

**S.THARUN**

**19MID0031**

**CSI3010 – DATA WAREHOUSING AND DATA MINING**

**FACULTY : CHELLATAMILAN T**

PRE-PROCESSING TECHNIQUES USING WEKA

# CREATING WEATHER DATASET IN .ARFF FORMAT

1) Open Start ->Programs -> Accessories ->Notepad

2) **Type the following** training data set with the help of Notepad for Weather Table.

Text

Description automatically generated

# LOAD AND VIEW DATA

3) After that the file is saved with .arff file format.

4) Minimize the arff -3-4.

5) Click on weka-3-4, then Weka dialog box is displayed on the screen.

6) In that dialog box there are four modes, click on explorer.

7) Explorer shows many options. In that click on ‘open file’ and select the arff file

8) **Click on edit button which shows weather table on weka**

Table

Description automatically generated

# ADD ATTRIBUTE

1. Click “Choose” in filter
2. Select weka.filters.unsupervised.attribute.Add
3. Enter details as given below and give “ok” and then apply the filter

Graphical user interface, application

Description automatically generated

**Entering each instance values manually**

A screenshot of a computer

Description automatically generated

# ADD INSTANCES

1. Click “Edit”
2. Click “Add Instance”

Table

Description automatically generated

1. Another instance automatically adds in the view
2. Modify instance value accordingly

Graphical user interface, table

Description automatically generated

# REMOVE ATTRIBUTE

1. Click “Edit”
2. Right click the attribute which you want to delete
3. Select delete attribute

A screenshot of a computer

Description automatically generated

1. We can also do this by filters. Click “Choose” in filters
2. Select weka.filters.unsupervised.attribute.Remove
3. Give the index of the column to be deleted
4. Click “OK” and then click “Apply” in filter

Graphical user interface, text, application

Description automatically generated

**Data after removing attribute**

Table

Description automatically generated

# REMOVING INSTANCE

1. Click Edit and select the instance to be deleted
2. Right click the instance and select “Delete selected Instance”

Graphical user interface, application, Word

Description automatically generated

**4TH attribute removed**

Table

Description automatically generated

# NORMALIZE

Graphical user interface, text, application, email

Description automatically generated

**Output window**

Table

Description automatically generated

# HANDLING MISSING VALUES

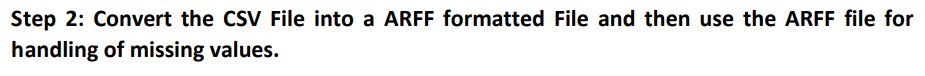
Graphical user interface, text

Description automatically generated

**The original .csv file**

Text

Description automatically generated



**Altered weather .csv file**

Text

Description automatically generated

Graphical user interface, text, application, letter

Description automatically generated

**Zero values in BMI Column**

Table

Description automatically generated

**Numerical cleaner filter**

Graphical user interface, application

Description automatically generated

**Clicked BMI column to sort in ascending order**

(Nan values are placed at last in ascending order)

Table

Description automatically generated

Text

Description automatically generated

**Removewithvalues Filter**

Graphical user interface, application

Description automatically generated

**BMI column Sorted in ascending order**

(But no missing values left)

Graphical user interface, table

Description automatically generated

**Before removing**

Graphical user interface, text, application

Description automatically generated

**After removing**

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

**Replaced NaN instances with mean**

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated