Experiment-1.3

Student Name: Ashish Kumar
UID: 23MAI10008
Branch: CSE AIML
Section/Group: 23MAI-1
Semester: 01
Date of Performance:
Subject Name: Artificial Intelligence Lab
Subject Code: 23CSH-621

Aim of the Experiment :

Apply the following Pre-Processing techniques to the training data set and report/analyze the changes in the dataset.

a) Normalization

b) Class balance

c) Add

d) Remove

e) Discretization

f) Sampling

Objective of the Experiment :

Task to be done for this experiment is that we have to create a Weather Table which includes attributes like outlook, temperature, humidity, windy and play. Then we will open the weather dataset in WEKA Tool and apply different pre processing techniques on the dataset to improve the quality of data and analyze the changes happened in the dataset.

Procedure / Steps for Experiment :

Step 1: Open Start → Programs → Accessories → Notepad.

Step 2: Write the dataset of **Weather Table** in Notepad with all attributes and their value.

Step 3: After writing the dataset, save the file with **.arff** format.

Step 3: Now open the Weather dataset in the WEKA Tool using **'Open file'** option.

Step 4: Now Choose one of the attributes as the Class attribute in bottom right section.

A) Normalization:

1) Open WEKA and load the dataset.
2) In **Filter** Section, Click on **'Choose'** Button.
3) Under the **filters** section, select the **Normalize** filter from the list.
4) Click on Normalize again and set **scale** as 2 and **translation** as -1.
5) Click on **'Apply'** Button to normalize the dataset.

Selected attribute

Name: temperature Type: Numeric
Missing: 0 (0%) Distinct: 12 Unique: 10 (71%)

Statistic	Value
Minimum	64
Maximum	85
Mean	73.571
StdDev	6.572

Selected attribute

Name: temperature Type: Numeric
Missing: 0 (0%) Distinct: 12 Unique: 10 (71%)

Statistic	Value
Minimum	-1
Maximum	1
Mean	-0.088
StdDev	0.626

Viewer

Relation: weather-weka.filters.unsupervised.attribute.Normalize-S2.0-T-1.0

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	1.0	0.29032258064516125	FALSE	no
2	sunny	0.5238095238095237	0.6129032258064515	TRUE	no
3	overcast	0.8095238095238095	0.35483870967741926	FALSE	yes
4	rainy	-0.4285714285714286	1.0	FALSE	yes
5	rainy	-0.6190476190476191	-0.032258064516129004	FALSE	yes
6	rainy	-0.9047619047619048	-0.6774193548387097	TRUE	no
7	overcast	-1.0	-1.0	TRUE	yes
8	sunny	-0.23809523809523814	0.935483870967742	FALSE	no
9	sunny	-0.5238095238095238	-0.6774193548387097	FALSE	yes
10	rainy	0.04761904761904767	-0.032258064516129004	FALSE	yes
11	sunny	0.04761904761904767	-0.6774193548387097	TRUE	yes
12	overcast	-0.23809523809523814	0.6129032258064515	TRUE	yes
13	overcast	0.6190476190476191	-0.3548387096774194	FALSE	yes
14	rainy	-0.33333333333333337	0.6774193548387097	TRUE	no

B) ClassBalancer:

1) Open WEKA and load the dataset.
2) In **Filter** Section, Click on **'Choose'** Button.
3) Under the **filters** section, select the **ClassBalancer** from the list.
4) Click on **'Apply'** Button to apply ClassBalancer on the dataset.

Selected attribute

Name: outlook Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

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

Selected attribute

Name: outlook Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.756
2	overcast	4	3.111
3	rainy	5	5.133

