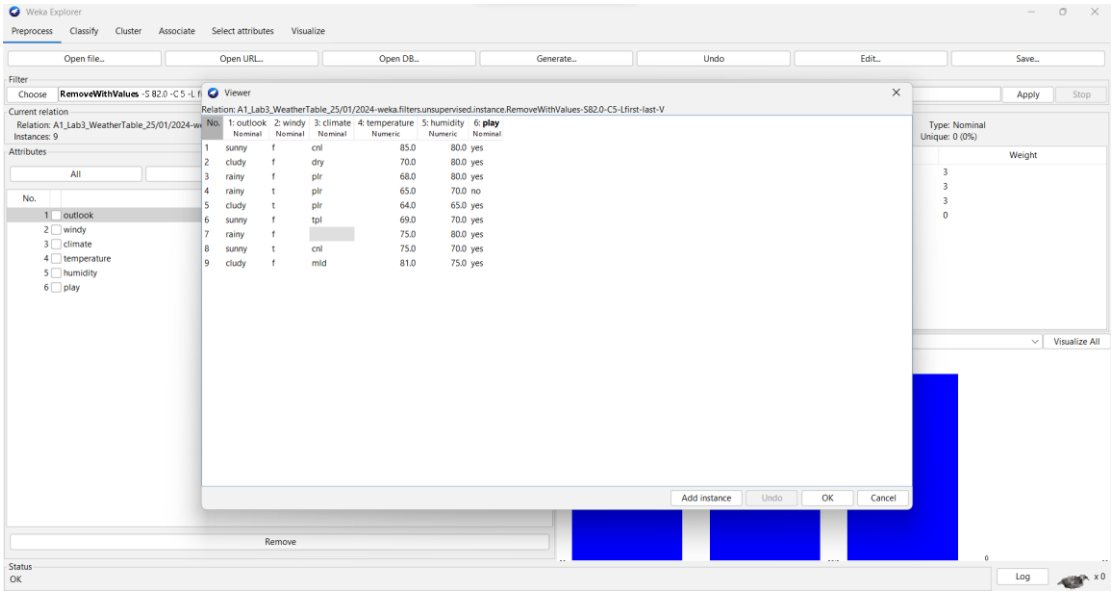
Assignment 2

Name – Ayush Kumar
Reg No. – 20214284
CSE – 6B1

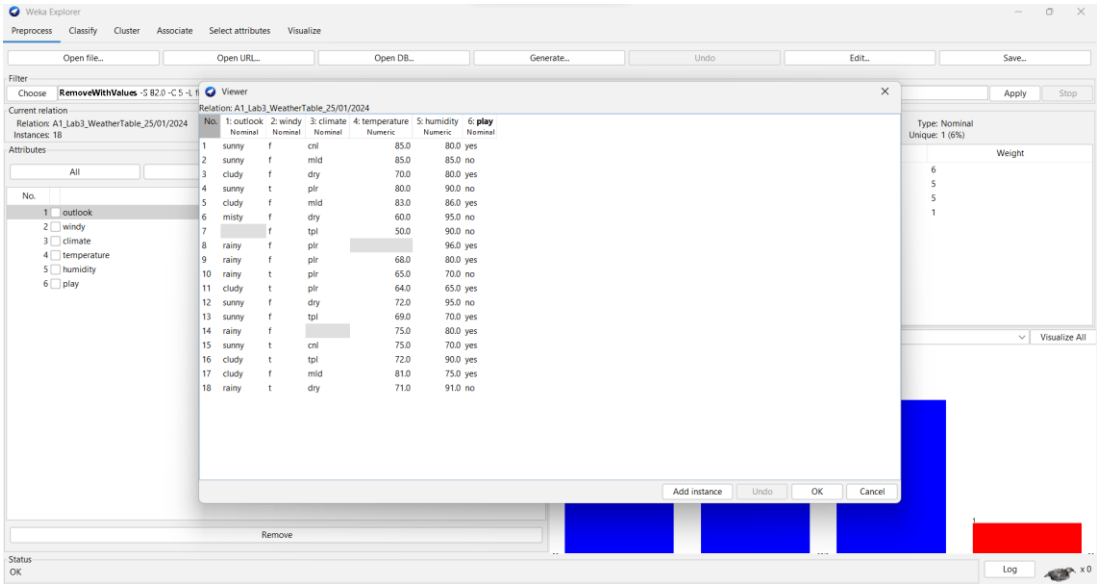
WEKA.FILTER.UNSUPERVISED.INSTANCE:

Q1. Load the weather.nominal dataset. Use the filter weka.unsupervised.instance and do analyze following preprocessing operations. From the Object Editor window, figure out changes appropriately after each preprocessing. Undo the change to the dataset that you just performed, and verify that the data has reverted to its original state.

1. Remove all instances in which the humidity attribute has the value high.

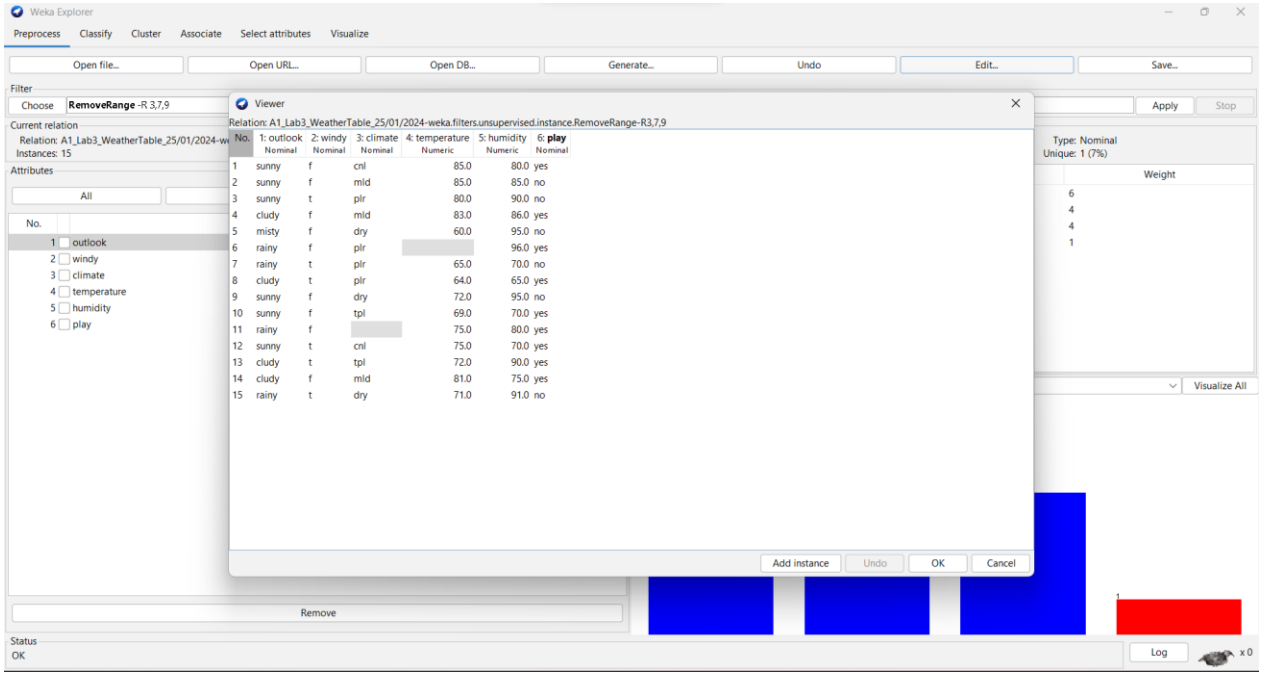


after removing all instances with the values high (i.e.greater than average), we get the above dataset.



