S. B. JAIN INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NAGPUR.

Practical No. 1

Aim: Create a Weather Table with the help of Data Mining Tool WEKA andApply Pre-Processing techniques(Add, Remove & Normalization) to the training data set of Weather Table.

Name of Student: Shrutika Pradeep Bagdi

Roll No.: CS22130

Semester/Year: 6th / 3rd

Academic Session: 2024-2025

Date of Performance:

Date of Submission:

AIM: Create a Weather Table with the help of Data Mining Tool WEKA and Apply Pre-Processing techniques(Add, Remove & Normalization) to the training data set of Weather Table.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- Key concepts and techniques in data preprocessing.
- Importance of data preprocessing in data mining.
- Define and understand data cleaning, data integration, data transformation, and feature selection.
- Implement data preprocessing in machine learning.

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement:

Software Requirement:

THEORY:

Data Preprocessing andit's Important.

Data preprocessing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying machine learning or data mining algorithms.

Applying data mining algorithms on this noisy data would not give quality results as they would fail to identify patterns effectively. Data Processing is, therefore, important to improve the overall data quality.

- Duplicate or missing values may give an incorrect view of the overall statistics of data.
- Outliers and inconsistent data points often tend to disturb the model's overall learning, leading to false predictions.

Description:

We need to create a Weather table with training data set which includes attributes like outlook, temperature, humidity, windy, play.

Procedure:

Steps 1) Type the following training data set with the help of Notepad for Weather Table.

@relation weather
@attribute outlook {sunny,rainy,overcast}
@attribute temparature numeric
@attribute humidity numeric
@attribute windy {true,false}
@attribute play {yes,no**}**

@data
sunny,85.0,85.0,false,no
overcast,80.0,90.0,true,no
sunny,83.0,86.0,false,yes
rainy,70.0,86.0,false,yes
rainy,68.0,80.0,false,yes
rainy,65.0,70.0,true,no
overcast,64.0,65.0,false,yes
sunny,72.0,95.0,true,no
sunny,69.0,70.0,false,yes
rainy,75.0,80.0,false,yes

Add ->Pre-Processing Technique:
Step

- Start ->Programs ->Weka
- Click on explorer.
- Click on open file.
- Select Weather.arff file and click on open.
- Click on Choose button and select the Filters option.
- In Filters, we have Supervised and Unsupervised data.
- Click on Unsupervised data.
- Select the attribute **Add.**
- In that we enter attribute index, type, data format, nominal label values for **Climate**.
- Click on OK.
- Press the Apply button, then a new attribute is added to the Weather Table.
- Click on the Edit button, it shows a new Weather Table on Weka.

Remove Pre-Processing Technique:

Procedure:

1) Start ->Programs ->Weka-

2) Click on explorer.

3) Click on open file.

4) Select Weather.arff file and click on open.

5) Click on Choose button and select the Filters option.

6) In Filters, we have Supervised and Unsupervised data.

7) Click on Unsupervised data.

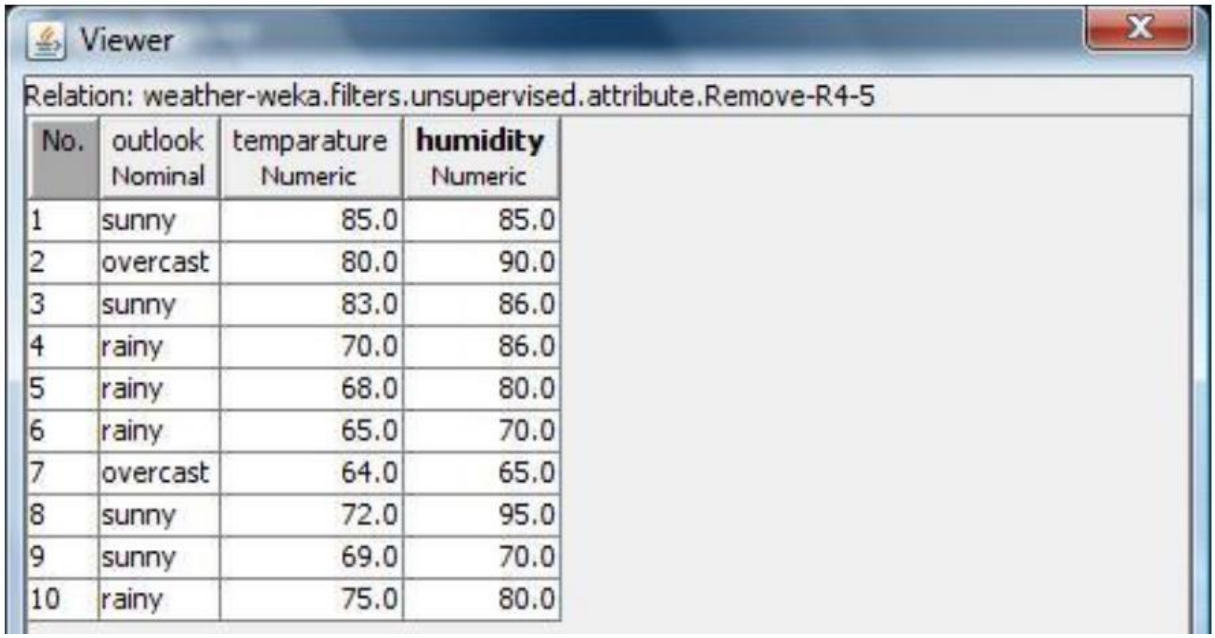
8) Select the attribute Remove.

