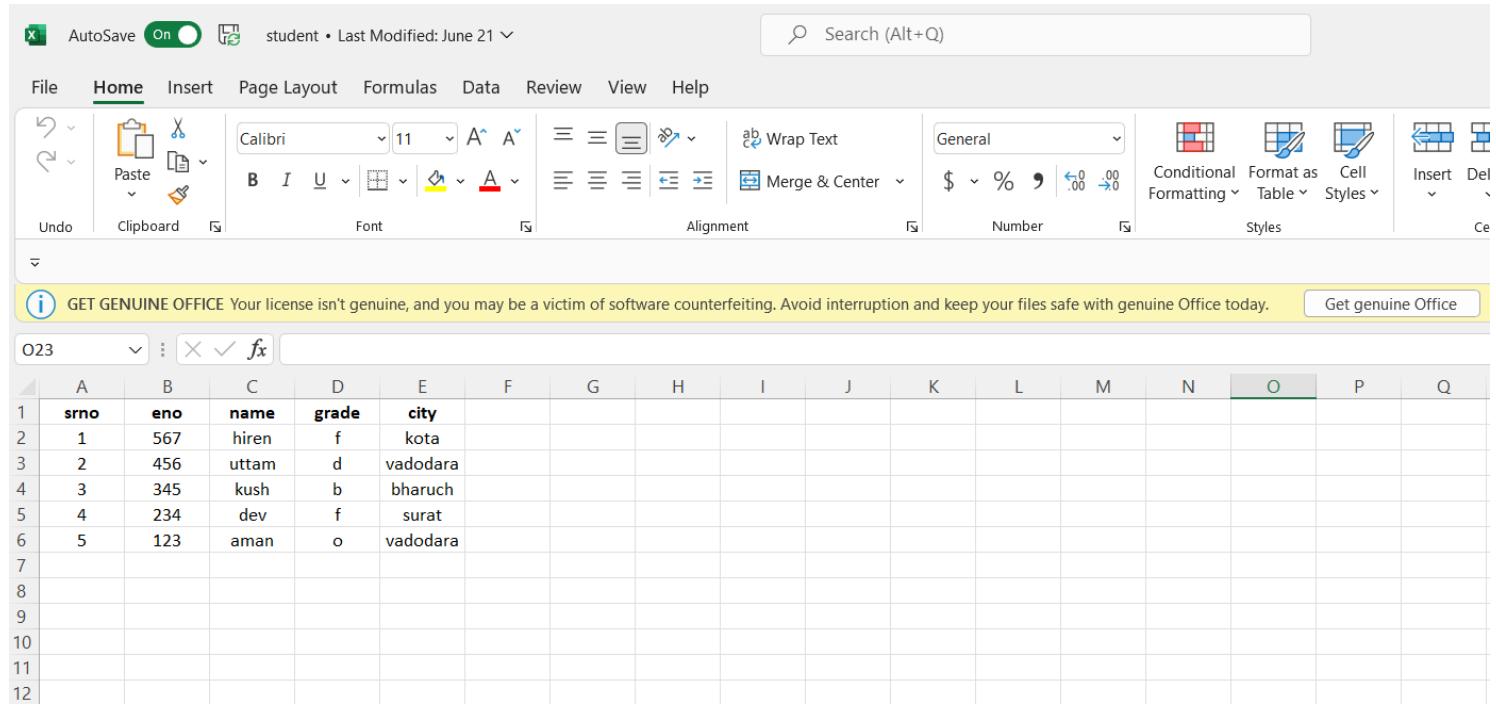


## Practical:1

**Aim:** Create an excel file convert it to .csv format and prepare ARFF files of it.

**Solution:**

Creating an excel file ‘Students.xls’ in Microsoft Excel

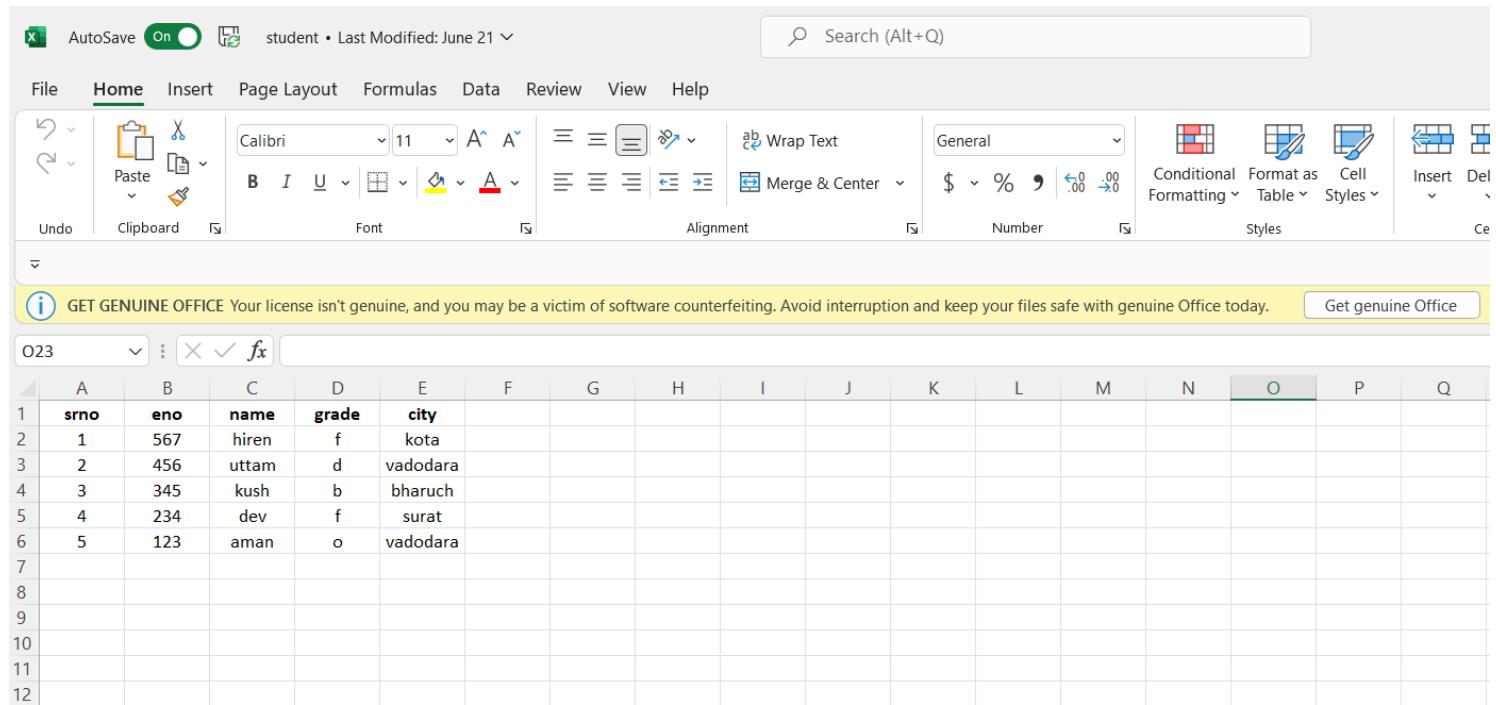


The screenshot shows a Microsoft Excel spreadsheet titled 'student' last modified on June 21. The table has columns labeled 'srno', 'eno', 'name', 'grade', and 'city'. The data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	srno	eno	name	grade	city												
2	1	567	hiren	f	kota												
3	2	456	uttam	d	vadodara												
4	3	345	kush	b	bharuch												
5	4	234	dev	f	surat												
6	5	123	aman	o	vadodara												
7																	
8																	
9																	
10																	
11																	
12																	

Now creating csv file. CSV stands for Comma Separated Values. Write this below file in notepad & save as ‘Students.csv’.

<b>Srno,</b>	<b>Eno,</b>	<b>Name,</b>	<b>Grade,</b>	<b>City</b>
1,	567,	Hiren,	F,	kota
2,	456,	Uttam,	D,	vadodara
3,	345,	Kush,	B,	bharuch
4,	234,	Dev,	F,	surat
5,	123,	Aman,	O,	vadodara



A screenshot of Microsoft Excel showing a table of student data. The table has columns labeled srno, eno, name, grade, and city. The data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	<b>srno</b>	<b>eno</b>	<b>name</b>	<b>grade</b>	<b>city</b>												
2	1	567	hiren	f	kota												
3	2	456	uttam	d	vadodara												
4	3	345	kush	b	bharuch												
5	4	234	dev	f	surat												
6	5	123	aman	o	vadodara												
7																	
8																	
9																	
10																	
11																	
12																	

Now creating arff file. ARFF stands for Attribute Relation File Format. Write this below file in notepad & save as ‘Students.arff’.

```
@RELATION Students
@ATTRIBUTE Sr. No. NUMERIC
@ATTRIBUTE E.NO NUMERIC
@ATTRIBUTE Name STRING
@ATTRIBUTE Division {B,D,F,O}
@ATTRIBUTE City STRING
@DATA
```

1,576,Hiren, F, Kota  
 2,456,Uttam D,Vadodara  
 3,345,Kush, B,Bharuch  
 4,234, Dev, F,Surat  
 5,123,Aman, O,Vadodara

api.txt Students.arff

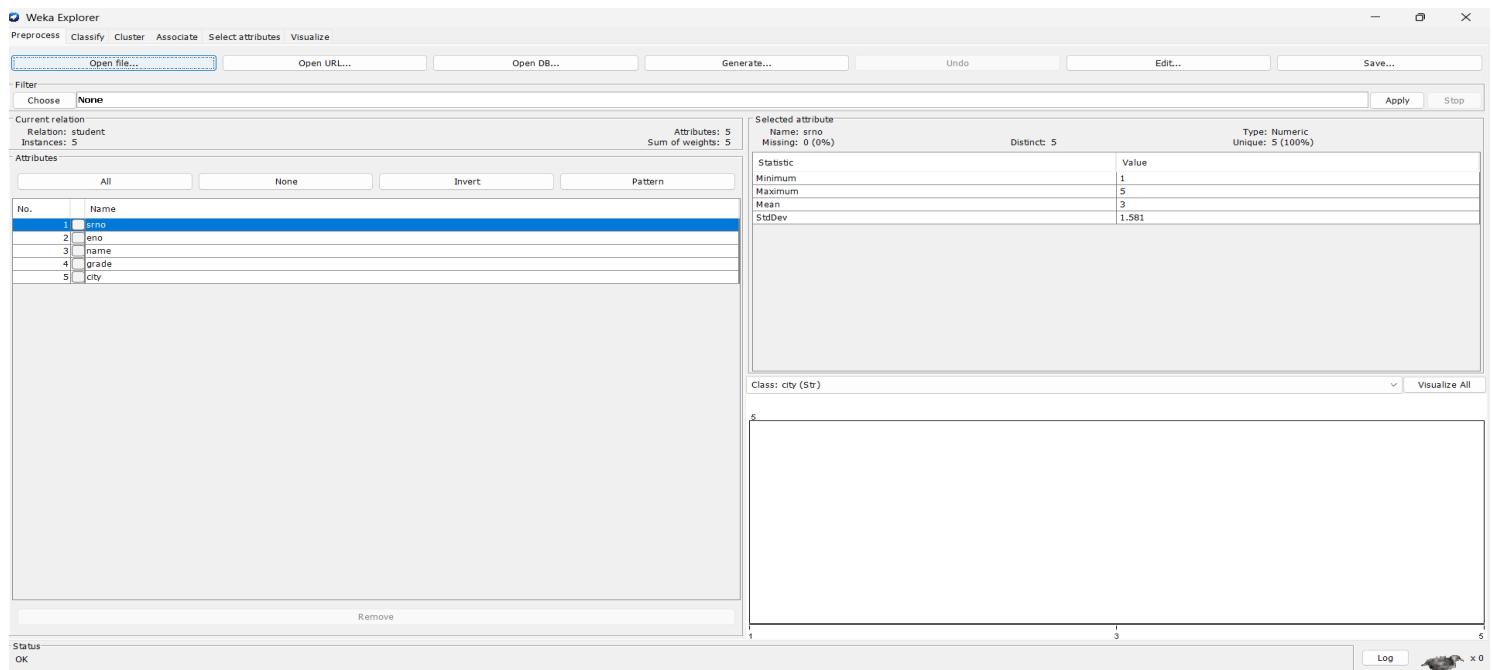
File Edit View

```

@RELATION Students|
@ATTRIBUTE Sr_No NUMERIC
@ATTRIBUTE E_NO NUMERIC
@ATTRIBUTE Name STRING
@ATTRIBUTE Division {B,D,F,O}
@ATTRIBUTE City STRING

@DATA
1,576,Hiren, F, Kota
2,456,Uttam D,Vadodara
3,345,Kush, B,Bharuch
4,234, Dev, F,Surat
5,123,Aman, O,Vadodara

```



## Practical: 2

**Aim:** Write the use of ARFF files taking input and display the output of the files.

**Solution:**

An ARFF (Attribute-Relation File Format) file is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software.

As the name suggests it described a list of instances sharing a set of attributes. these files are supported by WEKA machine Learning tool, arff files are used for the purpose of various operations related to data preprocessing, data cleaning etc.

**Structure of file.**

ARFF file contains 2 sections

1.Header Section

2.Data Section

All the keywords in ARFF file start with @ symbol.

**1.Header Section:**

This section contains various information related to the dataset like the name of the relation, columns, and type of columns. The header section contains 2 parts **Table/relation** and **attribute** part.

**@relation** : used to give the table name

**@attribute**: used to give a column name

**datatypes:**

- **nominal**: represented inside curly brackets (Like constants)
- **string** : data type which accepts only string value
- **numeric**: used to store numbers
- **date**: used to store date

**Syntax:**

    @relation tablename

    @attribute column\_name type

**Example:**

@RELATION Students

```
@ATTRIBUTE Sr. No. NUMERIC  
@ ATTRIBUTE E.NO NUMERIC  
@ATTRIBUTE Name STRING  
@ATTRIBUTE Division {B,D,F,O}  
@ATTRIBUTE City STRING
```

Here div column is having nominal data type so it can only accept above mentioned types of data only.

**2. Data section:**

Data section is used to represents the data or entries for available columns. (according to the order in header section data would be inserted).

data section starts with @data, and this section must be added after Header section. only single record can be written in single line.

**@data:** Used to start data section

%: % sign is used to represent the comment in file.

**Syntax:**

@data

<record 1>

<record 2>

.

