

# Assignment 2

Name – Ayush Kumar

Reg No. – 20214284

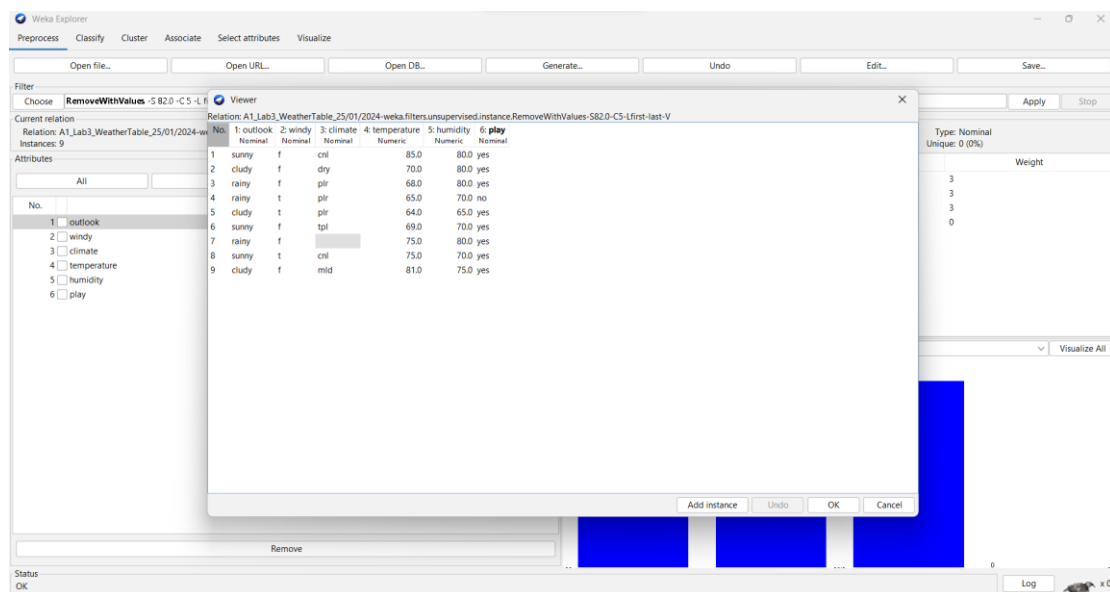
CSE – 6B1

---

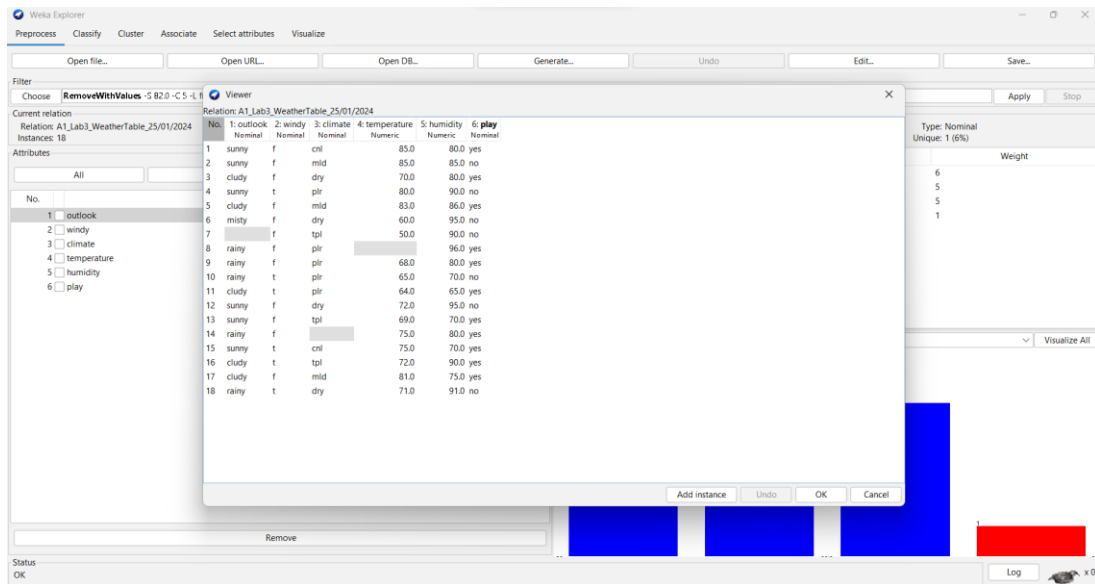
WEKA.FILTER.UNSUPERVISED.INSTANCE:

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.