9) Select the attributes windy, play to Remove.

10) Click Remove button

11) Click on the Edit button, it shows a new Weather Table on Weka

Weather Table after removing attributes WINDY, PLAY



No.	outlook Nominal	temparature Numeric	**humidity** Numeric
1	sunny	85.0	85.0
2	overcast	80.0	90.0
3	sunny	83.0	86.0
4	rainy	70.0	86.0
5	rainy	68.0	80.0
6	rainy	65.0	70.0
7	overcast	64.0	65.0
8	sunny	72.0	95.0
9	sunny	69.0	70.0
10	rainy	75.0	80.0

Normalize ->Pre-Processing Technique:

Procedure:

1) Start ->Programs ->Weka

2) Click on explorer.

3) Click on open file.

4) Select Weather.arff file and click on open.

5) Click on Choose button and select the Filters option.

6) In Filters, we have Supervised and Unsupervised data.

7) Click on Unsupervised data.

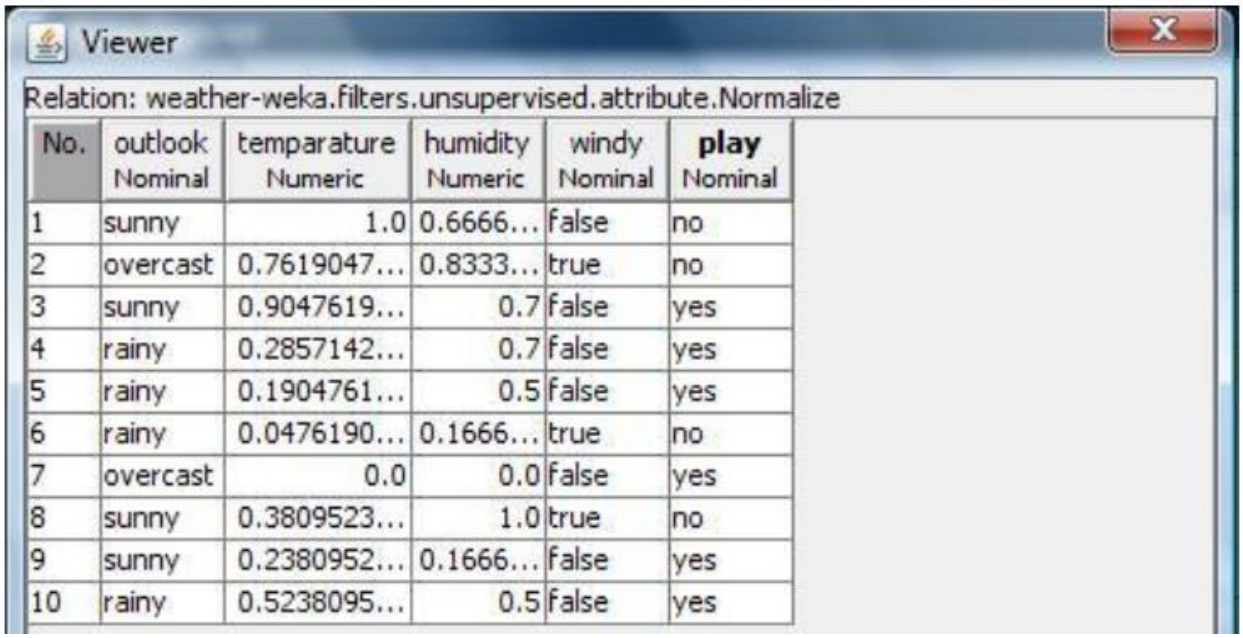
8) Select the attribute Normalize.

