**Ex.No.1 Build Data Warehouse and Explore WEKA**

Aim:

To build data warehouse and explore using WEKA tool.

**Introduction**

Weka (pronounced to rhyme with Mecca ) is a work bench that contains a collectionofvisualizationtoolsandalgorithmsfordataanalysisandpredictivemodeling,togetherwithgraphical user interfaces for easy access to these functions. The original non-Java version ofWeka was a Tcl/Tk front-end to (mostly third-party) modeling algorithms implemented in otherprogramming languages, plus data preprocessing utilities in C, and Make file-based system forrunning machine learning experiments. This original version was primarily designed as a tool foranalyzing data from agricultural domains, but the more recent fully Java-based version (Weka 3),for which development started in 1997, is now used in many different application areas, inparticularforeducational purposesandresearch.AdvantagesofWekainclude:

* FreeavailabilityundertheGNUGeneralPublicLicense.
* Portability,sinceitisfullyimplementedintheJavaprogramminglanguageandthusrunsonalmost anymodern computingplatform
* Acomprehensivecollectionofdatapreprocessingandmodelingtechniques
* Easeofuseduetoitsgraphicaluserinterfaces

## Description:

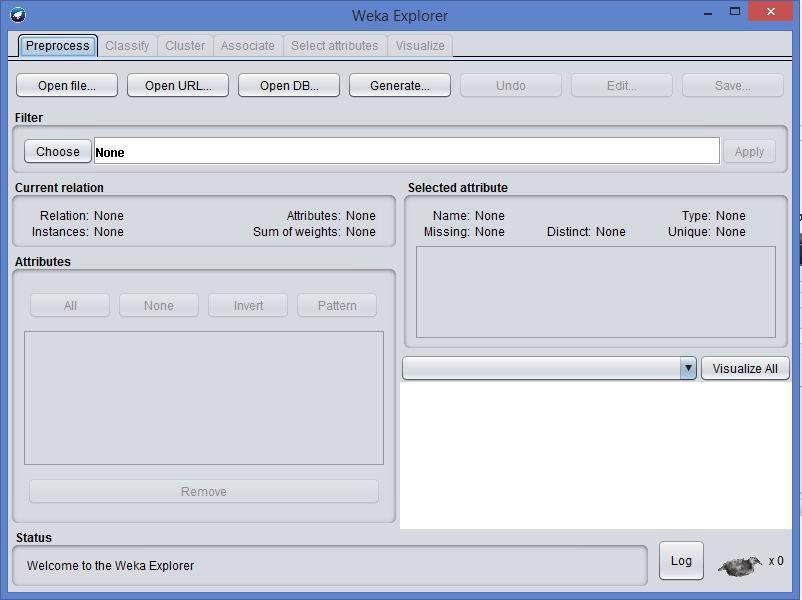
Open the program. Once the program has been loaded on the user‟s machine it is opened bynavigating to the programs start option and that will depend on the user‟s operating system.Figure1.1 isanexampleoftheinitialopeningscreen onacomputer.

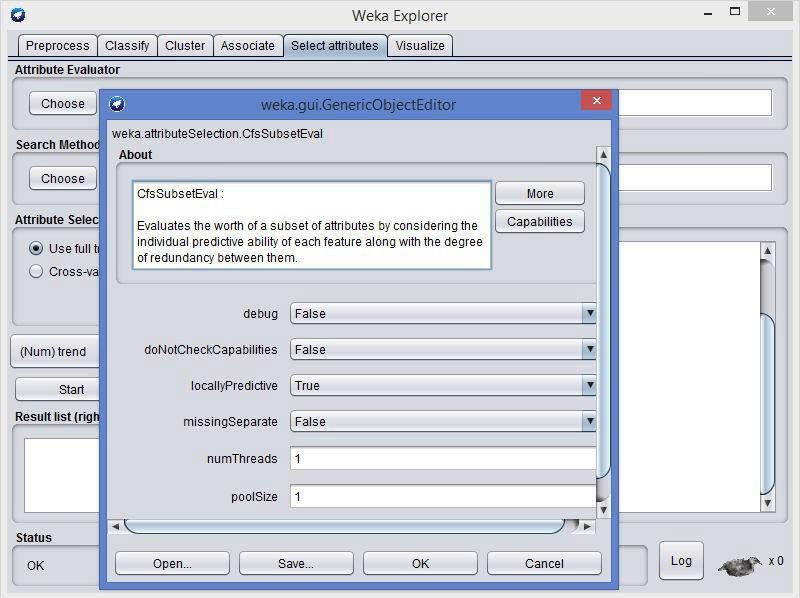
Therearefouroptionsavailableonthisinitialscreen:

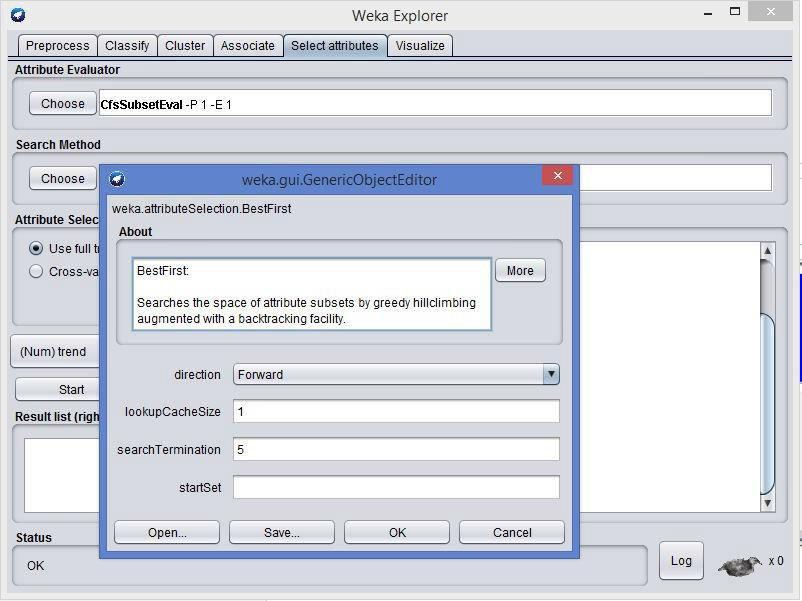
****

Fig:1.1WekaGUI

**1.Explorer**-thegraphicalinterfaceusedtoconductexperimentationonrawdataAfterclickingtheExplorerbuttonthewekaexplorerinterfaceappears.

Fig:1.2Pre-processor

****

****Insidethewekaexplorerwindowtherearesixtabs:

1. **Preprocess-**usedtochoosethedatafiletobeusedbytheapplication.

