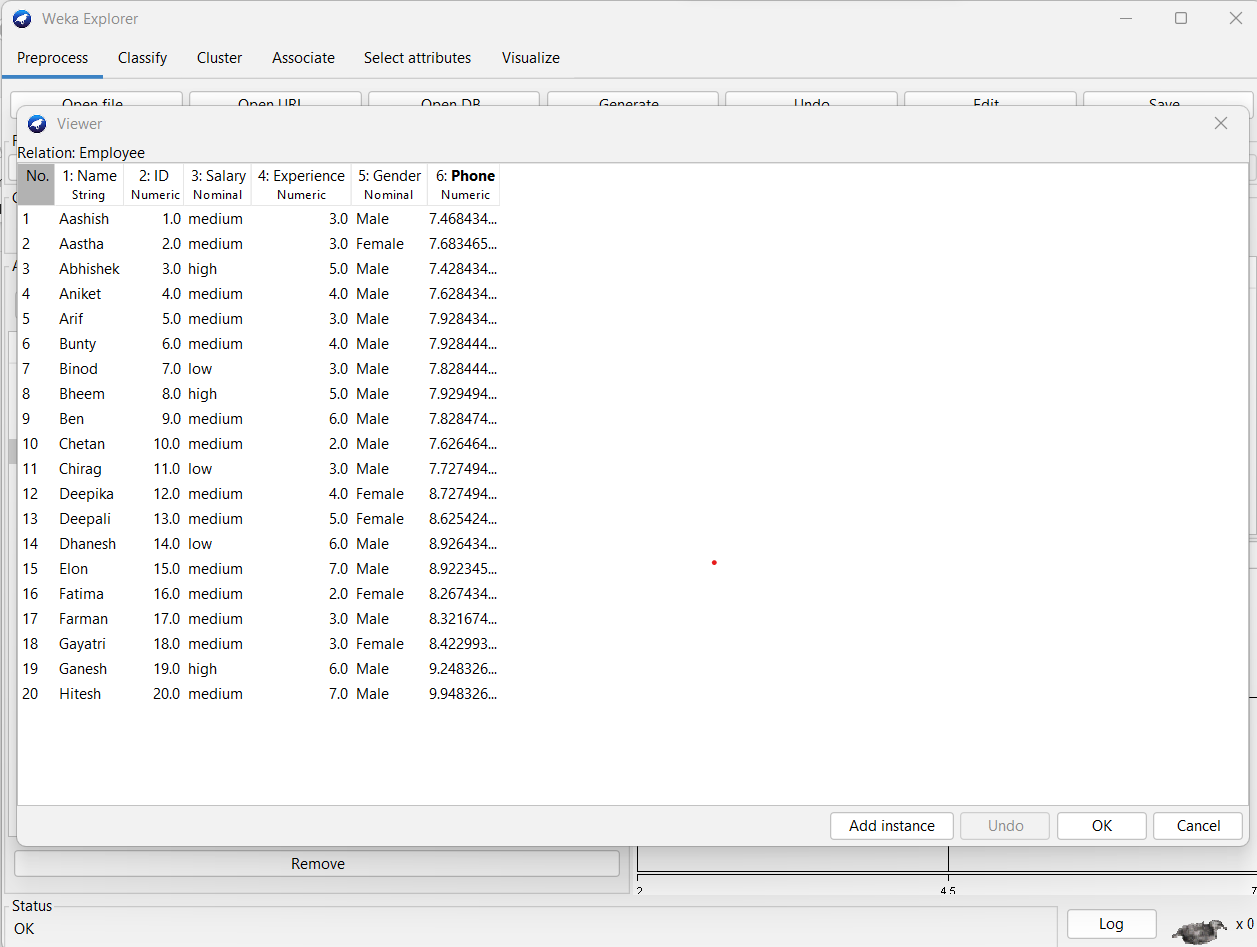
|  |  |  |
| --- | --- | --- |
| **S.No.** | **Topic** | **Remarks** |
| 1. | Create a relation named “Employee” with the help of data mining tool WEKA, which include attributes given below: -  |  |  | | --- | --- | | Attribute | DataType | | Name | String | | ID | Numeric | | Salary | {low, medium, high} | | Experience | Numeric | | Gender | {Male, Female} | | Phone | Numeric | |  |
| 2. | 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  |  |  | | --- | --- | | **Attribute Name** | **Data Type** | | Outlook | {Sunny, Rainy, Overcast} | | Temperature | Numeric | | Humidity | Numeric | | Windy | {true,false} | | Play | {yes,no} | |  |
| 3. | Implement Association Mining on dataset, “labor.arff” using a priory algorithm using Explorer interface. |  |
| 4. | Remove attribute “Age” and “Skin” from “diabetes” dataset using knowledge flow interface of WEKA |  |
| 5. | Apply “Association Mining” on dataset “Diabetes” using Knowledge Flow. Perform all requisite steps on the given data set. |  |

# **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:

Record in WEKA:



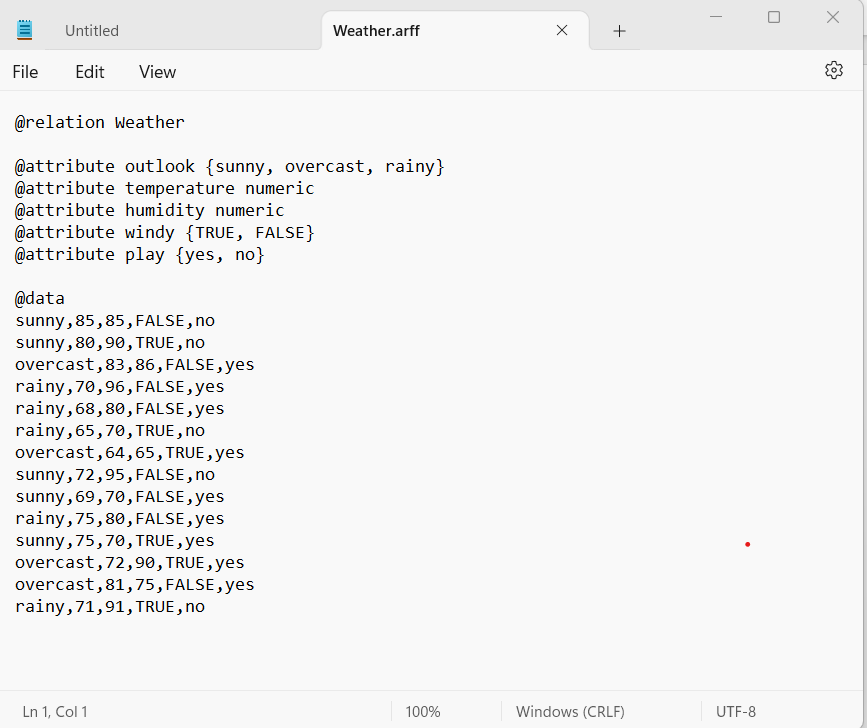
# **Program 2:** **Create a relation named “Weather” with the help of data mining tool WEKA, which include attributes given below, then apply pre-processing techniques.**

