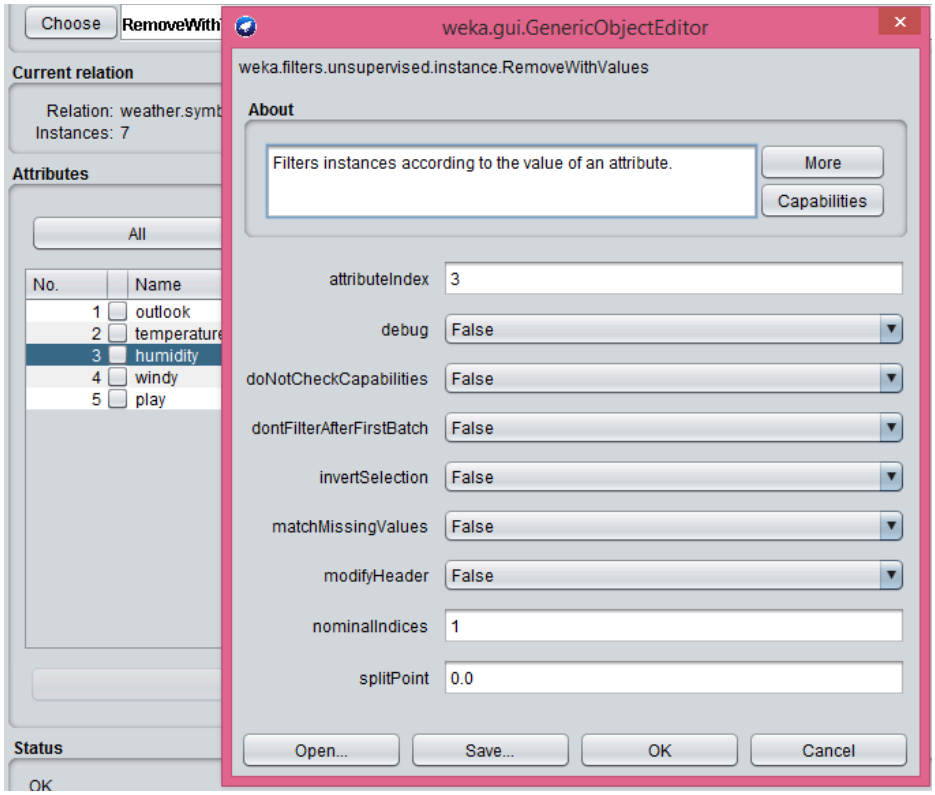
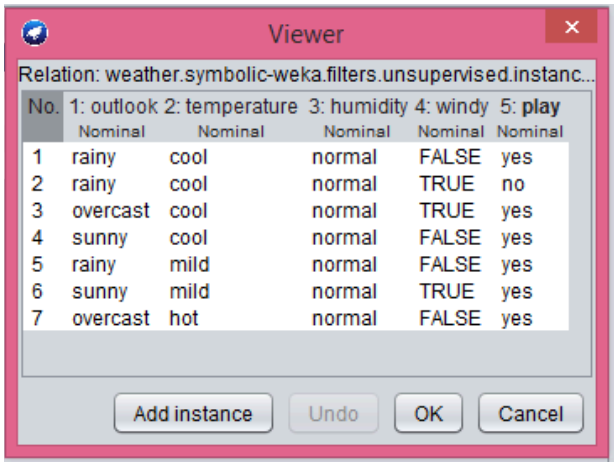
Assignment 2

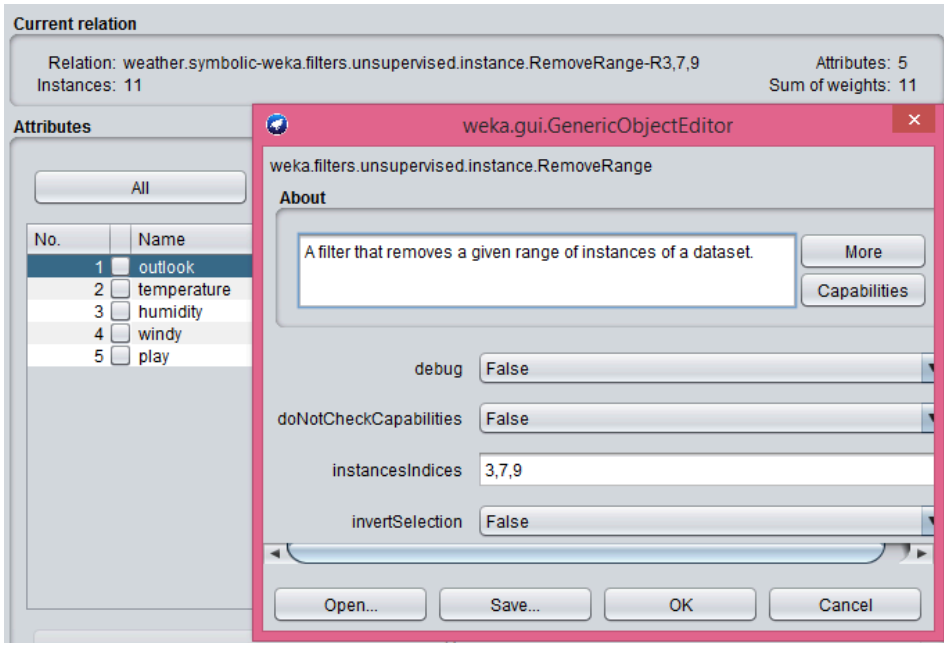
Q1. Load the weather.nominal dataset. Use the filter weka.unsupervised.instance and do analyze following preprocessing operations. From the Object Editor window, figure out changes appropriately after each preprocessing.

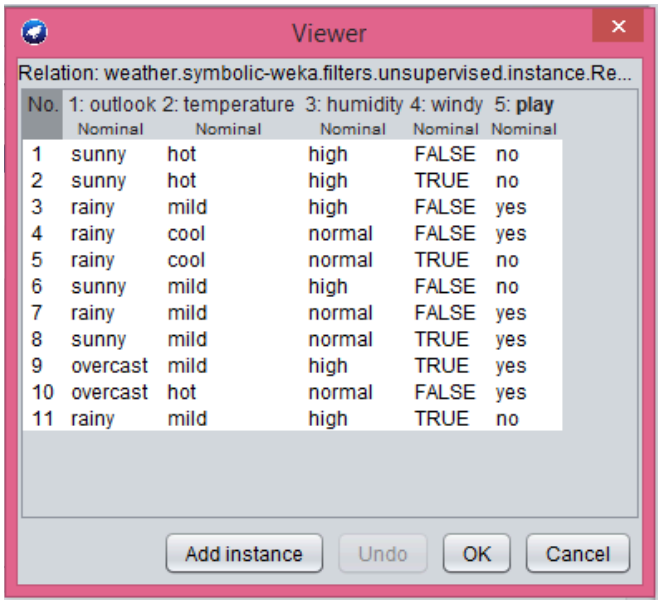
1. Remove all instances in which the humidity attribute has the value high.