**Open File**- allows for the user to select files residing on the local machine or recorded medium**Open URL**- provides a mechanism to locate a file or data source from a different locationspecifiedbytheuser

**OpenDatabase**-allowstheusertoretrievefilesordatafromadatabasesourceprovidedbyuser

1. **Classify-** used to test and train different learning schemes on the preprocessed data file underexperimentation

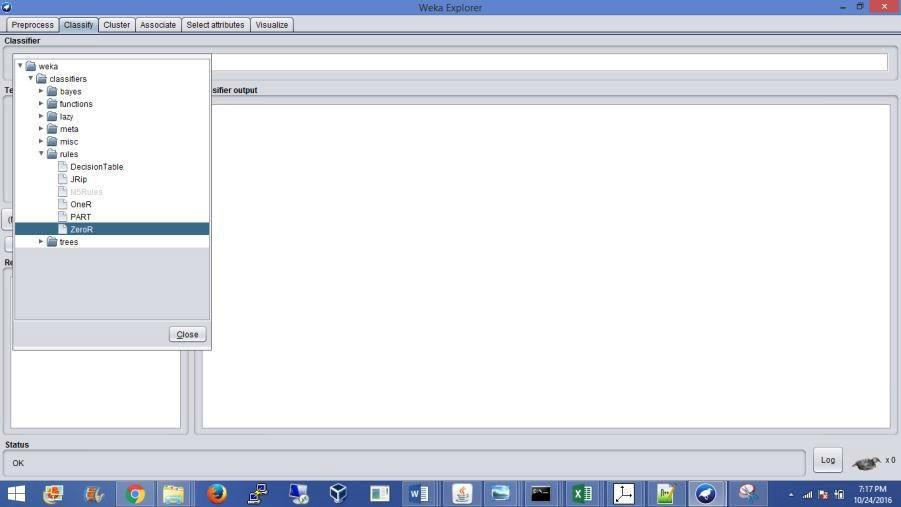
****

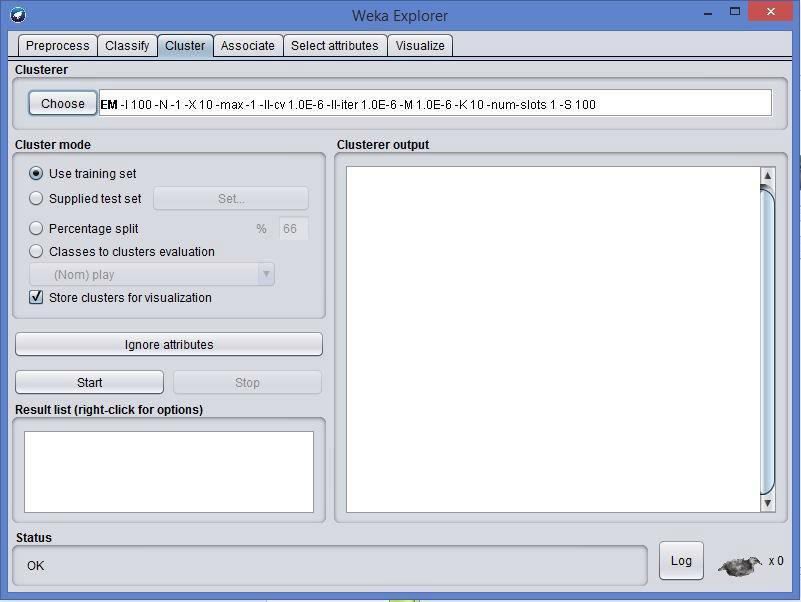
Fig:1.3choosingZerosetfromclassify

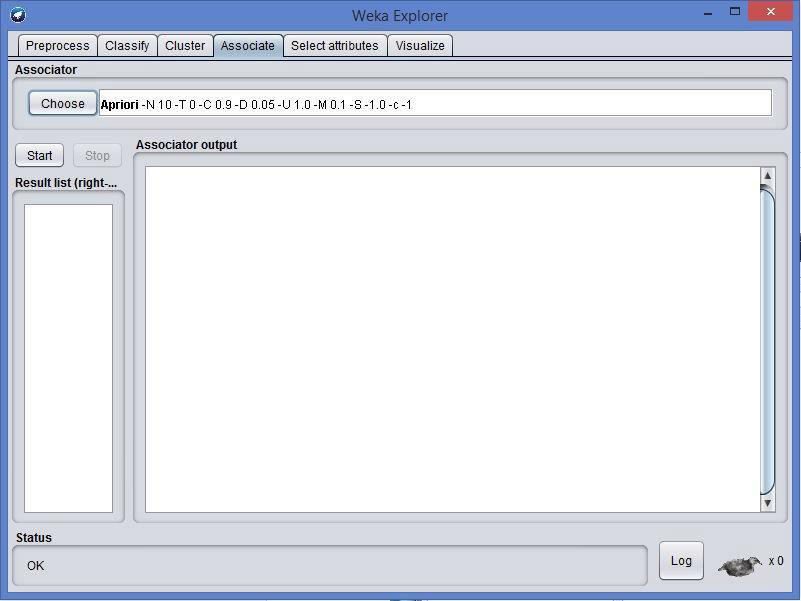
Again there are several options to be selected inside of the classify tab. Test option gives the userthechoiceofusingfourdifferenttest modescenarioson thedataset.

* 1. Usetrainingset
  2. Suppliedtrainingset
  3. Crossvalidation
  4. Splitpercentage

1. **Cluster-**usedtoapplydifferenttoolsthatidentifyclusterswithinthedatafile.

TheClustertab opens theprocess thatisusedtoidentifycommonalties orclustersofoccurrenceswithinthedataset andproduceinformationfortheuserto analyze.

****

1. **Association-**usedtoapplydifferentrulestothedatafilethatidentifyassociationwithinthedata.Theassociatetabopensawindowtoselecttheoptionsforassociationswithinthedataset.
2.  **Selectattributes-**usedtoapplydifferentrulestorevealchangesbasedonselectedattributesinclusionorexclusionfrom theexperiment
3. **Visualize-**usedtoseewhatthevariousmanipulationproducedonthedatasetina2Dformat,inscatterplot andbargraph output.
4. **Experimenter** - this option allows users to conduct different experimental variations on datasets and perform statistical manipulation. The Weka Experiment Environment enables the user tocreate, run, modify, and analyze experiments in a more convenient manner than is possible whenprocessing the schemes individually. For example, the user can create an experiment that runsseveral schemes against a series of datasets and then analyze the results to determine if one of theschemesis(statistically)betterthantheotherschemes.