# **Add attribute name, Climate. Add index number 1, having data type nominal {good,bad}**

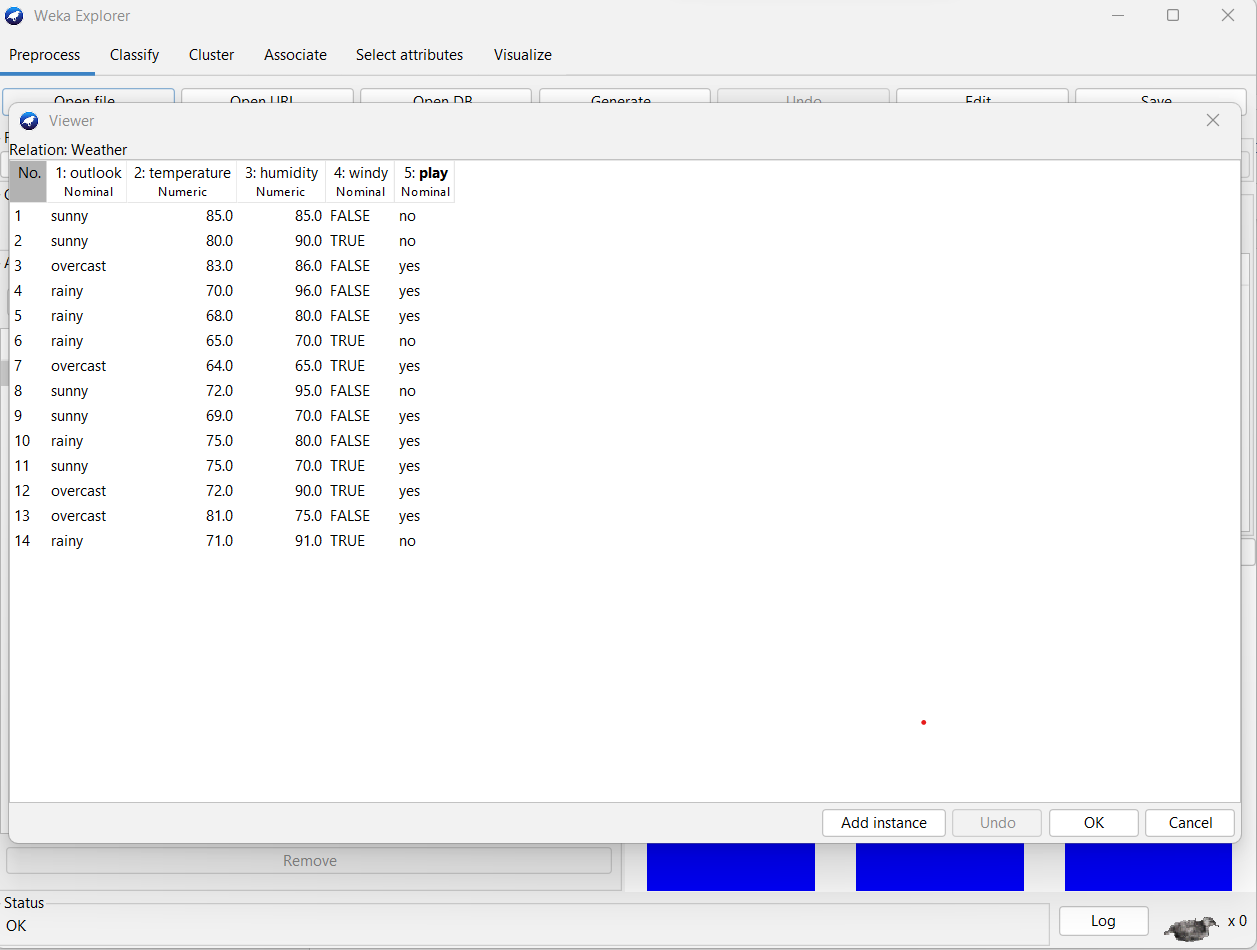
# **Remove “windy” attribute using remove filter.**

# **Normalization**

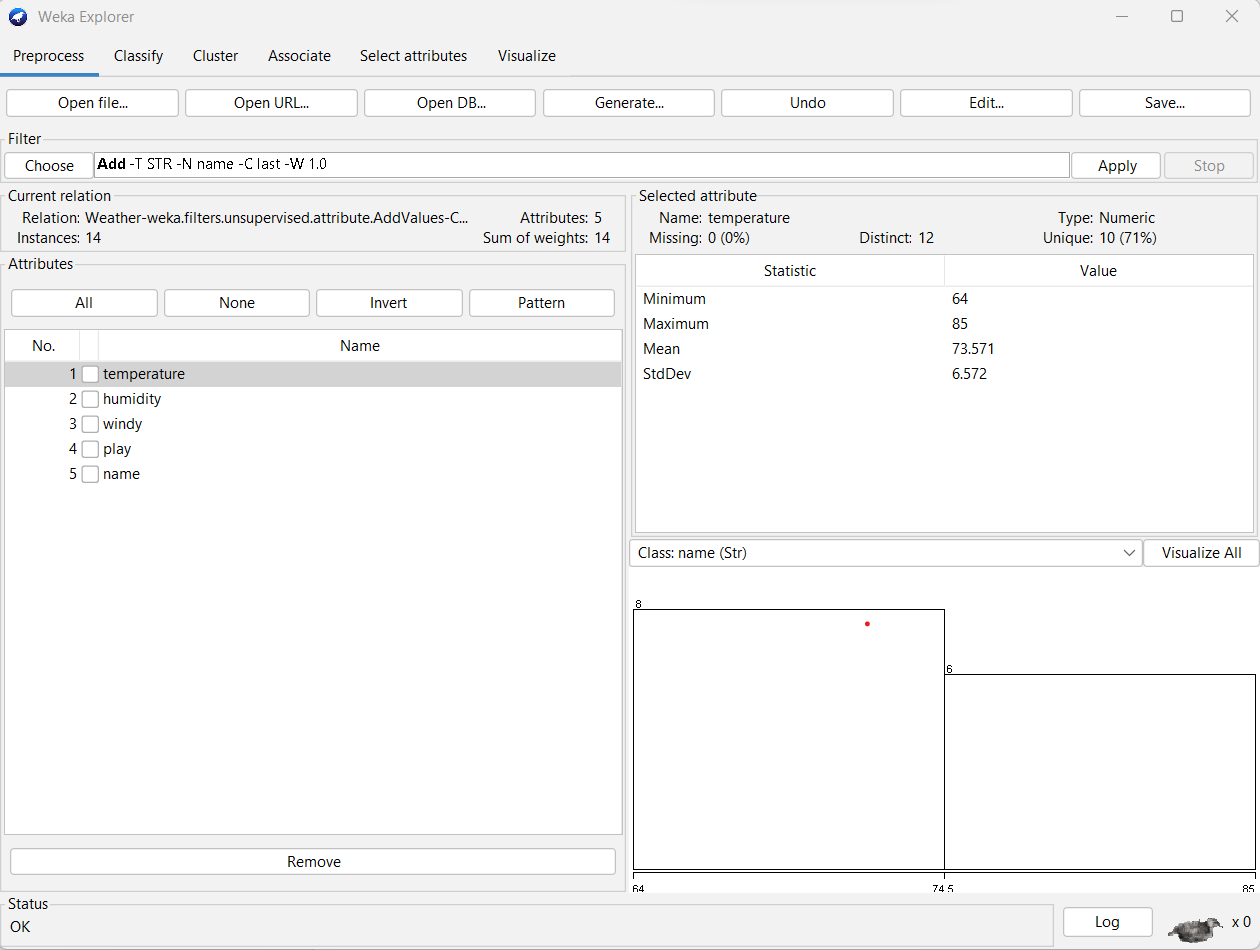
|  |  |
| --- | --- |
| **Attribute Name** | **Data Type** |
| Outlook | {Sunny, Rainy, Overcast} |
| Temperature | Numeric |
| Humidity | Numeric |
| Windy | {true,false} |
| Play | {yes,no} |