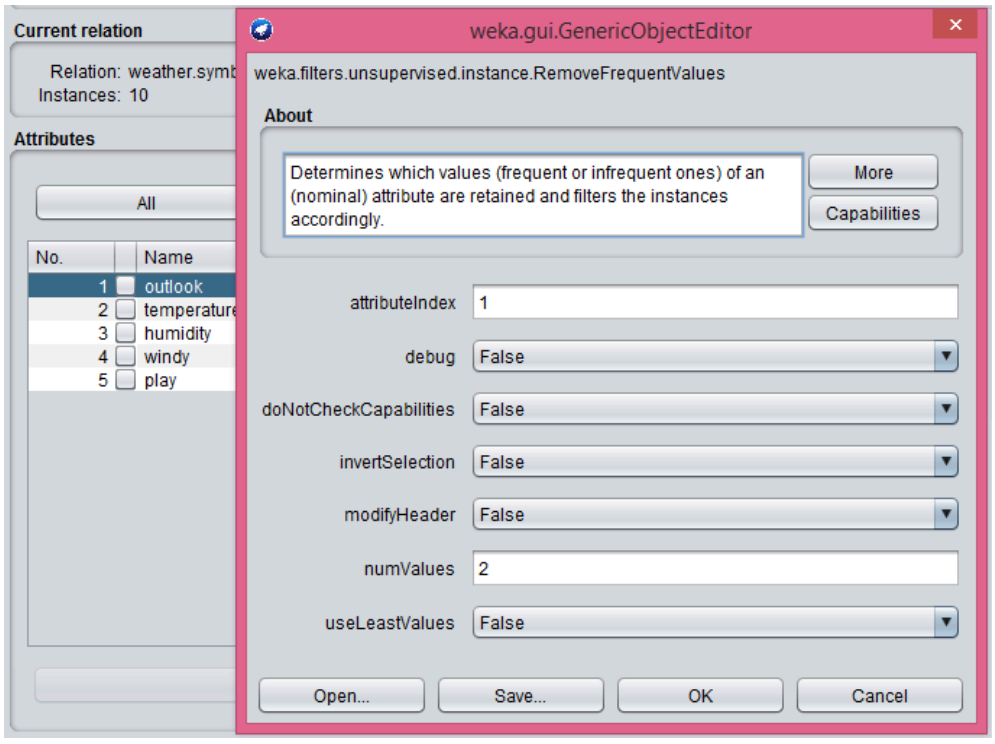


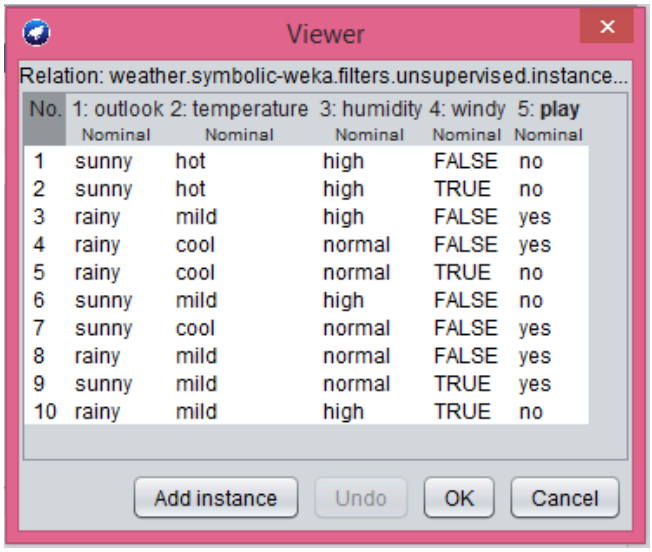
2. Remove instances with indices number 3, 7 and 9.



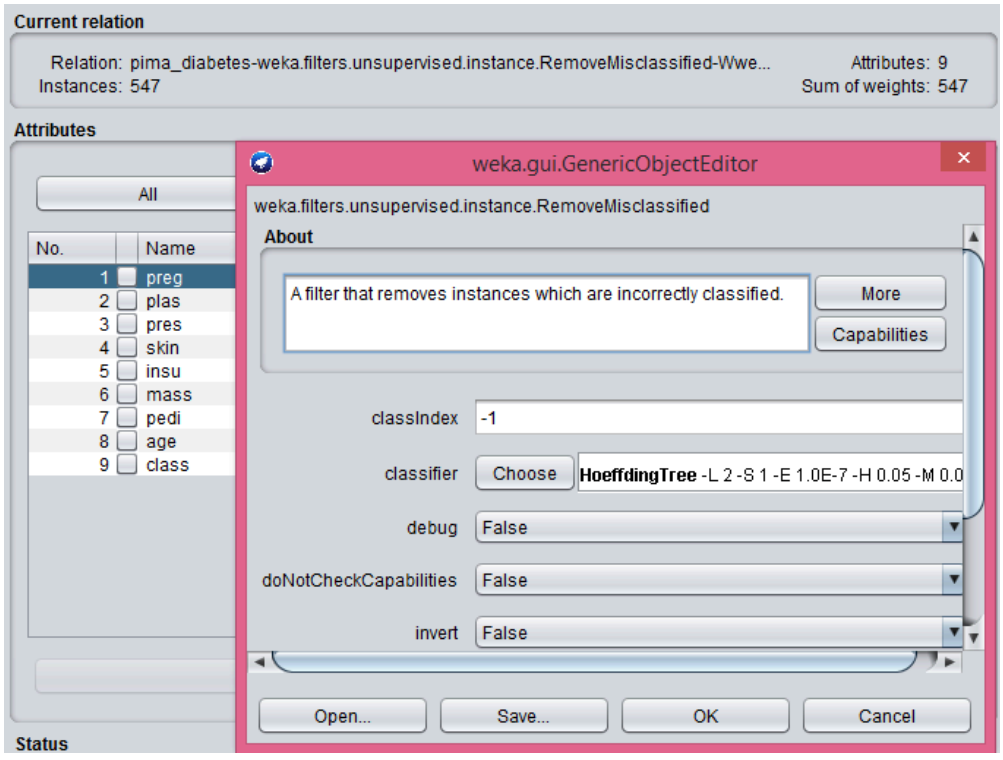


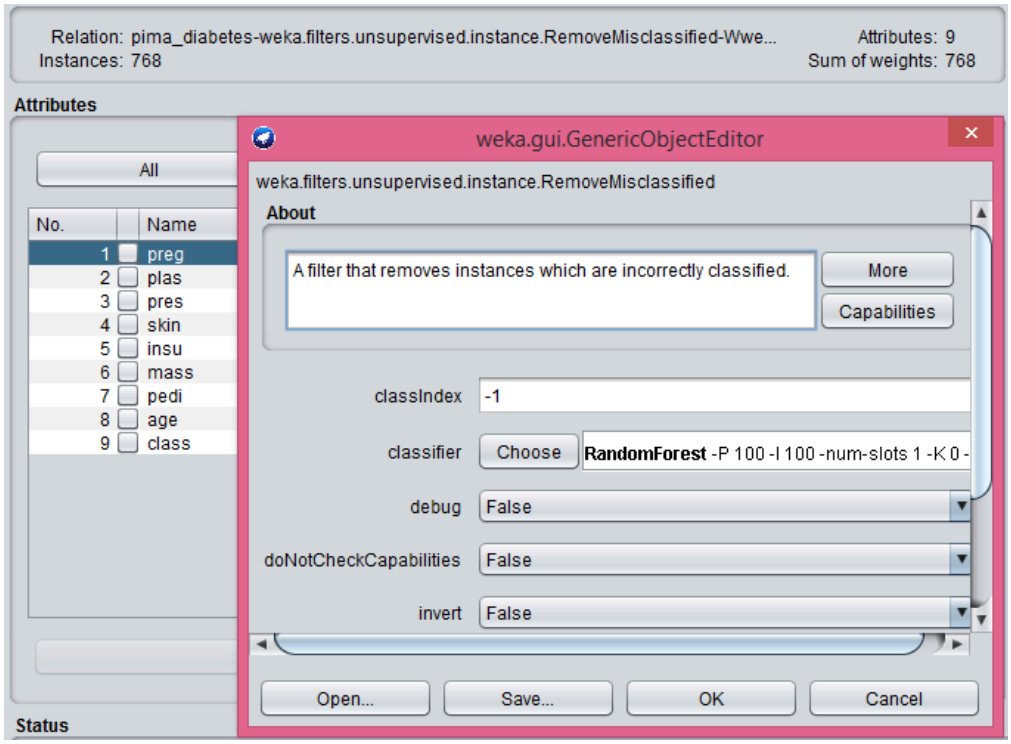
3. Determine two most frequent values with attribute outlook and retain it and filter all other remaining instances.