9) Select the attributes temparature, humidity to Normalize.

10) Click on Apply button

11) Click on the Edit button, it shows a new Weather Table with normalized values on Weka.

Weather Table after Normalizing TEMPARATURE, HUMIDITY

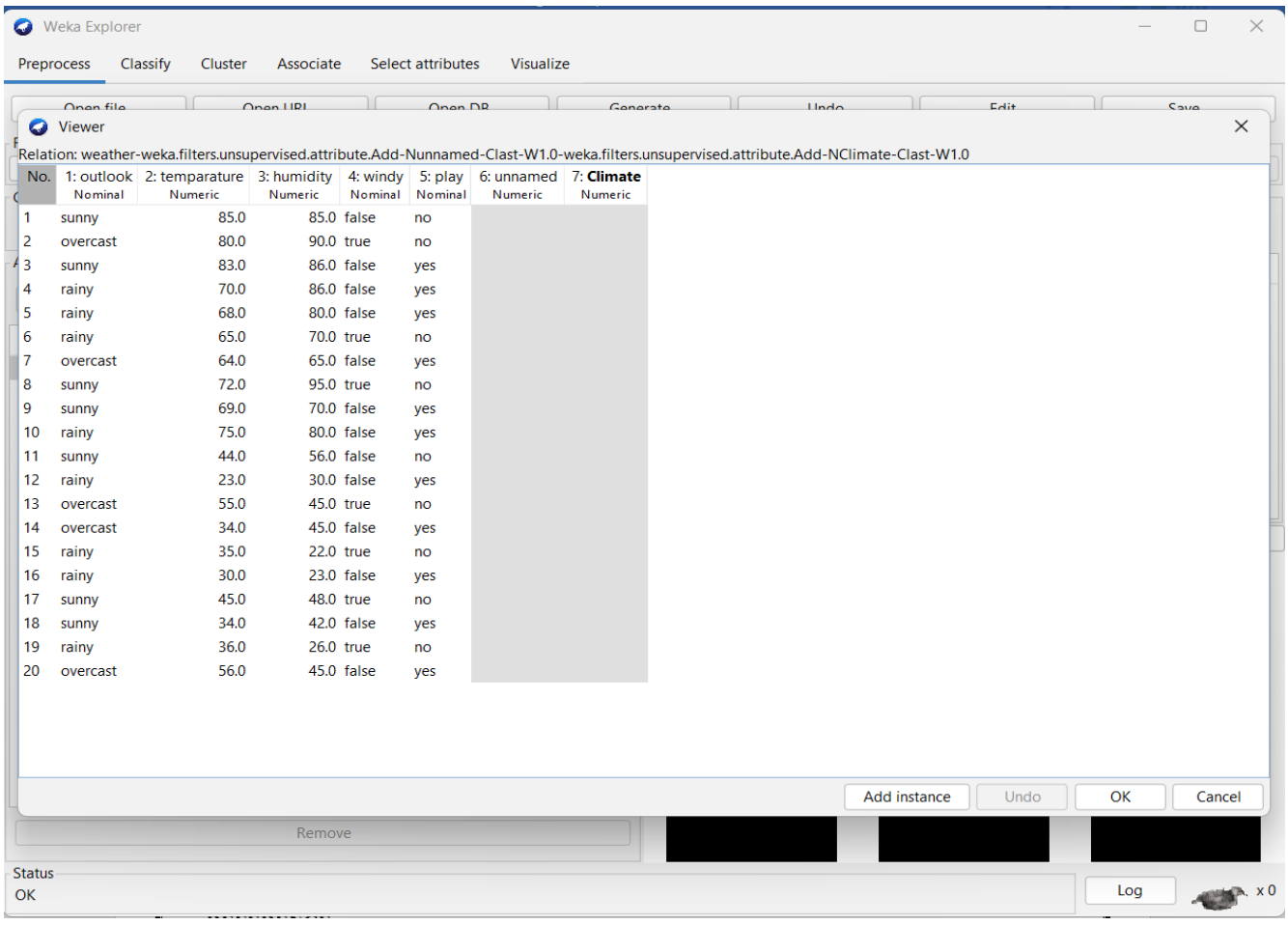


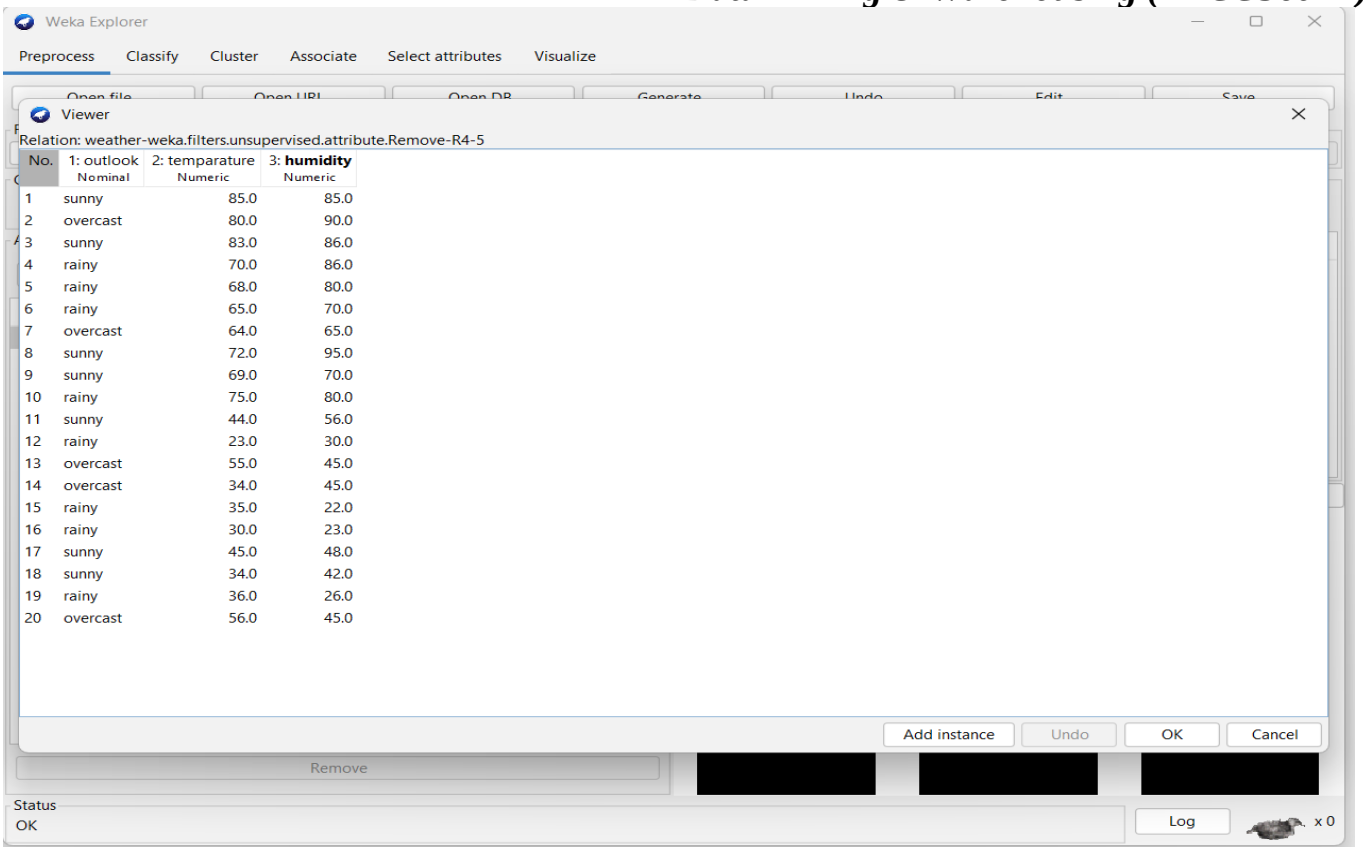
No.	outlook Nominal	temparature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	1.0	0.6666...	false	no
2	overcast	0.7619047...	0.8333...	true	no
3	sunny	0.9047619...	0.7	false	yes
4	rainy	0.2857142...	0.7	false	yes
5	rainy	0.1904761...	0.5	false	yes
6	rainy	0.0476190...	0.1666...	true	no
7	overcast	0.0	0.0	false	yes
8	sunny	0.3809523...	1.0	true	no
9	sunny	0.2380952...	0.1666...	false	yes
10	rainy	0.5238095...	0.5	false	yes