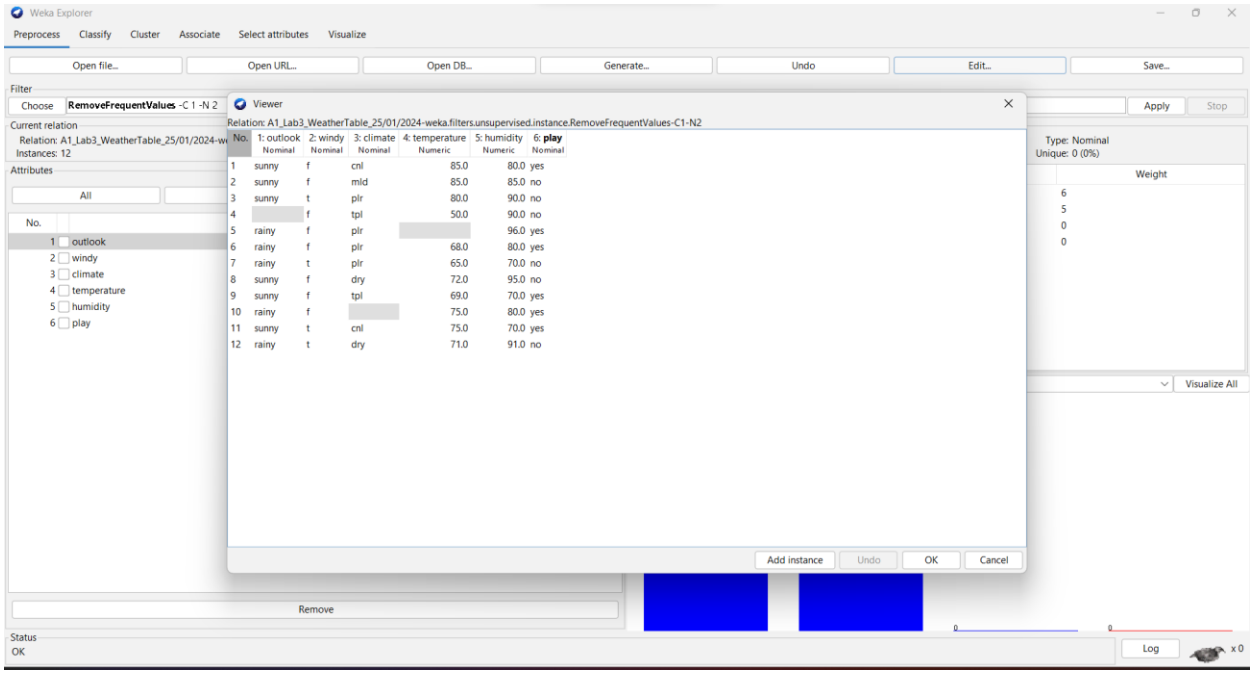
after undo, it returns back to the original dataset