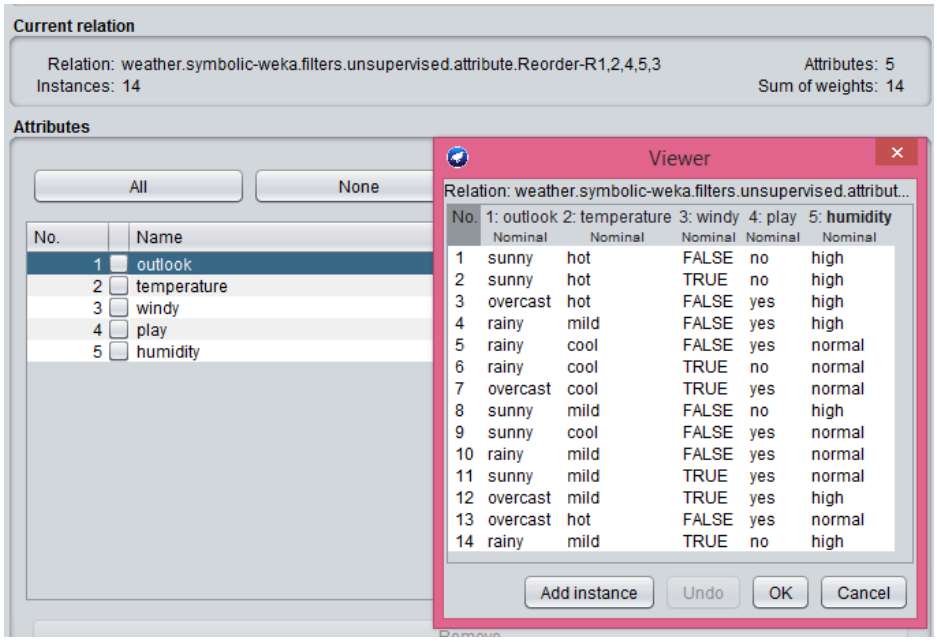


4. Find out classifiers in weka.classifier.tree with maximum RemoveMissclassified and minimum RemoveMissclassified instances (use diabetes.arff dataset).



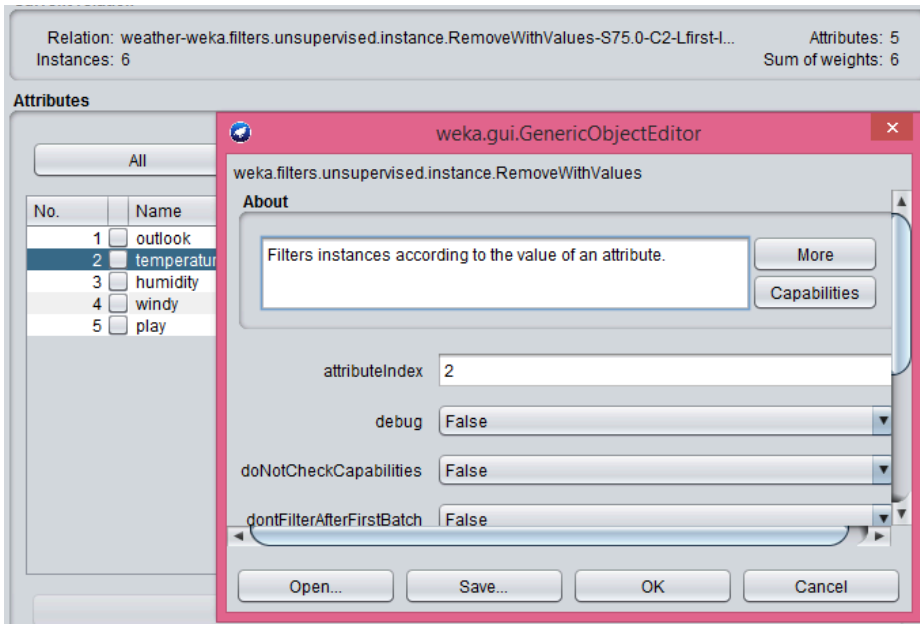


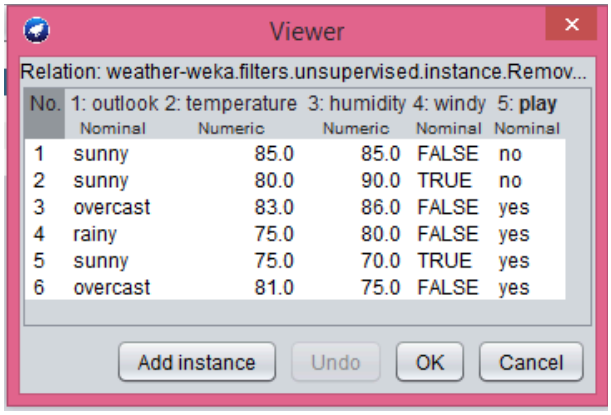
5. Reorder attributes such a way that humidity act as class attribute. (hint: By default last attribute treated as class attribute and use reorder unsupervised attribute filter)



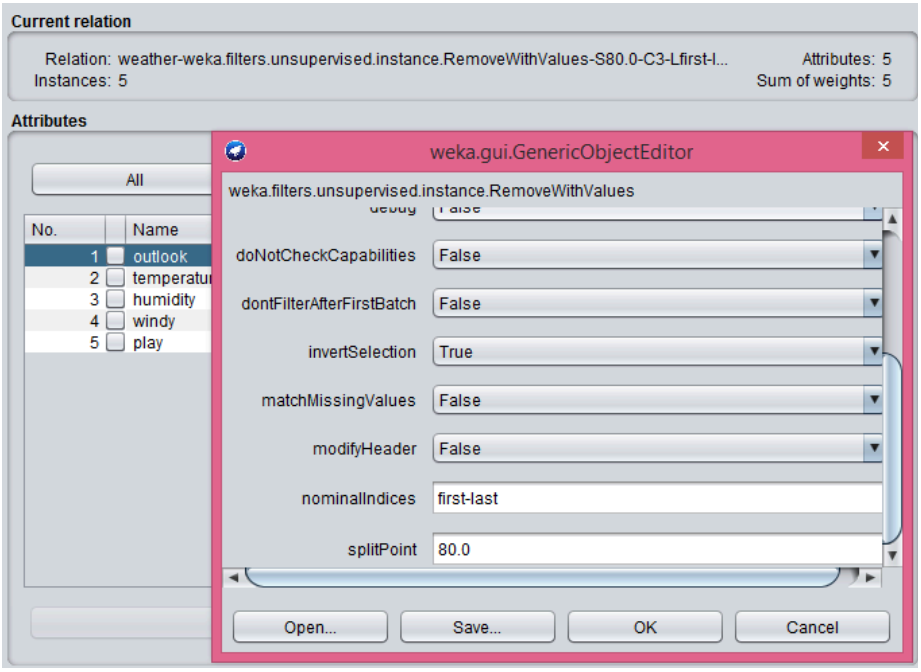
Q2. Load the weather.numeric dataset. Use the filter weka.unsupervised.instance and do analyze following preprocessing operations. From the Object Editor window, figure out changes appropriately after each preprocessing.