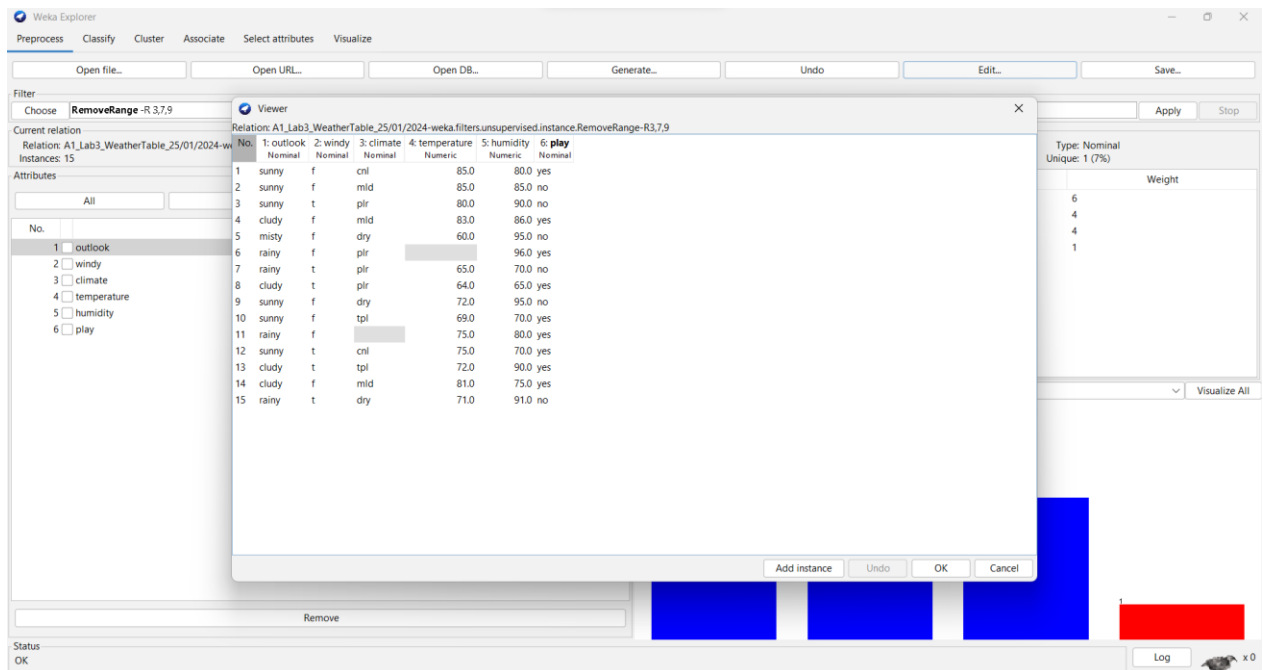


after removing all instances with the values high (i.e.greater than average), we get the above dataset.



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

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

FilterChooseRemove frequent values - C1 - N2

Current relationRelation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.Remove frequent values-C1-N2Instances: 12

Attributes

All

No.1outlook2windy3climate4temperature5humidity6play

No.1outlook2windy3climate4temperature5humidity6play

1sunnyfcln85.080.0yes2sunnyfmid85.085.0no3sunnytplr80.090.0no4f tpl50.090.0no5rainyfplr96.0yes6rainyfplr68.080.0yes7rainytplr65.070.0no8sunnyfdry72.095.0no9sunnyftpl69.070.0yes10rainyf75.080.0yes11sunnytcnl75.070.0yes12rainyt dry71.091.0no

Add instanceUndoOKCancel

Type: NominalUnique: 0 (0%)

