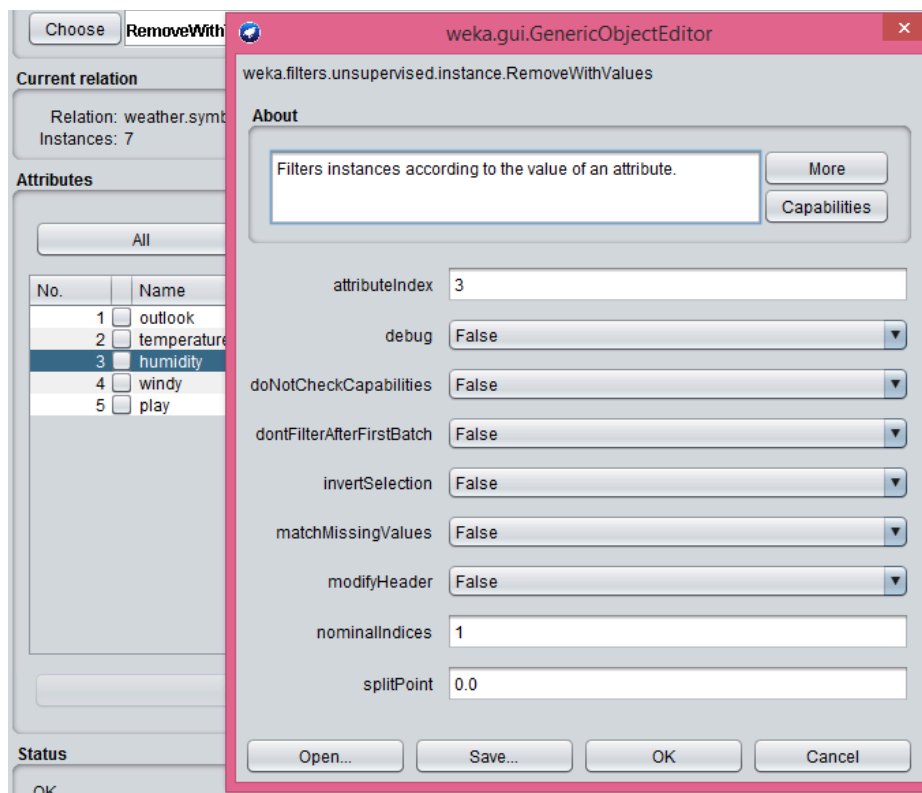


## Assignment 2

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.



Viewer

Relation: weather.symbolic-weka.filters.unsupervised.instance...

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	rainy	cool	normal	FALSE	yes
2	rainy	cool	normal	TRUE	no
3	overcast	cool	normal	TRUE	yes
4	sunny	cool	normal	FALSE	yes
5	rainy	mild	normal	FALSE	yes
6	sunny	mild	normal	TRUE	yes
7	overcast	hot	normal	FALSE	yes

Add instance Undo OK Cancel

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

Current relation

Relation: weather.symbolic-weka.filters.unsupervised.instance.RemoveRange-R3,7,9  
Instances: 11

Attributes: 5  
Sum of weights: 11

Attributes

All

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input checked="" type="checkbox"/> temperature
3	<input checked="" type="checkbox"/> humidity
4	<input checked="" type="checkbox"/> windy
5	<input checked="" type="checkbox"/> play

weka.gui.GenericObjectEditor

weka.filters.unsupervised.instance.RemoveRange

About

A filter that removes a given range of instances of a dataset.

More

Capabilities

debug False

doNotCheckCapabilities False

instancesIndices 3,7,9

invertSelection False

Open... Save... OK Cancel

Viewer

Relation: weather.symbolic-weka.filters.unsupervised.instance.Re...

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	rainy	mild	high	FALSE	yes
4	rainy	cool	normal	FALSE	yes
5	rainy	cool	normal	TRUE	no
6	sunny	mild	high	FALSE	no
7	rainy	mild	normal	FALSE	yes
8	sunny	mild	normal	TRUE	yes
9	overcast	mild	high	TRUE	yes
10	overcast	hot	normal	FALSE	yes
11	rainy	mild	high	TRUE	no

Add instance Undo OK Cancel

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

Current relation

Relation: weather.symbolic-weka.filters.unsupervised.instance.RemoveFrequentValues

Instances: 10

Attributes

All

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

weka.gui.GenericObjectEditor

weka.filters.unsupervised.instance.RemoveFrequentValues

About

Determines which values (frequent or infrequent ones) of an (nominal) attribute are retained and filters the instances accordingly.