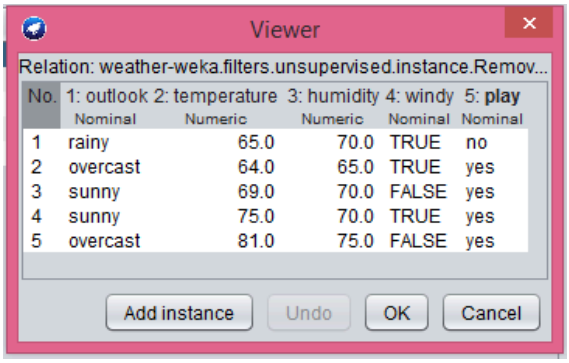
1. Filter the instances of the following expressions:
    a. Temperature >=75



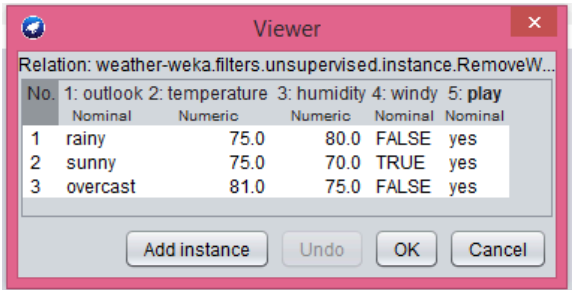


b. Humidity<=80

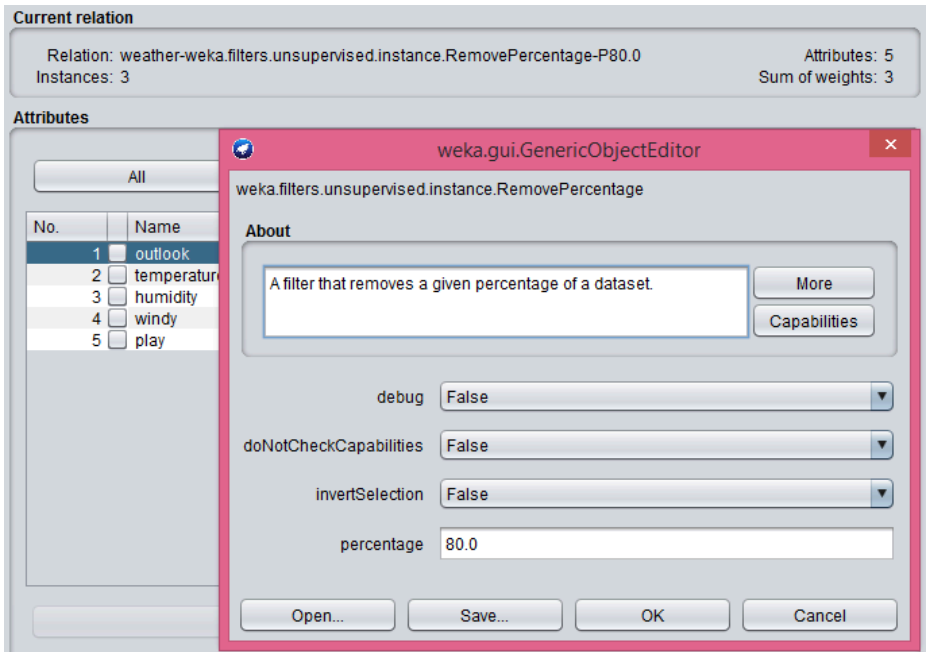


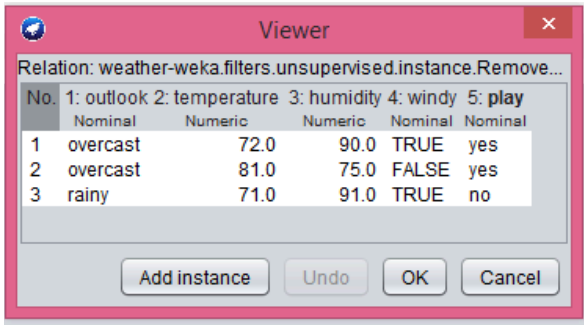


c. Temperature >=75 and Humidity<=80

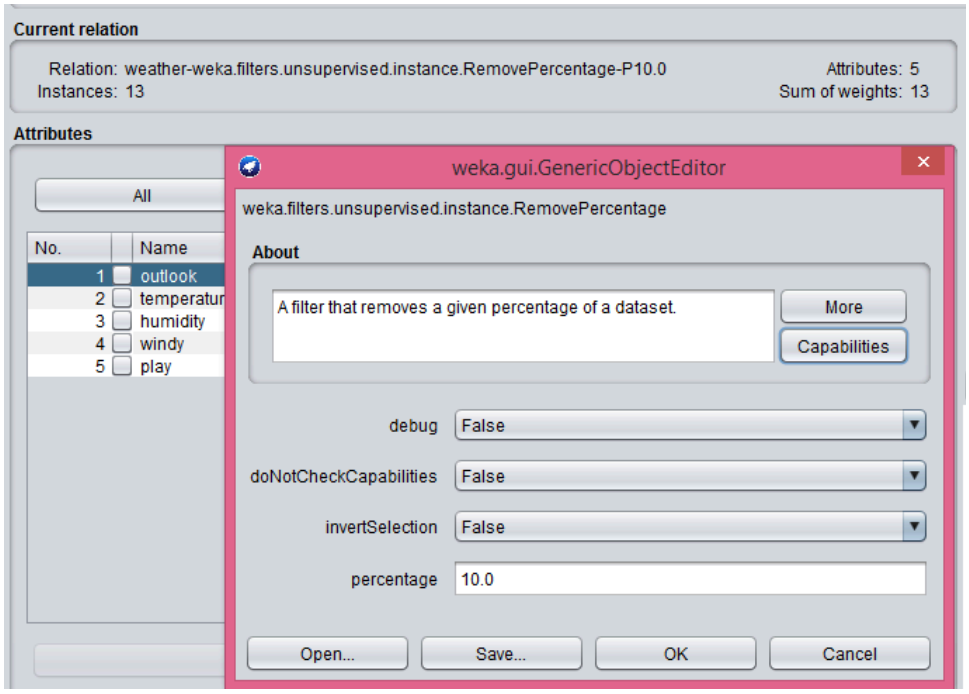


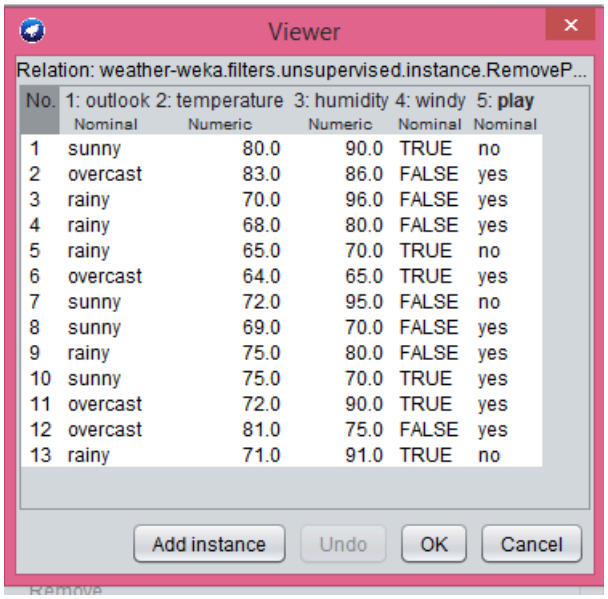
2. Filter 80% samples of the instances.