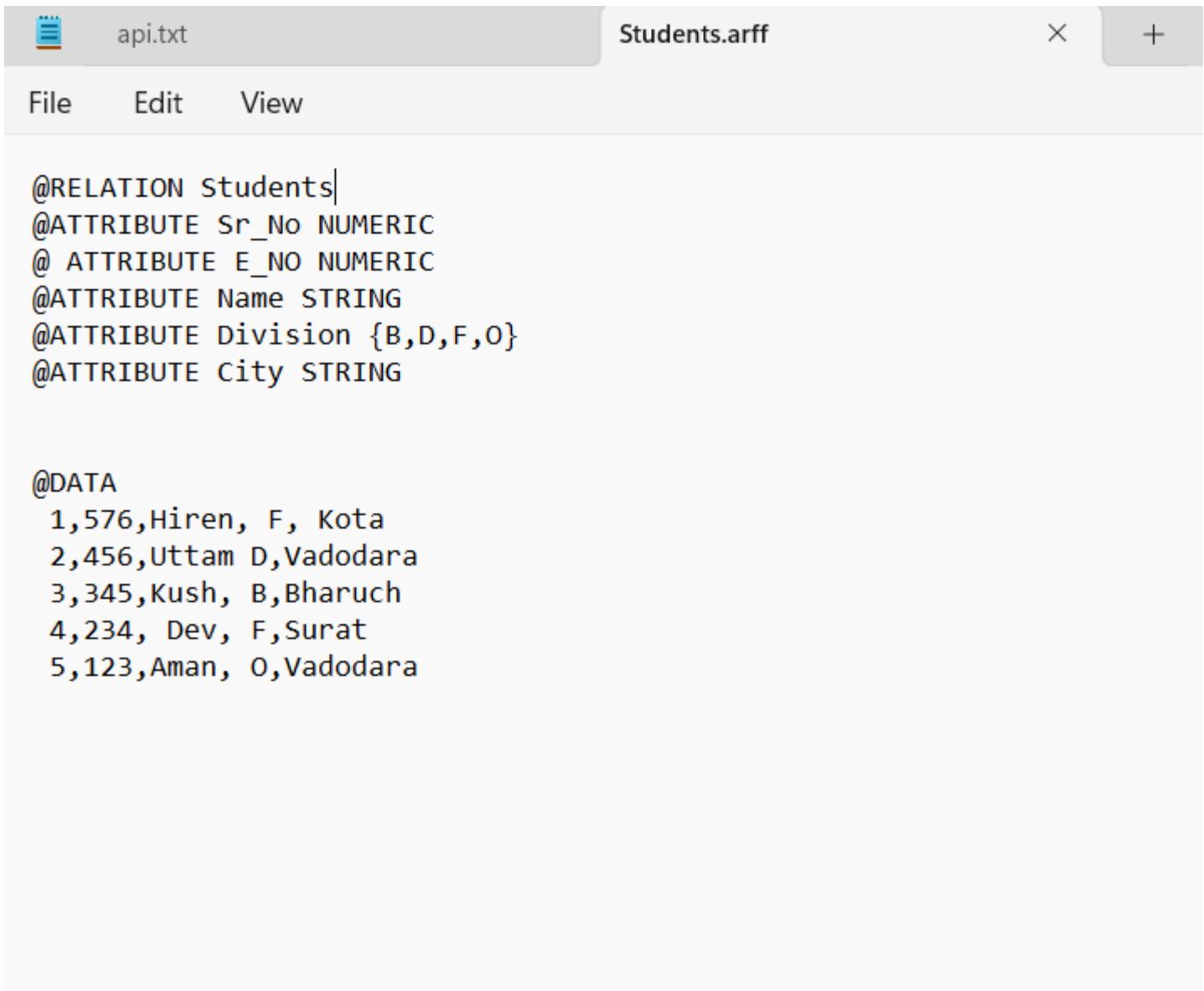
.

<record N>

all the Records must be in the same format as their attributes are defined in Header section Like

Example:

Entire file would look like this:



The screenshot shows a text editor window with the following details:

- File tabs: "api.txt" (active) and "Students.arff".
- Toolbar: "File", "Edit", "View".
- Content area (ARFF code):

```
@RELATION Students
@ATTRIBUTE Sr_No NUMERIC
@ATTRIBUTE E_NO NUMERIC
@ATTRIBUTE Name STRING
@ATTRIBUTE Division {B,D,F,O}
@ATTRIBUTE City STRING

@DATA
1,576,Hiren, F, Kota
2,456,Uttam D,Vadodara
3,345,Kush, B,Bharuch
4,234, Dev, F,Surat
5,123,Aman, O,Vadodara
```

**student.arff file:**

We separate values by comma(,) and to represent the empty or missing value for a particular column we use the (?)sign.

**How to Create and open arff file:**

you need to have weka tool install on your machine. you can check this [How to install Weka.](#)

**Step 1:** Open any text editor and paste the above code.

**Step 2:** Save the file with student.arff file extension

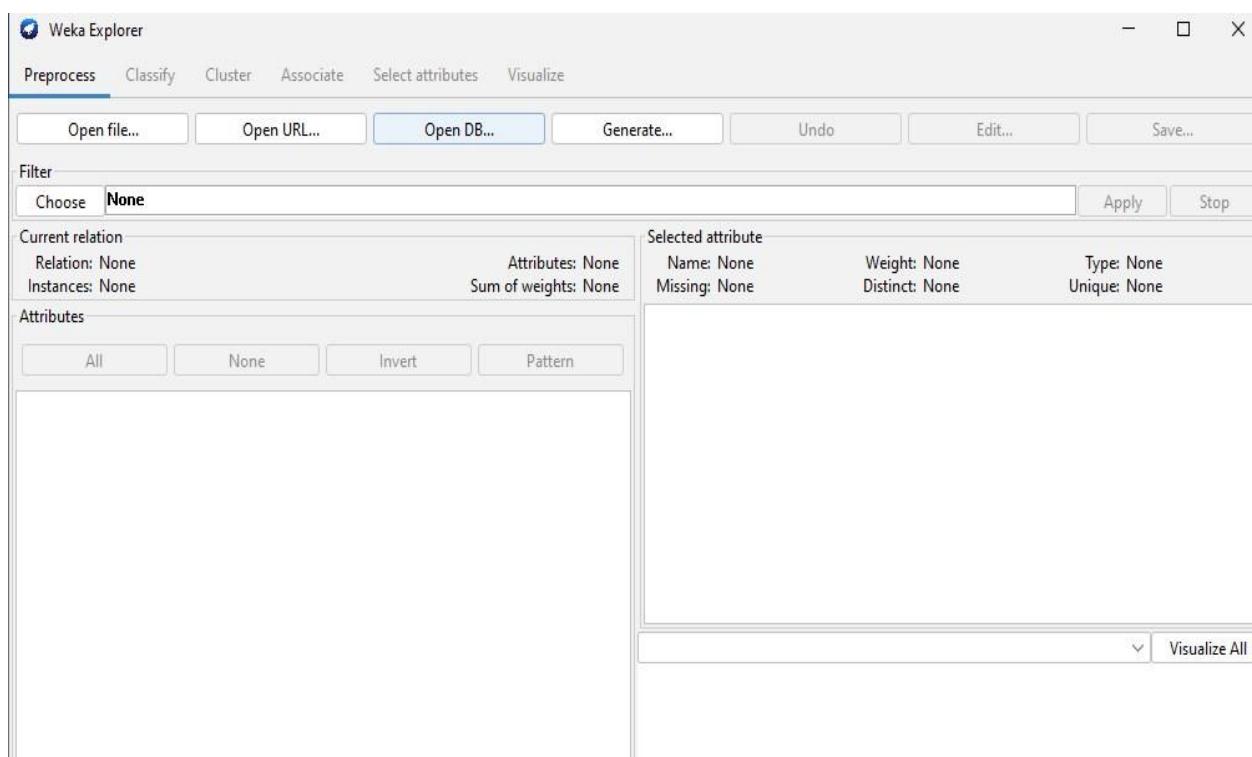
**Step 3:** Open weka tool

**Step 4:** Click on Explorer

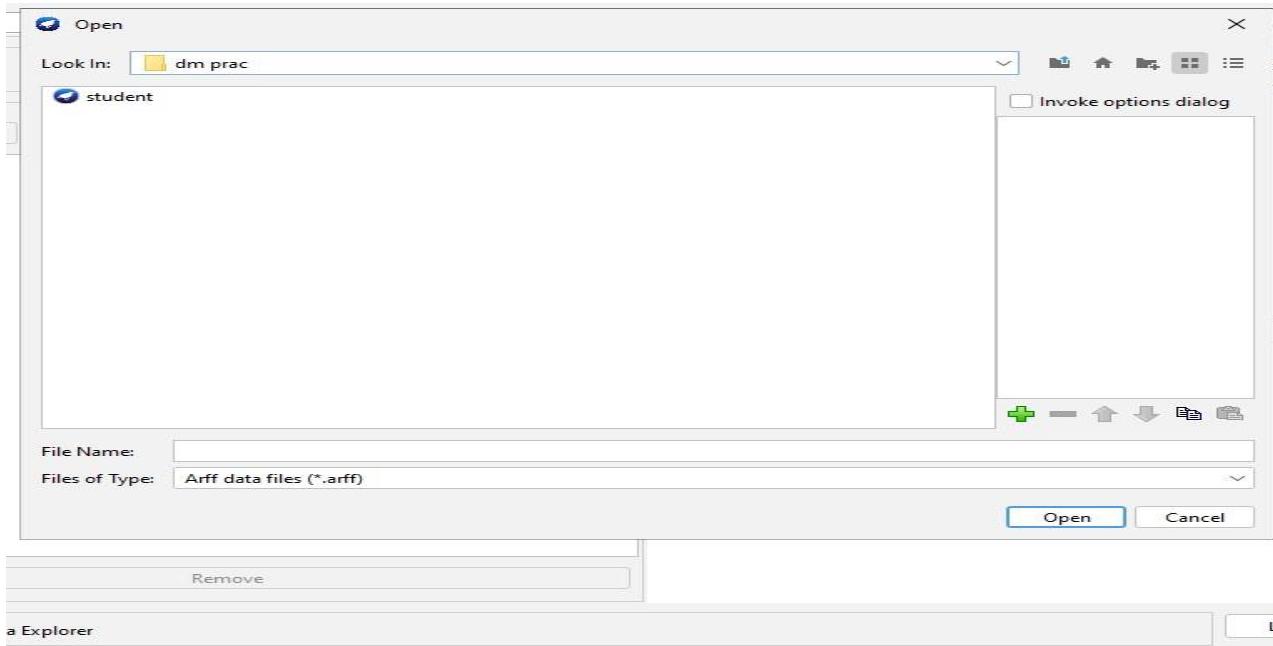




Then click on Open file



Select/Locate arff file from disk then click On Open.



**Step 6:** file is now Loaded now click on Edit from Preprocess Tab

No.	Name	
1	srno	<input checked="" type="checkbox"/>
2	eno	<input type="checkbox"/>
3	name	<input type="checkbox"/>
4	grade	<input type="checkbox"/>
5	city	<input type="checkbox"/>

Selected attribute  
Name: srno  
Missing: 0 (0%)  
Distinct: 5  
Type: Numeric  
Unique: 5 (100%)

Statistic	Value
Minimum	1
Maximum	5
Mean	3
StdDev	1.581

Class: city (Str)  
Visualize All

Viewer

Relation: student

No.	1: srno Numeric	2: eno Numeric	3: name String	4: grade Nominal	5: city String
1	1.0	567.0	hiren	f	kota
2	2.0	456.0	uttam	d	Vado...
3	3.0	345.0	kush	b	bharuch
4	4.0	234.0	dev	f	surat
5	5.0	123.0	aman	o	Vado...

Add instance Undo

So this is how you can work with arff file. with weka tool, various operations can be done on the Available Dataset.

## **Practical: 3**

**Aim: Study various Data Mining tool and their applications.**

**Solution:**

### **1.ORANGE:**

- Orange is component-based visual programming software for data mining, machine learning and data analysis.
- Components are called widgets and they range from simple data visualization, subset selection and preprocessing, to empirical evaluation of learning algorithms and predictive modeling.
- Visual programming is implemented through an interface in which workflows are created by linking predefined or user-designed widgets, while advanced users can use Orange as a Python library for data manipulation and widget alteration.
- Orange consists of a canvas interface onto which the user places widgets and creates a data analysis workflow. Widgets offer basic functionalities such as reading the data, showing a data table, features, training predictors, comparing learning algorithms, visualizing data elements, etc. The user can interactively explore visualizations or feed the selected subset into other widgets.

### **2.WEKA:**

- WEKA is an open-source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems.
- Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.
- Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions.
- Weka supports several standard data mining tasks, more specifically, data preprocessing,

clustering, classification, regression, visualization, and feature selection. Input to Weka is expected to be formatted according the Attribute- Relational File Format and with the filename bearing the .arff extension. All of Weka's techniques are predicated on the assumption that the data is available as one flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported).

### **3.XLMiner:**

- It is a Data mining software tool for business analytics, predictive model creation and testing through Excel, SQL or PowerPivot.
- With the XLMiner Analysis ToolPak App, you can perform statistical analyses of the same types available in the Analysis ToolPak in desktop Microsoft Excel. The App provides all 19 interactive functions included in the Excel Analysis ToolPak, plus a 20th function often requested by users, logistic regression. The App's Task Pane input fields and worksheet output will be very familiar to Excel Analysis ToolPak users.

### **4.RapidMiner:**

- Rapid Miner is a data science platform designed for enterprises that analyses the collective impact of organizations' employees, expertise and data. Rapid Miner's data science platform is intended to support many analytics users across a broad AI lifecycle. Rapid Miner claims that more than 1 million people use the platform.
- Rapid Miner, formerly known as YALE (Yet Another Learning Environment).
- Rapid Miner uses a client/server model with the server offered either on-premises or in public or private cloud infrastructures.
- Rapid Miner provides 99% of an advanced analytical solution through template- based frameworks that speed delivery and reduce errors by nearly eliminating the need to write code.

- Rapid Miner provides data mining and machine learning procedures including: data loading and transformation (ETL), data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment.
  
- Rapid Miner is written in the Java programming language. Rapid Miner provides a GUI to design and execute analytical workflows.

## Practical 4

### **AIM: Install and configure WEKA Tool**

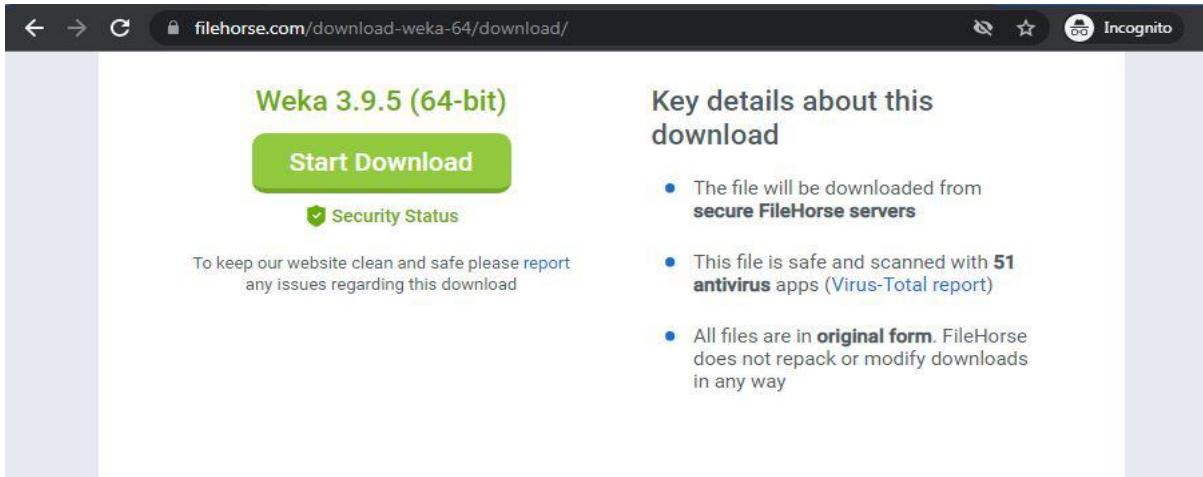
#### **Installing Weka on Windows:**

Follow the below steps to install Weka on Windows:

**Step 1:** Visit this [website](#) using any web browser. Click on Free Download.



**Step 2:** It will redirect to a new webpage, click on Start Download. Downloading of the executable file will start shortly. It is a big 118 MB file that will take some minutes.