Weight6500

Visualize All

Remove

StatusOK

Logx0

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

Weka Explorer - Filter: RemoveMissclassified -W "weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1" -C -1 -F 0 -T 0.1 -I 0

Current relation: pima\_diabetes-weka.filters.unsupervised.RemoveMissclassified-Wweka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -C -1 -F 0 -T 0.1 -I 0

Instances: 768

Attributes:

- 1 ☐ preg
- 2 ☐ plas
- 3 ☐ pres
- 4 ☐ skin
- 5 ☐ insu
- 6 ☐ mass
- 7 ☐ pedi
- 8 ☐ age
- 9 ☐ class

Viewer - Relation: pima\_diabetes-weka.filters.unsupervised.RemoveMissclassified-Wweka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -C -1 -F 0 -T 0.1 -I 0

No.	1: preg	2: plas	3: pres	4: skin	5: insu	6: mass	7: pedi	8: age	9: class
1	6.0	148.0	72.0	35.0	0.0	33.6	0.627	50.0	tested...
2	1.0	85.0	66.0	29.0	0.0	26.6	0.351	31.0	tested...
3	8.0	183.0	64.0	0.0	0.0	23.3	0.672	32.0	tested...
4	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21.0	tested...
5	0.0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	tested...
6	5.0	116.0	74.0	0.0	0.0	25.6	0.201	30.0	tested...
7	3.0	78.0	50.0	32.0	88.0	31.0	0.248	26.0	tested...
8	10.0	115.0	0.0	0.0	0.0	35.3	0.134	29.0	tested...
9	2.0	197.0	70.0	45.0	543.0	30.5	0.158	53.0	tested...
10	8.0	125.0	96.0	0.0	0.0	0.0	0.232	54.0	tested...
11	4.0	110.0	92.0	0.0	0.0	37.6	0.191	30.0	tested...
12	10.0	168.0	74.0	0.0	0.0	38.0	0.537	34.0	tested...
13	10.0	139.0	80.0	0.0	0.0	27.1	1.441	57.0	tested...
14	1.0	189.0	60.0	23.0	846.0	30.1	0.398	59.0	tested...
15	5.0	166.0	72.0	19.0	175.0	25.8	0.587	51.0	tested...
16	7.0	100.0	0.0	0.0	0.0	30.0	0.484	32.0	tested...
17	0.0	118.0	84.0	47.0	230.0	45.8	0.551	31.0	tested...
18	7.0	107.0	74.0	0.0	0.0	29.6	0.254	31.0	tested...
19	1.0	103.0	30.0	38.0	83.0	43.3	0.183	33.0	tested...
20	1.0	115.0	70.0	30.0	96.0	34.6	0.529	32.0	tested...
21	3.0	126.0	88.0	41.0	235.0	39.3	0.704	27.0	tested...
22	8.0	99.0	84.0	0.0	0.0	35.4	0.388	50.0	tested...
23	7.0	196.0	90.0	0.0	0.0	39.8	0.451	41.0	tested...
24	9.0	119.0	80.0	35.0	0.0	29.0	0.263	29.0	tested...

- 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)

Weka Explorer - Filter: Reorder -R 1-4,6,5

Current relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.attribute.Reorder-R1-4,6,5

Instances: 18

Attributes:

- 1 ☐ outlook
- 2 ☐ windy
- 3 ☐ climate
- 4 ☐ temperature
- 5 ☐ play
- 6 ☐ humidity

Viewer - Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.attribute.Reorder-R1-4,6,5

No.	1: outlook	2: windy	3: climate	4: temperature	5: play	6: humidity
1	sunny	f	cnl	85.0	yes	80.0
2	sunny	f	mld	85.0	no	85.0
3	cludy	f	dry	70.0	yes	80.0
4	sunny	t	plr	80.0	no	90.0
5	cludy	f	mld	83.0	yes	86.0
6	misty	f	dry	60.0	no	95.0
7		f	tpl	50.0	no	90.0
8	rainy	f	plr		yes	96.0
9	rainy	f	plr	68.0	yes	80.0
10	rainy	t	plr	65.0	no	70.0
11	cludy	t	plr	64.0	yes	65.0
12	sunny	f	dry	72.0	no	95.0
13	sunny	f	tpl	69.0	yes	70.0
14	rainy	f		75.0	yes	80.0
15	sunny	t	cnl	75.0	yes	70.0
16	cludy	t	tpl	72.0	yes	90.0
17	cludy	f	mld	81.0	yes	75.0
18	rainy	t	dry	71.0	no	91.0