More

Capabilities

attributeIndex 1

debug False

doNotCheckCapabilities False

invertSelection False

modifyHeader False

numValues 2

useLeastValues False

Open... Save... OK Cancel

Viewer

Relation: weather.symbolic-weka.filters.unsupervised.instance...

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	rainy	mild	high	FALSE	yes
4	rainy	cool	normal	FALSE	yes
5	rainy	cool	normal	TRUE	no
6	sunny	mild	high	FALSE	no
7	sunny	cool	normal	FALSE	yes
8	rainy	mild	normal	FALSE	yes
9	sunny	mild	normal	TRUE	yes
10	rainy	mild	high	TRUE	no

Add instance Undo OK Cancel

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

Current relation

Relation: pima\_diabetes-weka.filters.unsupervised.instance.RemoveMissclassified-Wwe...  
Instances: 547

Attributes: 9  
Sum of weights: 547

Attributes

All

No.	Name
<input checked="" type="checkbox"/>	preg
<input type="checkbox"/>	plas
<input type="checkbox"/>	pres
<input type="checkbox"/>	skin
<input type="checkbox"/>	insu
<input type="checkbox"/>	mass
<input type="checkbox"/>	pedi
<input type="checkbox"/>	age
<input type="checkbox"/>	class

Status

weka.gui.GenericObjectEditor

weka.filters.unsupervised.instance.RemoveMissclassified

About

A filter that removes instances which are incorrectly classified.

More

Capabilities

classIndex -1

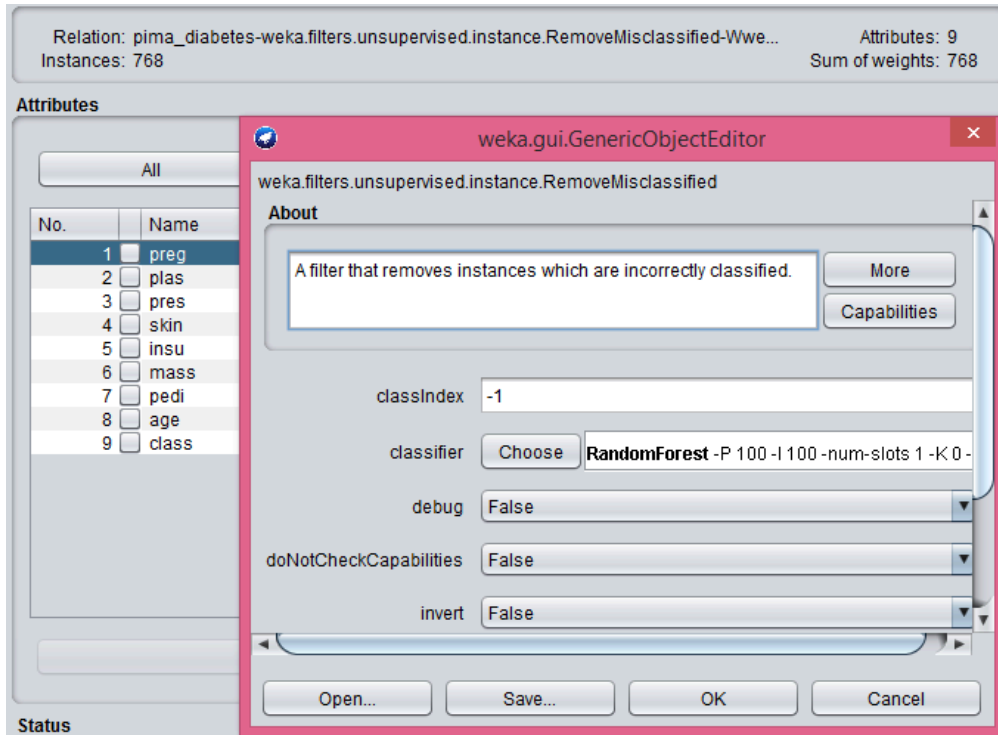
classifier Choose HoeffdingTree -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.0