**Step 3:** Now check for the executable file in downloads in your system and run it.



**Step 4:** It will prompt confirmation to make changes to your system. Click on Yes.



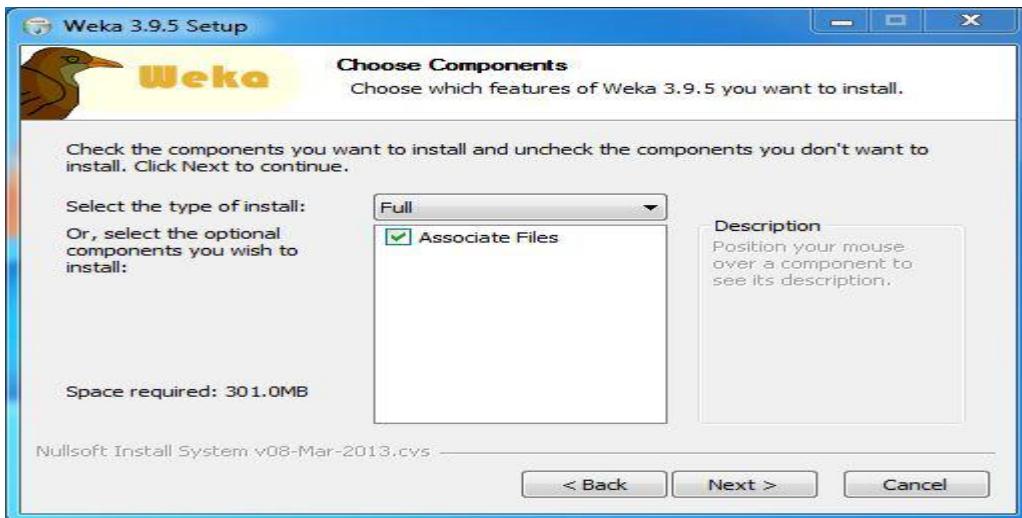
**Step 5:** Setup screen will appear, click on Next.



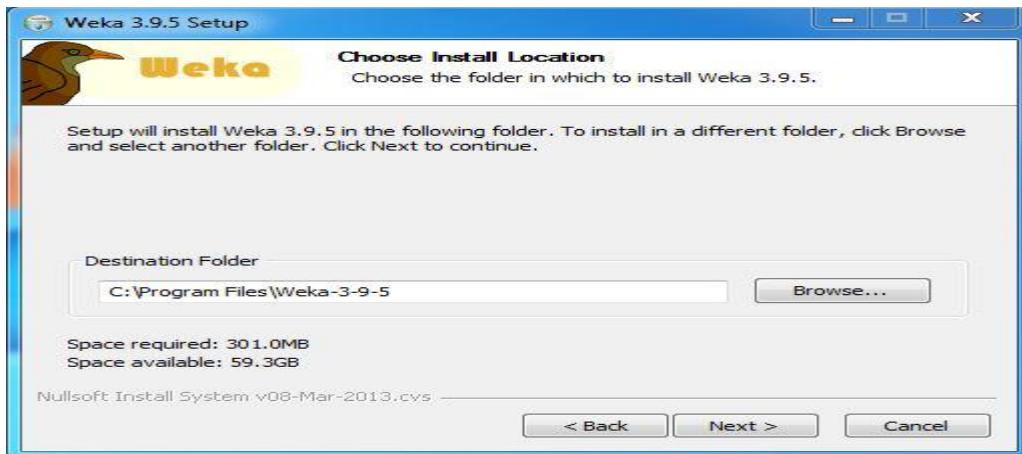
**Step 6:** The next screen will be of License Agreement, click on I Agree.



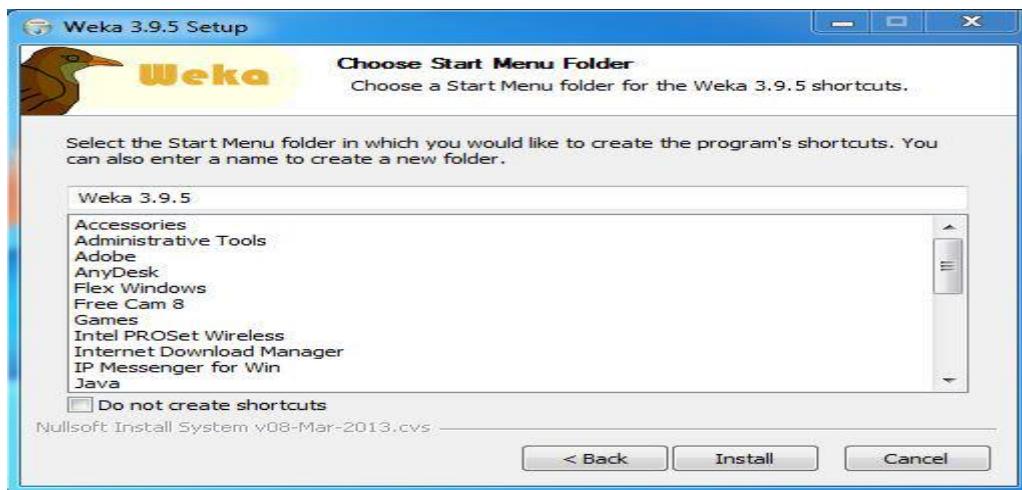
**Step 7:** Next screen is of choosing components, all components are already marked so don't change anything just click on the Install button.



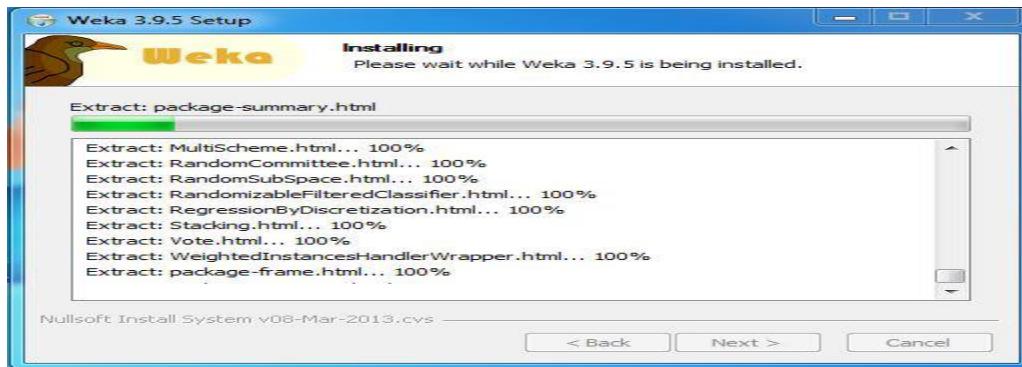
**Step 8:** The next screen will be of installing location so choose the drive which will have sufficient memory space for installation. It needed a memory space of 301 MB.



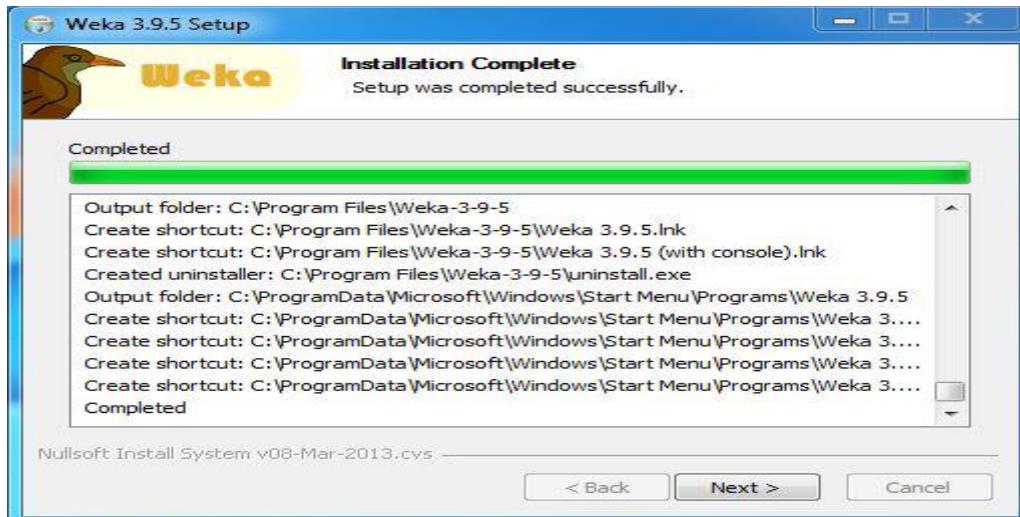
**Step 9:** Next screen will be of choosing the Start menu folder so don't do anything just click on Install Button.



**Step 10:** After this installation process will start and will hardly take a minute to complete the installation.



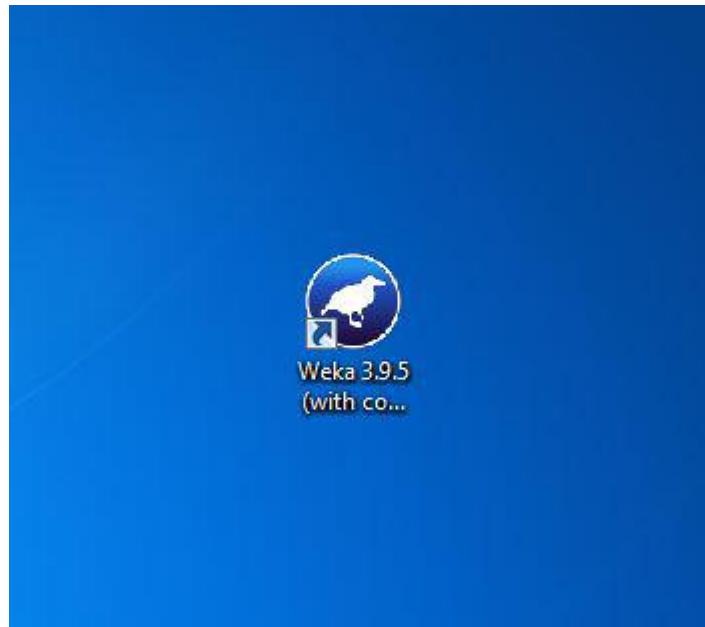
**Step 11:** Click on the Next button after the installation process is complete.



**Step 12:** Click on Finish to finish the installation process.



**Step 13:** Weka is successfully installed on the system and an icon is created on the desktop.



**Step 14:** Run the software and see the interface.



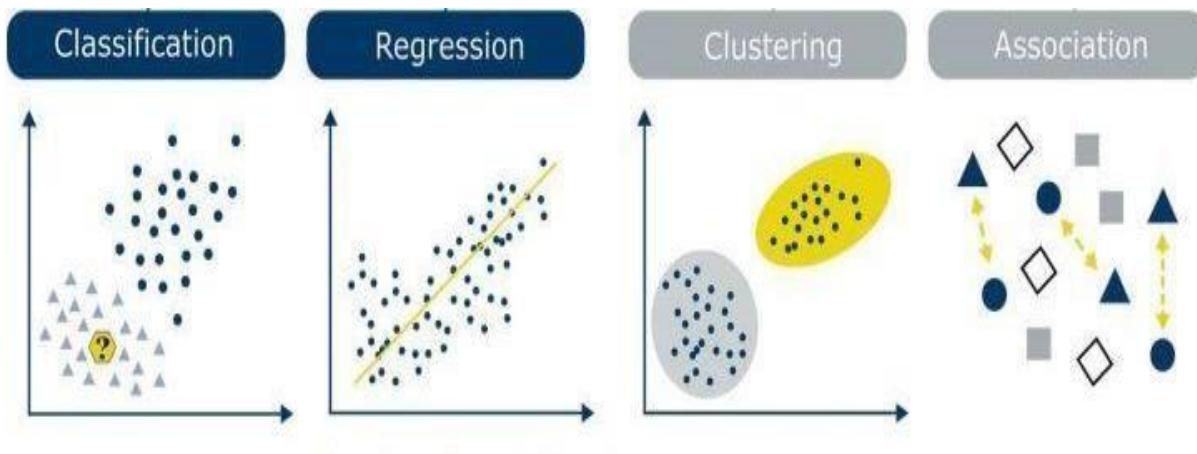
## Practical: 5

**Aim: Compare various Data Mining techniques available in WEKA.**

**Solution:**

CLASSIFICATION	CLUSTERING	REGRESSION	ASSOCIATION
Classification is a supervised learning approach where a specific label is provided to the machine to classify new observations. Here the machine needs proper testing and training for the label verification	Clustering is an unsupervised learning approach where grouping is done on similarities basis.	Regression is a supervised learning approach that is used to predict any continuous-valued attribute. Regression helps analyze the target variable and predictor variable relationships.	Association rule learning is a type of unsupervised learning technique that finds interesting associations and relationships among large business organization to sets of data items.
It uses a training dataset	It does not use a training dataset	It uses a train and test dataset	It uses a training dataset
It uses algorithm to categorize the new data as per the observation of the training set.	It uses statistical concepts in which the data set is divided into subset of the same features	It uses algorithm to find the new data as per the observation of the training set.	It uses algorithm to discover interesting patterns, associations, or relationships within large datasets.
Output is categorical quantity	Assigns data points into clusters	Output is continuous quantity	The output of association techniques in data mining typically includes discovered patterns, associations, or relationships within the dataset.
The main aim is to forecast or predict	The main aim is to group similar items into clusters	The main aim to forecast and predict.	The main aim of association analysis is to extract meaningful insights from data
Eg: classify mail as spam or non-spam	Eg: find all transaction which are fraudulent in nature	Eg: predict stock market price	Eg: the rule "If a customer buys bread, they are also likely to buy milk"

Algorithm: logistic Regression	Algorithm: k-means	Algorithm: Linear Regression	Algorithm :Apriori Algorithm
--------------------------------	--------------------	------------------------------	------------------------------



## Practical: 6

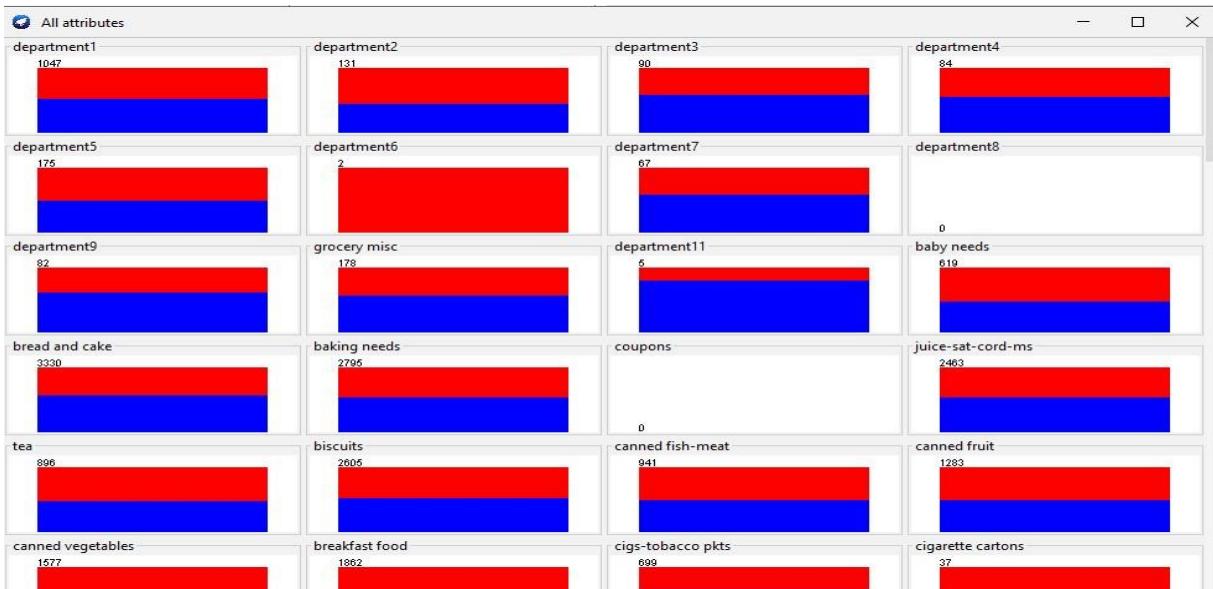
**Aim:** Apply filters on the customer dataset using WEKA.