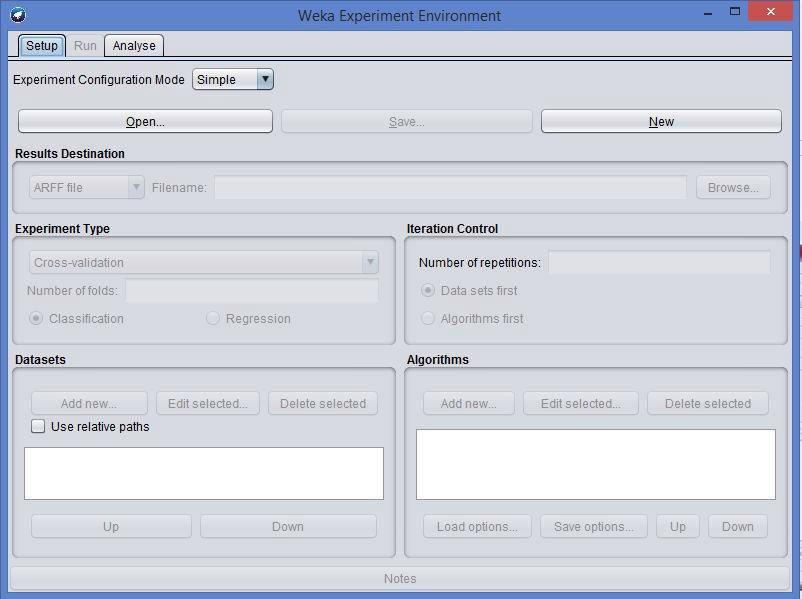


Fig:1.6Wekaexperiment

**Resultsdestination**:ARFFfile,CSVfile,JDBCdatabase.

**Experimenttype**:Cross-validation(default),Train/TestPercentageSplit(datarandomized).

**Iterationcontrol**:Numberofrepetitions,Datasetsfirst/Algorithmsfirst.

**Algorithms**:filters

1. **Knowledge Flow**-basically the same functionality as Explorer with drag and drop functionality. The advantage of this option is that it supports incremental learning from previousresults
2. **Simple CLI**-provides users without a graphic interface option the ability to execute commands from a terminal window.

## b. Explore the default datasets in weka tool.

Click the “***Open file…***” button to open a data set and double click on the “***data***” directory.Wekaprovidesanumberofsmallcommonmachinelearningdatasetsthatyoucanusetopracticeon.Selectthe“***iris.arff***”filetoloadtheIrisdataset.

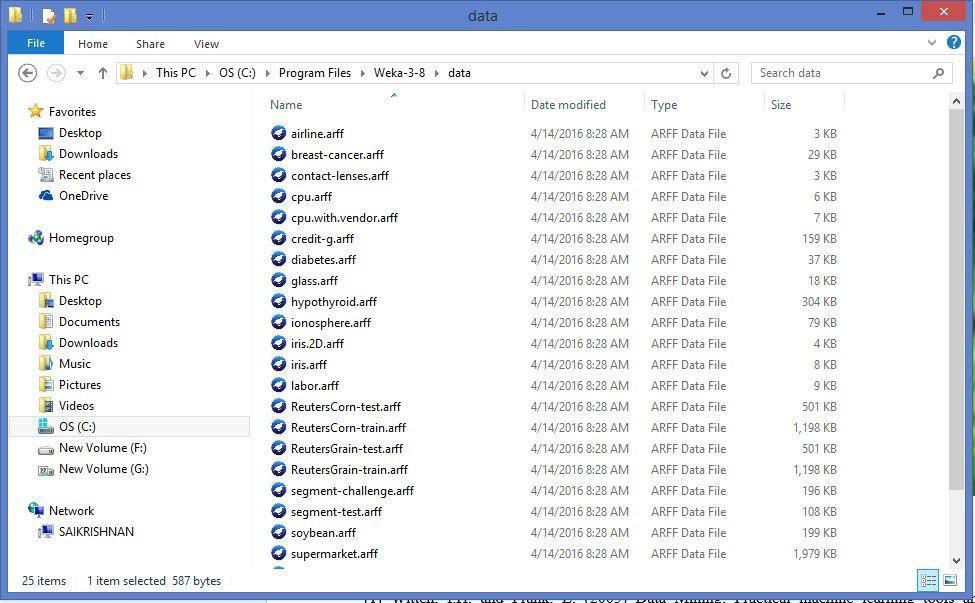


Fig:1.7DifferentDataSetsinweka

**Result:**

Thus the building data warehouse and explore using WEKA tool was executed successfully.

**Ex.No.2aPerform data preprocessing tasks and demonstrate**

**performing association rule mining on data sets.**

**Aim:**

To apply Pre-Processing techniques to the training data set of Employee Table

**Description:**

Real world databases are highly influenced to noise, missing and inconsistency due to their queue size so the data can be pre-processed to improve the quality of data and missing results and it also improves the efficiency.

There are 3 pre-processing techniques they are:

**1)** Add

**2)** Remove

**3)** Normalization

**Creation of Employee Table:**

**Procedure:**

**1)** Open Start Programs Accessories Notepad

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

@relation employee

@attribute name {x,y,z,a,b}

@attribute id numeric

@attribute salary {low,medium,high}

@attribute exp numeric

@attribute gender {male,female}

@attribute phone numeric

@data

x,101,low,2,male,250311

y,102,high,3,female,251665

z,103,medium,1,male,240238

a,104,low,5,female,200200

b,105,high,2,male,240240

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

**4)** Minimize the arff file and then open Start Programs weka-3-4.

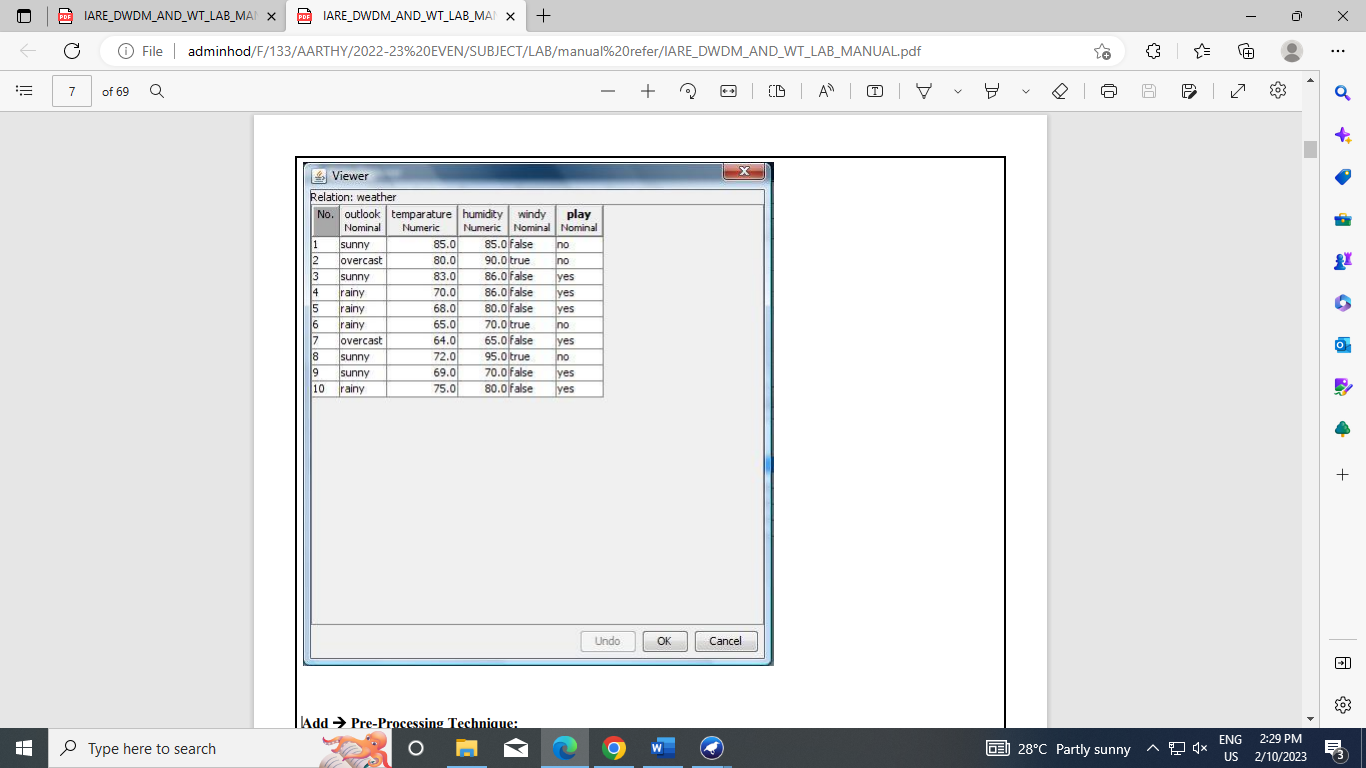
**5)** Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

**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 employee table on weka.

**Training Data Set Employee Table**



**AddPre-Processing Technique:**

**Procedure:**

**1)** Start Programs Weka-3-4 Weka-3-4

**2)** Click on **explorer.**

**3)** Click on **open file.**

**4)** Select **Employee.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 **Add**.

**9)** A new window is opened.

**10)** In that we enter attribute index, type, data format, nominal label values for **Address**.

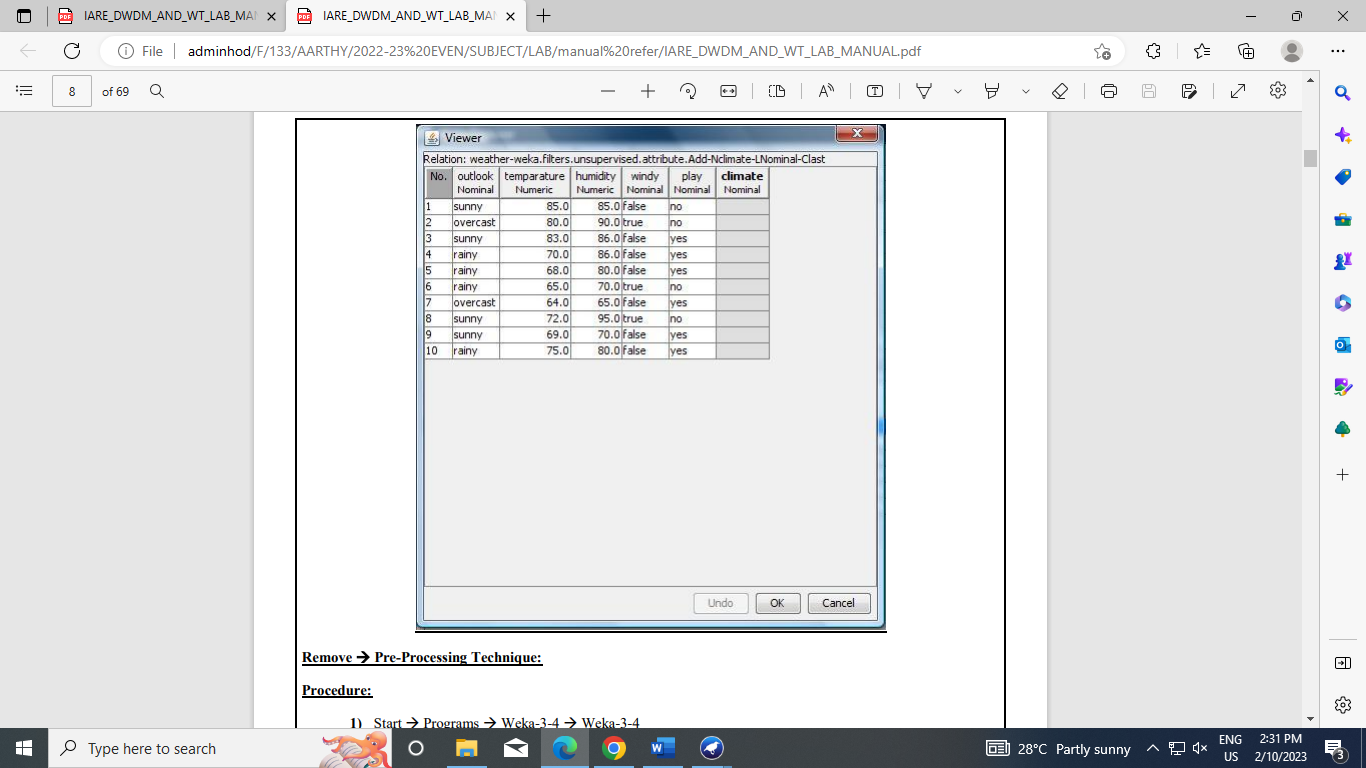
**11)** Click on **OK**.

**12)** Press the **Apply button**, then a new attribute is added to the Employee Table.

**13) Save** the file.

**14)** Click on the **Edit button**, it shows a new Employee Table on Weka.

**Weather Table after adding new attribute CLIMATE**:



Remove Pre-Processing Technique:

Procedure:

1) Start Programs Weka-3-4 Weka-3-4

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.