2. Remove instances with indices number 3, 7 and 9.

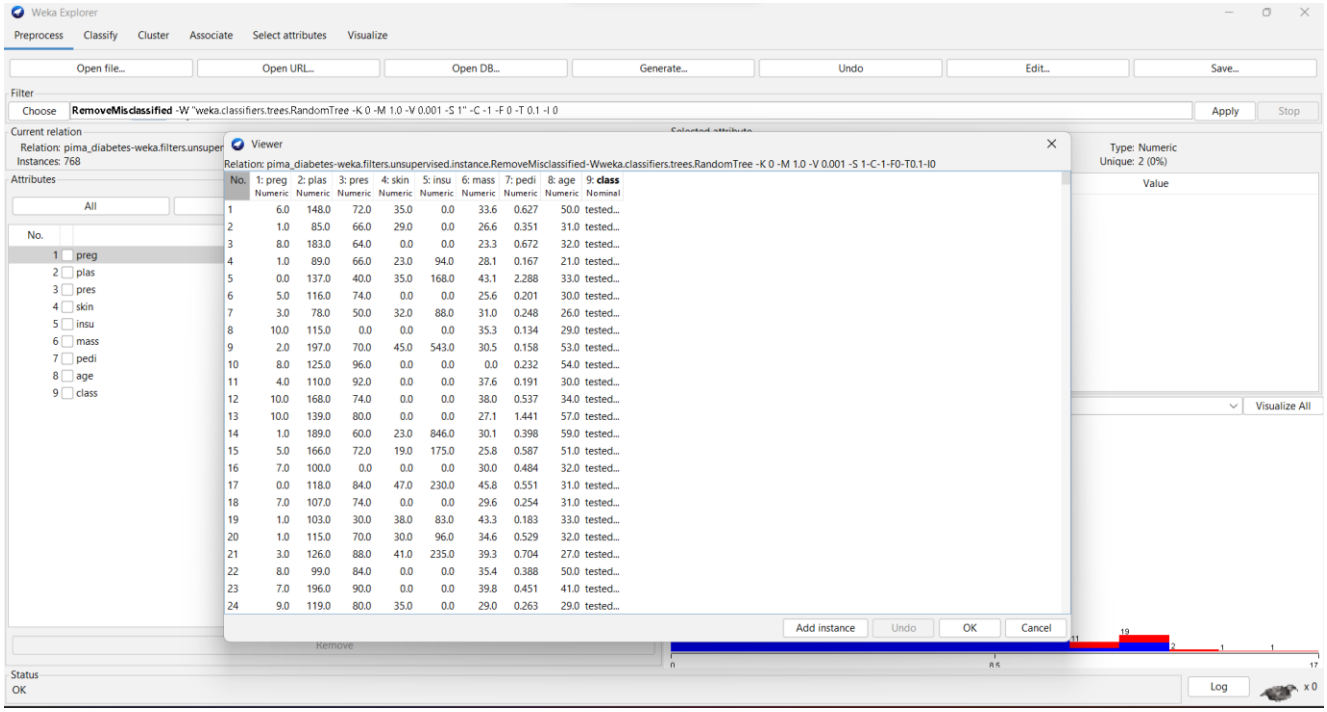


after removing indices 3,7 and 9

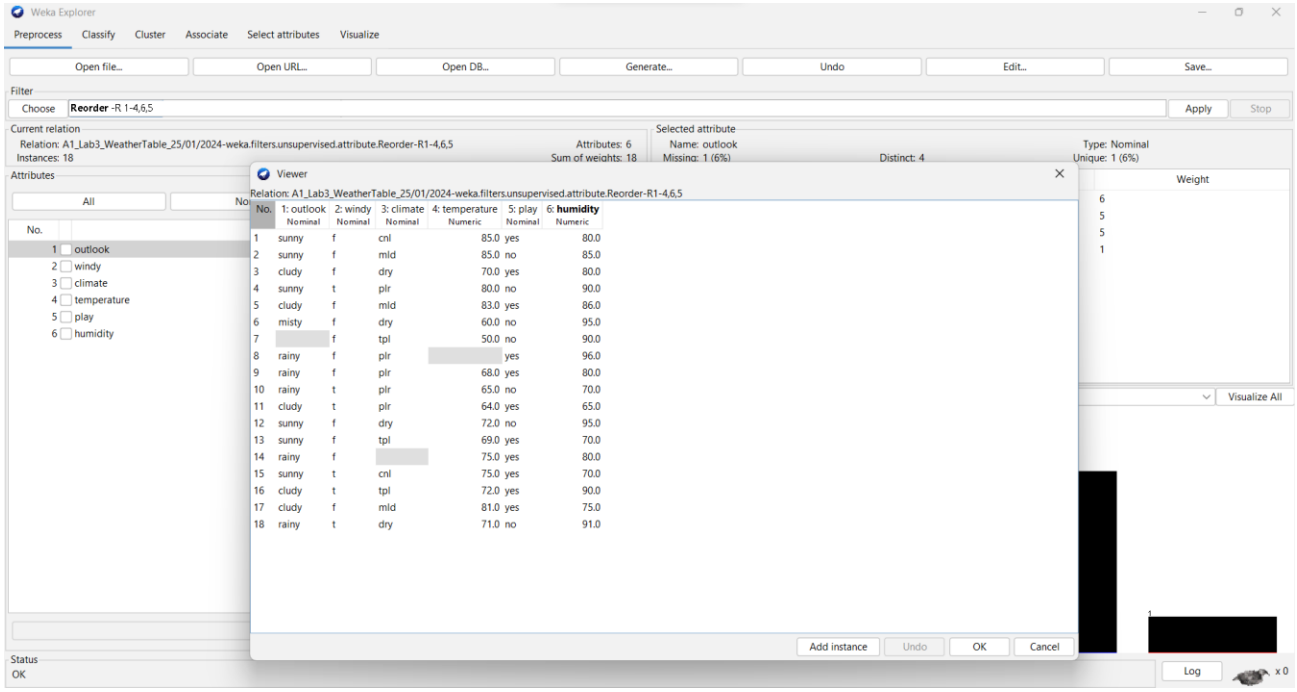
3. Determine two most frequent values with attribute outlook and retain it and filter all other remaining instances.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
RemoveFrequentValues -C 1 -N 2
Apply
Stop
Current relation
Relation: A1_Lab3_WeatherTable_25/01/2024-w
Instances: 12
Attributes
All
No.
1 outlook
2 windy
3 climate
4 temperature
5 humidity
6 play
Remove
Viewer
Relation: A1_Lab3_WeatherTable_25/01/2024-weka.filters.unsupervised.instance.RemoveFrequentValues-C1-N2
No. 1: outlook Nominal 2: windy Nominal 3: climate Nominal 4: temperature Numeric 5: humidity Numeric 6: play Nominal
1 sunny f cnl 85.0 80.0 yes
2 sunny f mld 85.0 85.0 no
3 sunny t plr 80.0 90.0 no
4 f tpl 50.0 90.0 no
5 rainy f plr 96.0 yes
6 rainy f plr 68.0 80.0 yes
7 rainy t plr 65.0 70.0 no
8 sunny f dry 72.0 95.0 no
9 sunny f tpl 69.0 70.0 yes
10 rainy f 75.0 80.0 yes
11 sunny t cnl 75.0 70.0 yes
12 rainy t dry 71.0 91.0 no
Add instance
Undo
OK
Cancel
Type: Nominal
Unique: 0 (0%)
Weight
6
5
0
0
Visualize All
Status
OK
Log
x 0

4. Find out classifiers in weka.classifier.tree with maximum RemoveMissclassified and minimum RemoveMissclassified instances (use diabetes.arff dataset).



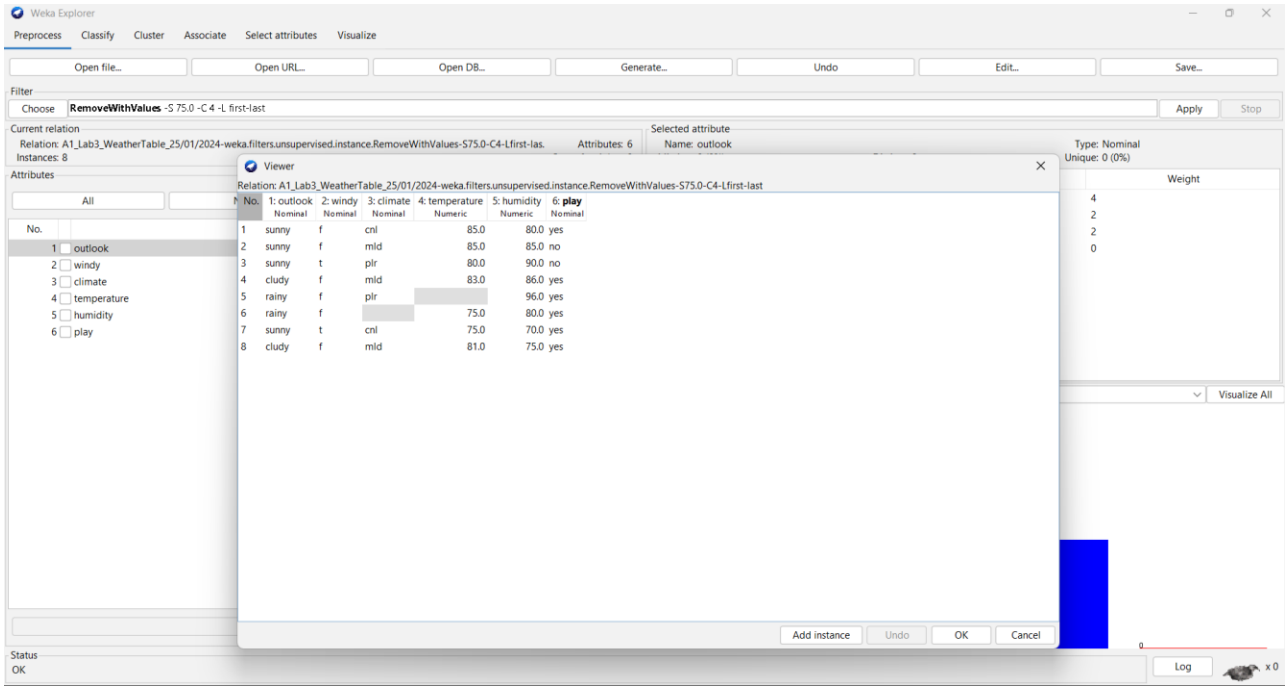
5. Reorder attributes such a way that humidity act as class attribute. (hint: By default last attribute treated as class attribute and use reorder unsupervised attribute filter)