**Solution:**

**Dataset:** SuperMarket.arff

❖ Steps:

1. First of all, download Customer data set in arff format.
2. Here we use SuperMarket.arff Dataset.
3. Open WEKA tool.
4. Then click on Explorer Button.
5. Now click on open file button & open SuperMarket.arff file.
6. Click on the Choose button in the Filter sub window and select the following filter  
– weka→filters→supervised→attribute→Discretize.
7. Click on the Apply button and examine the all attribute.



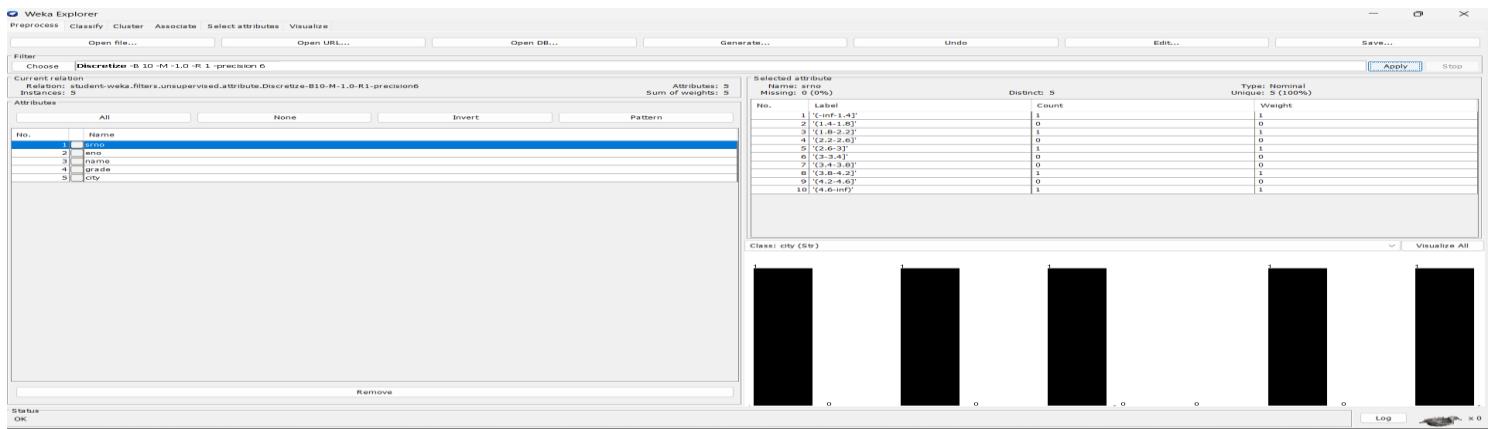
8. Let us look into another filter now.

Parameters	Supervised machine learning	Unsupervised machine learning
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data that is not labeled
Computational Complexity	Simpler method	Computationally complex
Accuracy	Highly accurate	Less accurate
No. of classes	No. of classes is known	No. of classes is not known
Data Analysis	Uses offline analysis	Uses real-time analysis of data
Algorithms used	Linear and Logistics regression, Random forest, Support Vector Machine, Neural Network, etc.	K-Means clustering, Hierarchical clustering, Apriori algorithm, etc.

## Dataset Name: SuperMarket

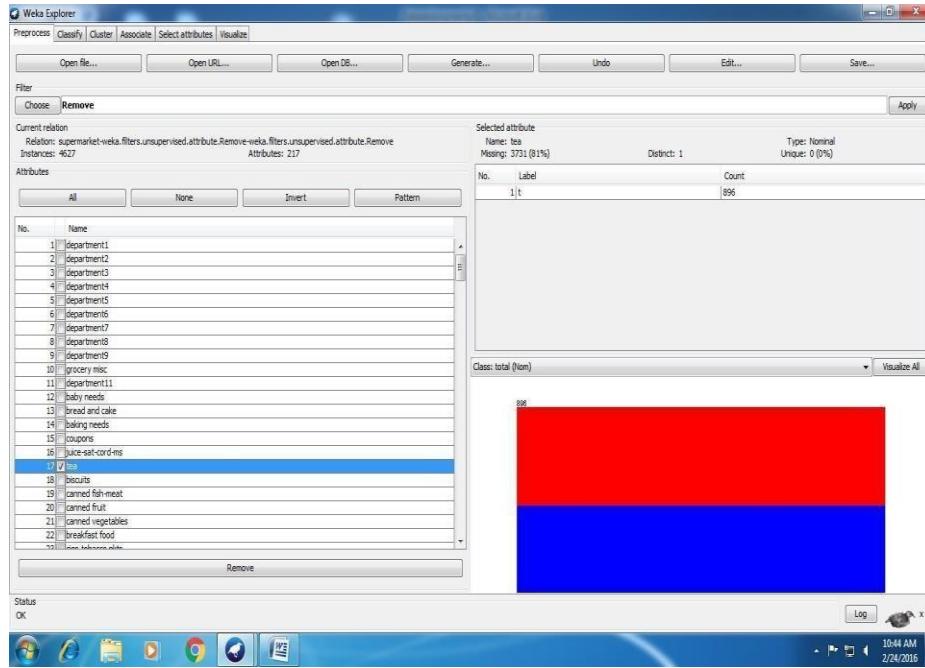
### Unsupervised attribute:

(1) Filter Name: **Discretize** After applying filter:

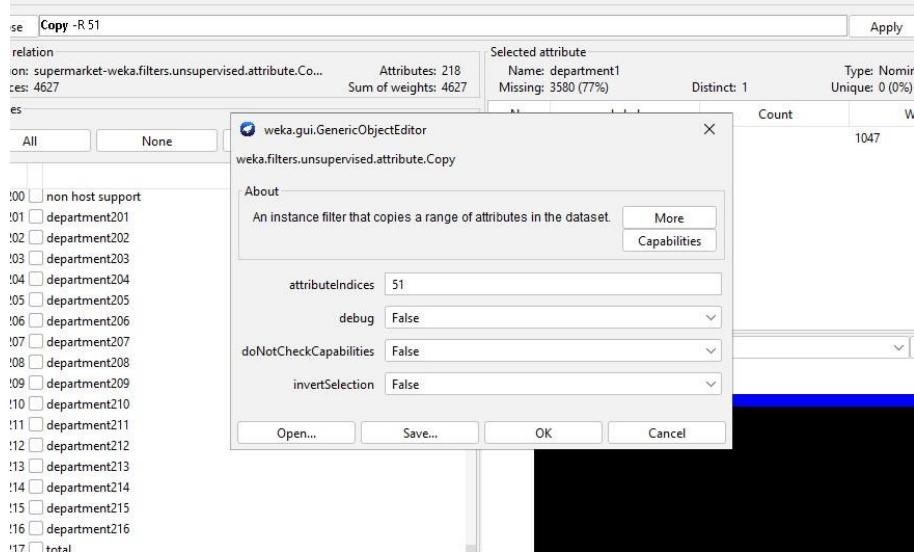


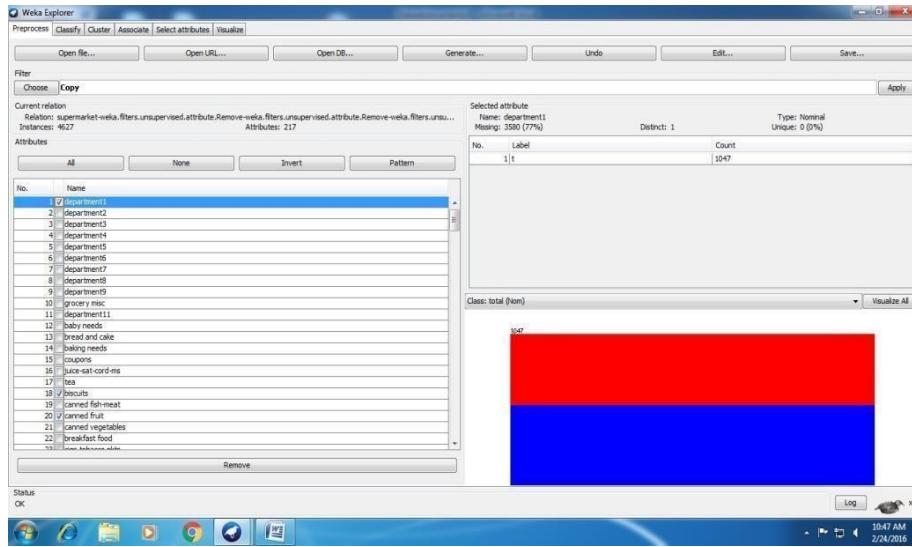
## (2) Filter Name :Remove

After applying filter :



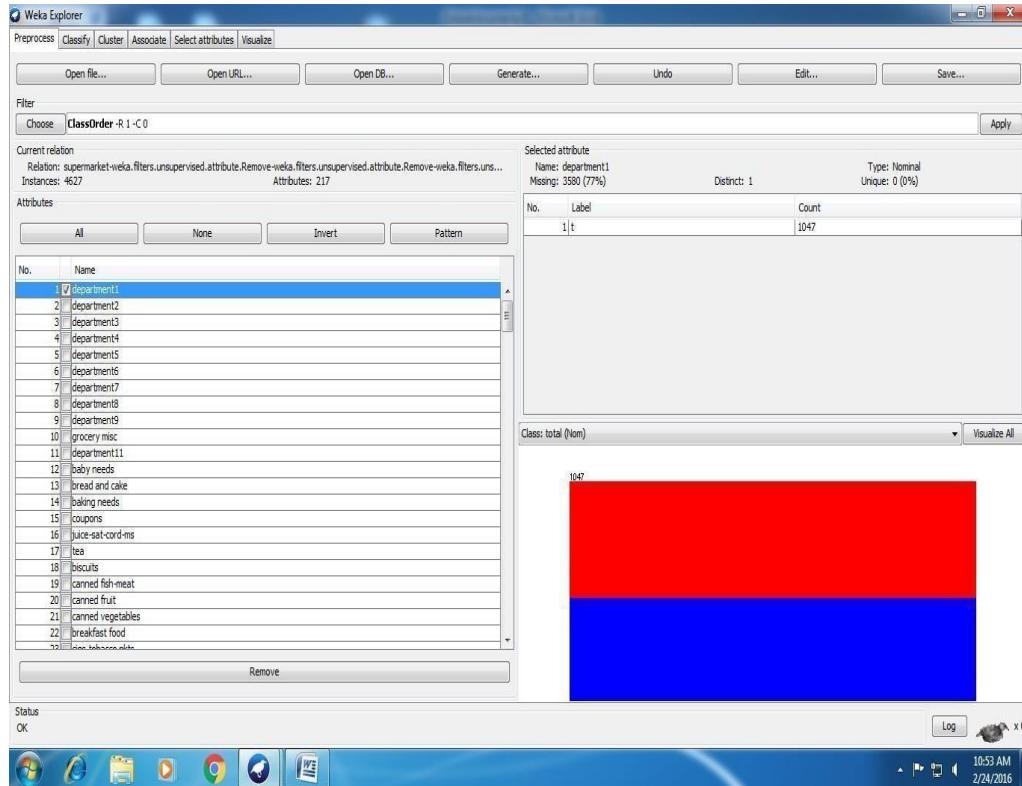
## 3) Filter Name: Copy After applying filter:





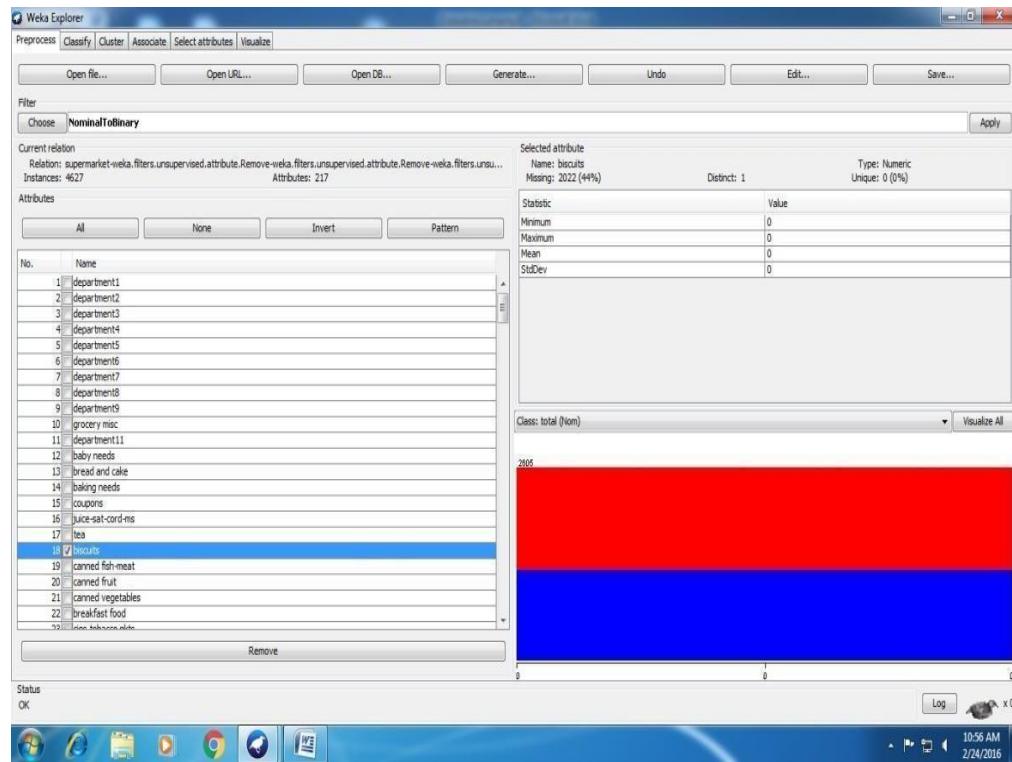
### Supervised Attribute:

(4) Filter Name:**Class-order** After applying filter:



(5) Filter Name: **Nominal To Binary**

After applying filter:



## Practical: 7

**Aim:** Preprocess the data using WEKA Mining Tool. Use at least two preprocessing methods on at least three datasets.

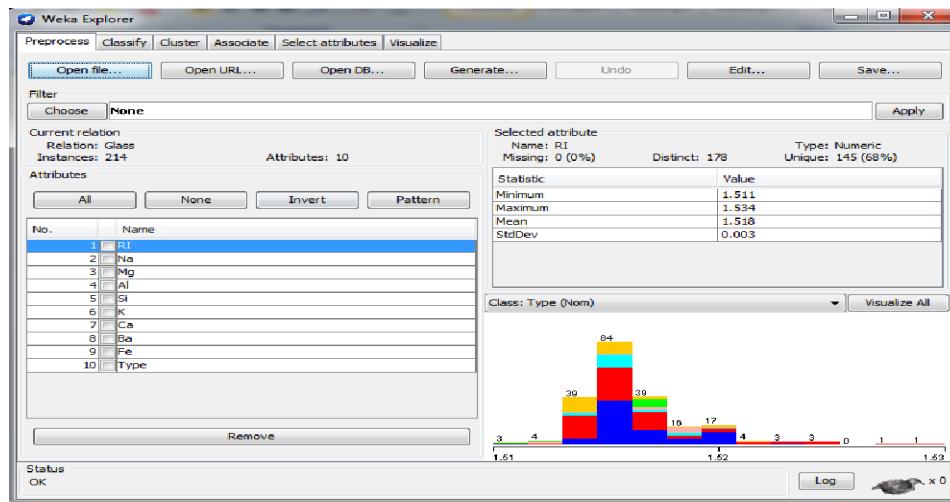
**Solution:**

### ⊕ PREPROCESS:

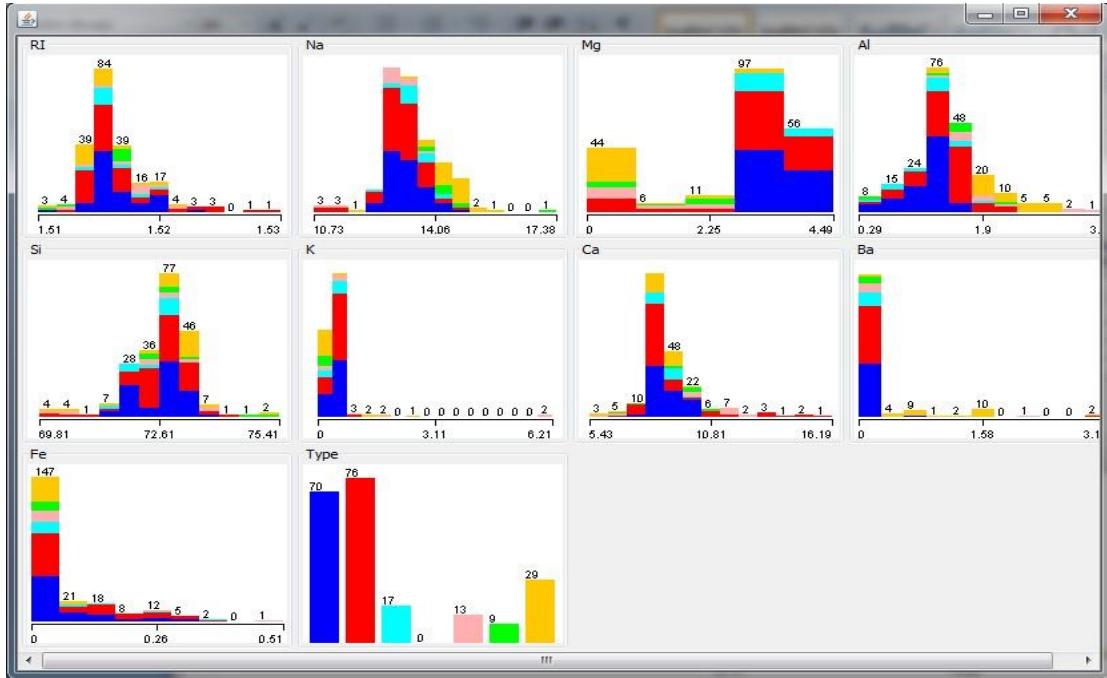
- ✓ In [machine learning](#), the greater the amount of data we have, the better models we can train.
- ✓ If you input such erroneous data, then the output will also be erroneous.
- ✓ We all know that Garbage in Garbage out (GIGO). So, first of all, it is required to remove all these anomalies from the raw data & then transforming it into a format that helps to increase the accuracy of the data model.
- ✓ Data pre-processing is the first step in data mining.
- ✓ It plays a vital role in data mining.
- ✓ This phase begins after the collection of the data.

### (1) Dataset : Glass

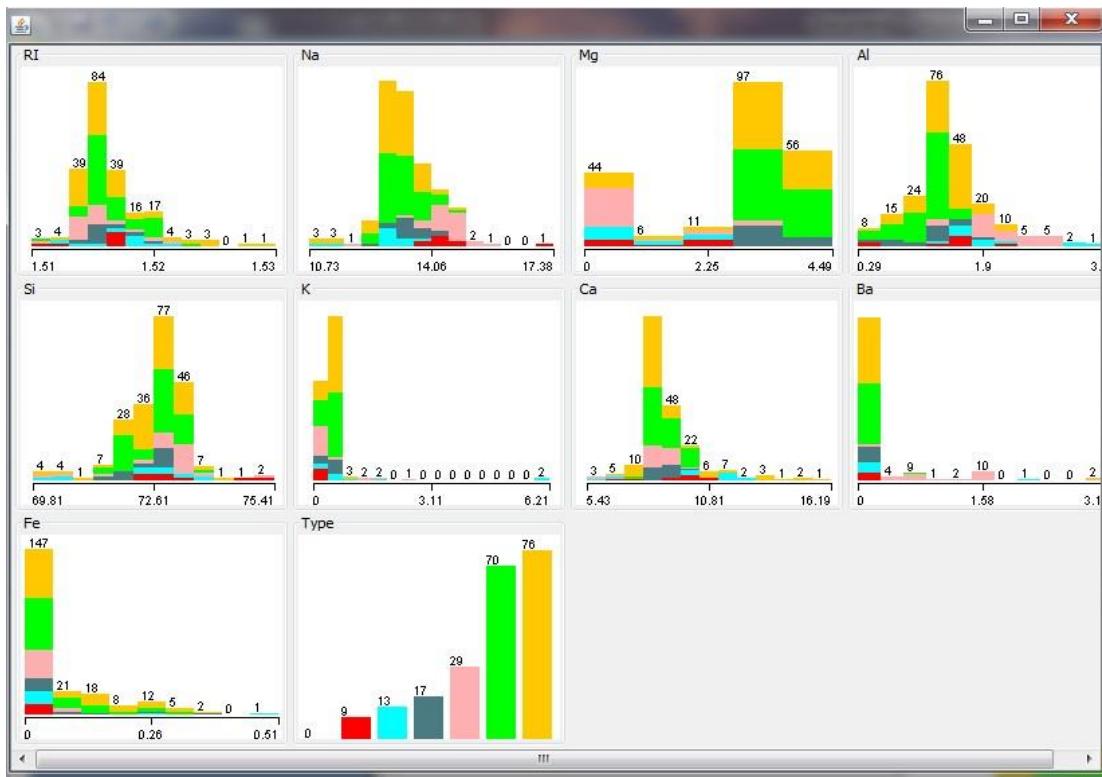
Before loading Datasets:



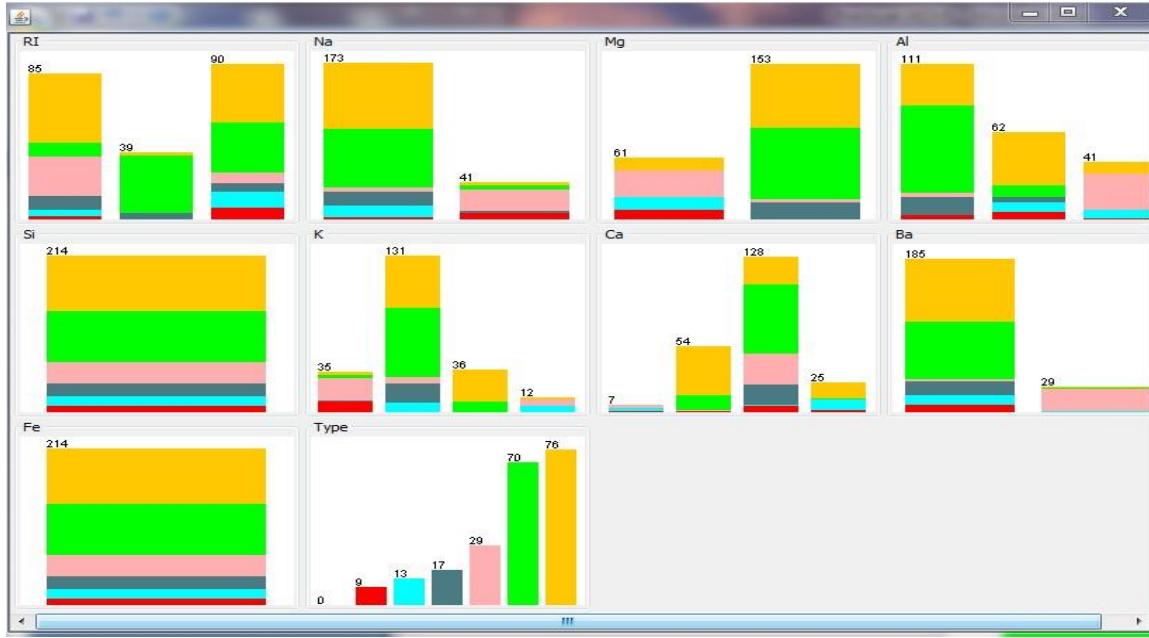
After Loading Datasets:



After Applying Filter:(1) Supervised: ClassOrder



## Applying Filter: (2) Supervised: Discretize

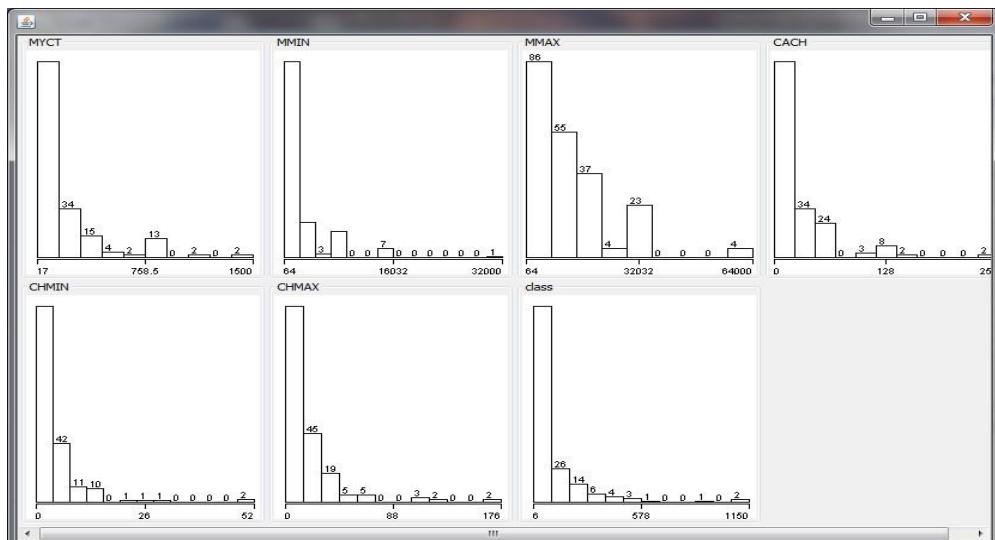


### Analysis:

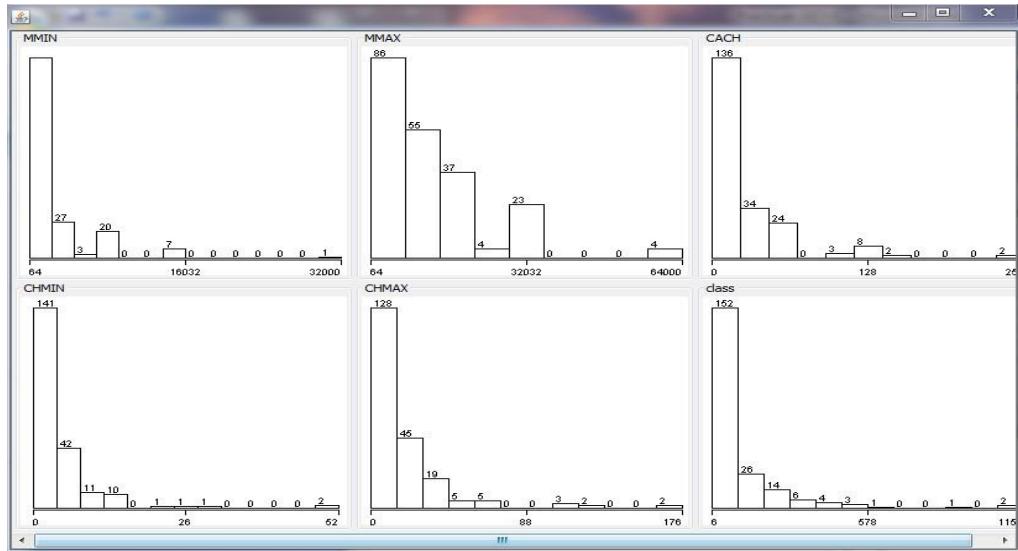
- In Glass Datasets we can conclude that after applying filter [Supervised: ClassOrder] the orientation of Type is high on right, while range and frequencies remains same.
- On applying another Filter [Supervised: Discretize] we can notice that the frequency has range of 90,while the orientation is also moved on to the right.