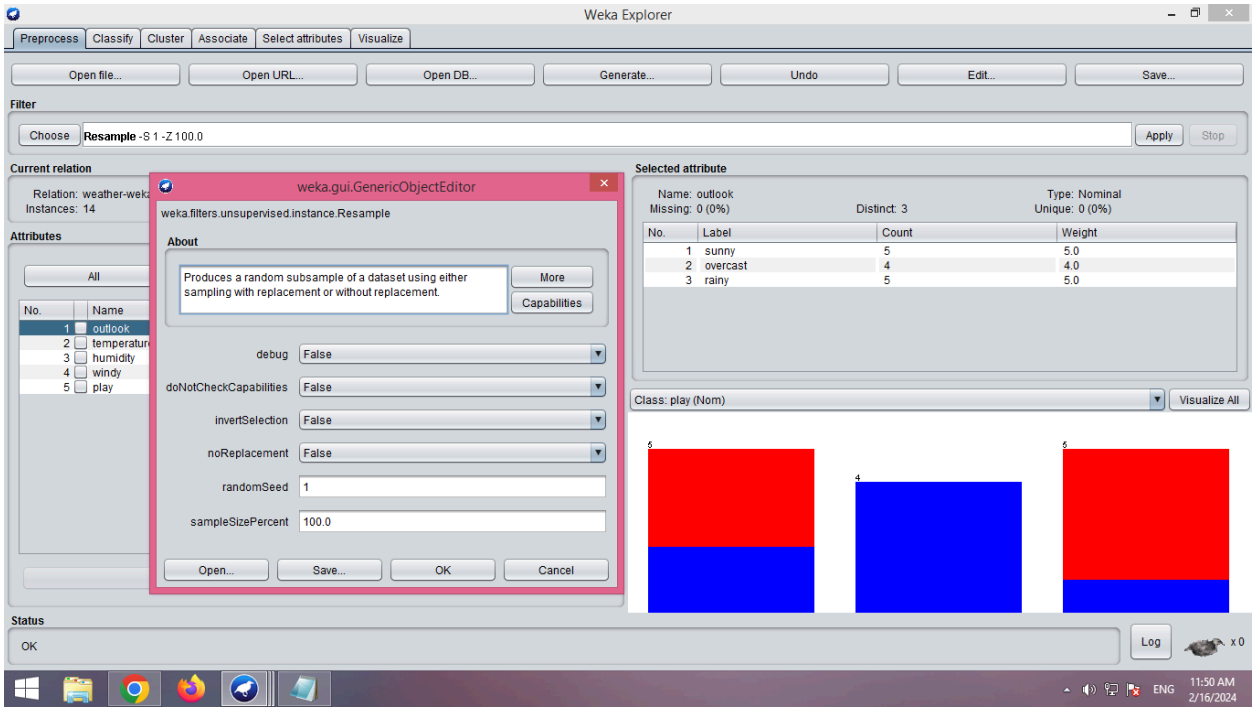


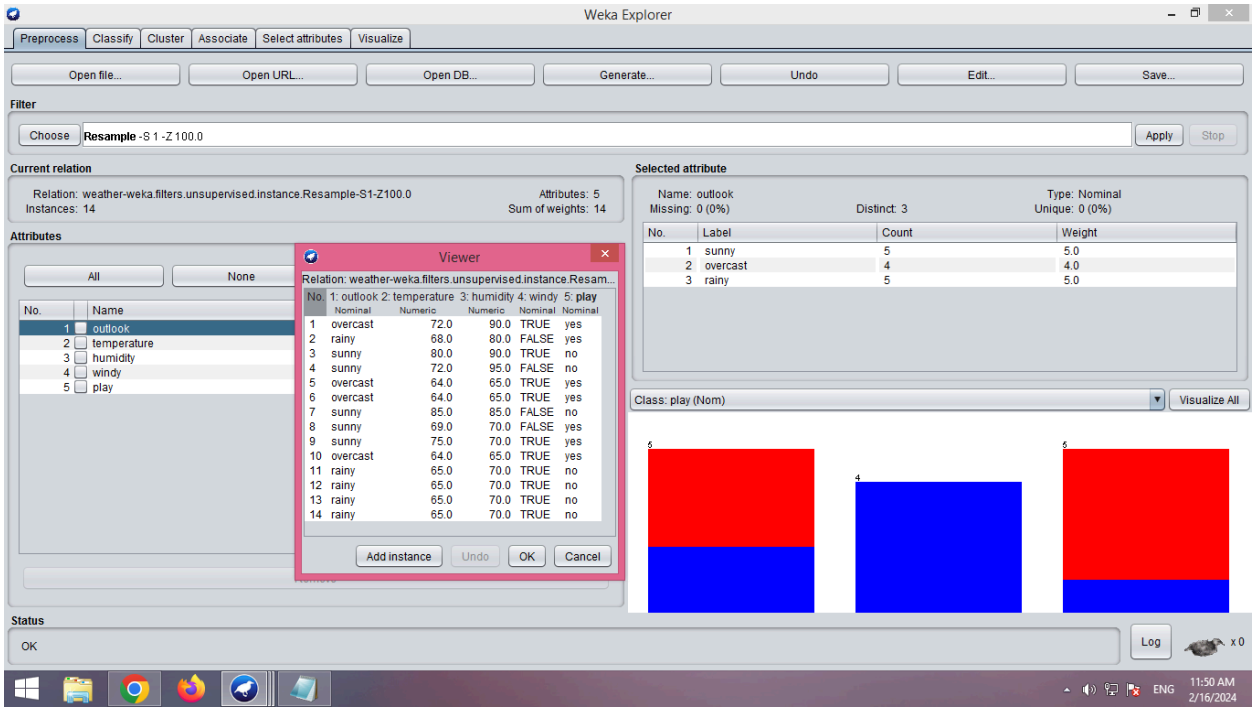
3. Remove 10% instances from dataset.