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.

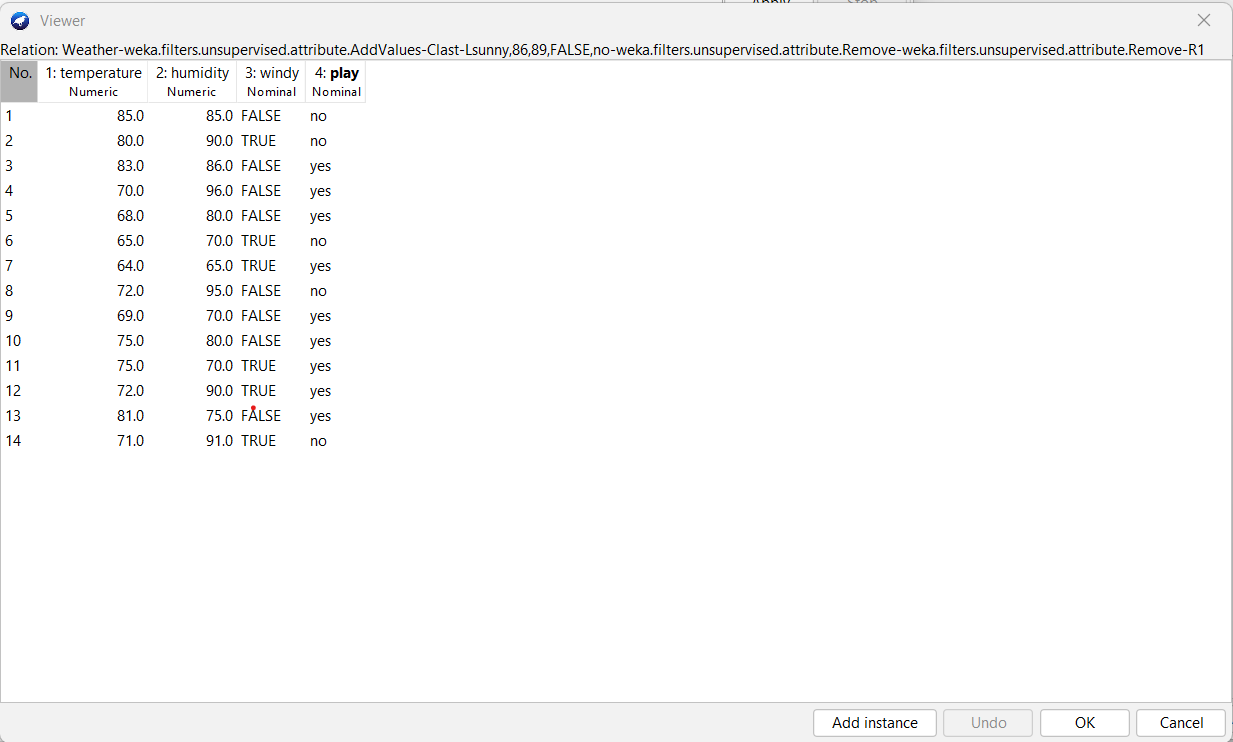


1. In Unsupervised filters, click on “Add filter” and fill the required values and apply it.

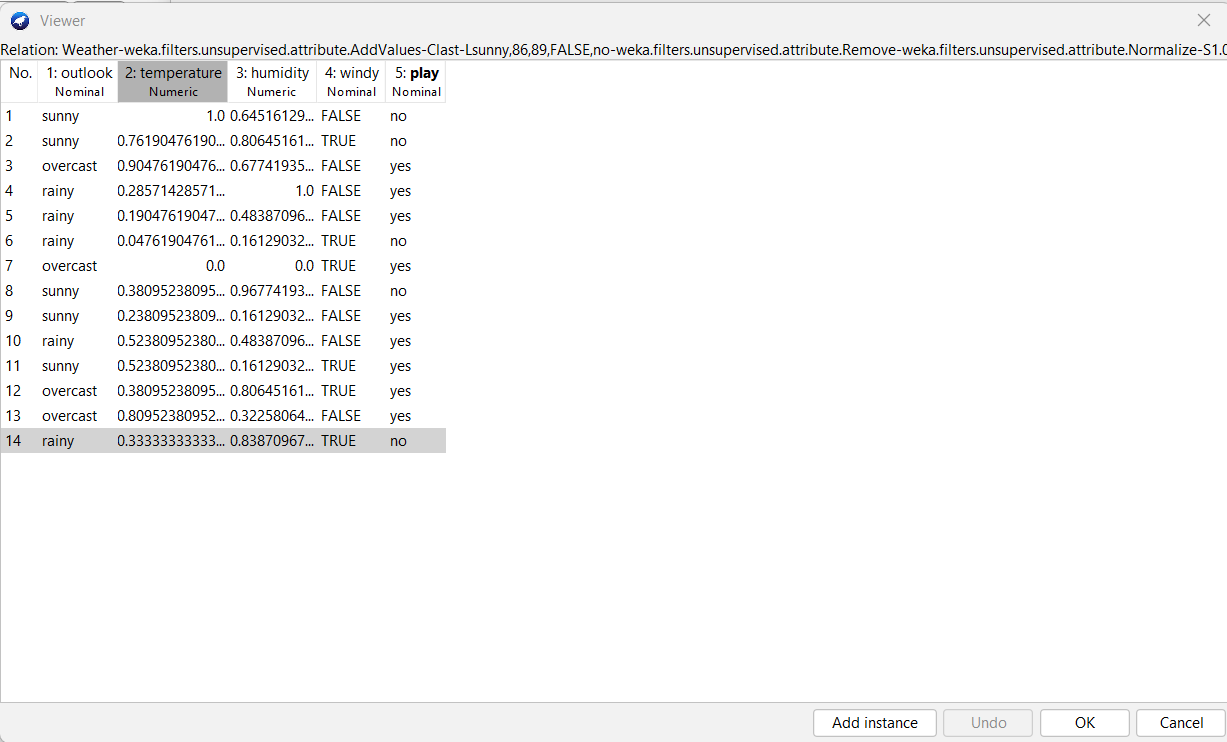