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  $\geq 75$

The screenshot shows the Weka Explorer interface. The 'Preprocess' tab is active. The 'Filter' section shows 'RemoveWithValues' with the expression '-S 75.0 -C 4 -L first-last'. The 'Current relation' is 'A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.RemoveWithValues-S75.0-C4-Lfirst-las'. The 'Attributes' list shows 6 attributes: outlook, windy, climate, temperature, humidity, and play. The 'Viewer' window displays the resulting dataset with 8 instances.

No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
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	cludy	f	mld	83.0	86.0	yes
5	rainy	f	plr		96.0	yes
6	rainy	f		75.0	80.0	yes
7	sunny	t	cnl	75.0	70.0	yes
8	cludy	f	mld	81.0	75.0	yes

b. Humidity  $\leq 80$

The screenshot shows the Weka Explorer interface. The 'Preprocess' tab is active. The 'Filter' section shows 'RemoveWithValues' with the expression '-S 80.0 -C 5 -L first-last -V'. The 'Current relation' is 'A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.RemoveWithValues-S80.0-C5-Lfirst-last-V'. The 'Attributes' list shows 6 attributes: outlook, windy, climate, temperature, humidity, and play. The 'Viewer' window displays the resulting dataset with 5 instances.

No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
1	rainy	t	plr	65.0	70.0	no
2	cludy	t	plr	64.0	65.0	yes
3	sunny	f	tpl	69.0	70.0	yes
4	sunny	t	cnl	75.0	70.0	yes
5	cludy	f	mld	81.0	75.0	yes

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

Weka Explorer interface showing the 'RemoveWithValues' filter applied to the 'A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.RemoveWithValues-S80.0-C4-L first-last-V' relation. The filter is set to remove instances where temperature is less than 75.0 or humidity is greater than 80.0. The 'Viewer' window shows the resulting 3 instances.

No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
1	rainy	t	plr	65.0	70.0	no
2	cludy	t	plr	64.0	65.0	yes
3	sunny	f	tpl	69.0	70.0	yes

### 2. Filter 80% samples of the instances.

Weka Explorer interface showing the 'RemovePercentage' filter applied to the 'A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.RemovePercentage-P80.0' relation. The filter is set to remove 80% of the instances. The 'Viewer' window shows the resulting 4 instances.

No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
1	sunny	t	cnl	75.0	70.0	yes
2	cludy	t	tpl	72.0	90.0	yes
3	cludy	f	mid	81.0	75.0	yes
4	rainy	t	dry	71.0	91.0	no

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

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **RemovePercentage -P 10.0** Apply Stop

Current relation: Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.RemovePercentage-P10.0  
Instances: 16

Attributes: All No

1 outlook 2 windy 3 climate 4 temperature 5 humidity 6 play

Viewer: Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.RemovePercentage-P10.0

No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
	Nominal	Nominal	Nominal	Numeric	Numeric	Nominal
1	cloudy	f	dry	70.0	80.0	yes
2	sunny	t	plr	80.0	90.0	no
3	cloudy	f	mld	83.0	86.0	yes
4	misty	f	dry	60.0	95.0	no
5		f	tpl	50.0	90.0	no
6	rainy	f	plr		96.0	yes
7	rainy	f	plr	68.0	80.0	yes
8	rainy	t	plr	65.0	70.0	no
9	cloudy	t	plr	64.0	65.0	yes
10	sunny	f	dry	72.0	95.0	no
11	sunny	f		69.0	70.0	yes
12	rainy	f		75.0	80.0	yes
13	sunny	t	cnl	75.0	70.0	yes
14	cloudy	t	tpl	72.0	90.0	yes
15	cloudy	f	mld	81.0	75.0	yes
16	rainy	t	dry	71.0	91.0	no