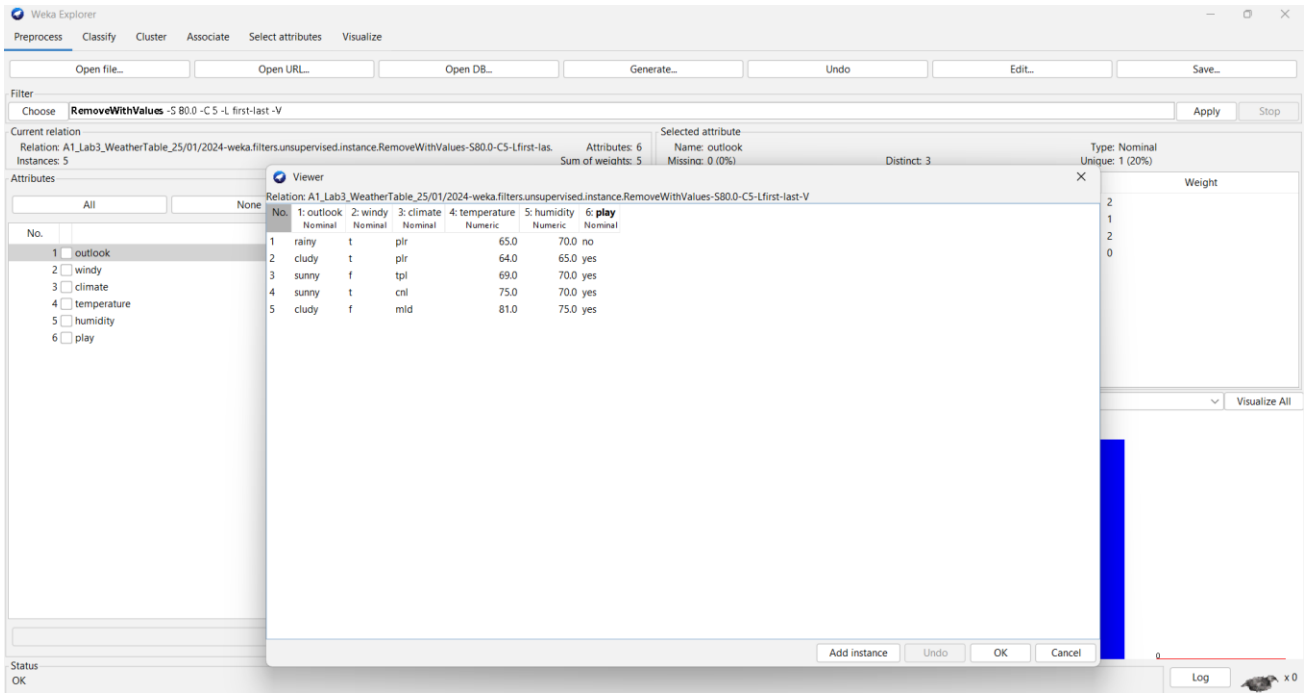
Q2. Load the weather.numeric dataset. Use the filter weka.unsupervised.instance and do analyze following preprocessing operations. From the Object Editor window, figure out changes appropriately after each preprocessing. Undo the change to the dataset that you just performed, and verify that the data has reverted to its original state.