Added name attribute.

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

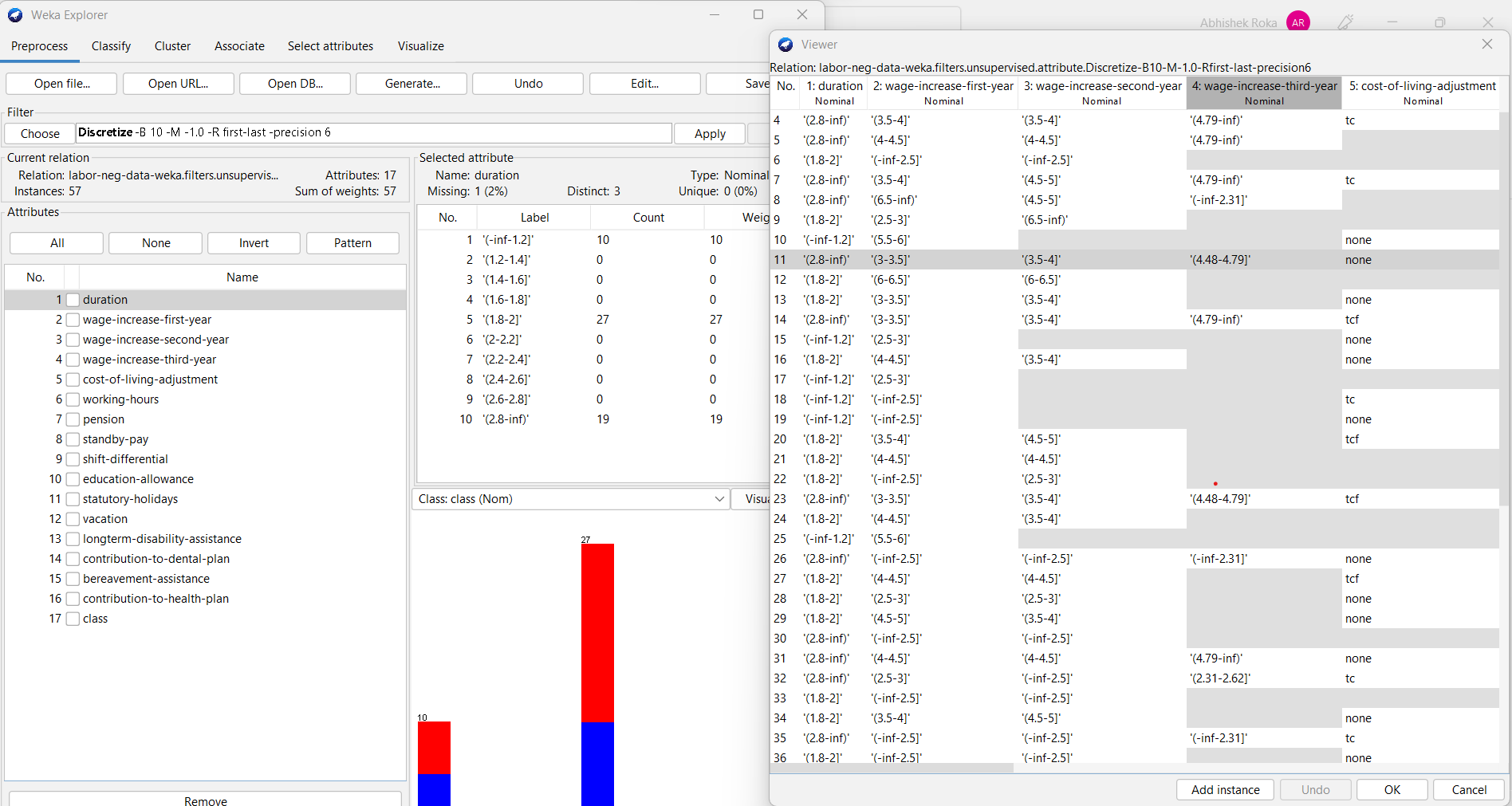


Step 3: Normalization  
Click the choose button to select the filter and select unsupervised attribute, normalize and apply it.