debug False

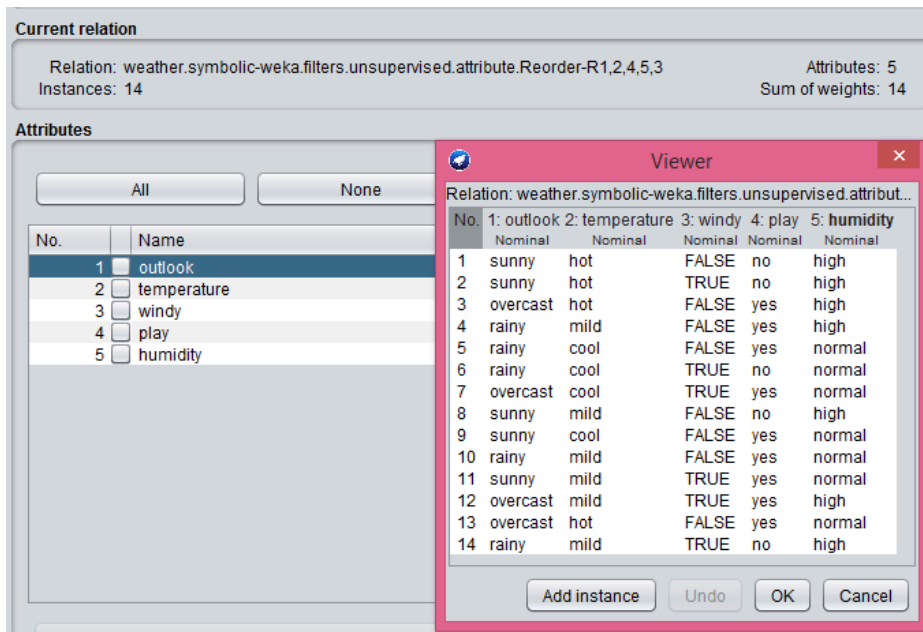
doNotCheckCapabilities False

invert False

Open... Save... OK Cancel



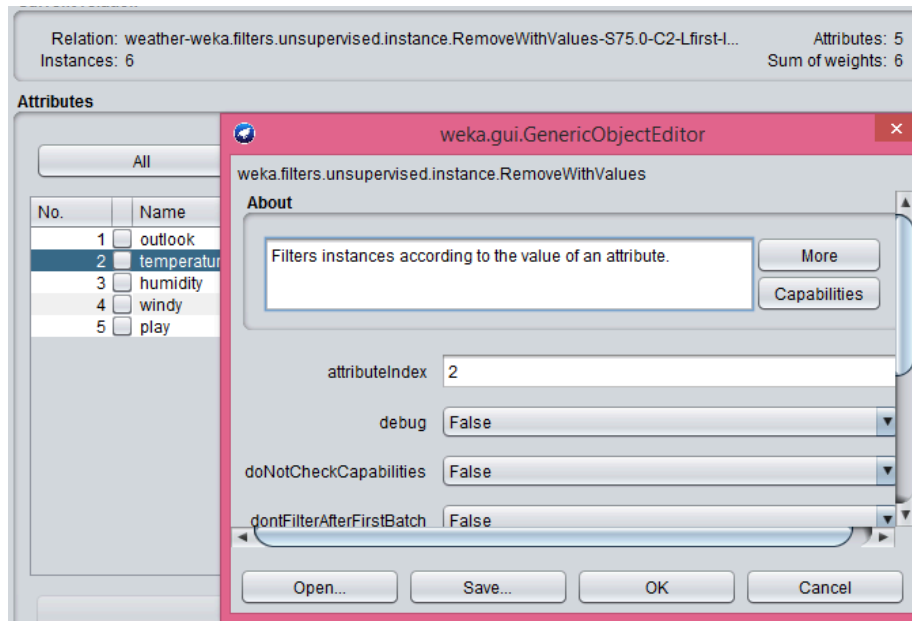
5. Reorder attributes such a way that humidity act as class attribute. (hint: By default last attribute treated as class attribute and use reorder unsupervised attribute filter )