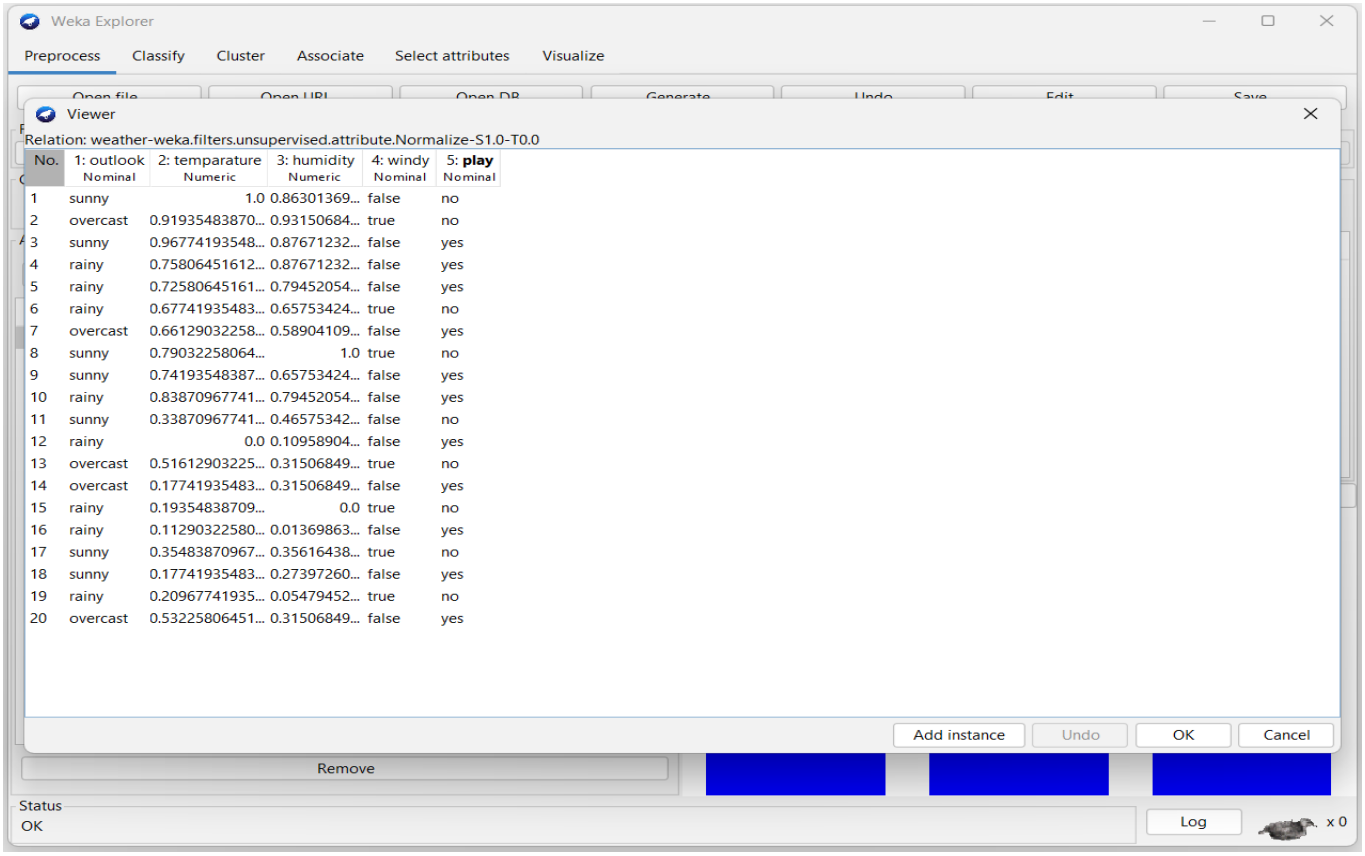
OUTPUT (SCREENSHOTS)

```
@relation weather
@attribute outlook {sunny,rainy,overcast}
@attribute temparature numeric
@attribute humidity numeric
@attribute windy {true,false}
@attribute play {yes,no}

@data
sunny,85.0,85.0,false,no
overcast,80.0,90.0,true,no
sunny,83.0,86.0,false,yes
rainy,70.0,86.0,false,yes
rainy,68.0,80.0,false,yes
rainy,65.0,70.0,true,no
overcast,64.0,65.0,false,yes
sunny,72.0,95.0,true,no
sunny,69.0,70.0,false,yes
rainy,75.0,80.0,false,yes
sunny,44.0,56.0,false,no
rainy,23.0,30.0,false,yes
overcast,55.0,45.0,true,no
overcast,34.0,45.0,flase,yes
rainy,35.0,22.0,true,no
rainy,30.0,23.0,false,yes
sunny,45.0,48.0,true,no
sunny,34.0,42.0,flase,yes
rainy,36.0,26.0,true,no
overcast,56.0,45.0,false,yes
```