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

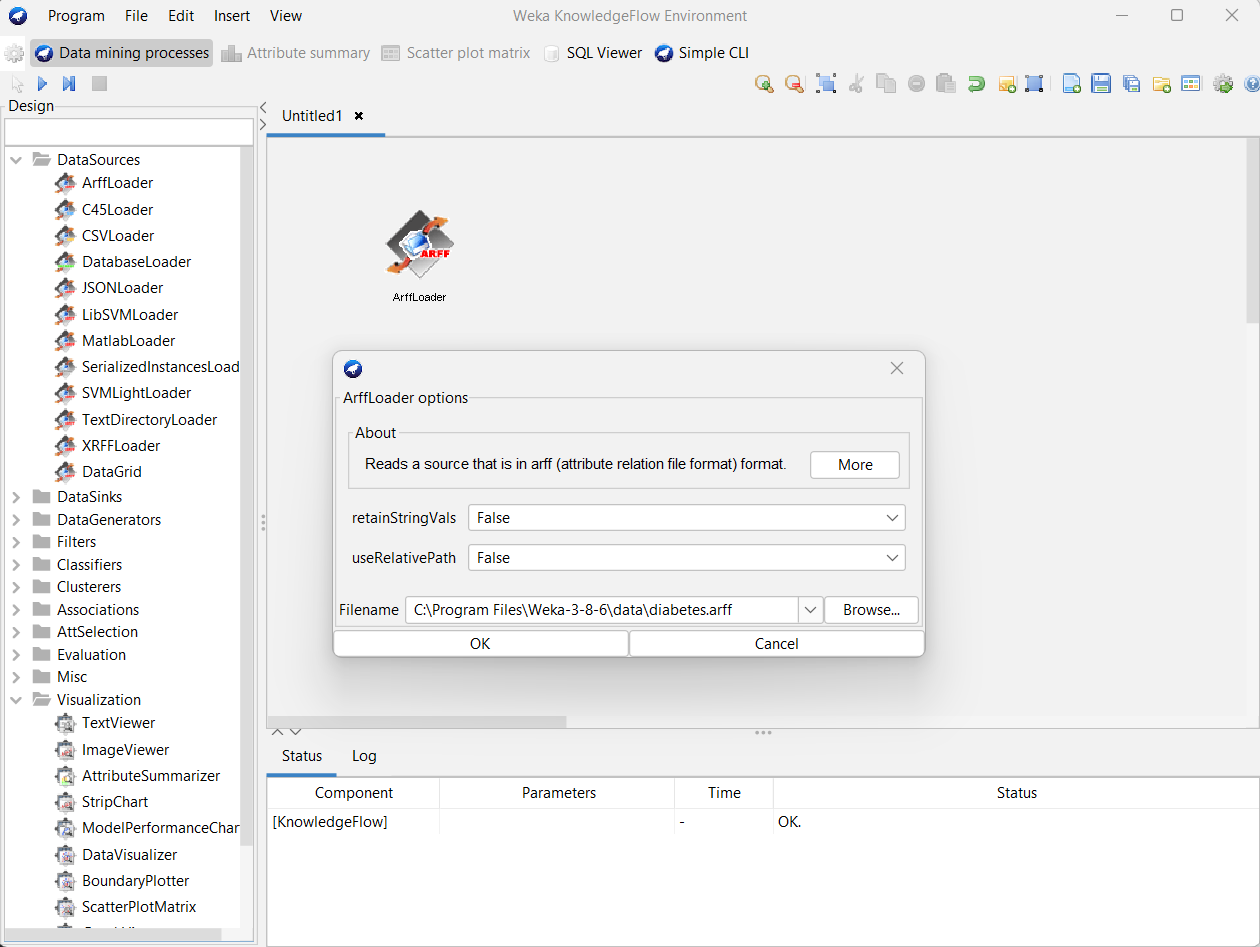
Solution: Click the “choose” button to select a filter and select unsupervised -> attribute -> discretize -> Apply it.



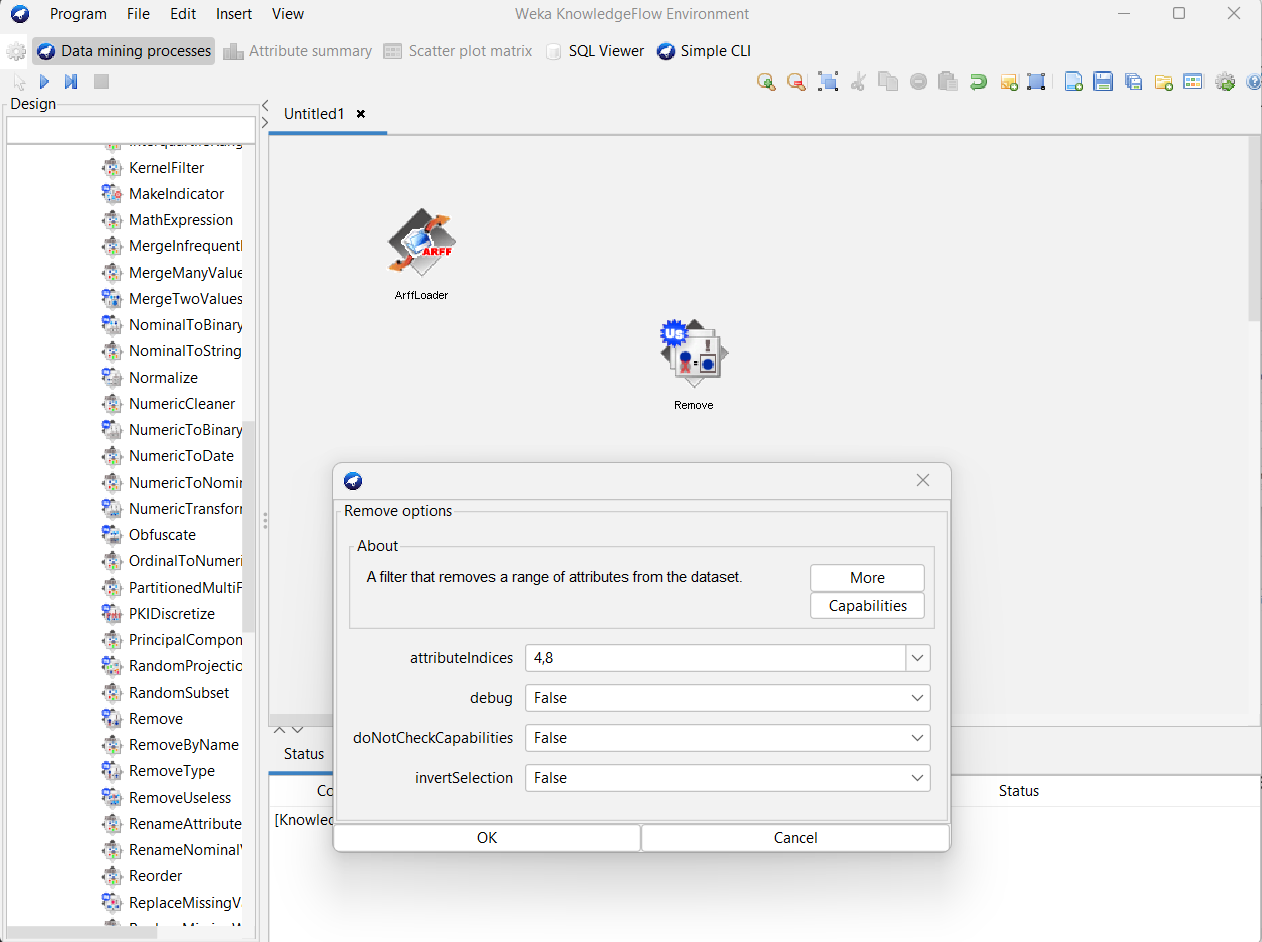
Program 4: Remove attribute “Age” and “Skin” from “diabetes” dataset using knowledge flow interface of WEKA