1. Filter the instances of the following expressions:

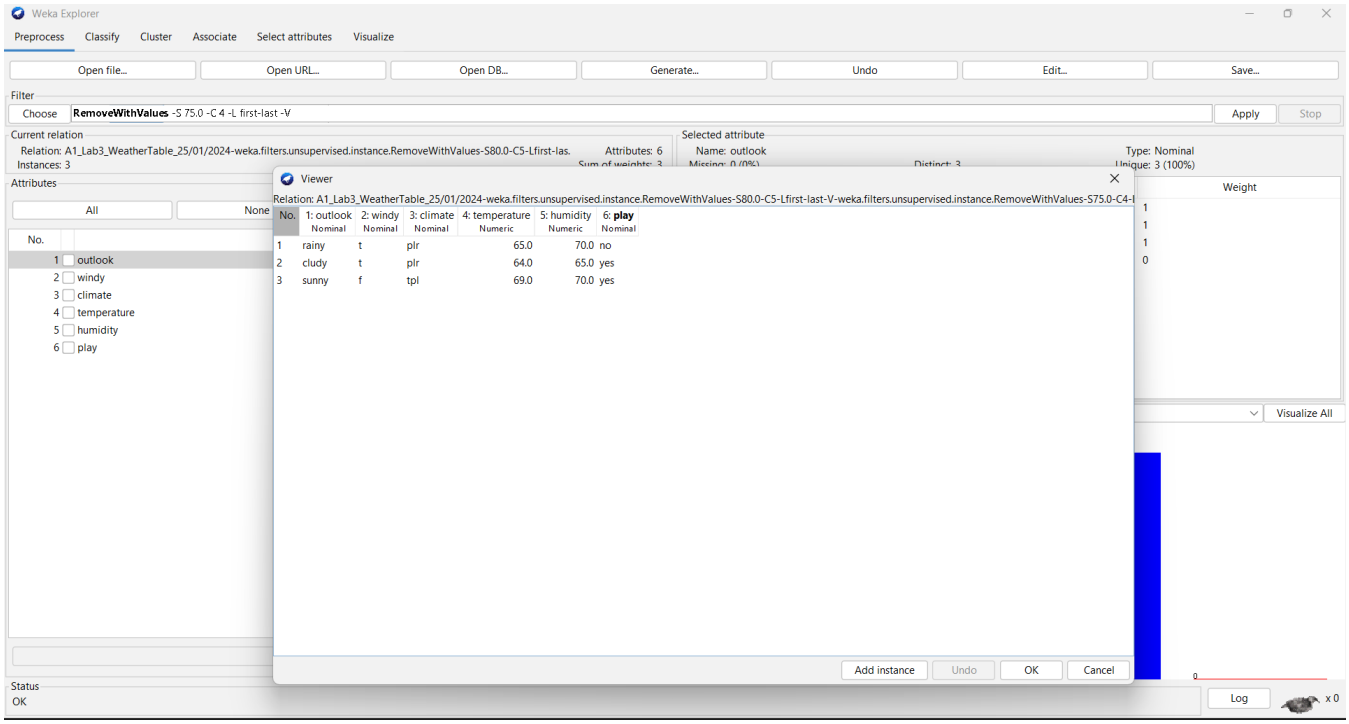
a. Temperature >=75



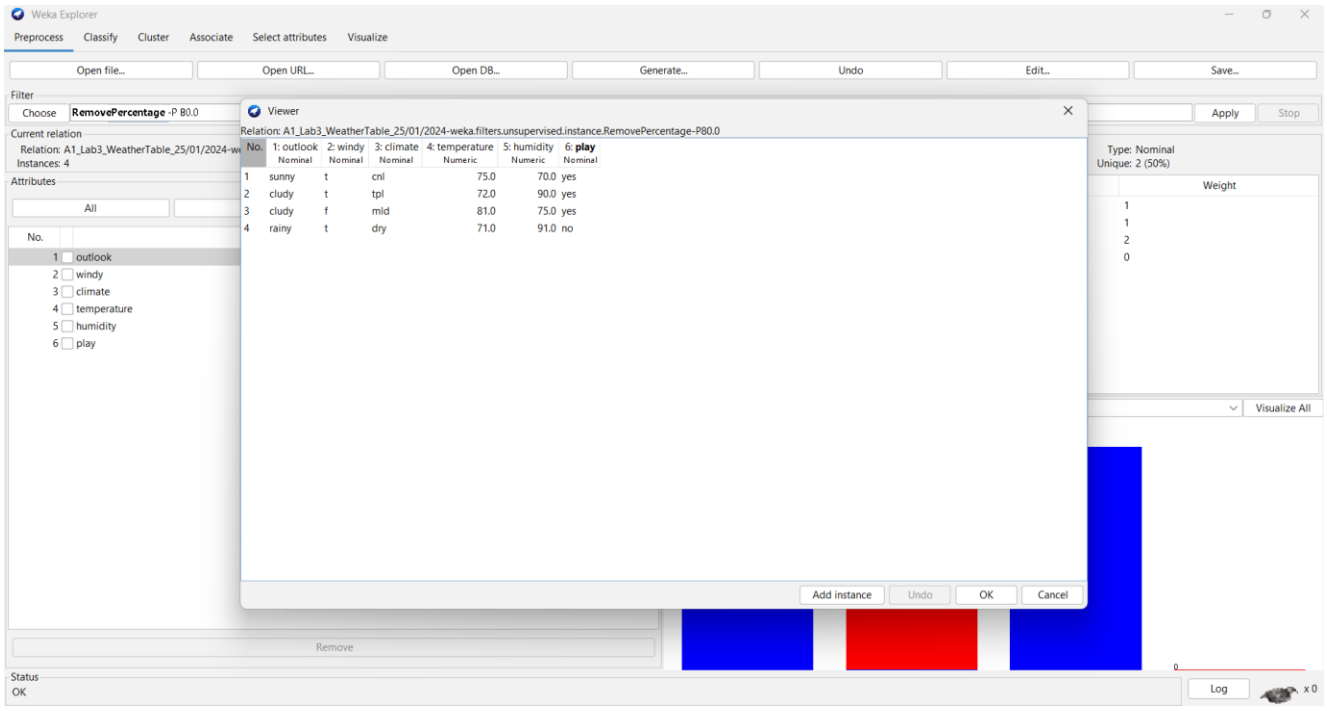
b. Humidity<=80



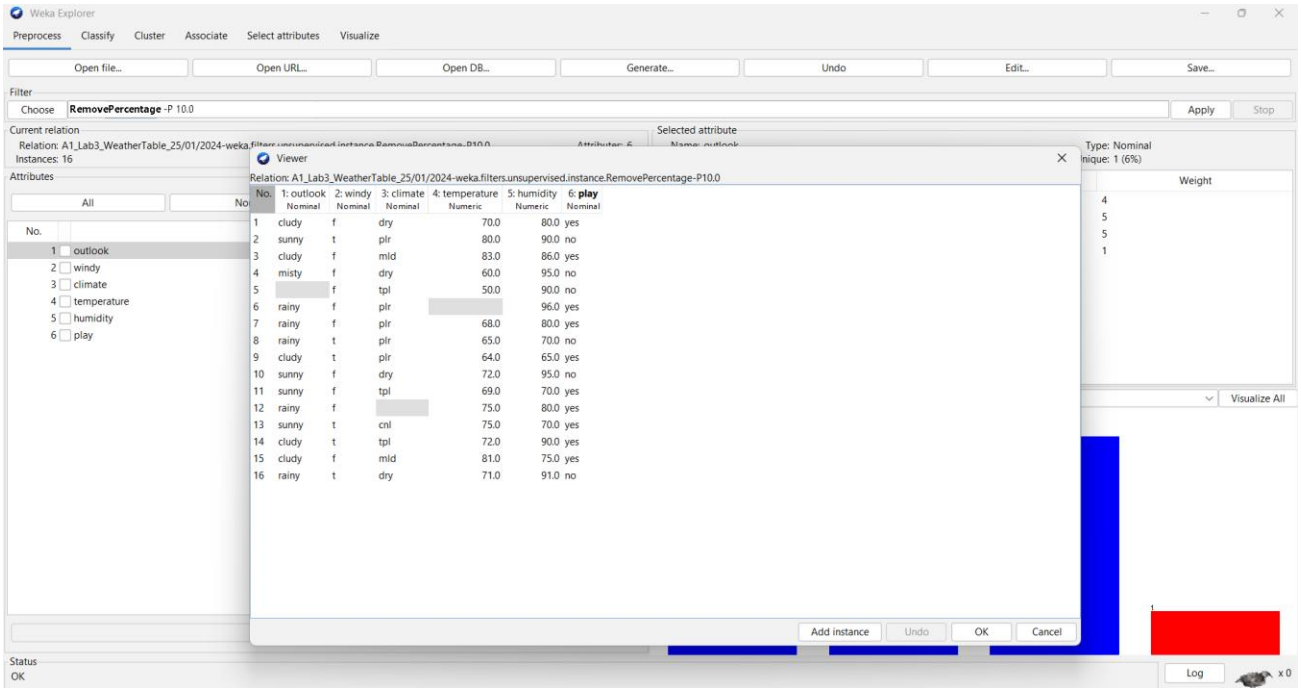
c. Temperature >=75 and Humidity<=80



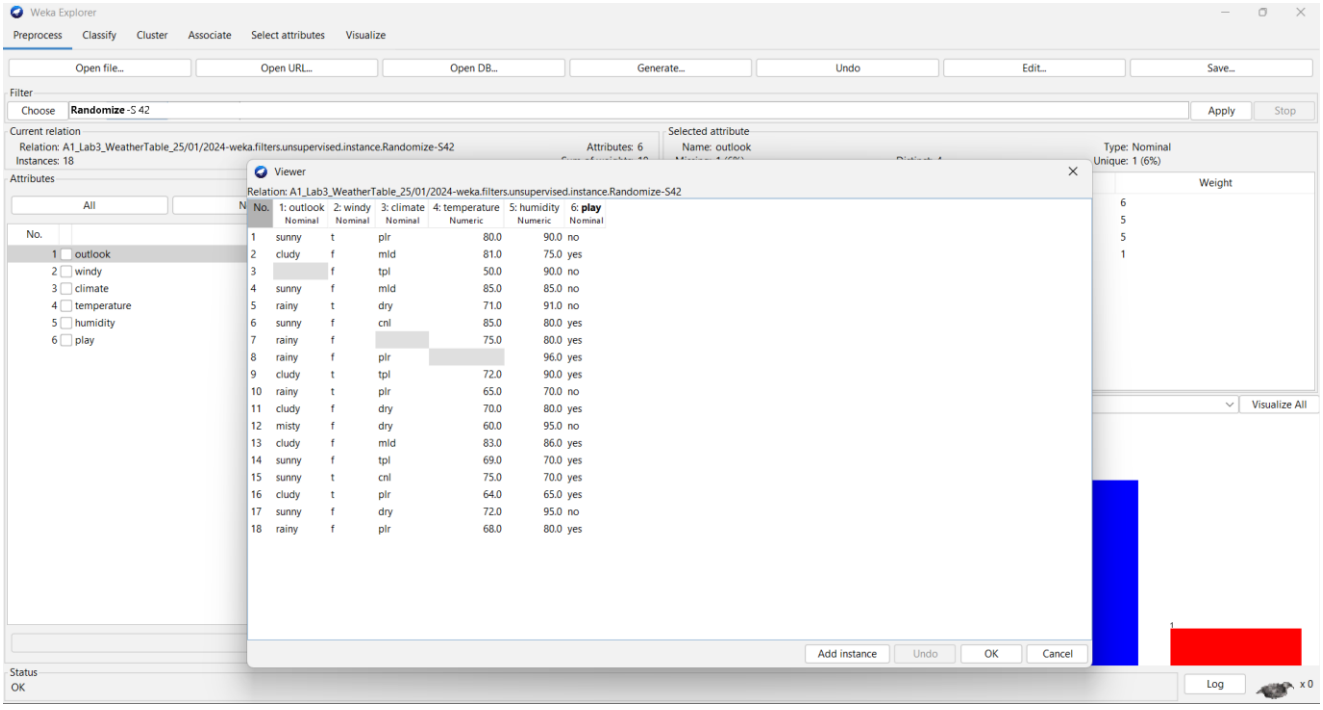
2. Filter 80% samples of the instances.