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  $\geq 75$



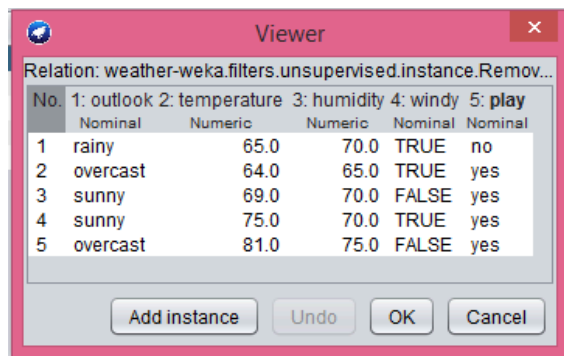
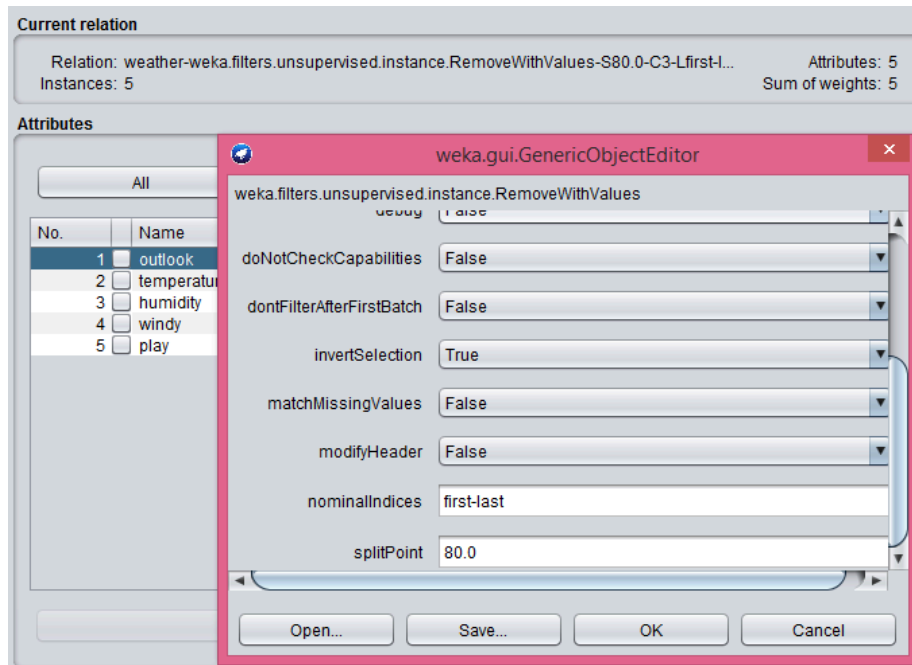
Viewer

Relation: weather-weka.filters.unsupervised.instance.Remov...

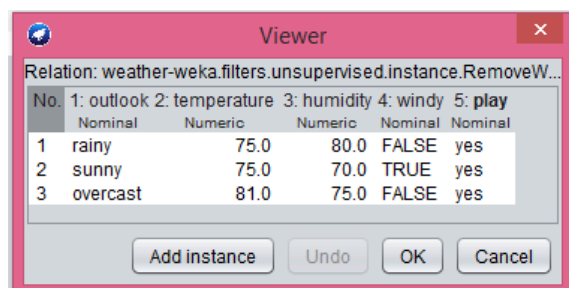
No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Numeric	Numeric	Nominal	Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	75.0	80.0	FALSE	yes
5	sunny	75.0	70.0	TRUE	yes
6	overcast	81.0	75.0	FALSE	yes

Buttons: Add instance, Undo, OK, Cancel

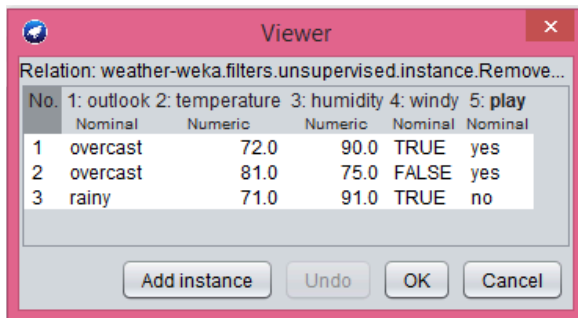
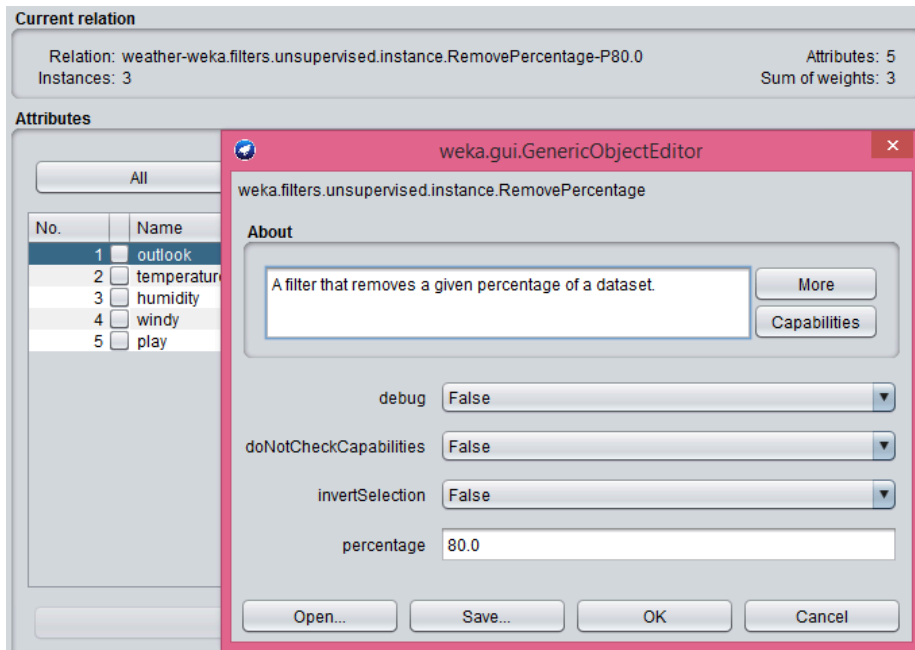
b. Humidity $\leq$ 80