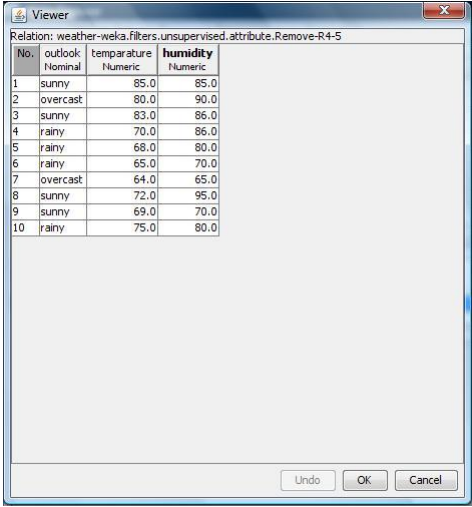
8) Select the attribute Remove.

9) Select the attributes windy, play to Remove.

10) Click Remove button and then Save.

11) Click on the Edit button, it shows a new Weather Table on Weka. 9 Weather Table after removing attributes

Weather Table after removing attributes WINDY, PLAY:



Normalize Pre-Processing Technique:

Procedure:

1) Start Programs Weka-3-4 Weka-3-4

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.

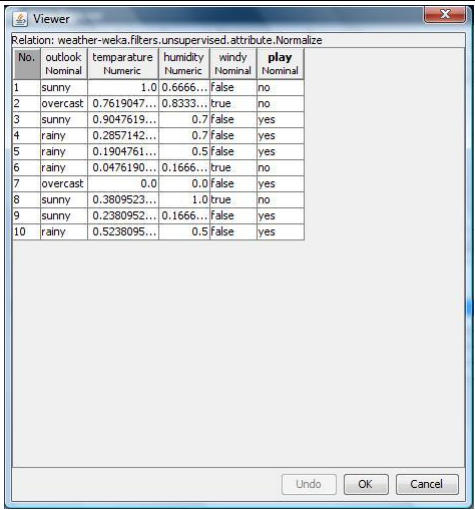
8) Select the attribute Normalize.

9) Select the attributes temparature, humidity to Normalize.

10) Click on Apply button and then Save.

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

Weather Table after Normalizing TEMPARATURE, HUMIDITY:



**Result:**

This program has been successfully executed.

**Ex.No.2b Association Rules for Buying data.**

**Aim:** ToFinding Association Rules for Buying data.

**Description:**

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It can be described as analyzing and presenting strong rules discovered in databases using different measures of interestingness. In market basket analysis association rules are used and they are also employed in many application areas including Web usage mining, intrusion detection and bioinformatics.

**Creation of Buying Table:**

Procedure:

1) Open Start Programs Accessories Notepad

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

@relation buying

@attribute age {L20,20-40,G40}

@attribute income {high,medium,low}

@attribute stud {yes,no}

@attribute creditrate {fair,excellent} \

@attribute buyscomp {yes,no}

@data

L20,high,no,fair,yes

20-40,low,yes,fair,yes

G40,medium,yes,fair,yes

L20,low,no,fair,no

G40,high,no,excellent,yes

L20,low,yes,fair,yes

20-40,high,yes,excellent,no

G40,low,no,fair,yes

L20,high,yes,excellent,yes

G40,high,no,fair,yes

L20,low,yes,excellent,no

G40,high,yes,excellent,no

20-40,medium,yes,excellent,yes

L20,medium,yes,fair,yes

G40,high,yes,excellent,yes

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

4) Minimize the arff file and then open Start Programs weka-3-4.

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

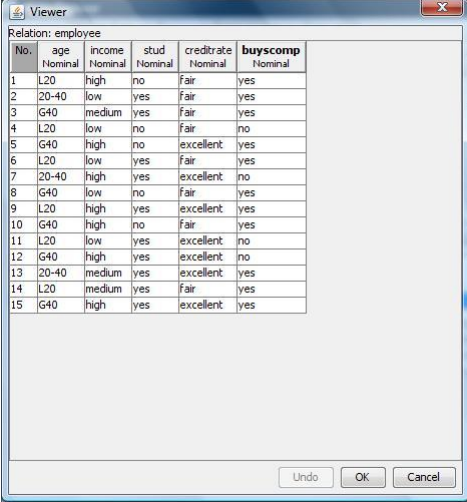
6) In that dialog box there are four modes, click on explorer.

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

8) Click on edit button which shows buying table on weka.

**Output:**

Training Data Set Buying Table



Procedure for Association Rules:

1) Open Start Programs Weka-3-4Weka-3-4

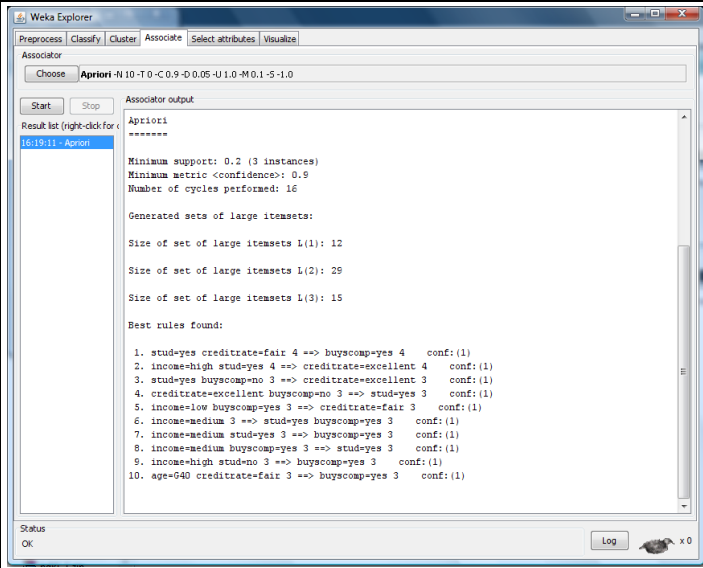
2) Open explorer.

3) Click on open file and select buying.arff

4) Select Associate option on the top of the Menu bar.

5) Select Choose button and then click on Apriori Algorithm.

6) Click on Start button and output will be displayed on the right side of the window

****

Result:

This program has been successfully executed.

**Ex.No.3 Demonstrate performing classification on data sets**

**Aim:**

To Construct Decision Tree for Weather data and classify it.

**Description:**

**Classification & Prediction:**

Classification is the process for finding a model that describes the data values and concepts for the purpose of Prediction.

**Decision Tree:**

A decision Tree is a classification scheme to generate a tree consisting of root node, internal nodes and external nodes.

Root nodes representing the attributes. Internal nodes are also the attributes. External nodes are the classes and each branch represents the values of the attributes

Decision Tree also contains set of rules for a given data set; there are two subsets in Decision Tree. One is a Training data set and second one is a Testing data set. Training data set is previously classified data. Testing data set is newly generated data.