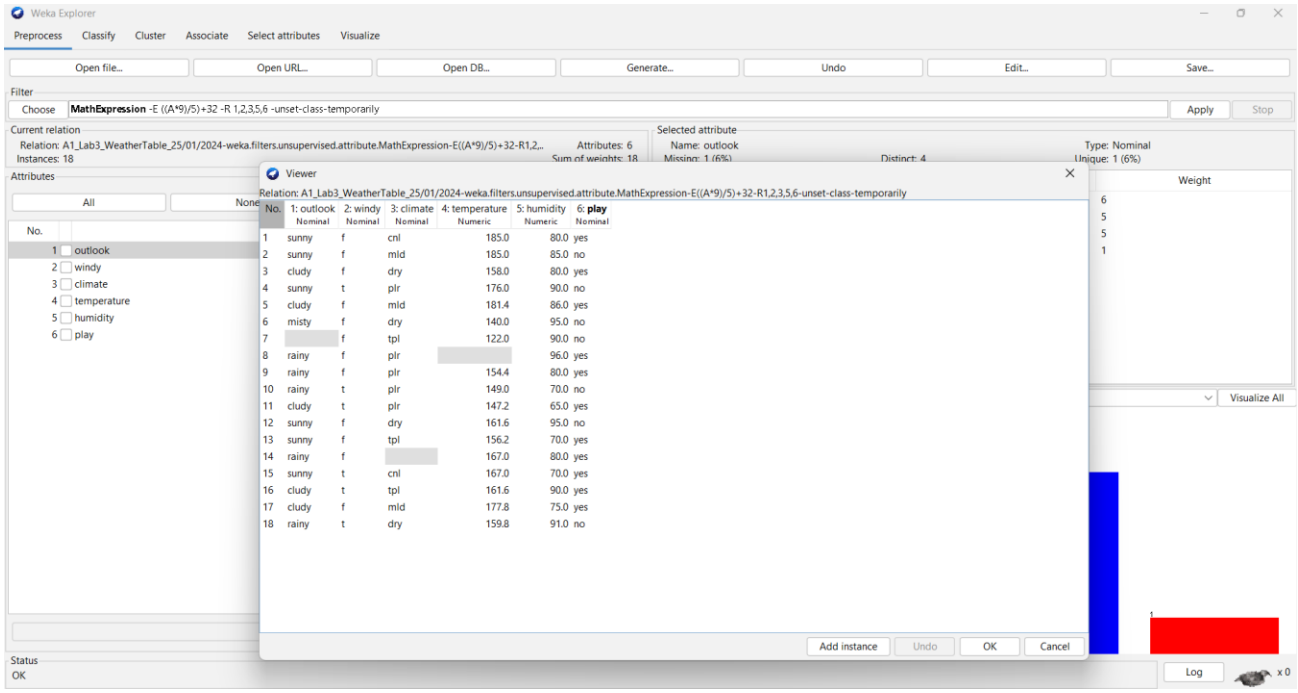
3. Remove 10% instances from dataset.



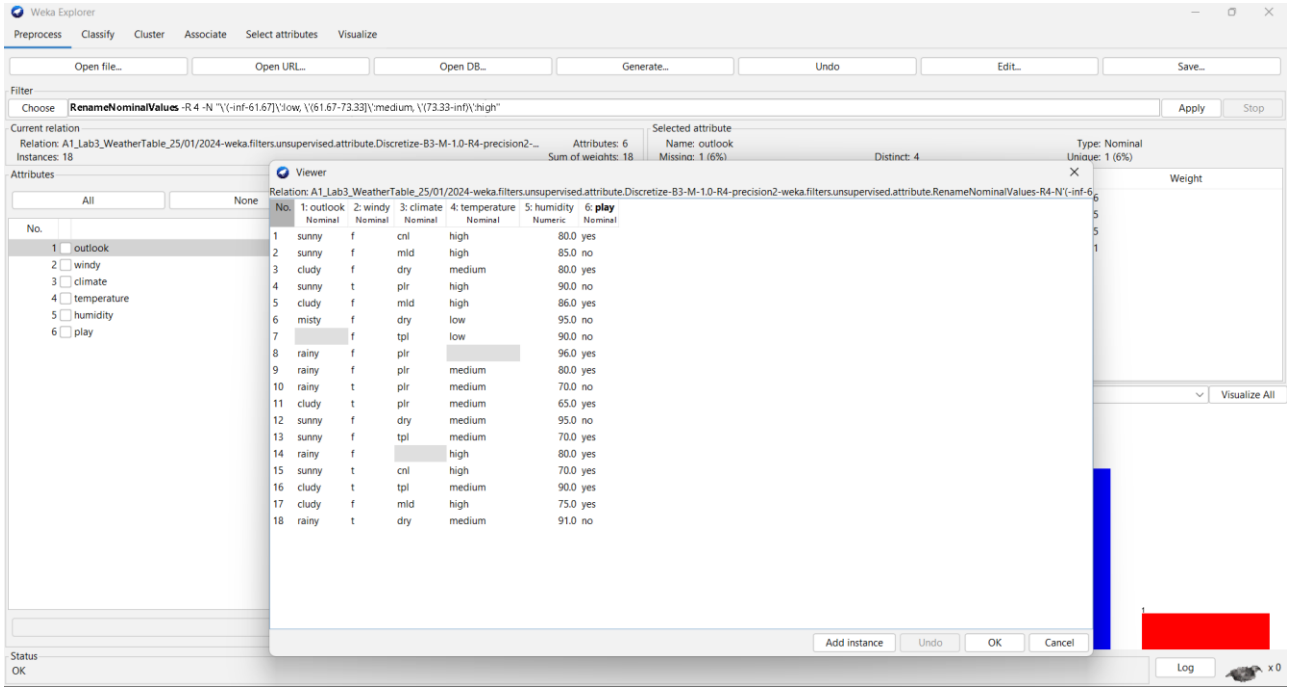
4. Reshuffle the original order of instances.