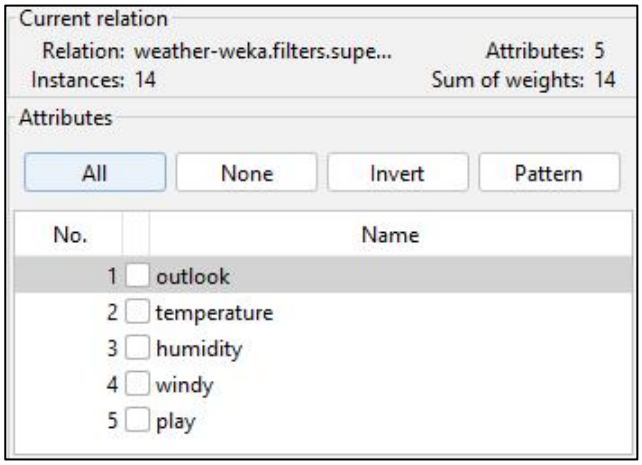
Viewer

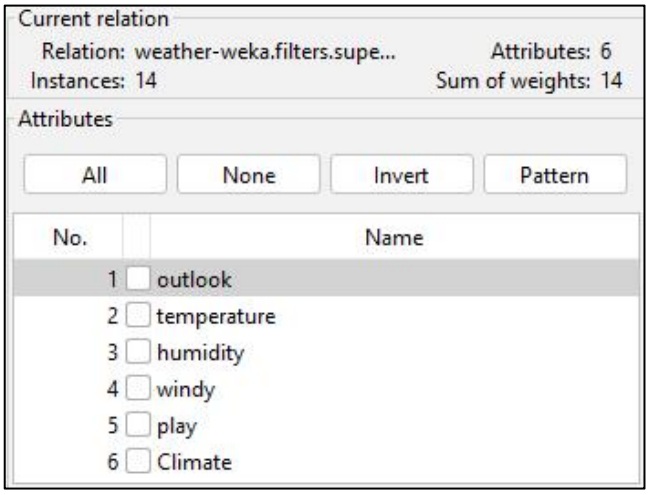
Relation: weather-weka.filters.supervised.instance.ClassBalancer-num-intervals10

No.	Weight	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	1.4	sunny	85.0	85.0	FALSE	no
2	1.4	sunny	80.0	90.0	TRUE	no
3	0.7777777777777778	overcast	83.0	86.0	FALSE	yes
4	0.7777777777777778	rainy	70.0	96.0	FALSE	yes
5	0.7777777777777778	rainy	68.0	80.0	FALSE	yes
6	1.4	rainy	65.0	70.0	TRUE	no
7	0.7777777777777778	overcast	64.0	65.0	TRUE	yes
8	1.4	sunny	72.0	95.0	FALSE	no
9	0.7777777777777778	sunny	69.0	70.0	FALSE	yes
10	0.7777777777777778	rainy	75.0	80.0	FALSE	yes
11	0.7777777777777778	sunny	75.0	70.0	TRUE	yes
12	0.7777777777777778	overcast	72.0	90.0	TRUE	yes
13	0.7777777777777778	overcast	81.0	75.0	FALSE	yes
14	1.4	rainy	71.0	91.0	TRUE	no

C) Add:

1) Open WEKA and load the dataset.
2) In **Filter** Section, Click on **'Choose'** Button.
3) Under the **filters** section, select the **Add** from the list.
4) Click on Add again and set **attributeName** to 'Climate' and **attributeType** to Nominal.
5) Click on **'Apply'** Button to apply Add on the dataset.
6) A new attribute Climate is added to the dataset.





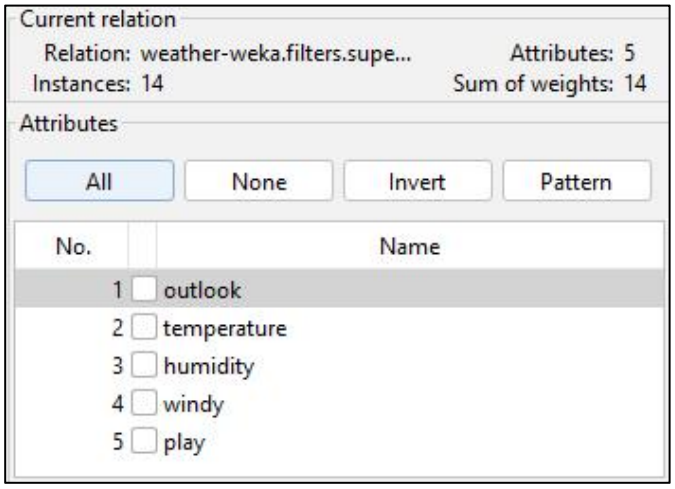
Viewer

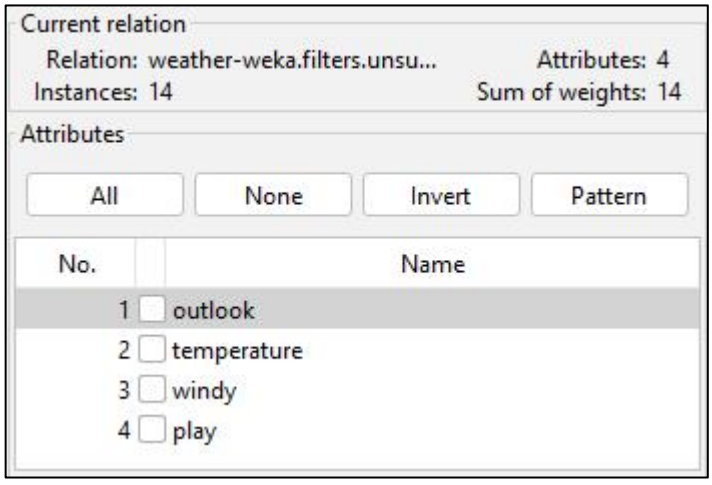
Relation: weather-weka.filters.unsupervised.attribute.Add-TNOM-NClimate-L-Clast-W1.0

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

D) <u>Remove:</u>

1) Open WEKA and load the dataset.
2) In **Filter** Section, Click on **'Choose'** Button.
3) Under the **filters** section, select the **Remove** from the list.
4) Click on Remove again and set **attributeIndices** to desired index value.
5) Click on **'Apply'** Button to apply Remove on the dataset.
6) The specified attribute is removed from the dataset.