Step 1:

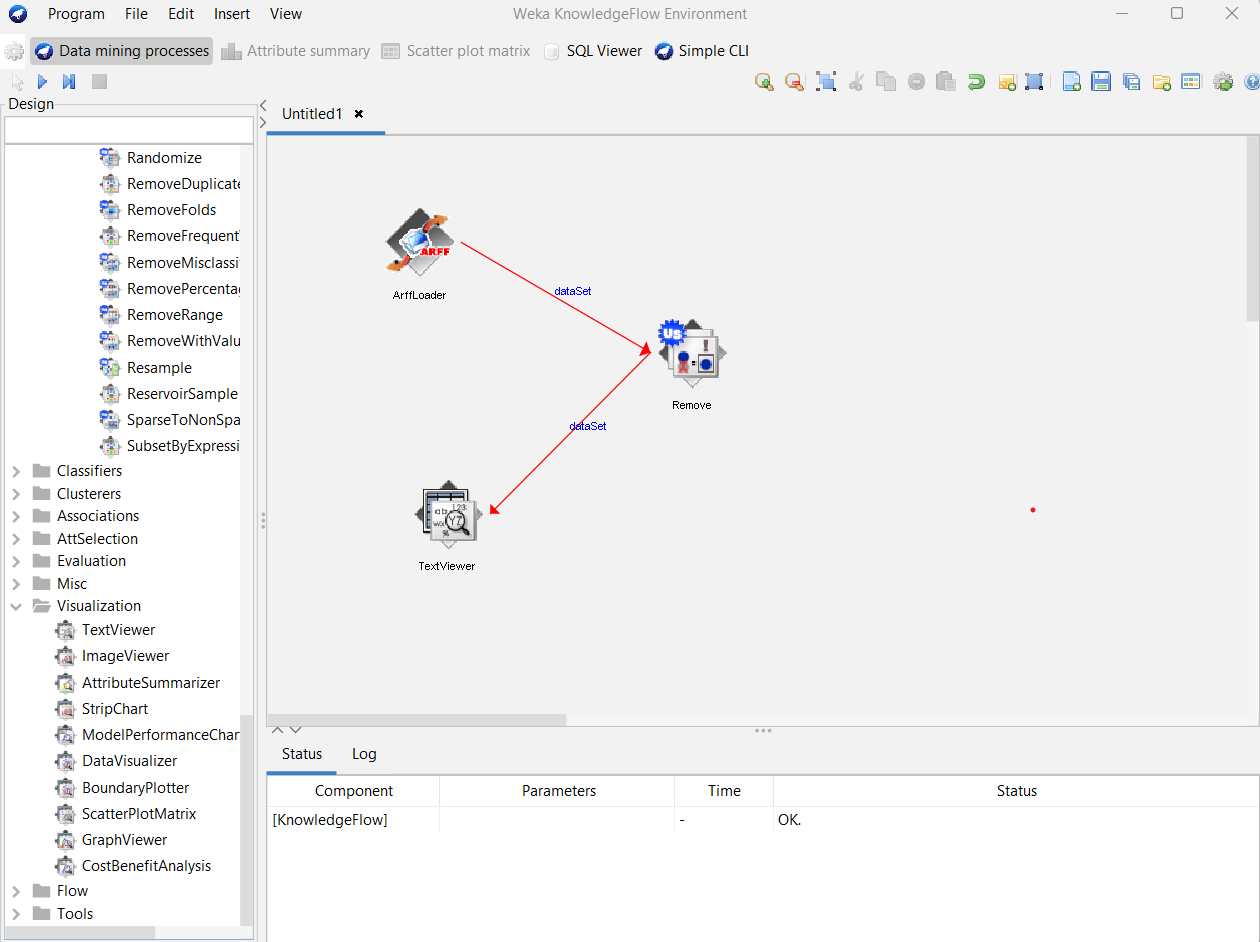
1. Open “Knowledge Flow” tab from WEKA main window.
2. Add arff loader in the “Knowledge Flow Environment”.
3. Select the “diabetes” data and Click “Ok”.



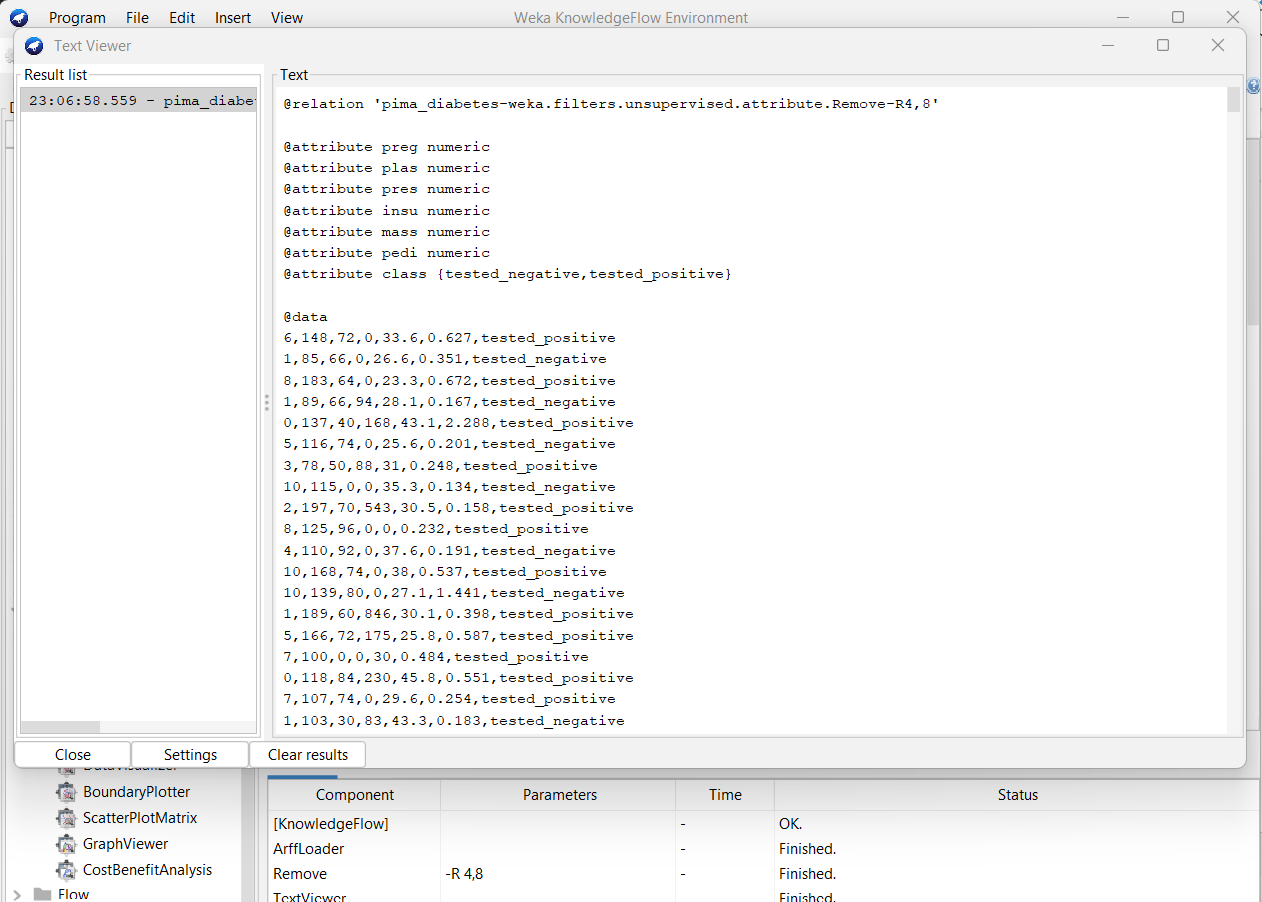
Step 2: Now, add “remove” filter from “Unsupervised” filters, connect it to “ArffLoader” and enter values (4,8) to remove age and skin attribute. Click “OK”