c. Temperature  $\geq$ 75 and Humidity $\leq$ 80

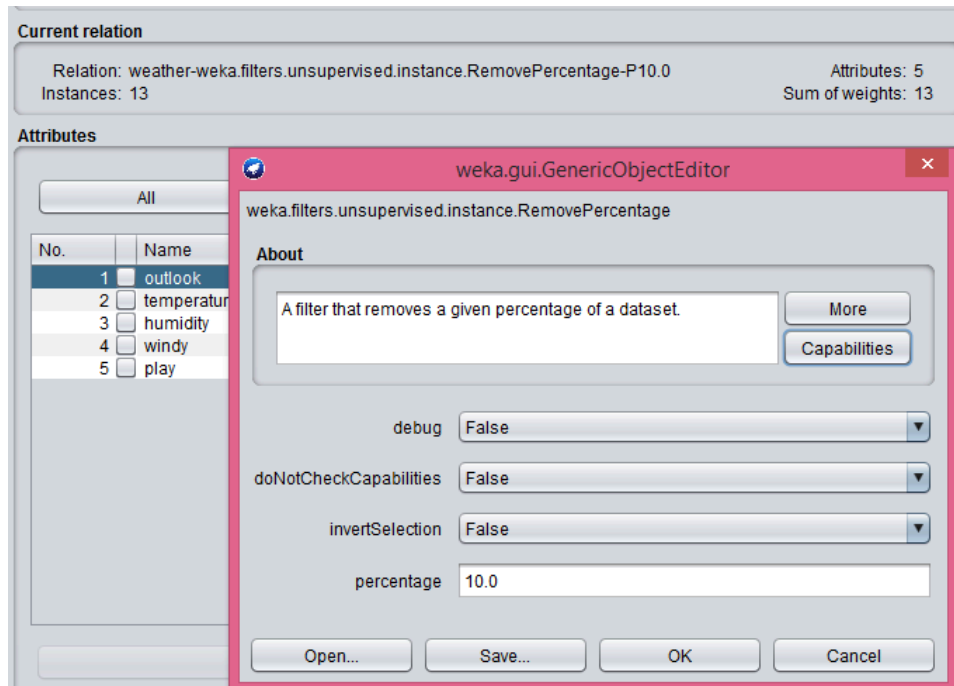


2. Filter 80% samples of the instances.





### 3. Remove 10% instances from dataset.



The screenshot shows the 'Viewer' window in Weka GUI. The title bar says 'Viewer'. The window displays 13 instances of the 'weather' dataset. The columns are: No., 1: outlook, 2: temperature, 3: humidity, 4: windy, 5: play. The data is as follows:

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

Relation: weather-weka.filters.unsupervised.instance.RemoveP...