Add instance Undo OK Cancel

Status: OK

Log x0

4. Reshuffle the original order of instances.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **Randomize -S 42** Apply Stop

Current relation: Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.Randomize-S42  
Instances: 18

Attributes: All No

1 outlook 2 windy 3 climate 4 temperature 5 humidity 6 play

Viewer: Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.instance.Randomize-S42

No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
	Nominal	Nominal	Nominal	Numeric	Numeric	Nominal
1	sunny	t	plr	80.0	90.0	no
2	cloudy	f	mld	81.0	75.0	yes
3		f	tpl	50.0	90.0	no
4	sunny	f	mld	85.0	85.0	no
5	rainy	t	dry	71.0	91.0	no
6	sunny	f	cnl	85.0	80.0	yes
7	rainy	f		75.0	80.0	yes
8	rainy	f	plr		96.0	yes
9	cloudy	t	tpl	72.0	90.0	yes
10	rainy	t	plr	65.0	70.0	no
11	cloudy	f	dry	70.0	80.0	yes
12	misty	f	dry	60.0	95.0	no
13	cloudy	f	mld	83.0	86.0	yes
14	sunny	f	tpl	69.0	70.0	yes
15	sunny	t	cnl	75.0	70.0	yes
16	cloudy	t	plr	64.0	65.0	yes
17	sunny	f	dry	72.0	95.0	no
18	rainy	f	plr	68.0	80.0	yes

Add instance Undo OK Cancel

Status: OK

Log x0

5. Convert temperature attribute in degree Fahrenheit.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **MathExpression** -E ((A\*9)/5)+32-R 1,2,3,5,6 -unset-class-temporarily Apply Stop

Current relation: Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.attribute.MathExpression-E((A\*9)/5)+32-R1,2,3,5,6 -unset-class-temporarily Instances: 18

Attributes: 6 Sum of weights: 18 Name: outlook Missing: 1 (6%) Distinct: 4 Type: Nominal Unique: 1 (6%)

Viewer

Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.attribute.MathExpression-E((A\*9)/5)+32-R1,2,3,5,6 -unset-class-temporarily

No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
	Nominal	Nominal	Nominal	Numeric	Numeric	Nominal
1	sunny	f	cnl	185.0	80.0	yes
2	sunny	f	mld	185.0	85.0	no
3	cludy	f	dry	158.0	80.0	yes
4	sunny	t	plr	176.0	90.0	no
5	cludy	f	mld	181.4	86.0	yes
6	misty	f	dry	140.0	95.0	no
7		f	tpl	122.0	90.0	no
8	rainy	f	plr		96.0	yes
9	rainy	f	plr	154.4	80.0	yes
10	rainy	t	plr	149.0	70.0	no
11	cludy	t	plr	147.2	65.0	yes
12	sunny	f	dry	161.6	95.0	no
13	sunny	f	tpl	156.2	70.0	yes
14	rainy	f		167.0	80.0	yes
15	sunny	t	cnl	167.0	70.0	yes
16	cludy	t	tpl	161.6	90.0	yes
17	cludy	f	mld	177.8	75.0	yes
18	rainy	t	dry	159.8	91.0	no

Status: OK

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

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **RenameNominalValues** -R 4 -N -N "\(-inf-61.67\)\low,\(61.67-73.33\)\medium,\(73.33-inf\)\high" Apply Stop

Current relation: Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision2-weka.filters.unsupervised.attribute.RenameNominalValues-R4-N(-inf-61.67)\low,\(61.67-73.33\)\medium,\(73.33-inf\)\high Instances: 18

Attributes: 6 Sum of weights: 18 Name: outlook Missing: 1 (6%) Distinct: 4 Type: Nominal Unique: 1 (6%)

Viewer

Relation: A1\_Lab3\_WeatherTable\_25/01/2024-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision2-weka.filters.unsupervised.attribute.RenameNominalValues-R4-N(-inf-61.67)\low,\(61.67-73.33\)\medium,\(73.33-inf\)\high