## (2) Dataset: CPU

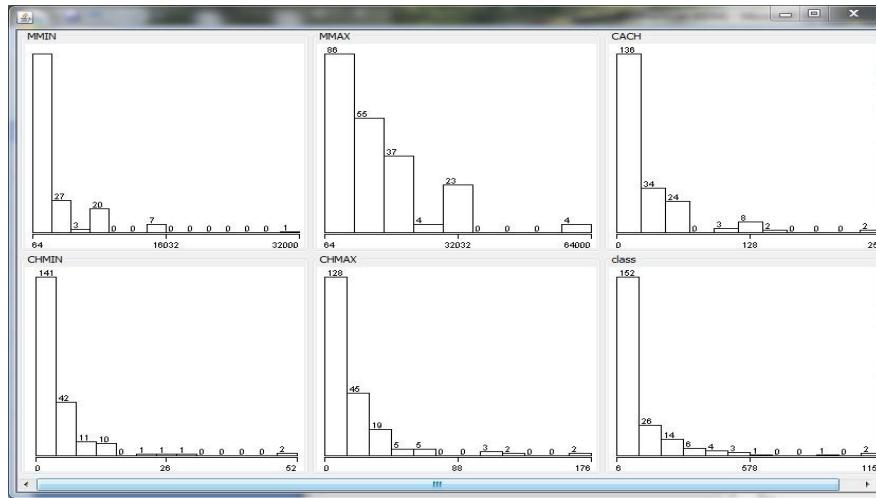
Before loading Datasets& Filters:



### After Applying Filter: (1) Supervised: AttributeSelection



### After Applying Filter: (2) Supervised: AddClassification



### Analysis:

- In CPU Datasets we can conclude that after applying filter [Supervised: AttributeSelection] frequency of MMIN has ranged to 27, while orientation and frequency remains same.
- On applying another Filter [Supervised:AddClassification]the orientation of MMIN & MMAX remains on the left hand side, same remains for range and frequency.

## **Practical -8**

**Aim: Perform Preprocessing, Classification techniques on Agriculture dataset.**

Solution:

1. Perform Data Pre-processing techniques on agriculture dataset (Soyabean) in data mining using WEKA tool

We will use below three filters for removing unwanted & unimportant attributes & instances that are not critical for analytic purpose.

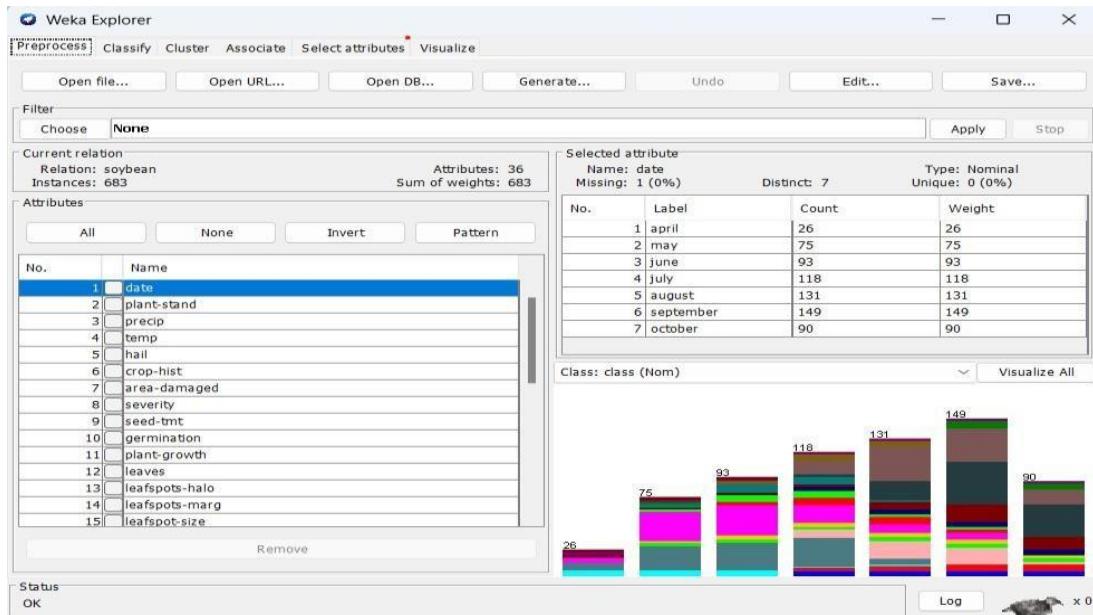
Filters Used:

1. Remove (Attribute)
2. ReplaceMissingWithUserConstant (Missing Values with Constant)
3. ReplaceMissing Values (Missing Values with mode & mean)

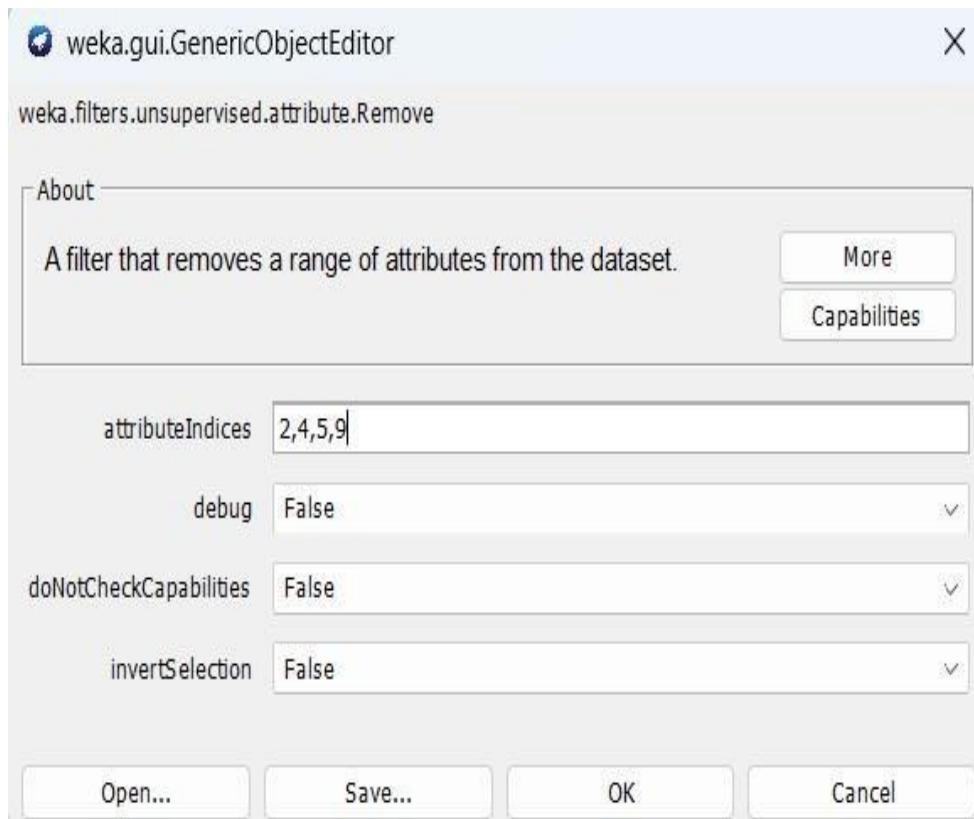
In this practical, we will use Dataset: soyabean.arff & Remove (Attribute) filter.

Here we use soyabean.arff Dataset.

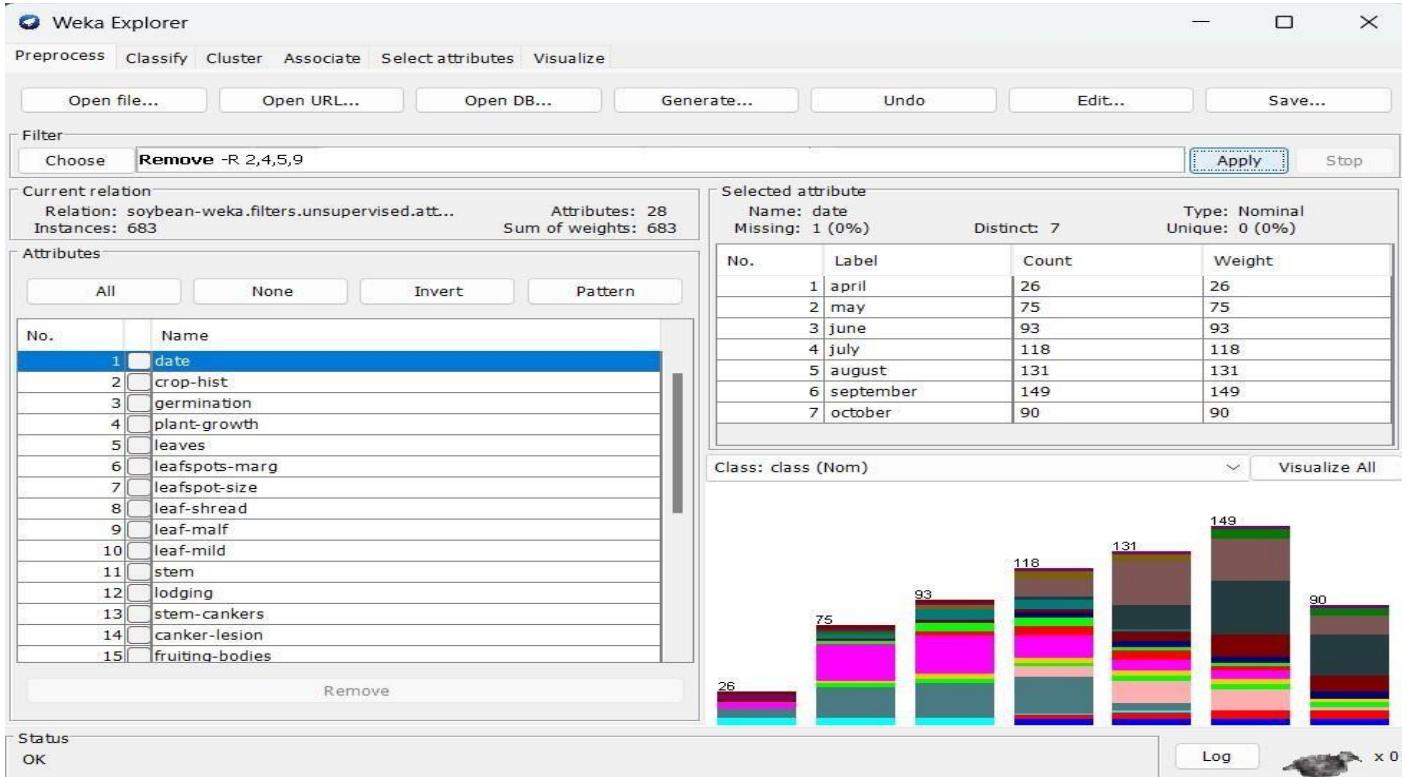
1. Open WEKA tool.
2. Then click on Explorer Button.
3. Now click on open file button & open soyabean.arff file.
4. Before Apply filter :



- Click on the Choose button in the Filter sub window and select the following filter  
Weka → filter— unsupervised— attribute Remove



- Click on the Apply button and examine the all attribute



- To implement Classification techniques on Agriculture data set (Soyabean) in Data Mining using WEKA Tool.

Classification is the process of grouping things according to shared characteristics, properties & structure.

We will use **Naive Bayes** classifier for our practical.

### Naive Bayes Algorithm:

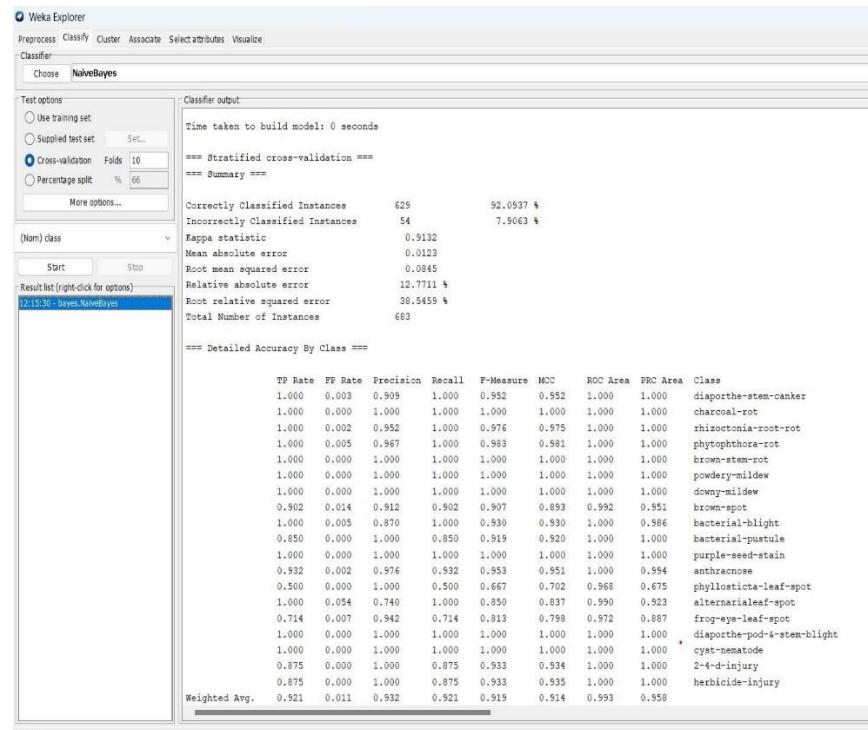
- Naive Bayes algorithm is a supervised learning algorithm & used for solving classification problems. • It is based on Bayes' Theorem.
- It is one of the simple & most effective Classification algorithms.
- It is extremely fast compared to other algorithms.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Some popular examples of Naive Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

### Steps:

- Here we use soybean.arff Dataset.

2. Open WEKA tool.
3. Then click on Explorer Button.
4. Now click on open file button & open soybean.arff file.
5. Click on the Choose button in the Classify tab and select the following Classifier Algorithm:  
o weka→classifiers→Bayes→NaiveBayes
6. Click on the Start button and examine the Result.