**Creation of Weather Table:**

**Procedure:**

**1)** Open Start Programs Accessories Notepad

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

@relation weather

@attribute outlook {sunny, rainy, overcast}

@attribute temperature numeric

@attribute humidity numeric

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

@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

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

**4)** Minimize the arff file and then open Start Programs weka-3-4.

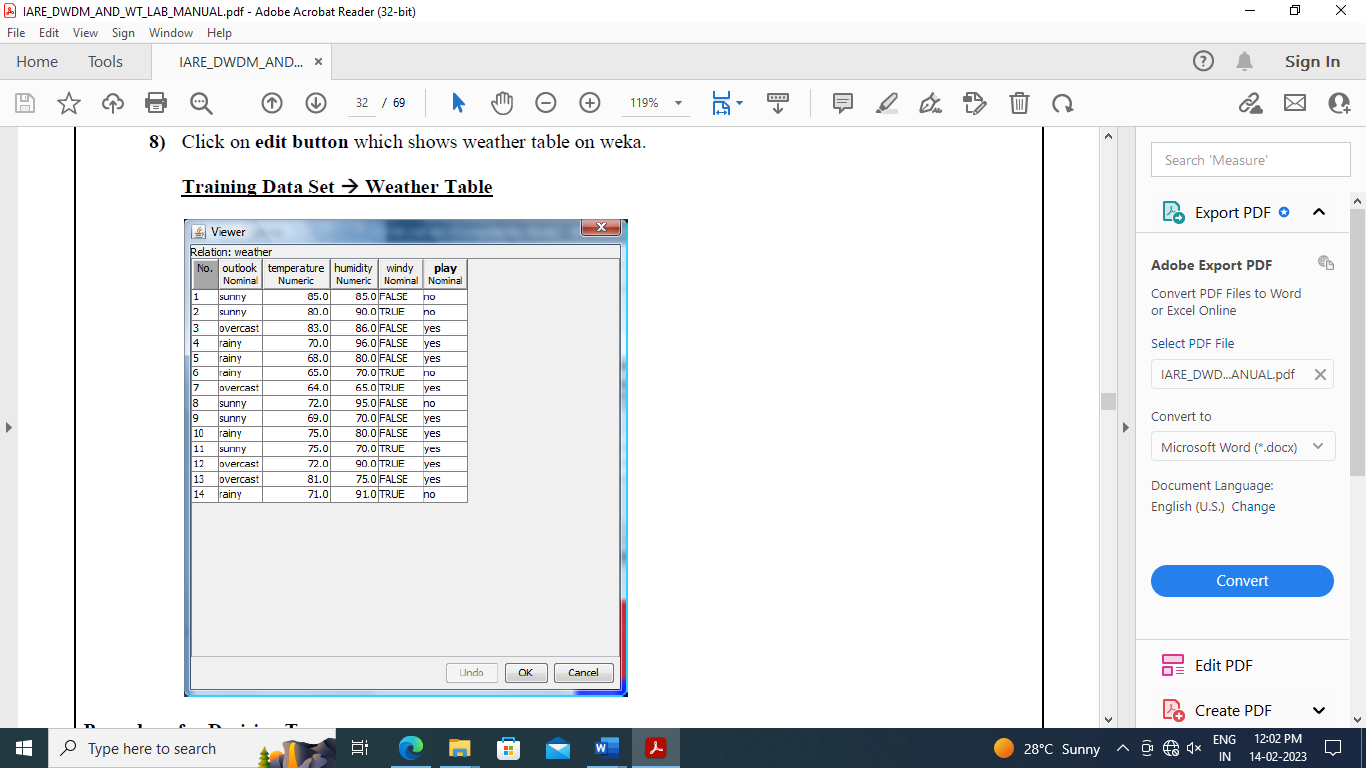
**5)** Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