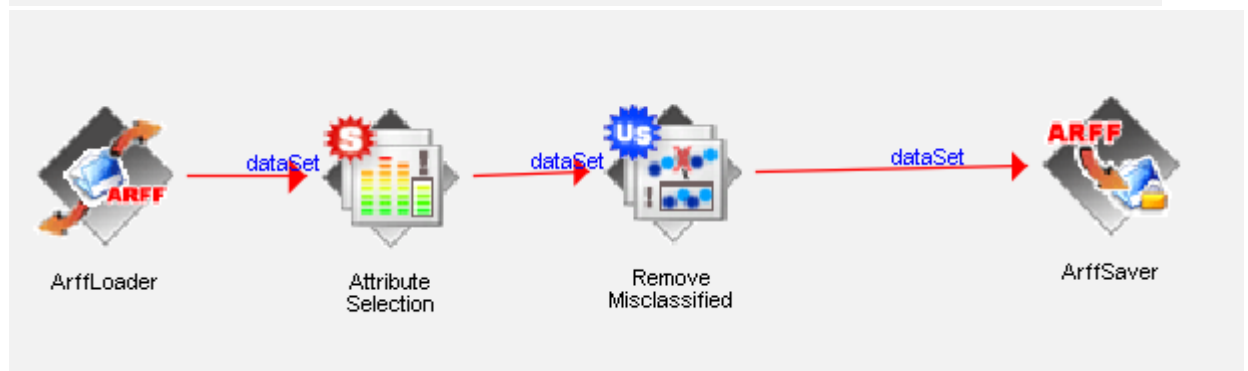
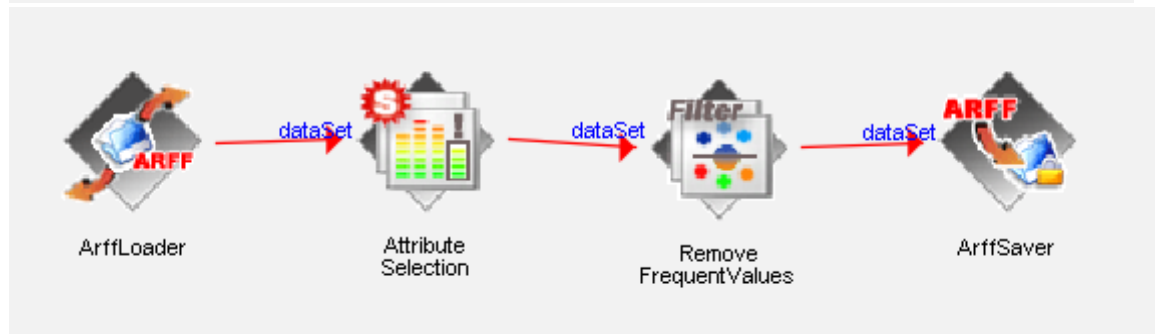
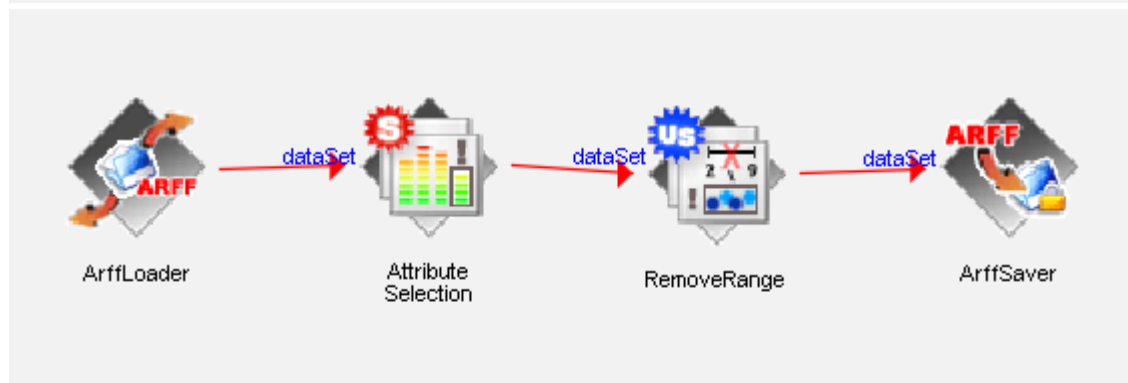
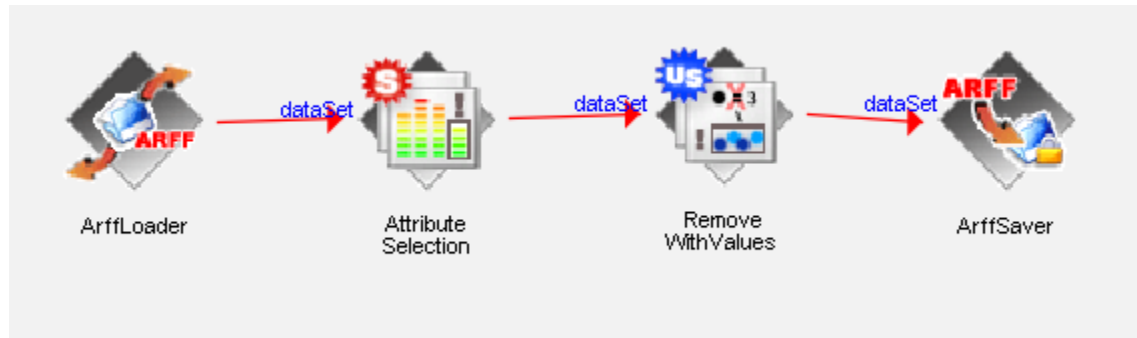
No.	1: outlook	2: windy	3: climate	4: temperature	5: humidity	6: play
	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal
1	sunny	f	cnl	high	80.0	yes
2	sunny	f	mld	high	85.0	no
3	cludy	f	dry	medium	80.0	yes
4	sunny	t	plr	high	90.0	no
5	cludy	f	mld	high	86.0	yes
6	misty	f	dry	low	95.0	no
7		f	tpl	low	90.0	no
8	rainy	f	plr		96.0	yes
9	rainy	f	plr	medium	80.0	yes
10	rainy	t	plr	medium	70.0	no
11	cludy	t	plr	medium	65.0	yes
12	sunny	f	dry	medium	95.0	no
13	sunny	f	tpl	medium	70.0	yes
14	rainy	f		high	80.0	yes
15	sunny	t	cnl	high	70.0	yes
16	cludy	t	tpl	medium	90.0	yes
17	cludy	f	mld	high	75.0	yes
18	rainy	t	dry	medium	91.0	no

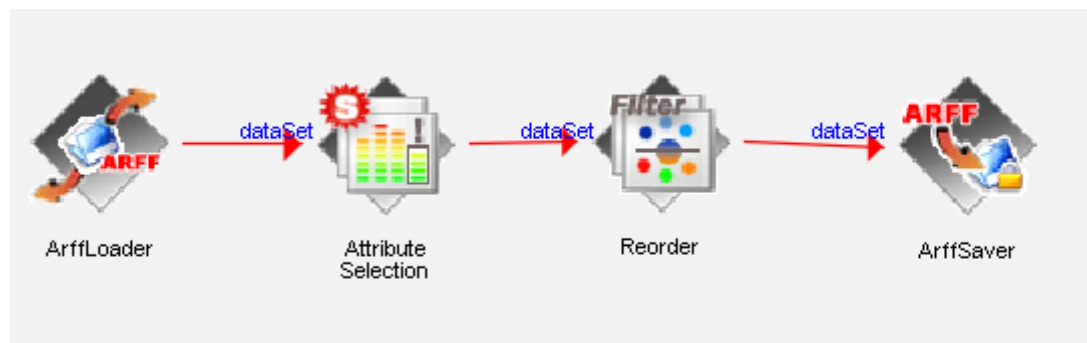
Status: OK



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

Q1





Q2