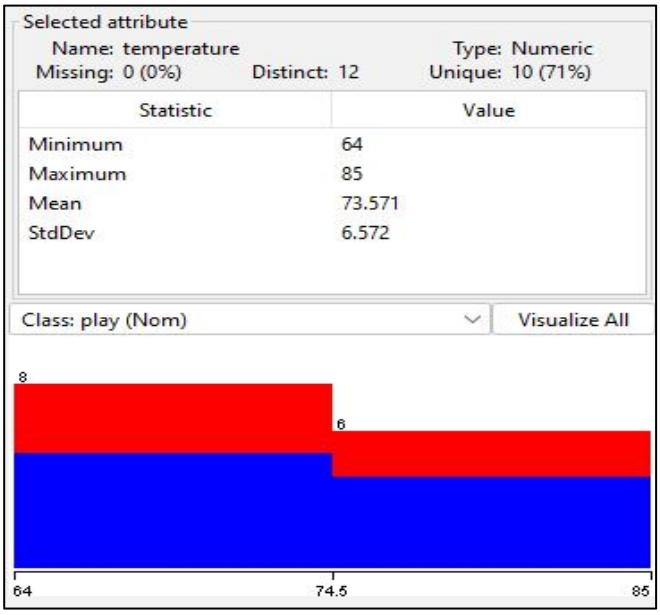
Viewer

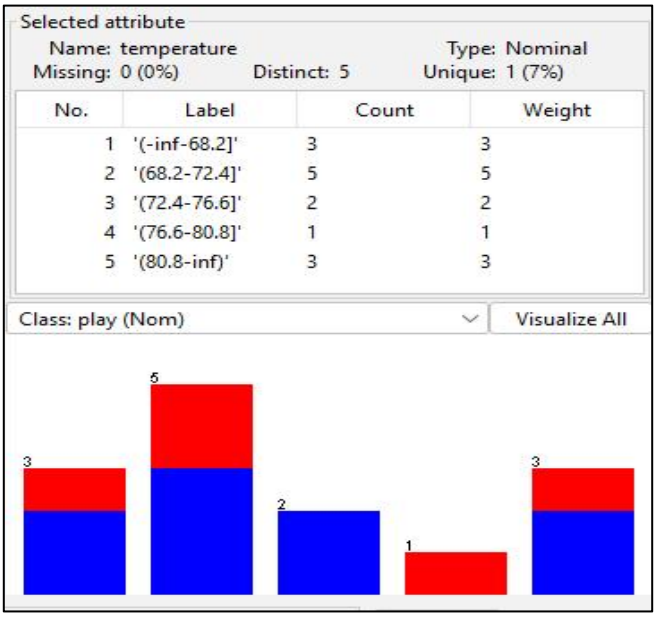
Relation: weather-weka.filters.unsupervised.attribute.Remove-R4

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: **play** Nominal
1	sunny	85.0	85.0	no
2	sunny	80.0	90.0	no
3	overcast	83.0	86.0	yes
4	rainy	70.0	96.0	yes
5	rainy	68.0	80.0	yes
6	rainy	65.0	70.0	no
7	overcast	64.0	65.0	yes
8	sunny	72.0	95.0	no
9	sunny	69.0	70.0	yes
10	rainy	75.0	80.0	yes
11	sunny	75.0	70.0	yes
12	overcast	72.0	90.0	yes
13	overcast	81.0	75.0	yes
14	rainy	71.0	91.0	no

E) <u>Discretize:</u>

1) Open WEKA and load the dataset.
2) In **Filter** Section, Click on **'Choose'** Button.
3) Under the **filters** section, select the **Discretize** from the list.
4) Click on **'Apply'** Button to apply Discretize on the dataset.
5) Data values are converted into the discrete values.