**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.

**Training Data Set** **Weather Table**



**Procedure for Decision Trees:**

**1)** Open Start Programs Weka-3-4 Weka-3-4

**2)** Open **explorer**.

**3)** Click on **open file** and select **weather.arff**

**4)** Select **Classifier option** on the top of the Menu bar.

**5)** Select **Choose button** and click on **Tree option**.

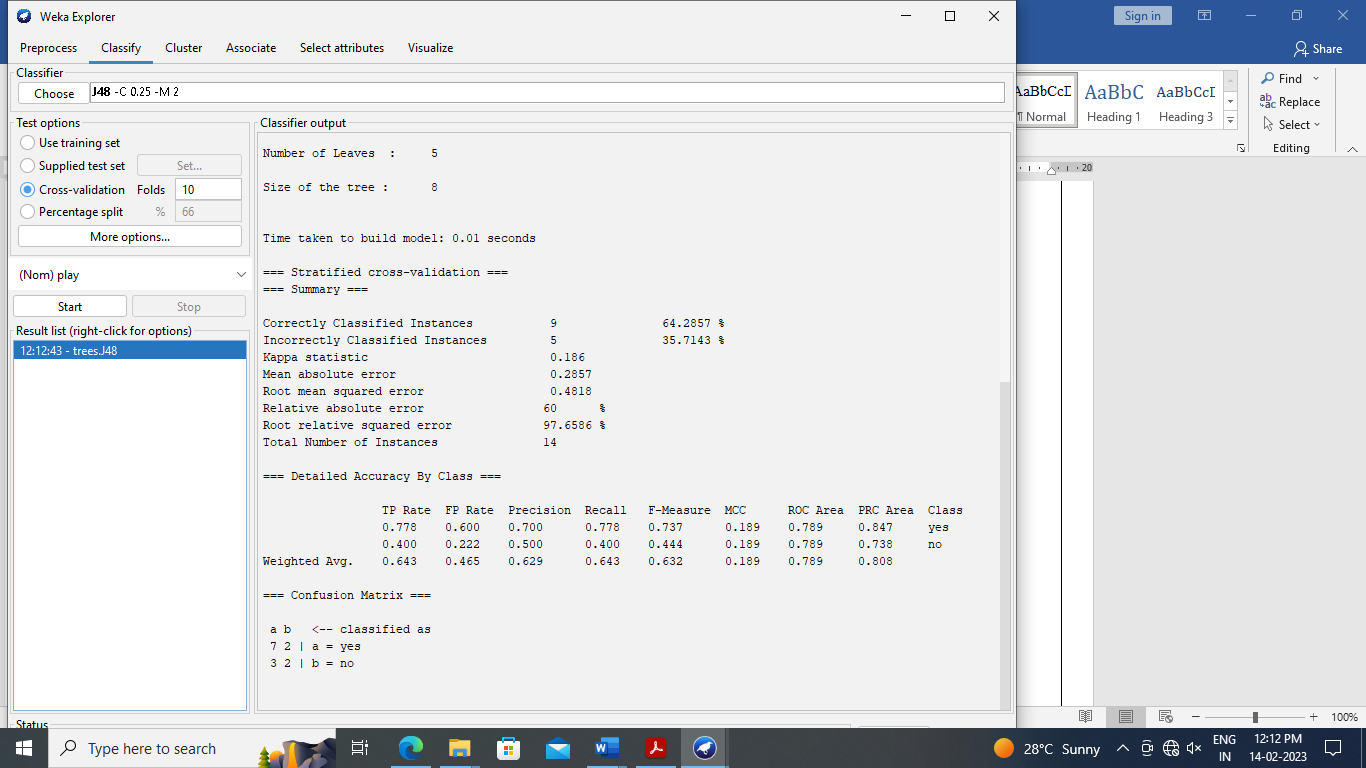
**6)** Click on **J48.**

**7)** Click on **Start button** and output will be displayed on the **right side** of the window.

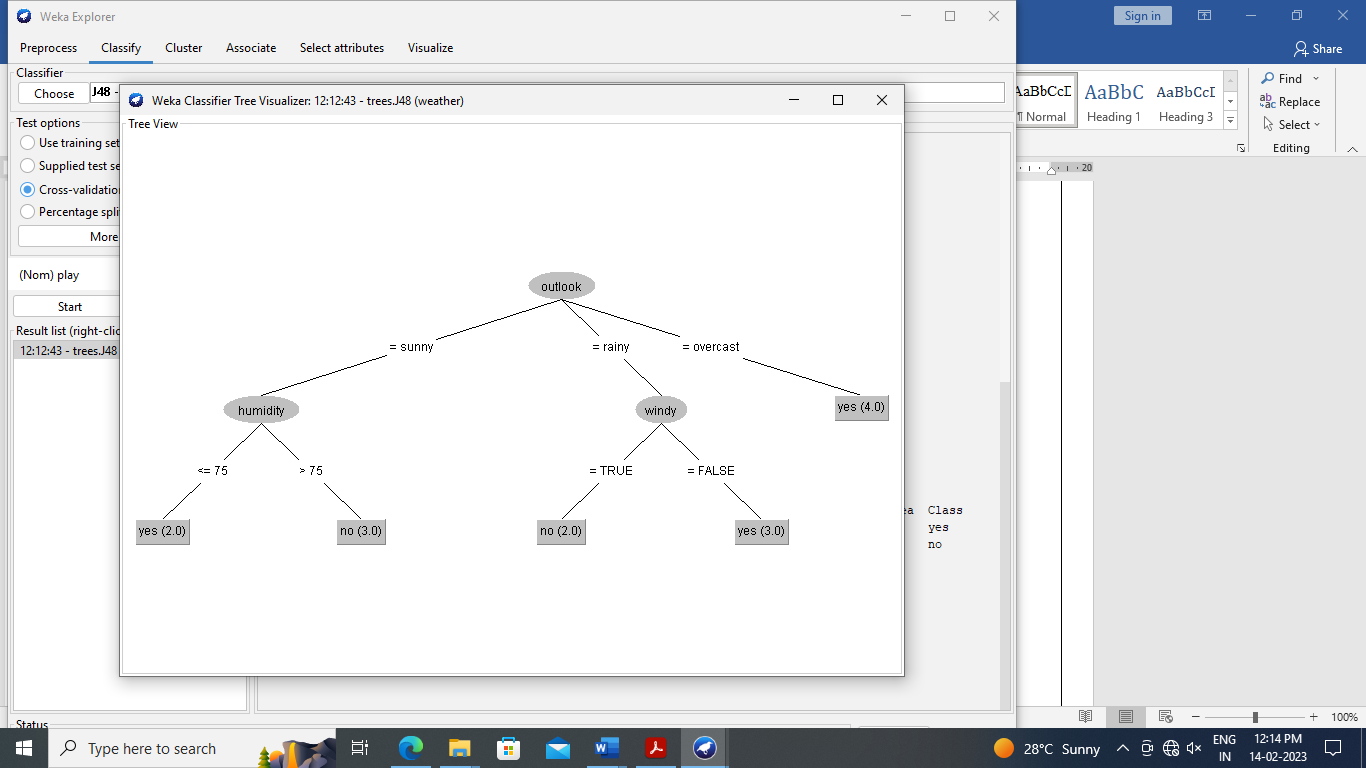
**8)** Select the **result list** and **right click** on result list and select **Visualize Tree option**.

**9)** Then **Decision Tree** will be displayed on **new window**.

**Output:**



**Decision Tree:**



**Result:**

This the program has been successfully executed.

**Ex.No.4 Demonstrate performing clustering on data sets**

Aim:

To write a procedure for Clustering Buying data using Cobweb Algorithm.