Add instance Undo OK Cancel

#### 4. Reshuffle the original order of instances.

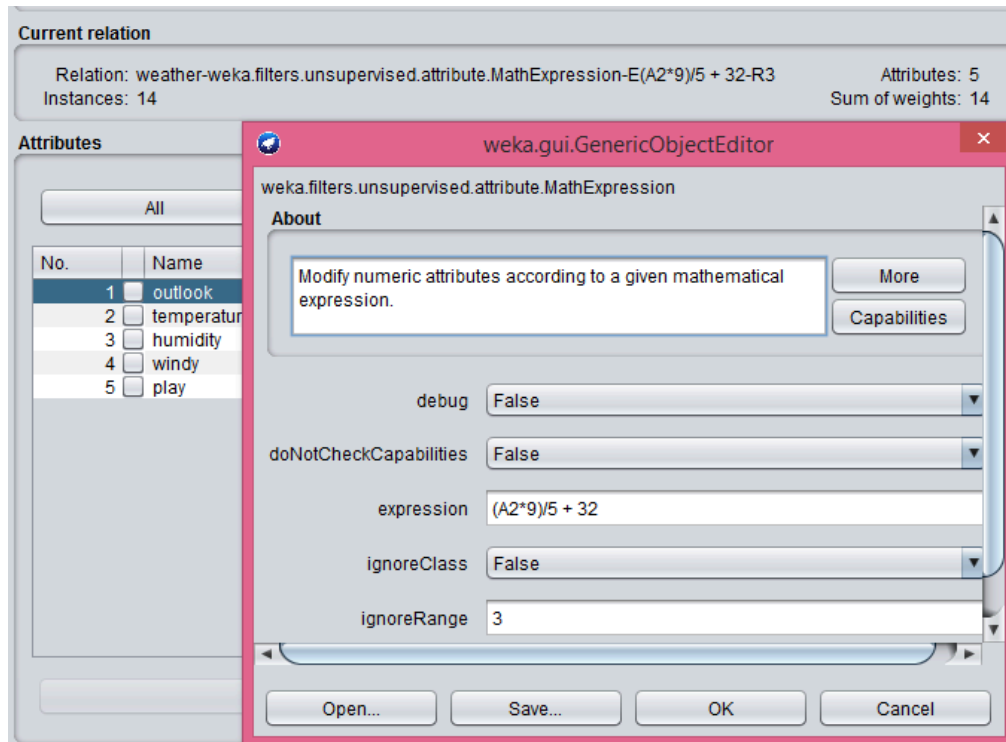
The screenshot shows the Weka Explorer interface with the 'Resample' filter selected in the Filter list. The 'Current relation' is 'weather-wk'. The 'Attributes' list on the left includes 'outlook', 'temperature', 'humidity', 'windy', and 'play'. A dialog box titled 'weka.gui.GenericObjectEditor' is open, showing the configuration for the 'Resample' filter. The 'About' tab is active, displaying the description: 'Produces a random subsample of a dataset using either sampling with replacement or without replacement.' The 'sampleSizePercent' is set to 100.0. The 'Selected attribute' table on the right shows the distribution of the 'outlook' attribute: sunny (5 instances), overcast (4 instances), and rainy (5 instances). The 'Class: play (Nom)' is selected at the bottom.

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

The screenshot shows the Weka Explorer interface after applying the 'Resample' filter. The 'Current relation' is 'weather-wk.filters.unsupervised.instance.Resample-S1-Z100.0'. The 'Attributes' list on the left includes 'outlook', 'temperature', 'humidity', 'windy', and 'play'. A 'Viewer' dialog box is open, displaying the data instances. The 'Selected attribute' table on the right shows the distribution of the 'outlook' attribute: sunny (5 instances), overcast (4 instances), and rainy (5 instances). The 'Class: play (Nom)' is selected at the bottom.

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

5. Convert temperature attribute in degree Fahrenheit.

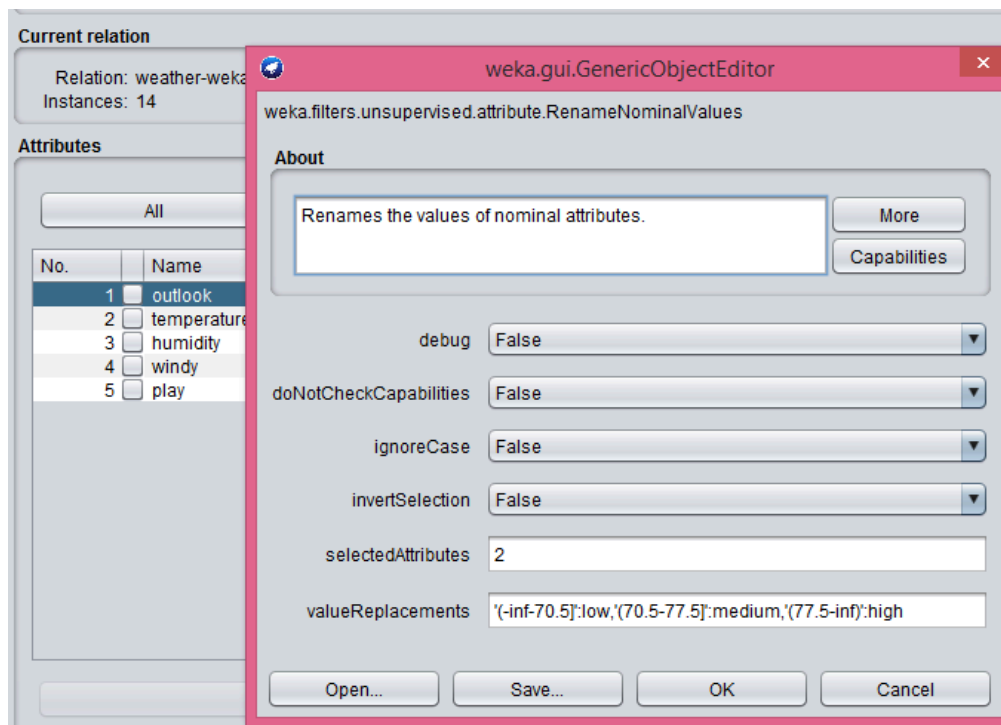


The screenshot shows the 'Viewer' window in Weka, displaying the transformed data. The title bar is 'Viewer'. The relation is 'weather-weka.filters.unsupervised.attribute.MathExp...'. The data is presented in a table with 14 instances and 5 attributes.