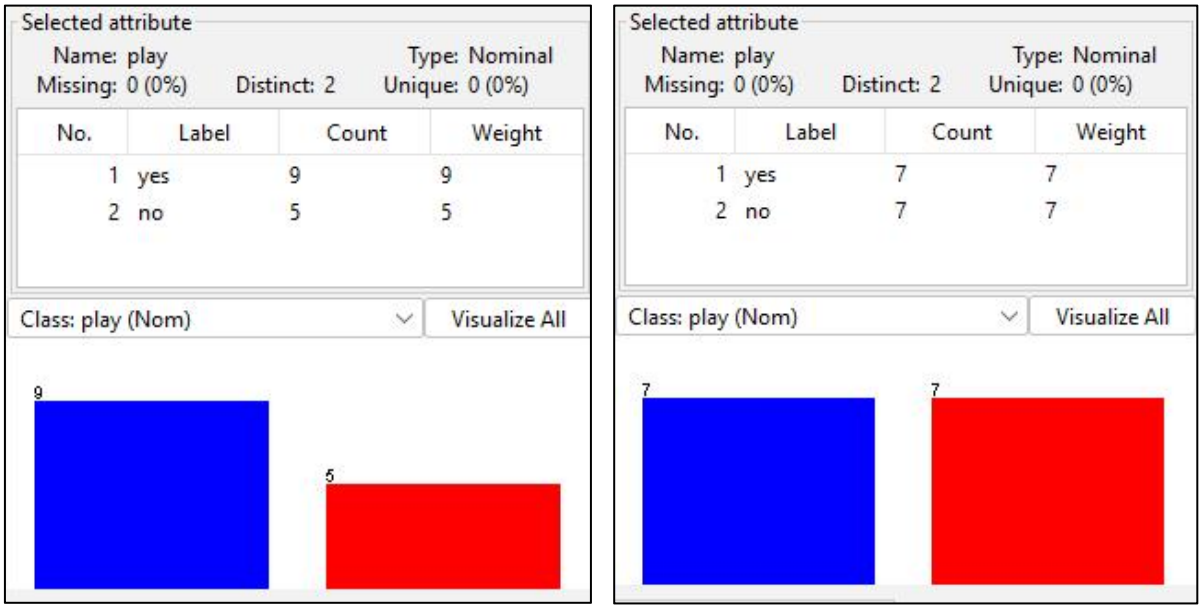
Viewer

Relation: weather-weka.filters.unsupervised.attribute.Discretize-B5-M-1.0-Rfirst-last-precision6

No.	1: outlook Nominal	2: temperature Nominal	3: humidity Nominal	4: windy Nominal	5: **play** Nominal
1	sunny	'(80.8-inf)'	'(83.6-89.8]'	FALSE	no
2	sunny	'(76.6-80.8]'	'(89.8-inf)'	TRUE	no
3	overcast	'(80.8-inf)'	'(83.6-89.8]'	FALSE	yes
4	rainy	'(68.2-72.4]'	'(89.8-inf)'	FALSE	yes
5	rainy	'(-inf-68.2]'	'(77.4-83.6]'	FALSE	yes
6	rainy	'(-inf-68.2]'	'(-inf-71.2]'	TRUE	no
7	overcast	'(-inf-68.2]'	'(-inf-71.2]'	TRUE	yes
8	sunny	'(68.2-72.4]'	'(89.8-inf)'	FALSE	no
9	sunny	'(68.2-72.4]'	'(-inf-71.2]'	FALSE	yes
10	rainy	'(72.4-76.6]'	'(77.4-83.6]'	FALSE	yes
11	sunny	'(72.4-76.6]'	'(-inf-71.2]'	TRUE	yes
12	overcast	'(68.2-72.4]'	'(89.8-inf)'	TRUE	yes
13	overcast	'(80.8-inf)'	'(71.2-77.4]'	FALSE	yes
14	rainy	'(68.2-72.4]'	'(89.8-inf)'	TRUE	no

F) <u>Resample:</u>

1) Open WEKA and load the dataset.
2) In **Filter** Section, Click on **'Choose'** Button.
3) Under the **filters** section, select the **supervised Resample** filter from the list.
4) Click on Resample again and set **biasToUniformClass** as 1.
5) Click on **'Apply'** Button to apply Resample on the dataset.



Viewer

Relation: weather-weka.filters.unsupervised.instance.Resample-S1-Z100.0

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	overcast	72.0	90.0	TRUE	yes
2	rainy	68.0	80.0	FALSE	yes
3	sunny	80.0	90.0	TRUE	no
4	sunny	72.0	95.0	FALSE	no
5	overcast	64.0	65.0	TRUE	yes
6	overcast	64.0	65.0	TRUE	yes
7	sunny	85.0	85.0	FALSE	no
8	sunny	69.0	70.0	FALSE	yes
9	sunny	75.0	70.0	TRUE	yes
10	overcast	64.0	65.0	TRUE	yes
11	rainy	65.0	70.0	TRUE	no
12	rainy	65.0	70.0	TRUE	no
13	rainy	65.0	70.0	TRUE	no
14	rainy	65.0	70.0	TRUE	no

Learning outcomes (What I have learnt):

1. I learnt about the Weka Tool and its applications.
2. I learnt about how to create dataset in .arff format.
3. I learnt about how to open .arff format file in Weka Tool.
4. I learnt about different pre processing techniques in Weka Tool.
5. I learnt about supervised and unsupervised data in Weka Tool.