## Practical -9

### **AIM: Preprocess and classify Customer dataset.**

**(1) Implementation of Data Preprocessing techniques on Customer data set (Telecommunication) in Data Mining using WEKA tool.**

In this practical, we will use

**Dataset : Soyabean.arff & Replace Missing With User Constant (Missing Values with Constant) filter.**

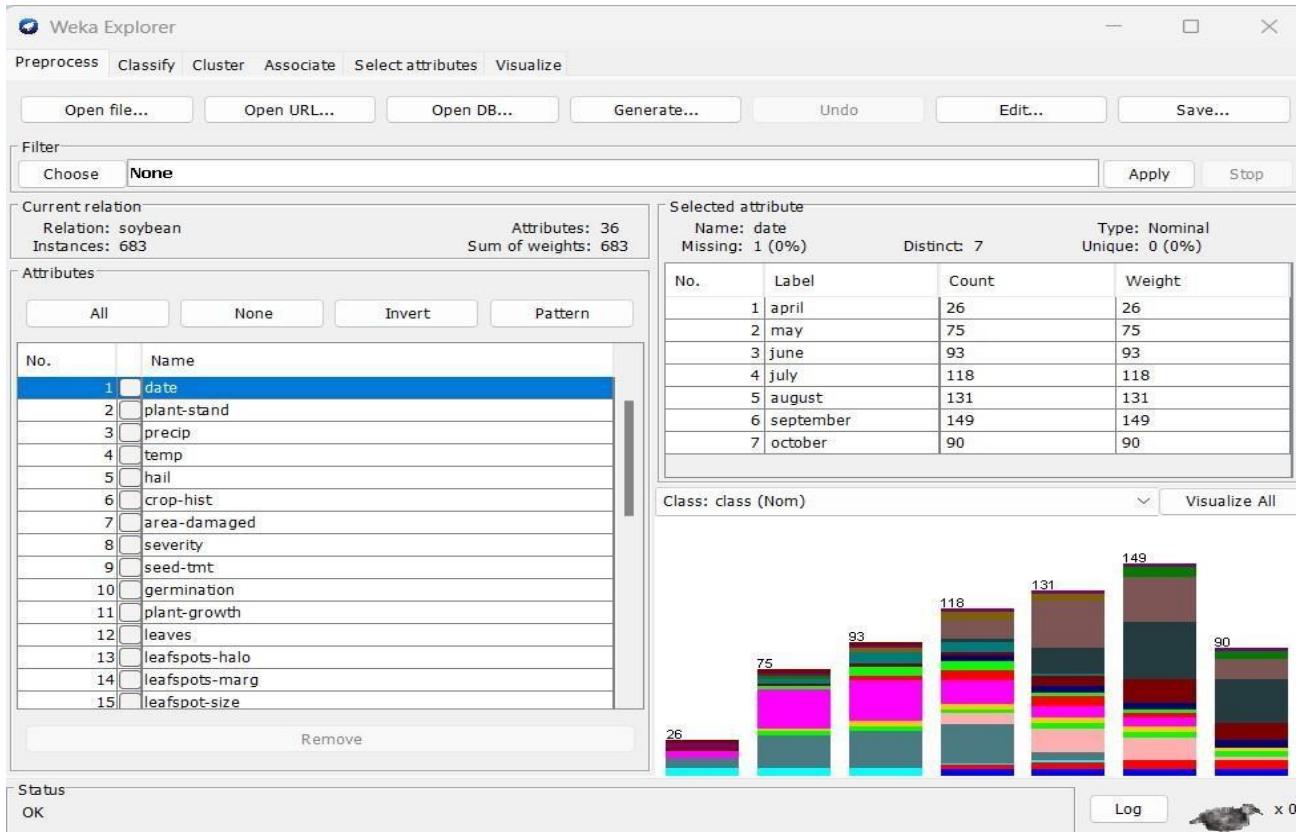
#### **How to Handle Missing Values**

- Ignore the record with missing values
- Fill in the missing values manually
- Use a global constant to fill in missing values (NULL, unknown, etc.)
- Use the attribute mean to fill in the missing values of that attribute
- Use the attribute mean for all samples belonging to the same class to fill in the missing values
- Use the most probable value to fill in the missing values

In this Practical, we will use **Soyabean.arff** Dataset.

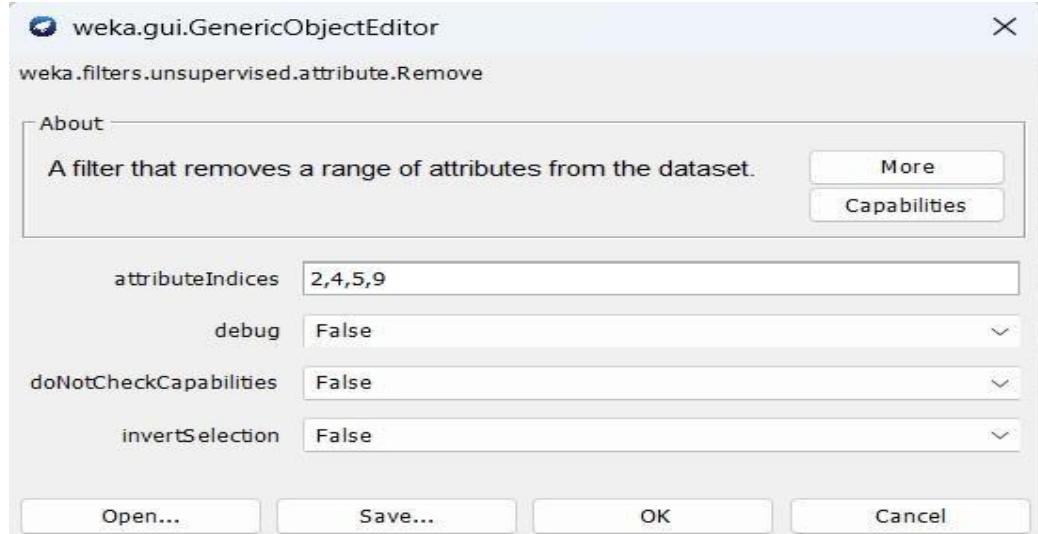
- Open WEKA tool.
- Then click on Explorer Button.
- Now click on open file button & open **Soyabean.arff** file.

Before Apply filter:

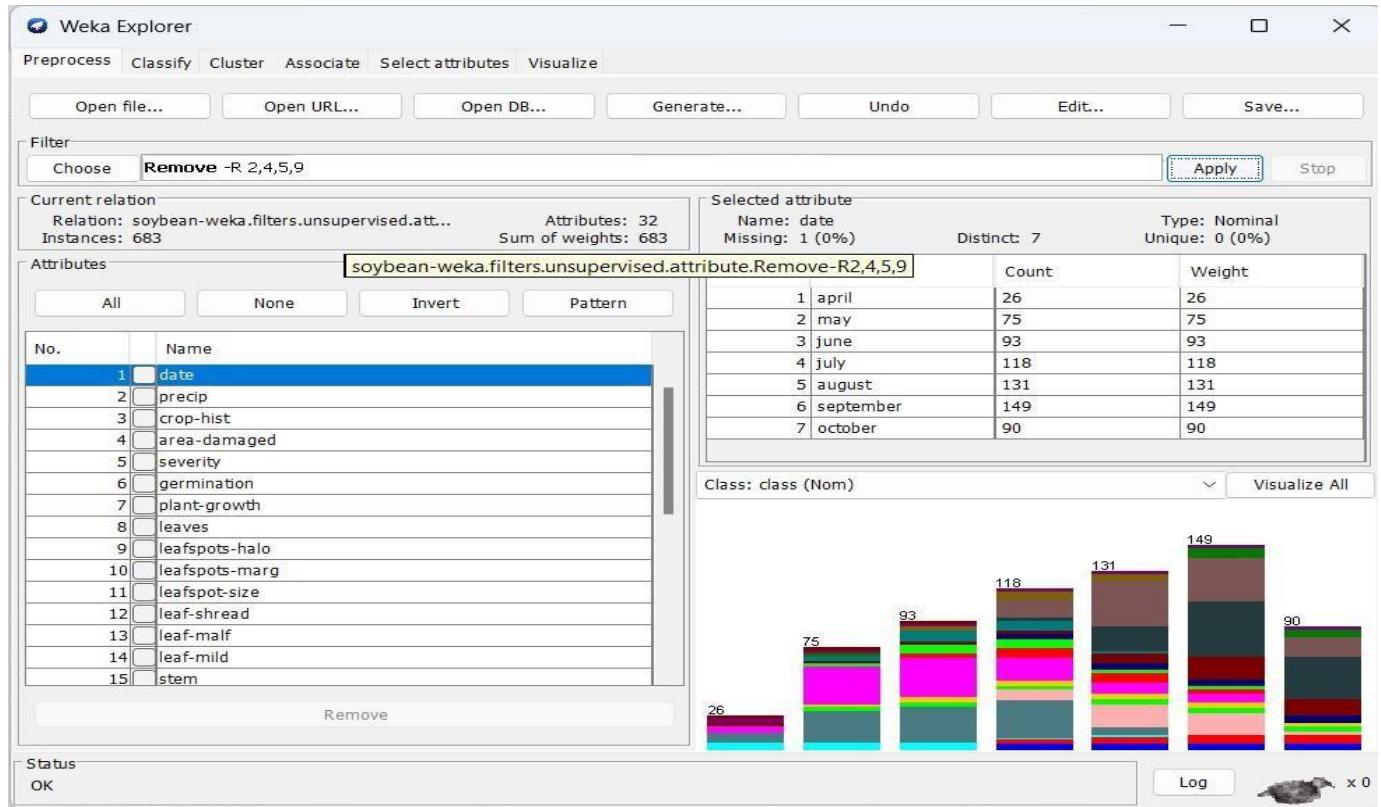


- 1) Click on the Choose button in the Filter sub window and select the following filter.

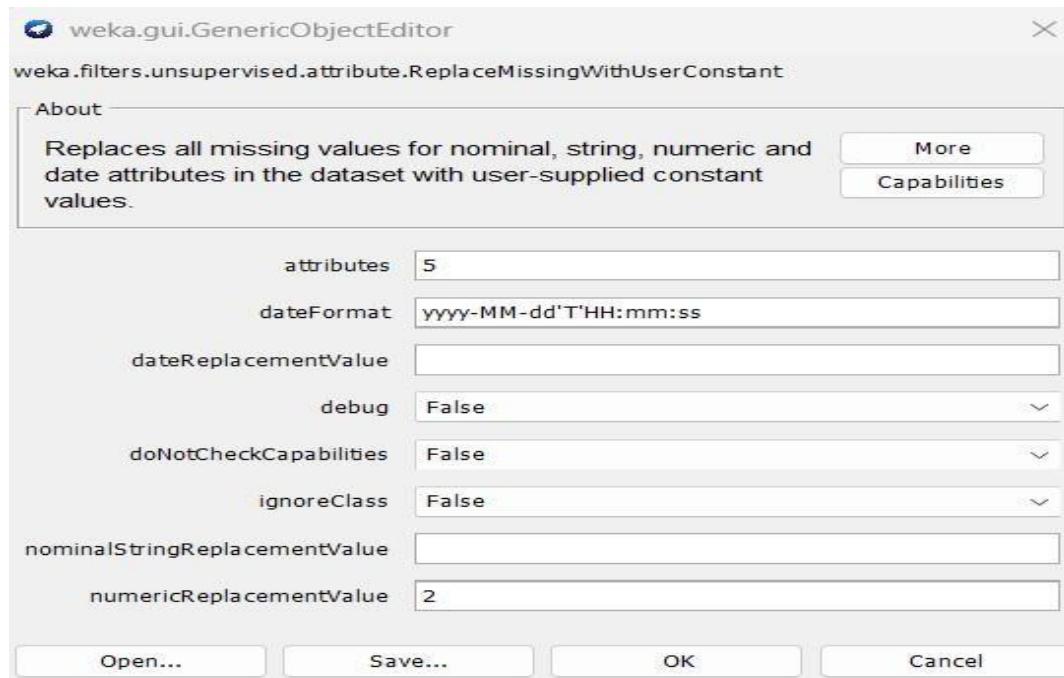
**weka→filter→unsupervised→attribute→attribute → ReplaceMissing With User Constant**



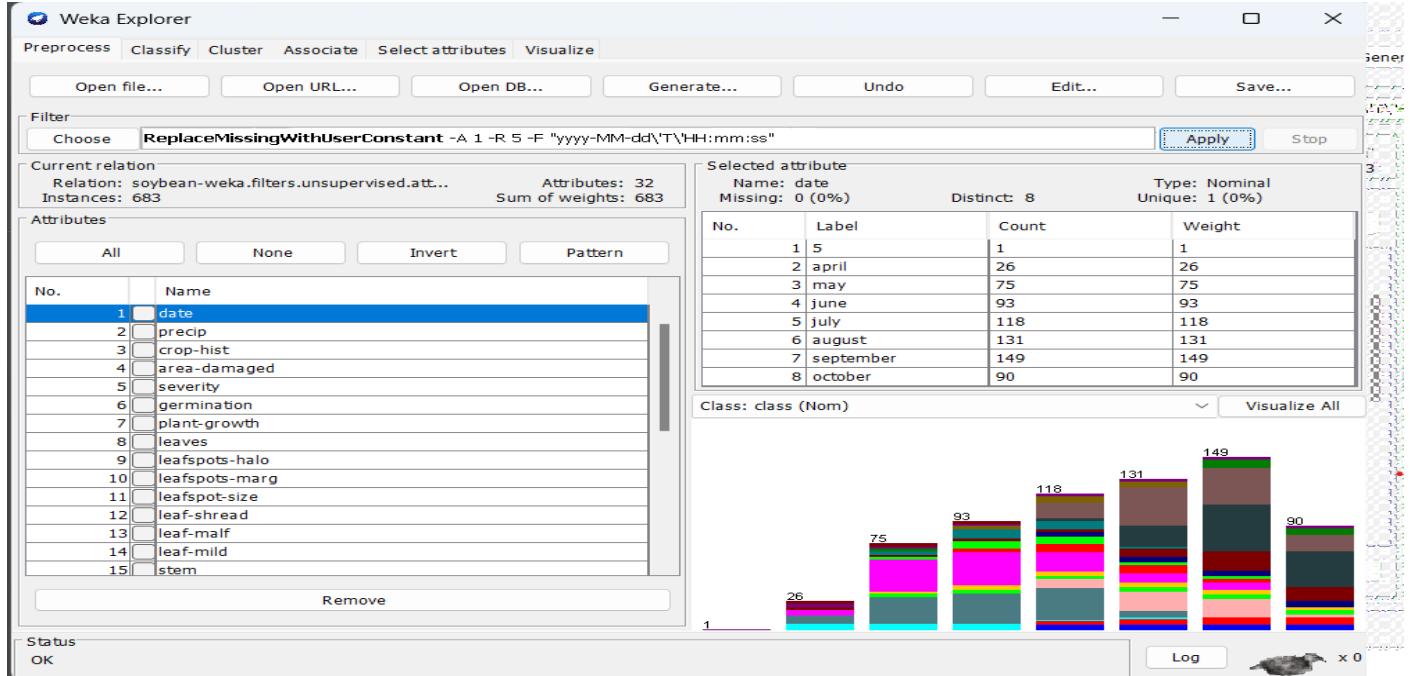
- 2) Click on apply button and examine all attributes.