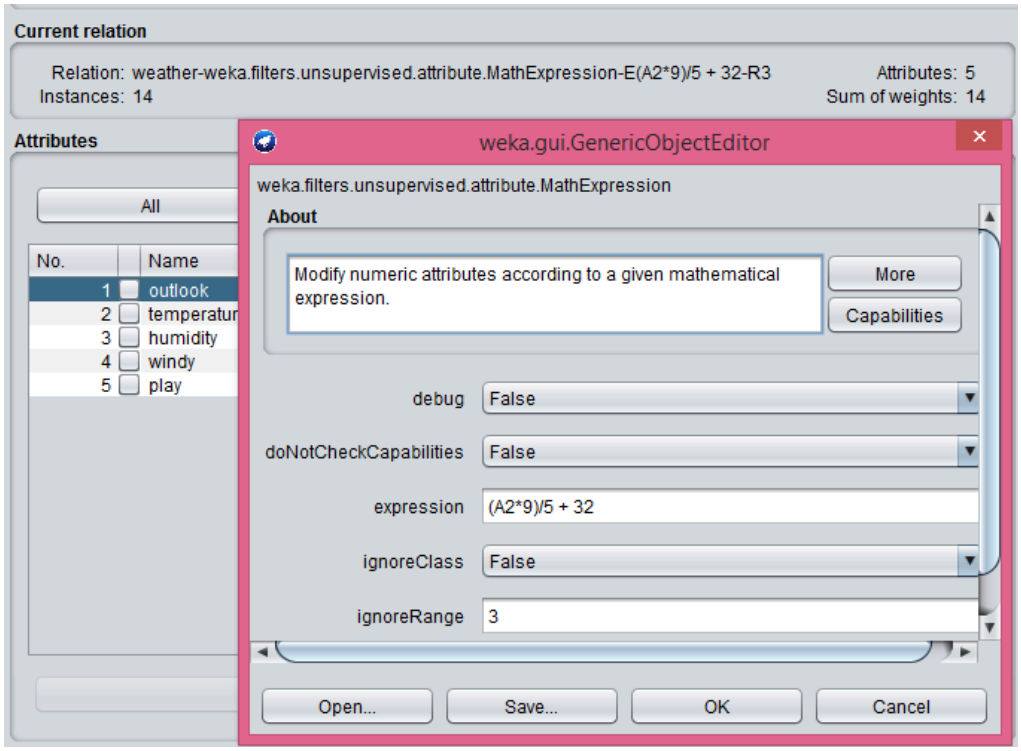


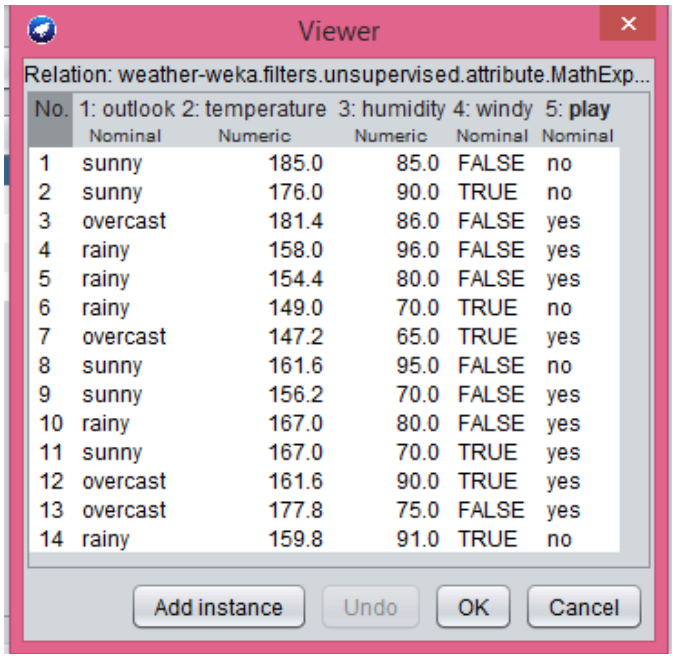
4. Reshuffle the original order of instances.



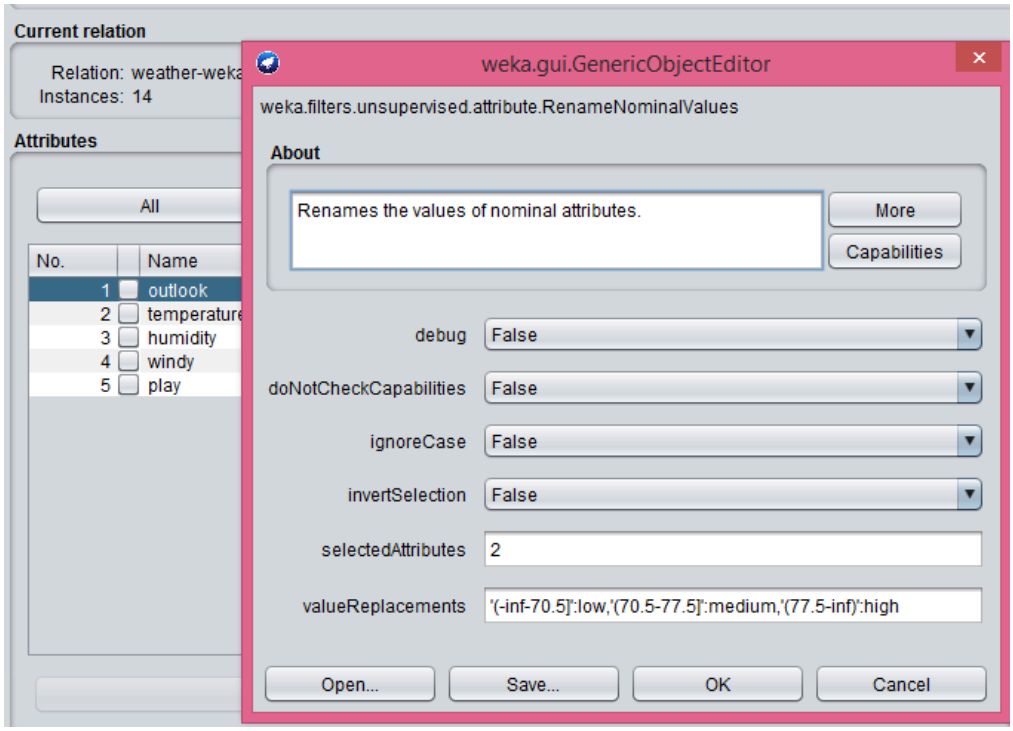


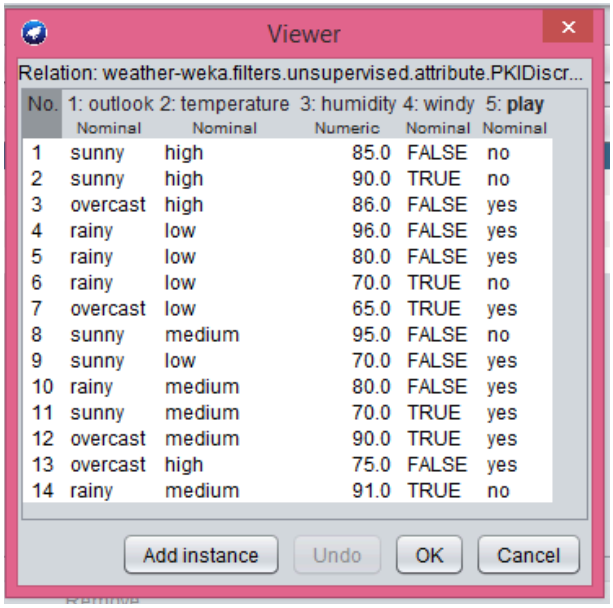
5. Convert temperature attribute in degree Fahrenheit.