5. Convert temperature attribute in degree Fahrenheit.

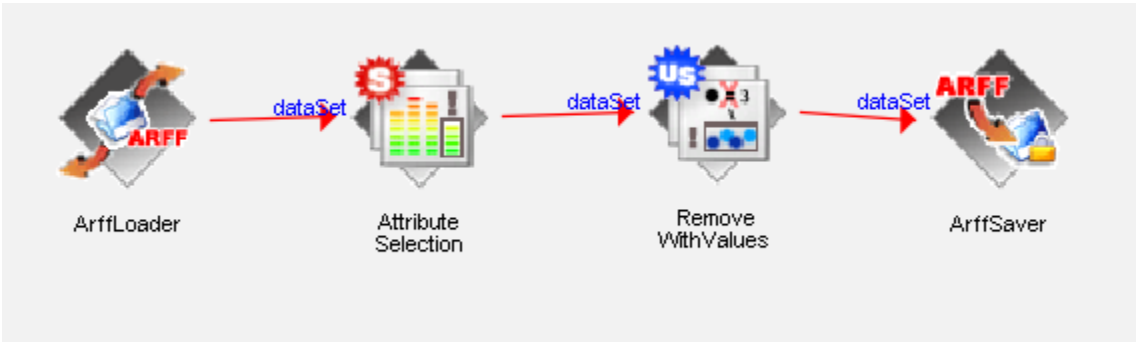


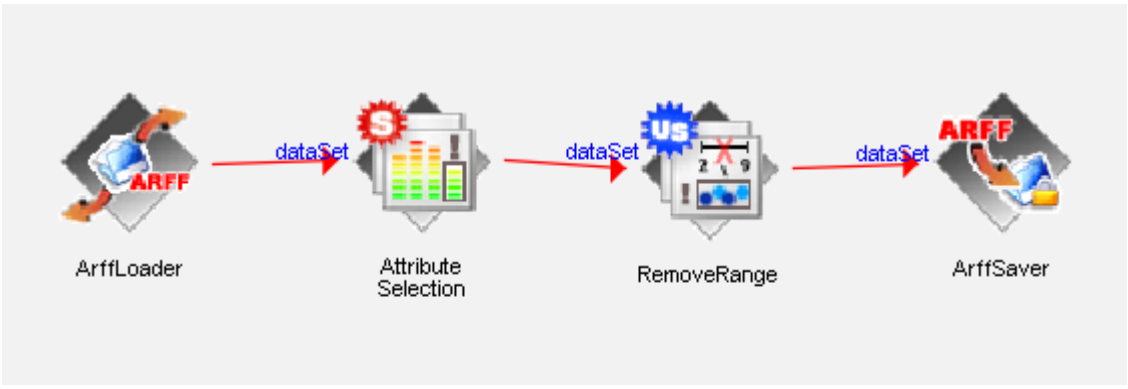
6. Convert temperature in three nominal values (low, medium and high). Use data discretization.

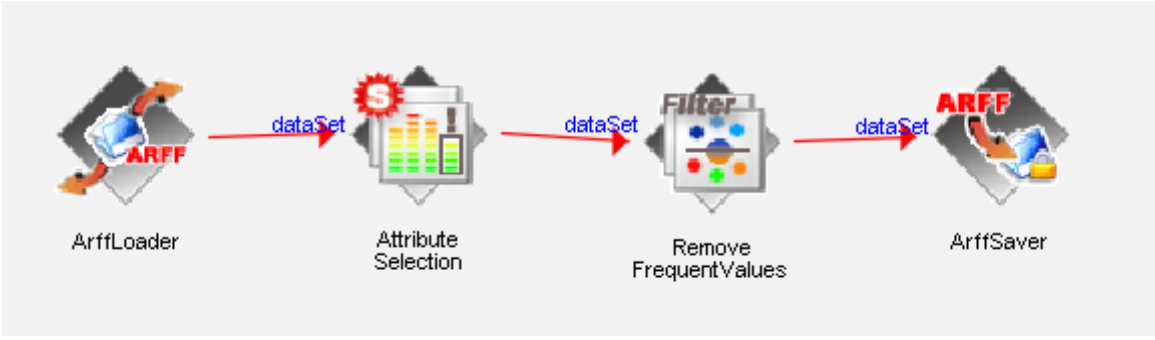


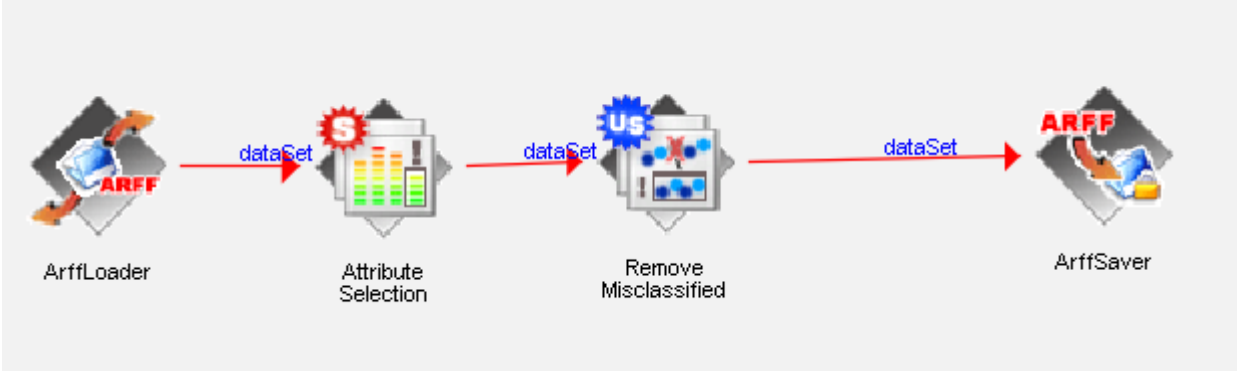
Q3. Design a knowledge flow network for the above given questions 1 & 2 and analyze your result. Use diabetes as target dataset.