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



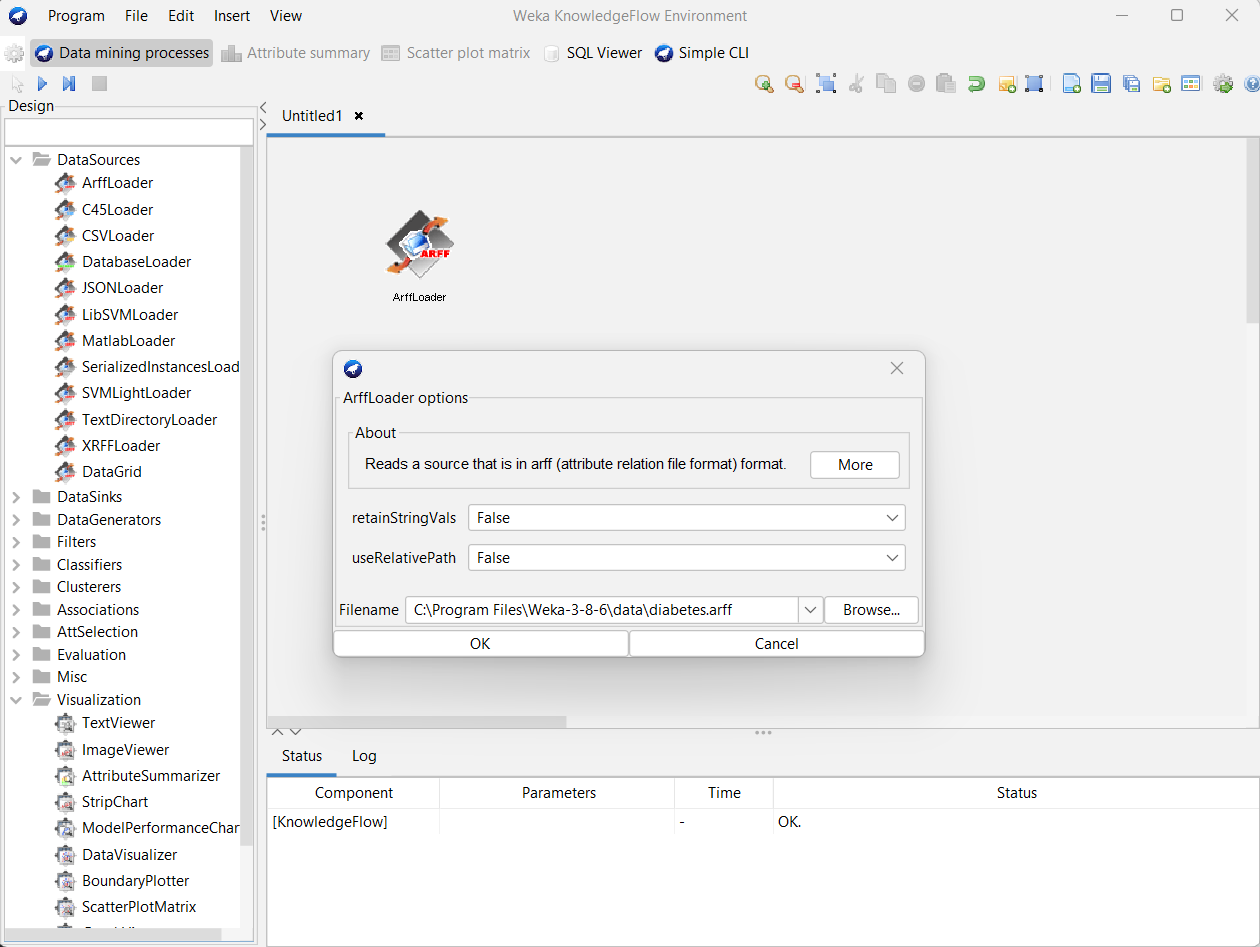
Step 4: Finally, run the “Knowledge Flow” and see the result in “TextViewer” by clicking on “Show Results” option.