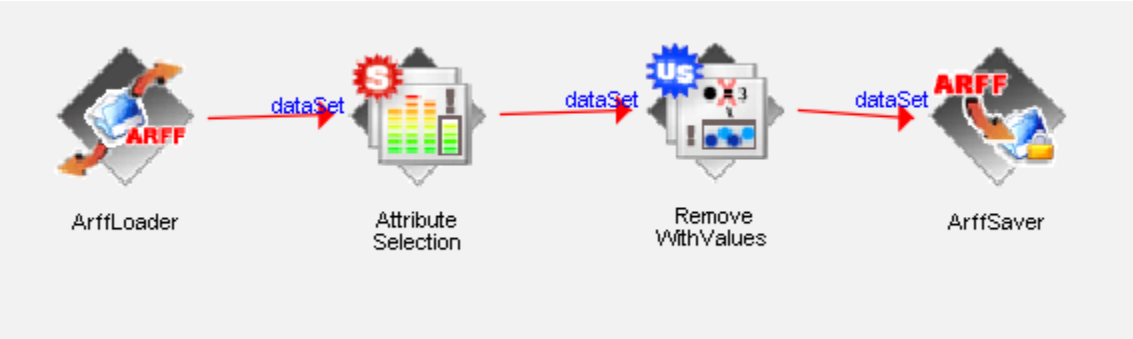
6. Convert temperature in three nominal values (low, medium and high). Use data discretization.

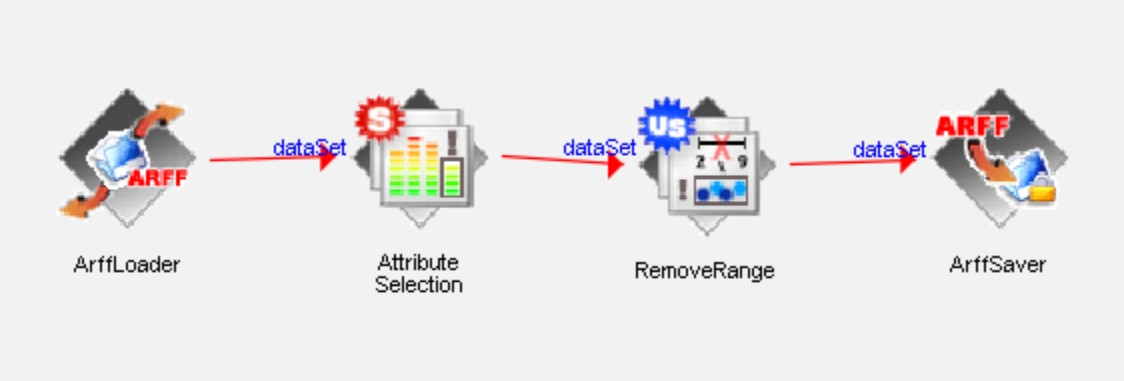


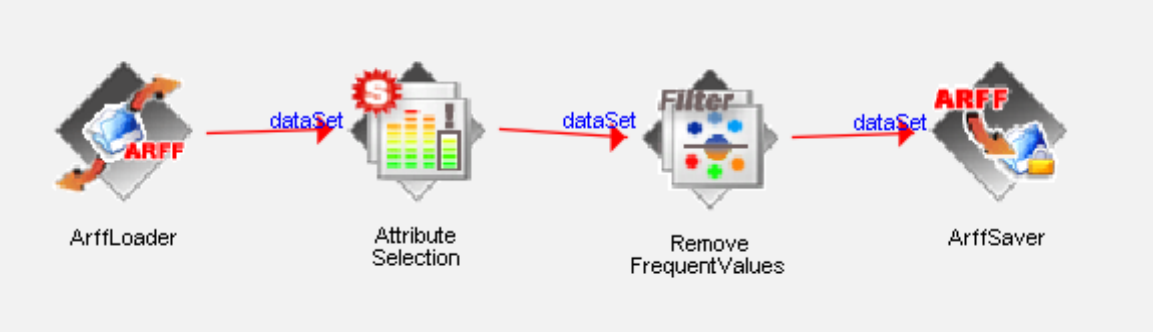


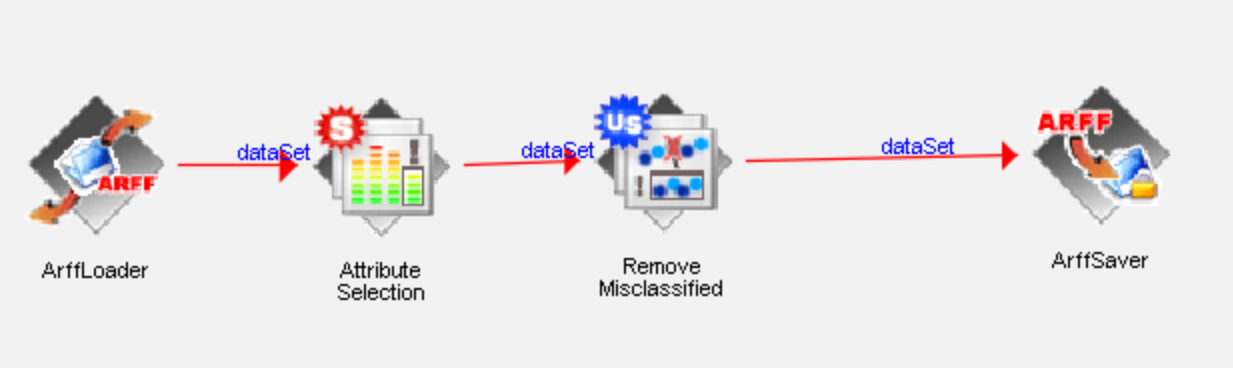
Q3. Design a knowledge flow network for the above given questions 1 & 2 and analyze your result. Use diabetes as target dataset.