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Viewer
Relation: weather-weka.filters.unsupervised.attribute.Remove-R4-5
No. 1: outlook Nominal 2: temparature Numeric 3: humidity Numeric
1 sunny 85.0 85.0
2 overcast 80.0 90.0
3 sunny 83.0 86.0
4 rainy 70.0 86.0
5 rainy 68.0 80.0
6 rainy 65.0 70.0
7 overcast 64.0 65.0
8 sunny 72.0 95.0
9 sunny 69.0 70.0
10 rainy 75.0 80.0
11 sunny 44.0 56.0
12 rainy 23.0 30.0
13 overcast 55.0 45.0
14 overcast 34.0 45.0
15 rainy 35.0 22.0
16 rainy 30.0 23.0
17 sunny 45.0 48.0
18 sunny 34.0 42.0
19 rainy 36.0 26.0
20 overcast 56.0 45.0
Add instance
Undo
OK
Cancel
Remove
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Viewer
Relation: weather-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0
No. 1: outlook Nominal 2: temparature Numeric 3: humidity Numeric 4: windy Nominal 5: play Nominal
1 sunny 1.0 0.86301369... false no
2 overcast 0.91935483870... 0.93150684... true no
3 sunny 0.96774193548... 0.87671232... false yes
4 rainy 0.75806451612... 0.87671232... false yes
5 rainy 0.72580645161... 0.79452054... false yes
6 rainy 0.67741935483... 0.65753424... true no
7 overcast 0.66129032258... 0.58904109... false yes
8 sunny 0.79032258064... 1.0 true no
9 sunny 0.74193548387... 0.65753424... false yes
10 rainy 0.83870967741... 0.79452054... false yes
11 sunny 0.33870967741... 0.46575342... false no
12 rainy 0.0 0.10958904... false yes
13 overcast 0.51612903225... 0.31506849... true no
14 overcast 0.17741935483... 0.31506849... false yes
15 rainy 0.19354838709... 0.0 true no
16 rainy 0.11290322580... 0.01369863... false yes
17 sunny 0.35483870967... 0.35616438... true no
18 sunny 0.17741935483... 0.27397260... false yes
19 rainy 0.20967741935... 0.05479452... true no
20 overcast 0.53225806451... 0.31506849... false yes
Add instance
Undo
OK
Cancel
Remove
Status
OK
Log
x 0

CONCLUSION:

DISCUSSION AND VIVA VOCE:

- What are the data preprocessing steps in order?
- What is the meaning of data cleansing?
- What is the difference between data mining and data preprocessing?

REFERENCE:

- www.cs.sfu.ca/~han/DMbook.html
- http://www.cs.waikato.ac.nz/ml/weka/documentation.html
- http://ai.fon.bg.ac.rs/wp-content/uploads/2015/04/ML-Attribute-Discretisation-and-Selection-Clustering-2014_eng.pdf

Observation book: (3)	**Viva-Voce (3)**	**`Quality of Submission and timely Evaluation (4)**
Total:		**Sign with date:**