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
1	sunny	185.0	85.0	FALSE	no
2	sunny	176.0	90.0	TRUE	no
3	overcast	181.4	86.0	FALSE	yes
4	rainy	158.0	96.0	FALSE	yes
5	rainy	154.4	80.0	FALSE	yes
6	rainy	149.0	70.0	TRUE	no
7	overcast	147.2	65.0	TRUE	yes
8	sunny	161.6	95.0	FALSE	no
9	sunny	156.2	70.0	FALSE	yes
10	rainy	167.0	80.0	FALSE	yes
11	sunny	167.0	70.0	TRUE	yes
12	overcast	161.6	90.0	TRUE	yes
13	overcast	177.8	75.0	FALSE	yes
14	rainy	159.8	91.0	TRUE	no

The 'Add instance' button is highlighted.

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



Viewer

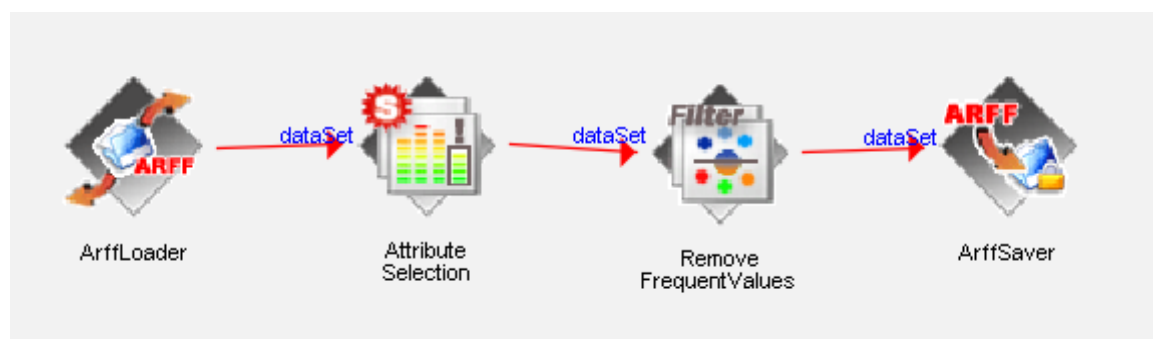
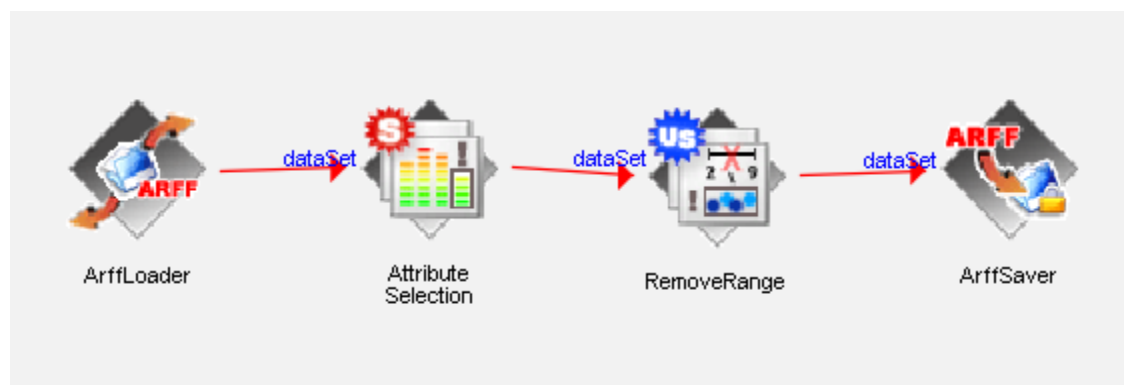
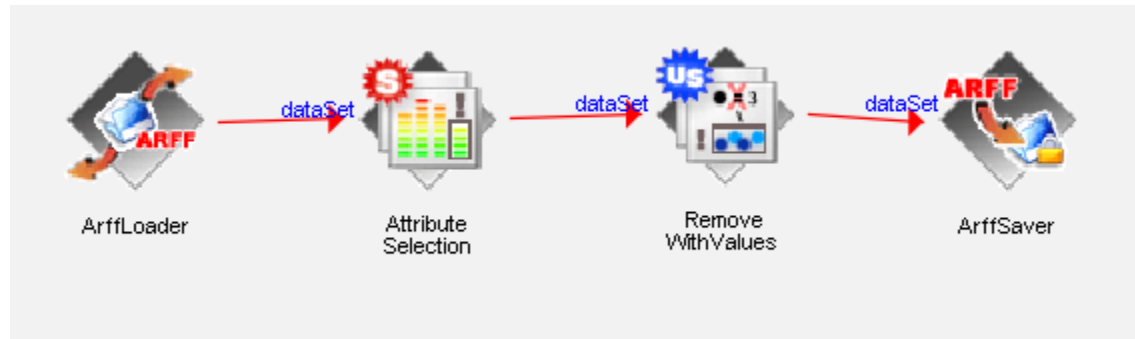
Relation: weather-weka.filters.unsupervised.attribute.PKIDiscr...