**Description:**

**Cluster analysis** or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

**Creation of Buying Table:**

**Procedure:**

**1)** Open Start Programs Accessories Notepad

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

@relation buying

@attribute age {L20,20-40,G40}

@attribute income {high,medium,low}

@attribute stud {yes,no}

@attribute creditrate {fair,excellent}

@attribute buyscomp {yes,no}

@data

L20,high,no,fair,yes

20-40,low,yes,fair,yes

G40,medium,yes,fair,yes

L20,low,no,fair,no

G40,high,no,excellent,yes

L20,low,yes,fair,yes

20-40,high,yes,excellent,no

G40,low,no,fair,yes

L20,high,yes,excellent,yes

G40,high,no,fair,yes

L20,low,yes,excellent,no

G40,high,yes,excellent,no

20-40,medium,yes,excellent,yes

L20,medium,yes,fair,yes

G40,high,yes,excellent,yes

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

**4)** Minimize the arff file and then open Start Programs weka-3-4.

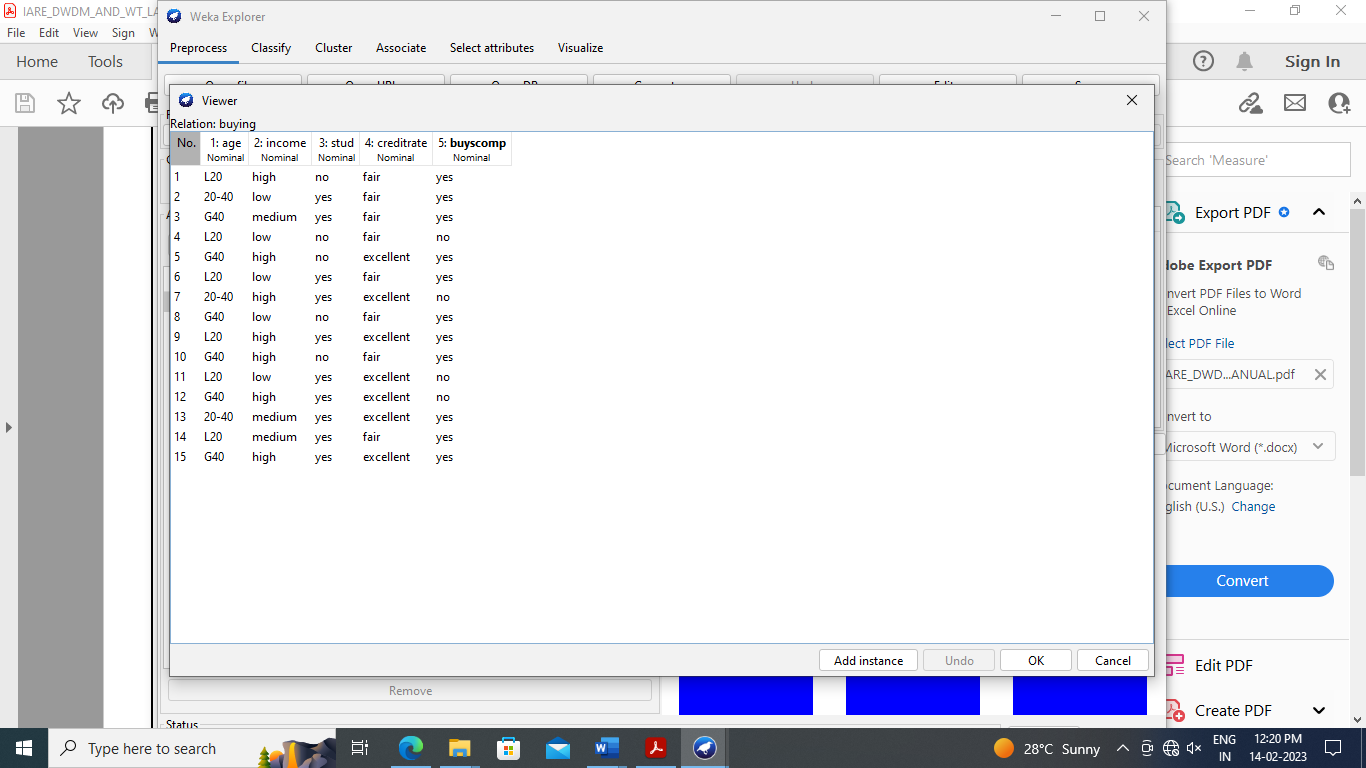
**5)** Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

**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 buying table on weka.

**Training Data Set** **Buying Table**



**Procedure:**

**1)** Click **Start** ->**Programs** ->**Weka 3.4**

**2)** Click on **Explorer**.

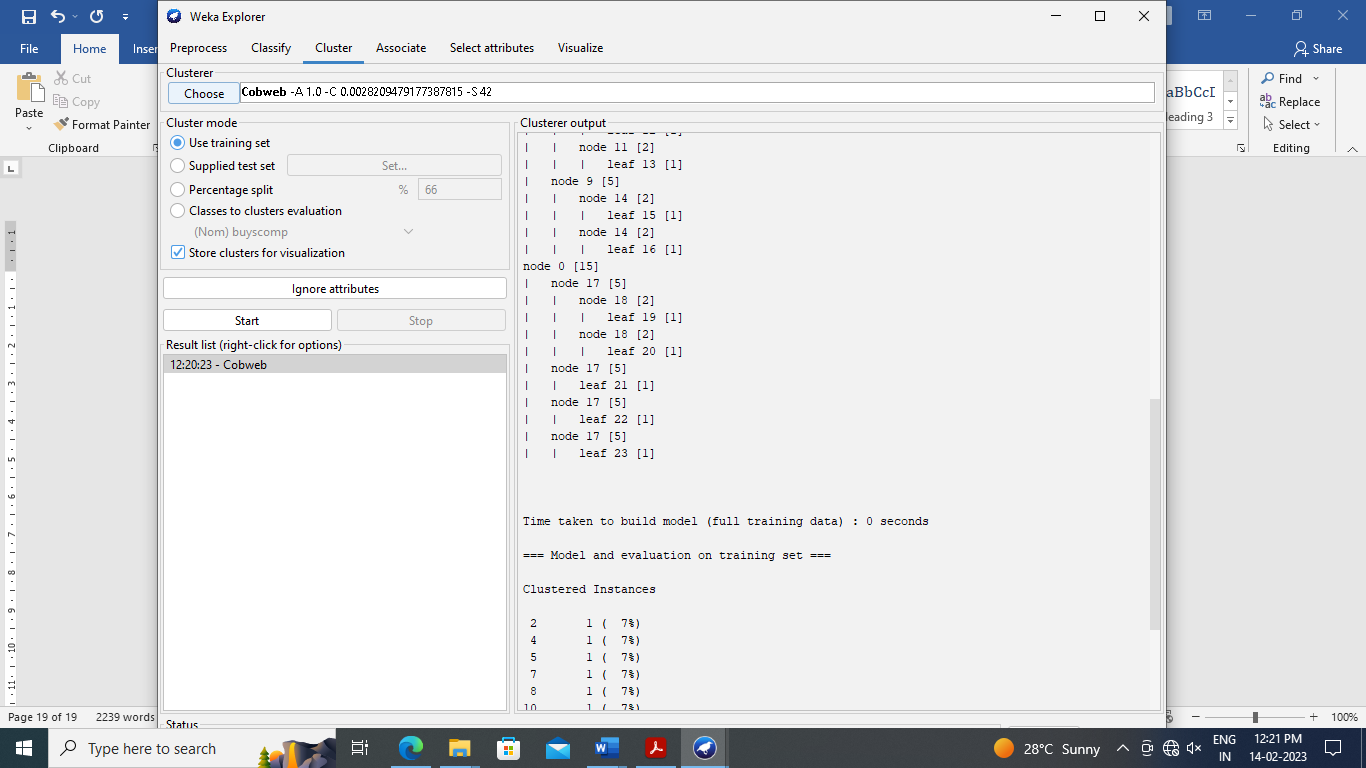
**3)** Click on **open file** & then select **Buying.arff**file.

**4)** Click on **Cluster menu**. In this there are different algorithms are there.

**5)** Click on **Choose button** and then select **cobweb** algorithm.

**6)** Click on **Start button** and then **output** will be displayed on the screen.

**Output:**



**Result:**

Thus, the program has been successfully executed.

**Ex.No.5 Demonstrate performing Regression on data sets**

**Ex.No.6 Credit Risk Assessment. Sample Programs using German Credit Data**

**Ex.No.7 Sample Programs using Hospital Management System**

**Ex.No.8 Beyond the Syllabus -Simple Project on Data Preprocessing**