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

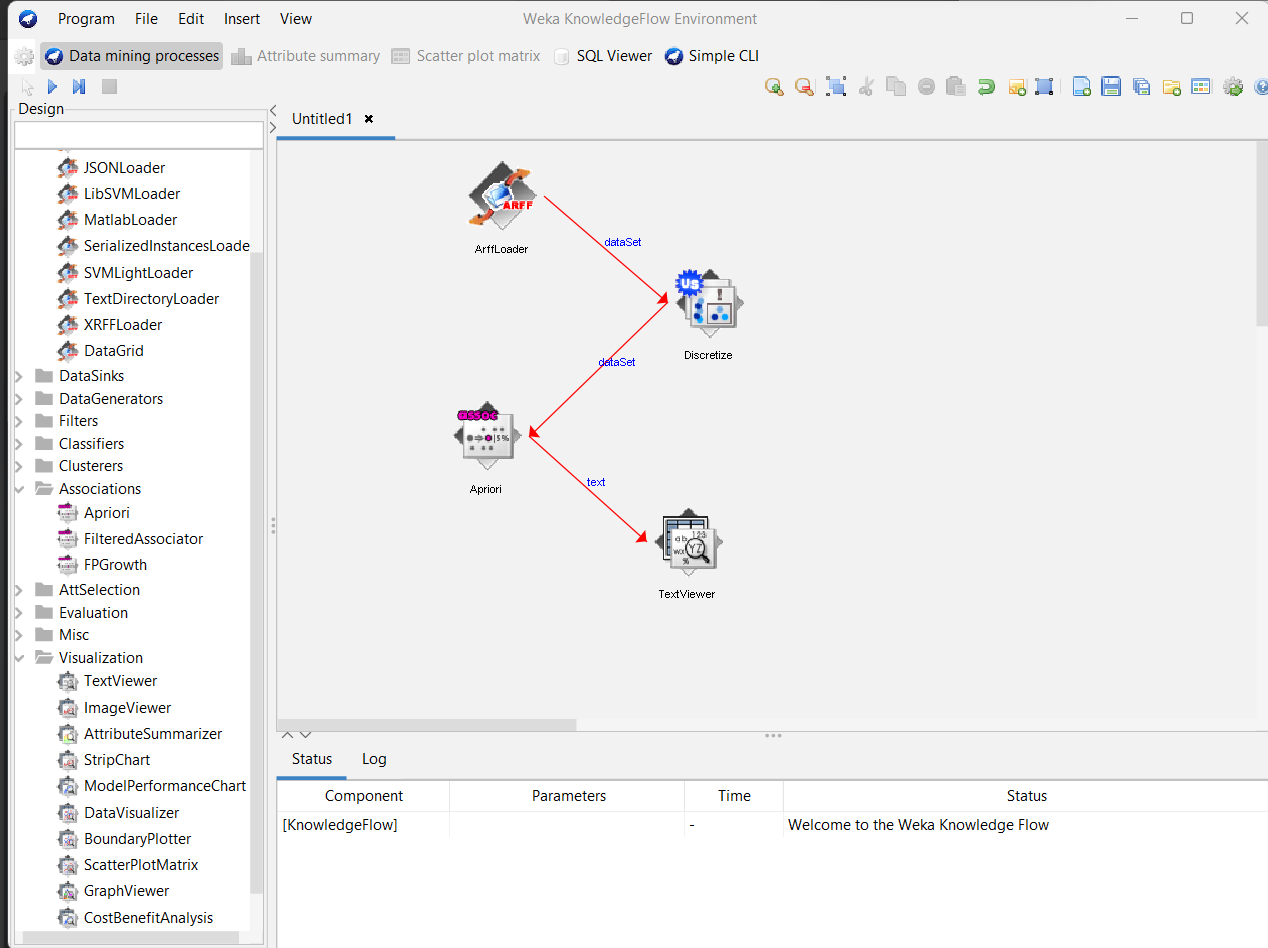
Step 1:

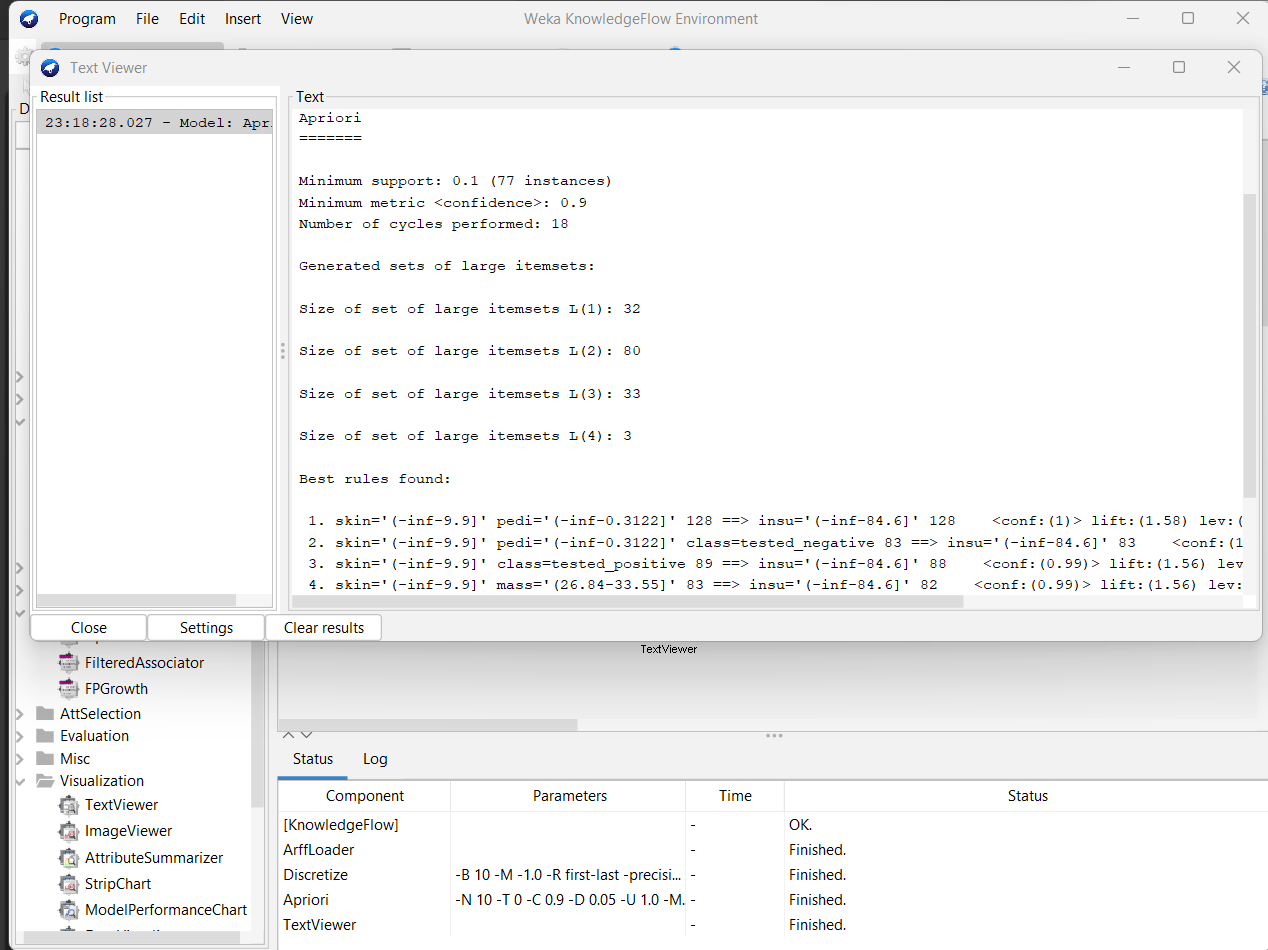
1. Open “Knowledge Flow” tab from WEKA main window.
2. Add “ArffLoader” in the Knowledge Flow Environment.
3. Select the Diabetes data and click OK button.



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.