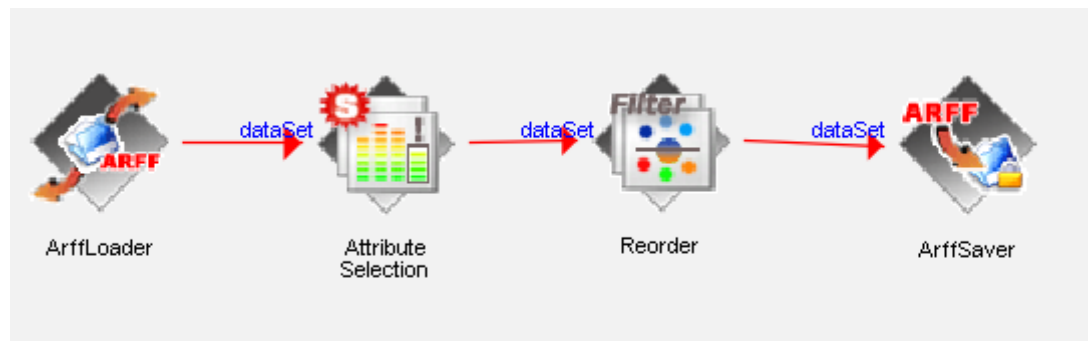
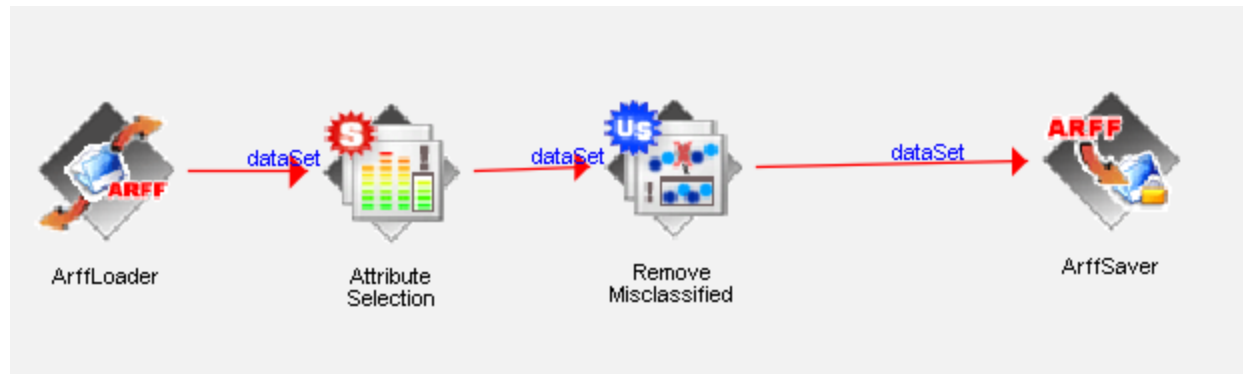
No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Numeric	Nominal	Nominal
1	sunny	high	85.0	FALSE	no
2	sunny	high	90.0	TRUE	no
3	overcast	high	86.0	FALSE	yes
4	rainy	low	96.0	FALSE	yes
5	rainy	low	80.0	FALSE	yes
6	rainy	low	70.0	TRUE	no
7	overcast	low	65.0	TRUE	yes
8	sunny	medium	95.0	FALSE	no
9	sunny	low	70.0	FALSE	yes
10	rainy	medium	80.0	FALSE	yes
11	sunny	medium	70.0	TRUE	yes
12	overcast	medium	90.0	TRUE	yes
13	overcast	high	75.0	FALSE	yes
14	rainy	medium	91.0	TRUE	no

Add instance Undo OK Cancel

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

Q1





Q2