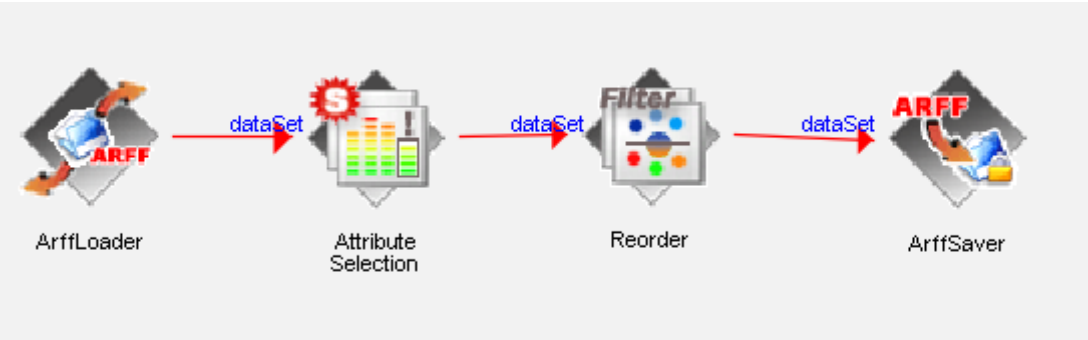
Q1



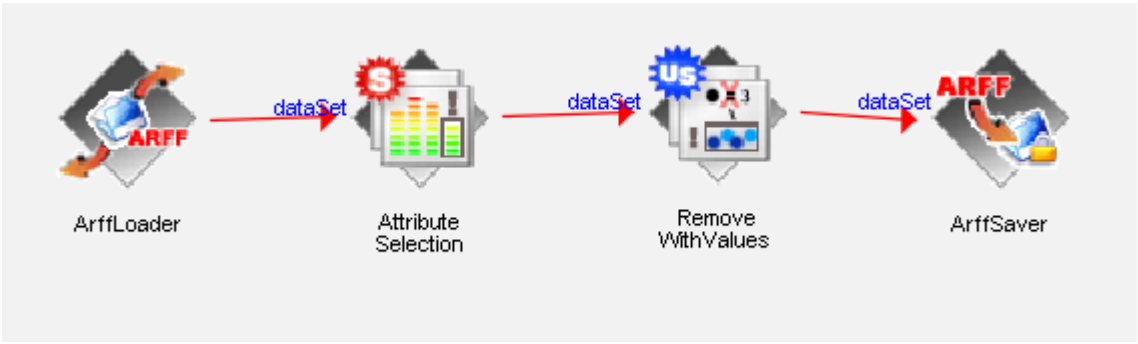


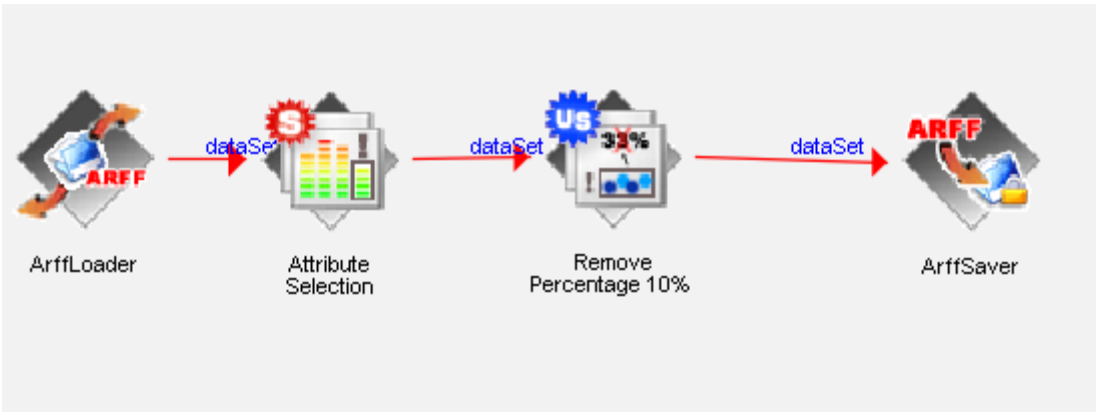


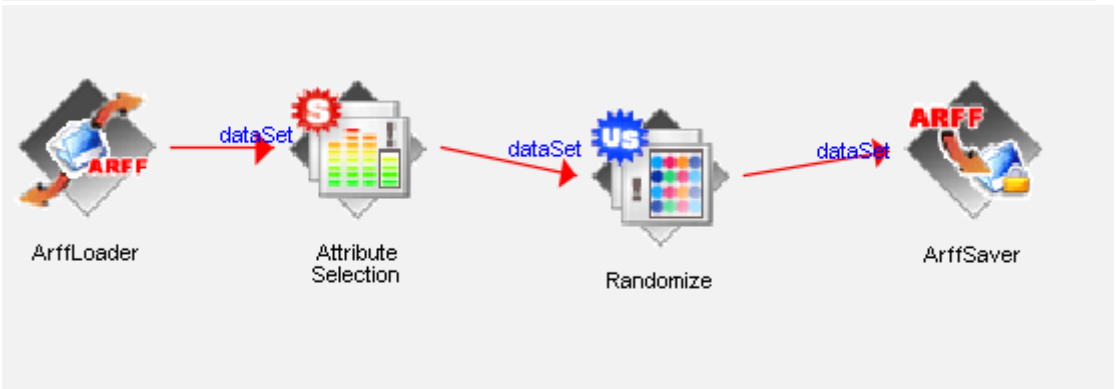


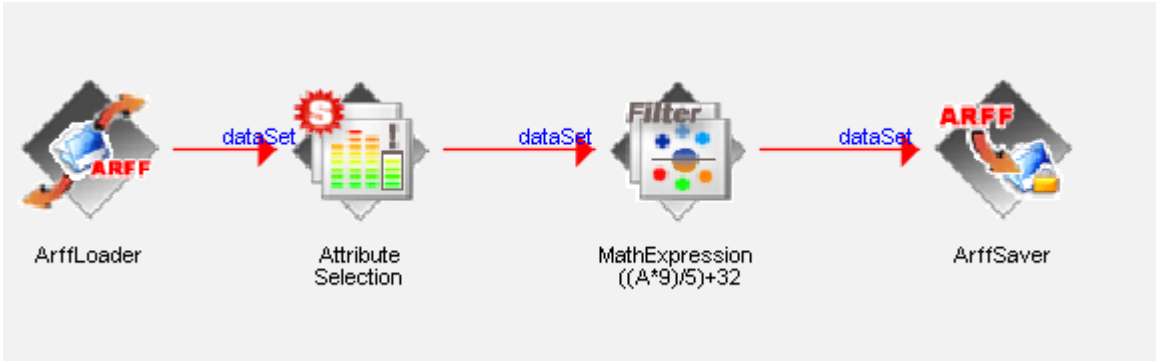

ARFF
dataSet
dataSet
Filter
dataSet
ARFF
ArffLoader
Attribute Selection
Reorder
ArffSaver

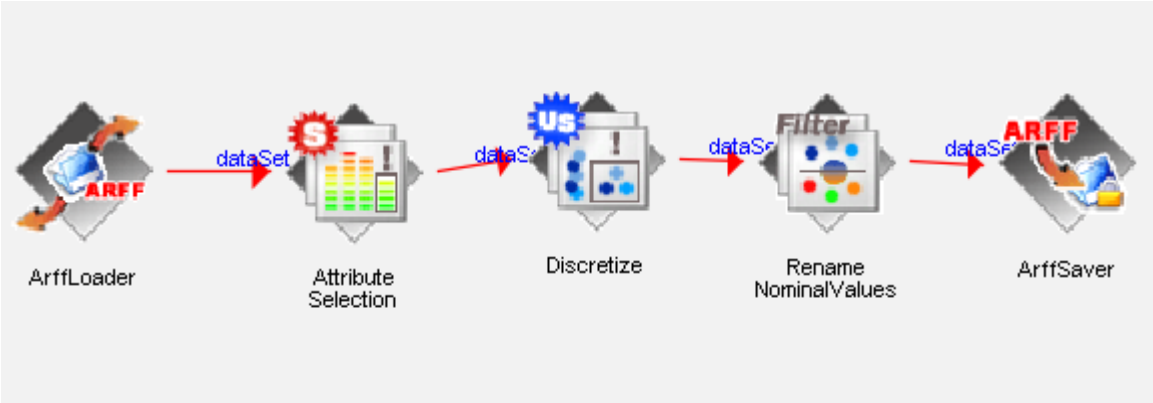
Q2


ARFF
dataSet
dataSet
dataSet
ARFF
ArffLoader
Attribute Selection
Remove WithValues
ArffSaver


ARFF
dataSet
dataSet
dataSet
ARFF
ArffLoader
Attribute Selection
Remove Percentage 10%
ArffSaver


ARFF
dataSet
dataSet
dataSet
ARFF
ArffLoader
Attribute Selection
Randomize
ArffSaver


ARFF
dataSet
dataSet
Filter
dataSet
ARFF
ArffLoader
Attribute Selection
MathExpression ((A*9)/5)+32
ArffSaver


ARFF
dataSet
dataSe
US
dataSe
Filter
dataSe
ARFF
ArffLoader
Attribute Selection
Discretize
Rename NominalValues
ArffSaver