

DMWA Lab -3

Teghdeep Kapoor

18104050

B12

Q1

a. Add Attribute(Date)

Before Adding the attribute

The screenshot shows the Weka Explorer interface. The 'Filter' tab is active, and the 'Add -N unname -C last -W 1.0' filter is applied. The 'Current relation' section shows 'Relation: weather' with 'Attributes: 5' and 'Instances: 14'. The 'Selected attribute' section shows 'Name: outlook' with 'Type: Nominal', 'Missing: 0 (0%)', 'Distinct: 3', and 'Unique: 0 (0%)'. The 'Attributes' list shows 'outlook' selected. The 'Class: play (Nom)' is selected. A bar chart at the bottom right shows the distribution of 'play' values for each 'outlook' category: 'sunny' has 5 instances (red bar), 'overcast' has 4 instances (blue bar), and 'rainy' has 5 instances (red bar).

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

After Adding the Attribute

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter

Choose Add -T DAT -N date -C last -W 1.0 Apply Stop

Current relation

Relation: weather-weka.filt... Attributes: 6
Instances: 14 Sum of weights: 14

Selected attribute

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

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

Class: date (Dat) Visualize All

Attributes

All None Invert Pattern

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
6	<input type="checkbox"/> date

Remove

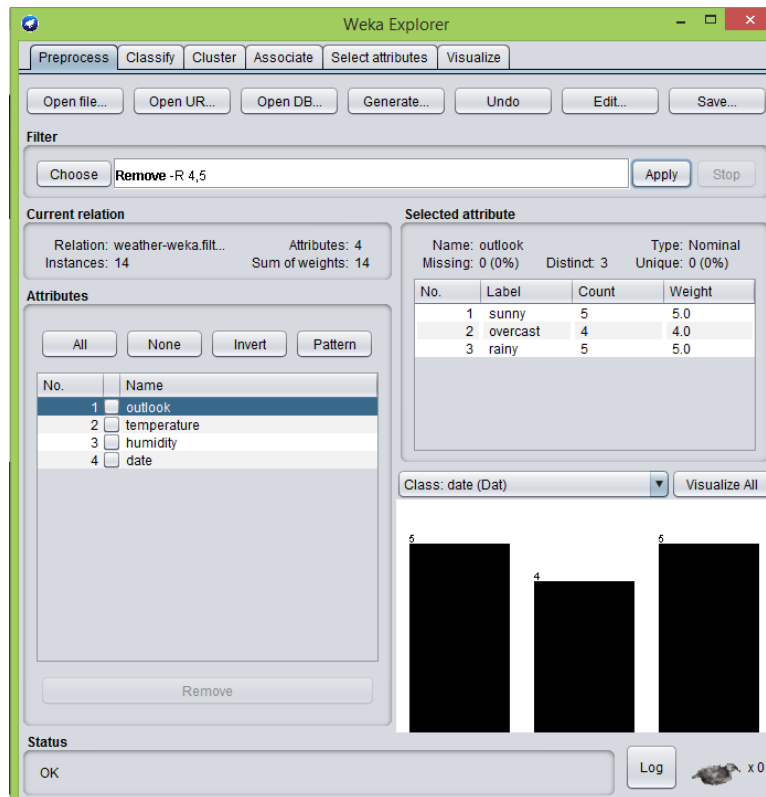
Status

OK Log x 0

File Edit Format View Help

```
@relation weather-weka.filters.unsupervised.attribute.Add-TDAT-Ndate-Clast-W1.0
@attribute outlook {sunny,overcast,rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE,FALSE}
@attribute play {yes,no}
@attribute date date 'yyyy-MM-dd'T\HH:mm:ss'
@data
sunny,85,85,FALSE,no,?
sunny,80,90,TRUE,no,?
overcast,83,86,FALSE,yes,?
rainy,70,96,FALSE,yes,?
rainy,68,80,FALSE,yes,?
rainy,65,70,TRUE,no,?
overcast,64,65,TRUE,yes,?
sunny,72,95,FALSE,no,?
sunny,69,70,FALSE,yes,?
rainy,75,80,FALSE,yes,?
sunny,75,70,TRUE,yes,?
overcast,72,90,TRUE,yes,?
overcast,81,75,FALSE,yes,?
rainy,71,91,TRUE,no,?
```

b. Removing the Attribute (Windy and Play)