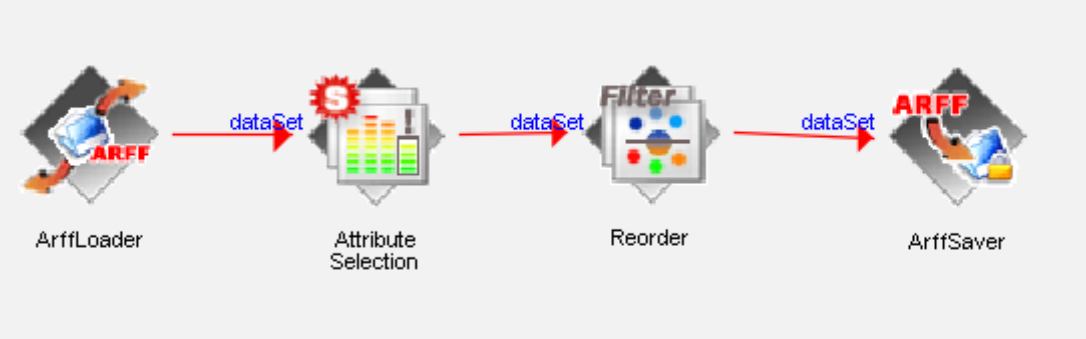
Q1



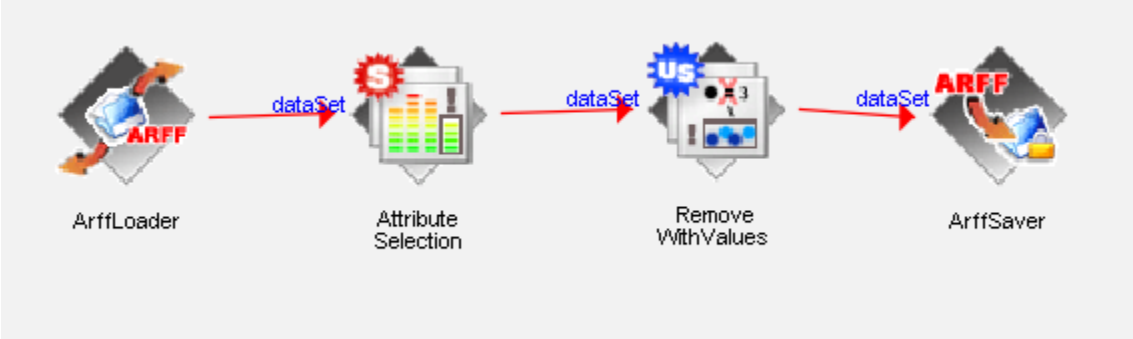


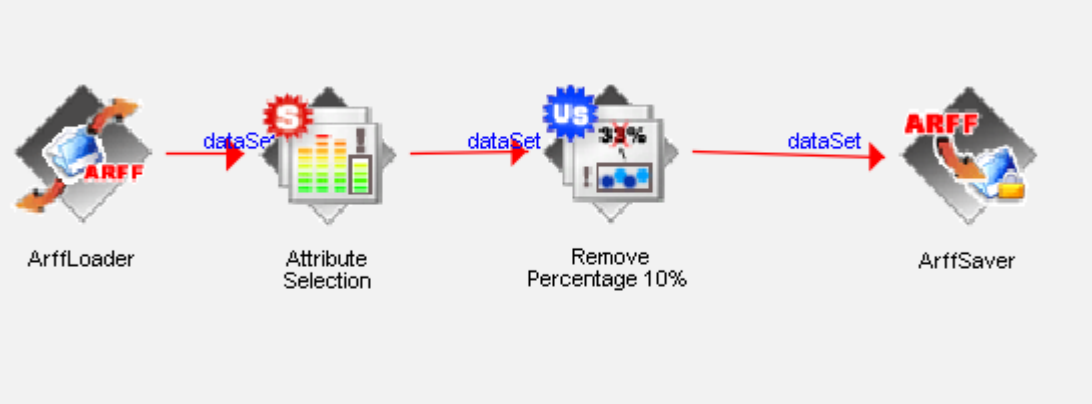


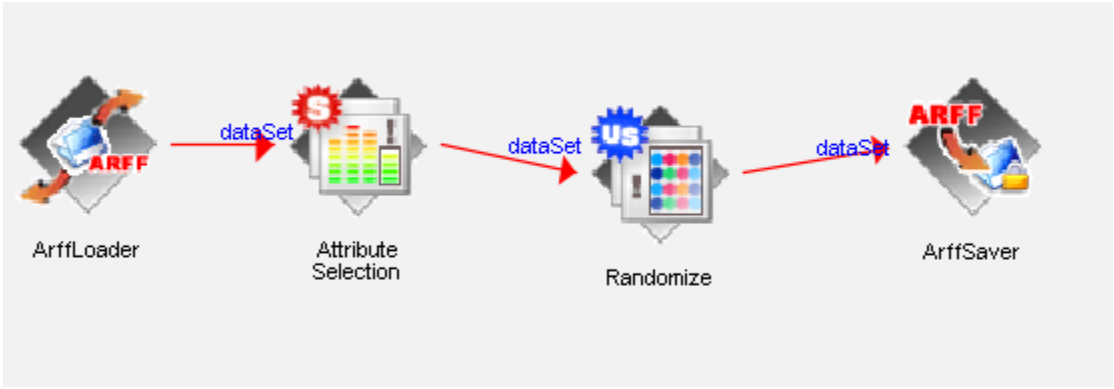

ARFF
dataSet
dataSet
dataSet
ARFF
ArffLoader
Attribute Selection
Remove Misclassified
ArffSaver

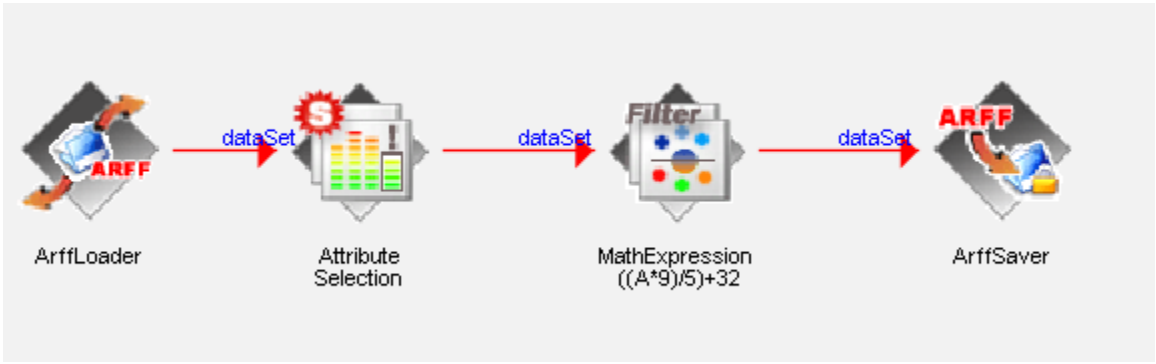

ARFF
dataSet
dataSet
dataSet
ARFF
Filter
ArffLoader
Attribute Selection
Reorder
ArffSaver

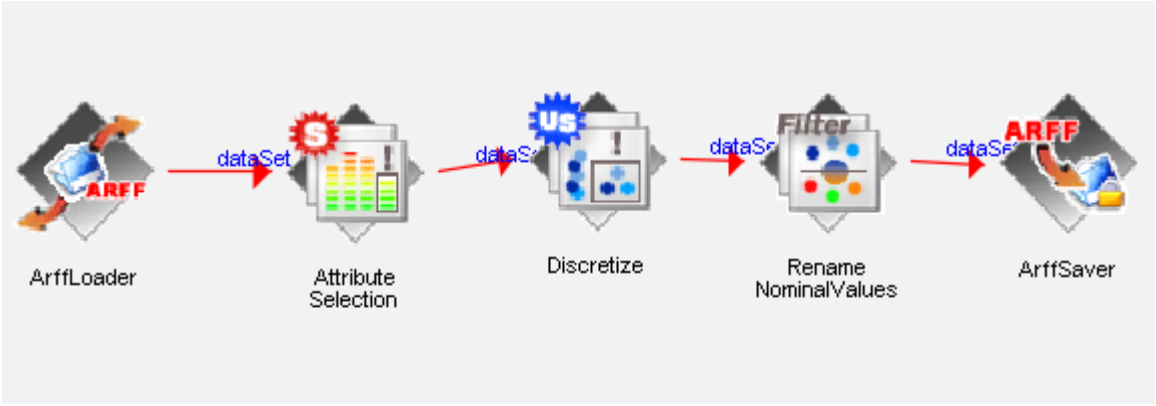
Q2


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
Filter
ARFF
ArffLoader
Attribute Selection
Discretize
Rename NominalValues
ArffSaver