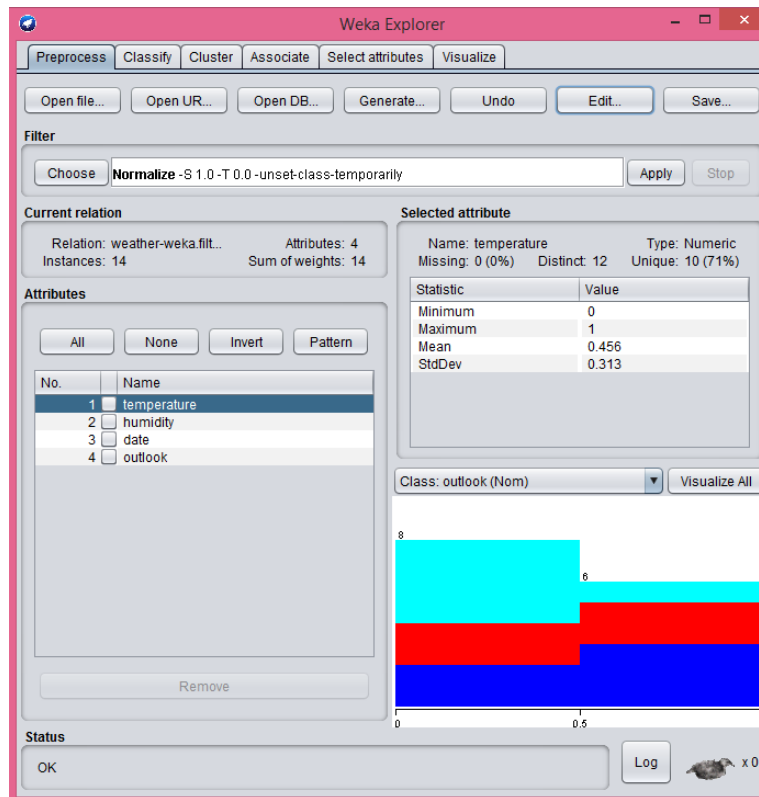


```

@relation 'weather-weka.filters.unsupervised.attribute.Add-TDAT-Ndate-Clast-W1.0-weka.filters.unsupervised.attribute.Remove-R4,5'
@attribute outlook {sunny,overcast,rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute date date 'yyyy-MM-dd\'T\'HH:mm:ss'
@data
sunny,85,85,?
sunny,80,90,?
overcast,83,86,?
rainy,70,96,?
rainy,68,80,?
rainy,65,70,?
overcast,64,65,?
sunny,72,95,?
sunny,69,70,?
rainy,75,80,?
sunny,75,70,?
overcast,72,90,?
overcast,81,75,?
rainy,71,91,?

```

c. Normalizing the Data (Temperature, and Humidity)



```

@relation 'weather-weka.filters.unsupervised.attribute.Add-TDAT-Ndate-Clast-W1.0-weka.filters.unsupervised.attribute.Remove-R4,5-weka.filters.unsupervised.attribute.No
@attribute temperature numeric
@attribute humidity numeric
@attribute date date 'yyyy-MM-dd\T\HH:mm:ss'
@attribute outlook {sunny,overcast,rainy}
@data
1,0.645161,?,sunny
0.761905,0.806452,?,sunny
0.904762,0.677419,?,overcast
0.285714,1,?,rainy
0.190476,0.483871,?,rainy
0.047619,0.16129,?,rainy
0,0,?,overcast
0.380952,0.967742,?,sunny
0.238095,0.16129,?,sunny
0.52381,0.483871,?,rainy
0.52381,0.16129,?,sunny
0.380952,0.806452,?,overcast
0.809524,0.322581,?,overcast
0.333333,0.83871,?,rainy

```

Q2

a. Downloading the IRIS dataset from online website

```
iris - Notepad
File Edit Format View Help
@RELATION iris

@ATTRIBUTE sepallength REAL
@ATTRIBUTE sepalwidth REAL
@ATTRIBUTE petallength REAL
@ATTRIBUTE petalwidth REAL
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3.0,1.4,0.1,Iris-setosa
4.3,3.0,1.1,0.1,Iris-setosa
5.8,4.0,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
5.4,3.9,1.3,0.4,Iris-setosa
5.1,3.5,1.4,0.3,Iris-setosa
5.7,3.8,1.7,0.3,Iris-setosa
5.1,3.8,1.5,0.3,Iris-setosa
5.4,3.4,1.7,0.2,Iris-setosa
5.1,3.7,1.5,0.4,Iris-setosa
4.6,3.6,1.0,0.2,Iris-setosa
5.1,3.3,1.7,0.5,Iris-setosa
4.8,3.4,1.9,0.2,Iris-setosa
5.0,3.0,1.6,0.2,Iris-setosa
5.0,3.4,1.6,0.4,Iris-setosa
```

b. Removing the Values in the dataset where petalWidth = 1.4

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose RemoveWithValues -S 1.41 -C 3 -L last Apply Stop

Current relation: Relation: iris-weka.filters.unsupervised.instance.RemoveWithValues-S1.4-C3-Lfirst-Hast... Instances: 127 Attributes: 5 Sum of weights: 127

Attributes:

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Remove

Selected attribute:

Name: sepallength
Missing: 0 (0%)
Distinct: 32
Type: Numeric
Unique: 9 (7%)

Statistic	Value
Minimum	4.6
Maximum	7.9
Mean	6.015
StdDev	0.767

Class: class (Nom)

Visualize All

Status: OK Log x 0

4:26 PM 09-Feb-21

b. The automatic generation of co-relation coefficient between two random variables.

```
def correlationCoefficient(X, Y, n) :  
    sum_X = 0  
    sum_Y = 0  
    sum_XY = 0  
    squareSum_X = 0  
    squareSum_Y = 0  
  
    i = 0  
    while i < n :  
        sum_X = sum_X + X[i]  
        sum_Y = sum_Y + Y[i]  
        sum_XY = sum_XY + X[i] * Y[i]  
        squareSum_X = squareSum_X + X[i] * X[i]  
        squareSum_Y = squareSum_Y + Y[i] * Y[i]  
  
        i = i + 1  
  
    corr = (float)(n * sum_XY - sum_X * sum_Y) /  
    (float)(math.sqrt((n * squareSum_X -  
    sum_X * sum_X) * (n * squareSum_Y -  
    sum_Y * sum_Y)))  
    return corr
```