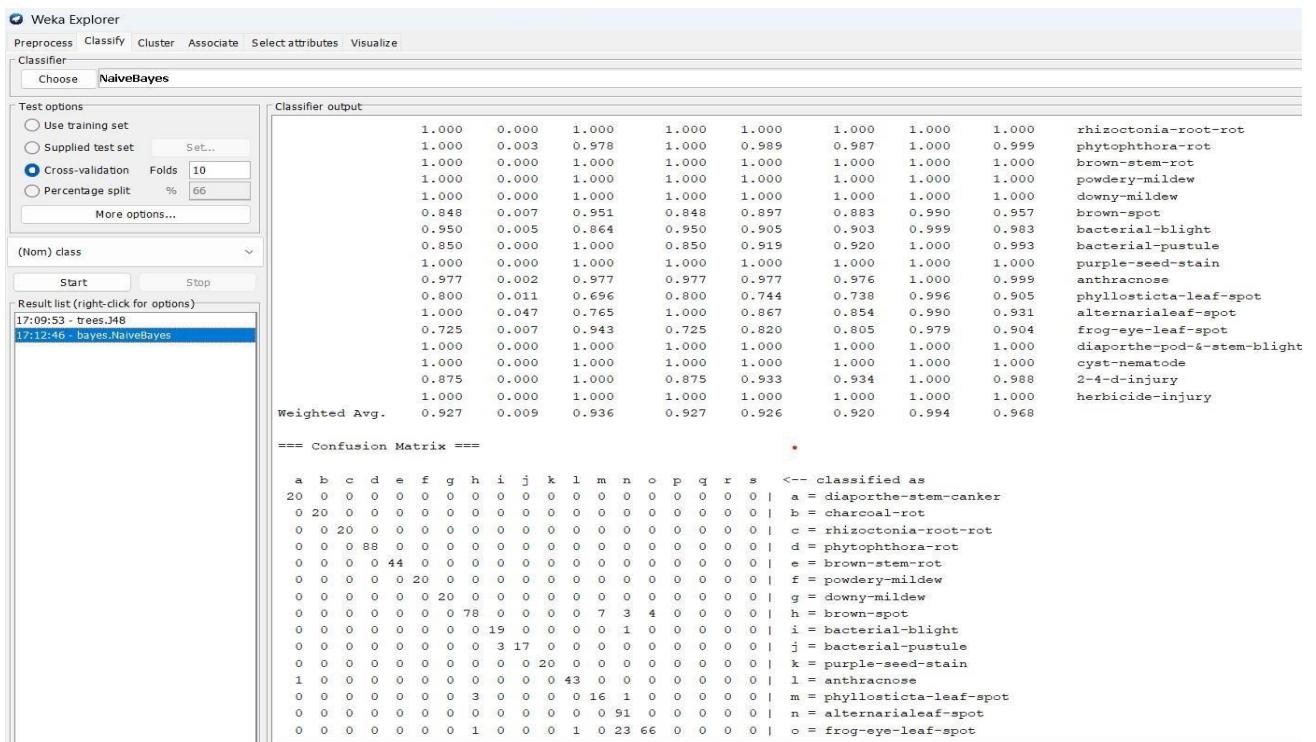
weka → filter → unsupervised → attribute → attribute → ReplaceMissing With User Constant



After click on Apply Button.



classify → choose → weka → bayes → NaïveBayes



**(2) Implementation of Classification techniques on Customer data set (Telecommunication) in Data Mining using WEKA tool.**

- Classification is the process of grouping things according to shared characteristics, properties & structure.
  - We will use **J48 classifier** for our practical.

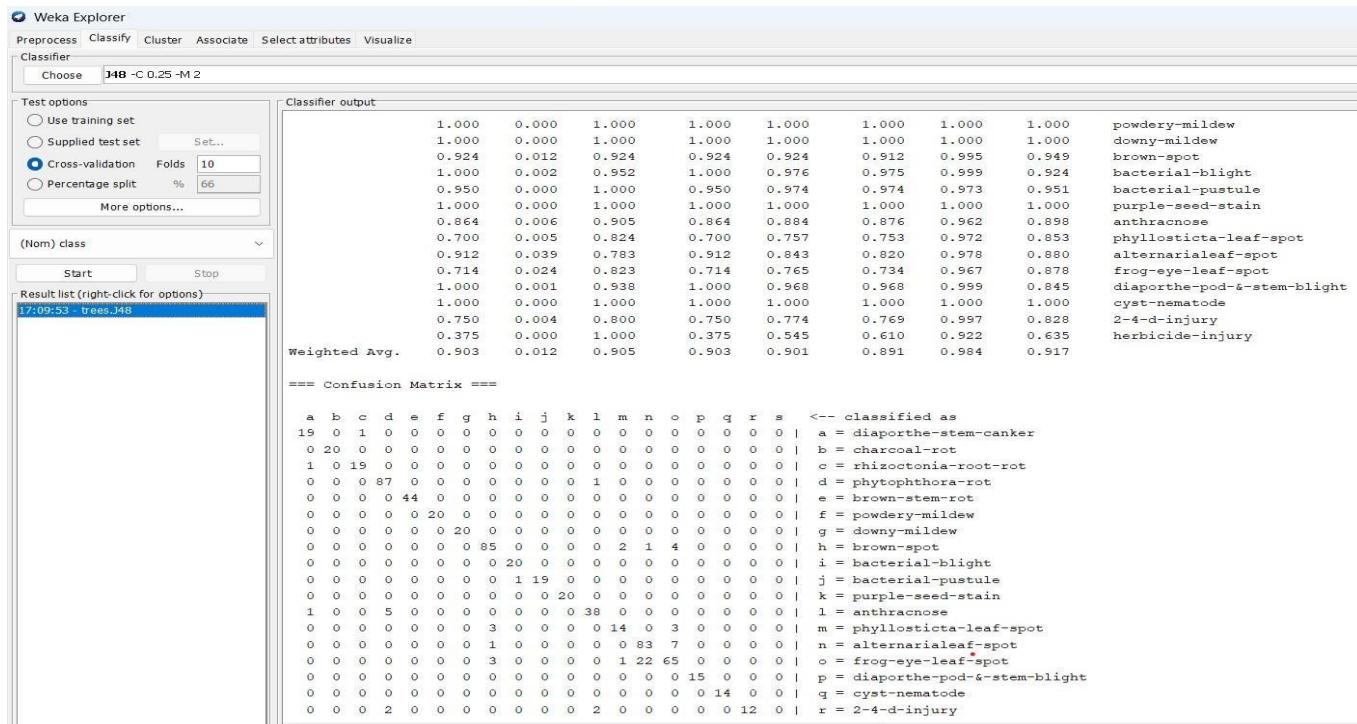
### J48 Algorithm:

- J48 is a **machine learning decision tree classification algorithm based on Iterative Dichotomiser 3**.
  - It is very helpful in examine the data categorically and continuously. It can help not only to make accurate predictions from the data but also to explain the patterns in it. **Steps:**

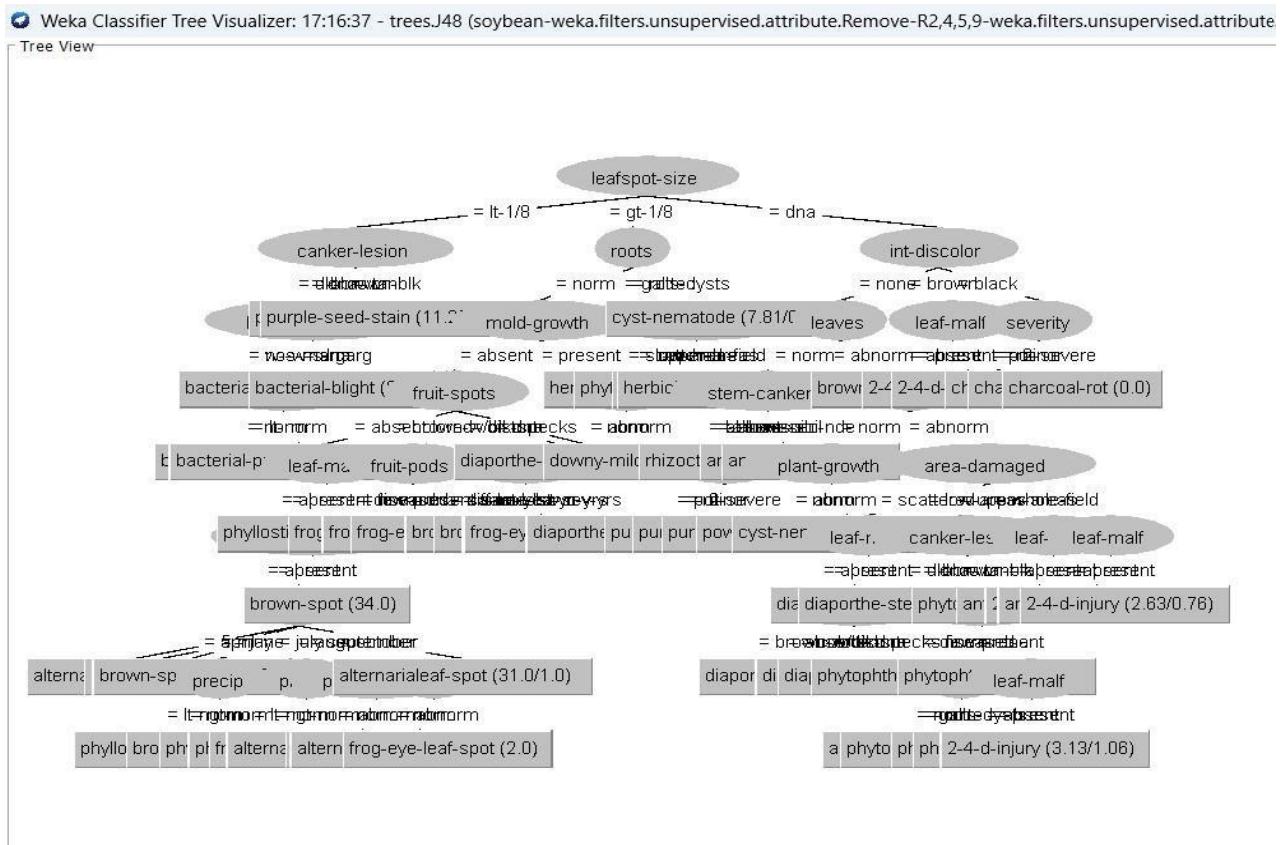
1. Here we use Soyabean.arff **Dataset**.
  2. Open WEKA tool.
  3. Then click on Explorer Button.
  4. Now click on open file button & open **Soyabean.arff** file.
  5. Click on the Choose button in the Classify tab and select the following Classifier Algo.

## **Weka-classifiers-trees-J48**

6. Click on the Start button and examine the Result.



7. Now, Right click on trees J48 & click on Visualize tree option.



## Practical: 10

**Aim: Preprocess and classify Weather dataset.**

**Solution:**

**(1) Implementation of Preprocessing techniques on Weather data set (seattle- weather2014-15) in Data Mining using WEKA tool**

In this practical, we will use **Dataset: seattle -weather-2014-15.arff & ReplaceMissingValues (Missing Values with mode & mean) Filter.**

1. In this Practical, we will use Telco **seattle -weather-2014-15.arff Dataset.**
2. Open WEKA tool.
3. Then click on Explorer Button.
4. Now click on open file button & open **seattle -weather-2014-15.arff** file.

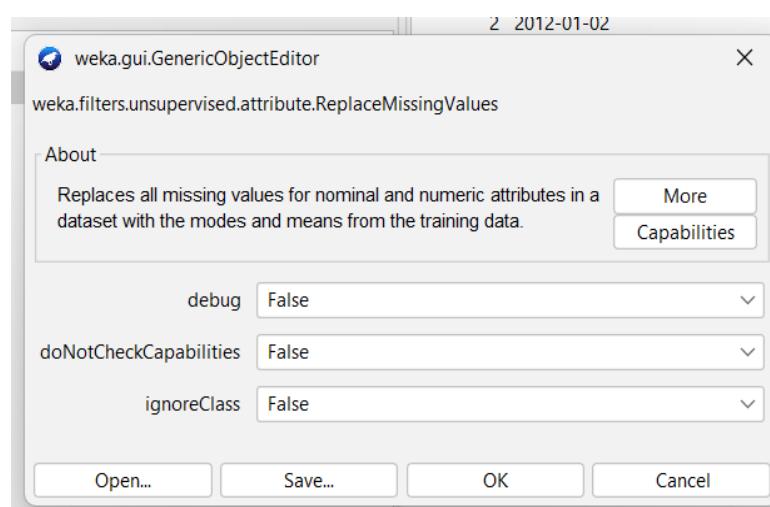
**Before Apply filter:**

No.	Label	Count	Weight
1	2012-01-01	1	1
2	2012-01-02	1	1
3	2012-01-03	1	1
4	2012-01-04	1	1
5	2012-01-05	1	1
6	2012-01-06	1	1
7	2012-01-07	1	1
8	2012-01-08	1	1
9	2012-01-09	1	1
10	2012-01-10	1	1
11	2012-01-11	1	1
12	2012-01-12	1	1

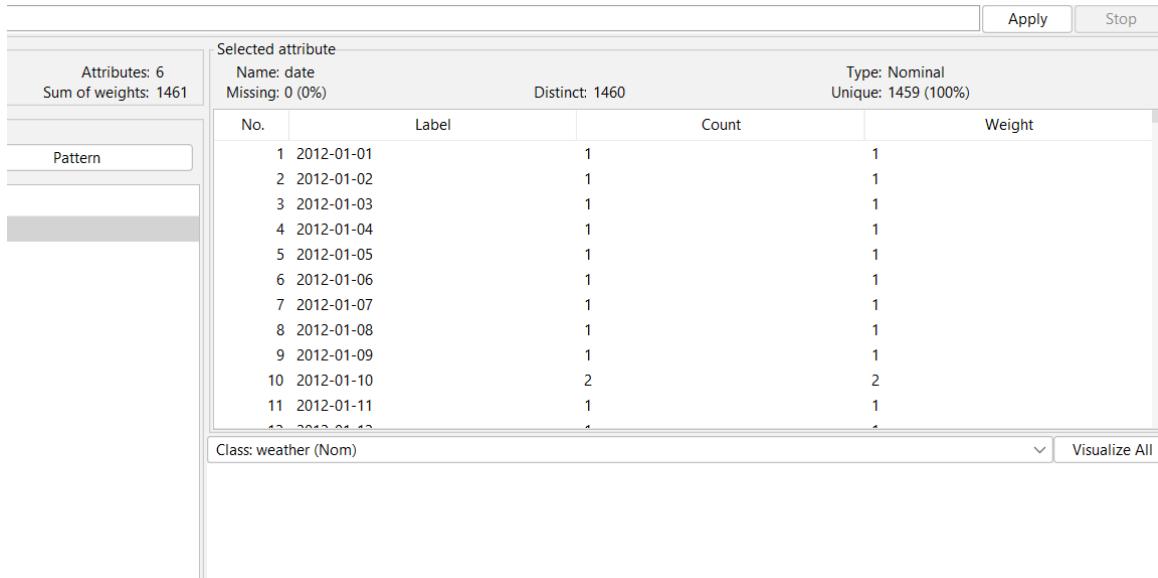
XThere are following 3 missing values in **seattle-weather-2014-15 dataset.**

No.	1: date	2: precipitation	3: temp_max	4: temp_min	5: wind	6: weather
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	2012-0... 0.0		12.8		4.7	
2	2012-0... 10.9			2.8	4.5	rain ▾
3	2012-0... 0.8		11.7	7.2		rain
4	2012-0... 20.3		12.2	5.6	4.7	rain
5	2012-0... 1.3		8.9	2.8	6.1	rain
6	2012-0... 2.5			2.2	2.2	rain
7	2012-0... 0.0		7.2	2.8	2.3	rain
8	2012-0... 0.0		10.0	2.8	2.0	sun
9	2012-0... 4.3		9.4	5.0	3.4	rain
10	2012-0... 1.0		6.1	0.6	3.4	rain
11	2012-0... 0.0		6.1	-1.1	5.1	sun
12	2012-0... 0.0		6.1	-1.7	1.9	sun
13	2012-0... 0.0		5.0	-2.8	1.3	sun
14	2012-0... 4.1		4.4	0.6	5.3	snow
15	2012-0... 5.3		1.1	-3.3	3.2	snow
16	2012-0... 2.5		1.7	-2.8	5.0	snow
17	2012-0... 8.1		3.3	0.0	5.6	snow
18	2012-0... 19.8		0.0	-2.8	5.0	snow
19	2012-0... 15.2		-1.1	-2.8	1.6	snow
20	2012-0... 13.5		7.2	-1.1	2.3	snow

- Click on the Choose button in the Filter sub window and select the following filter  
– weka→filters→unsupervised→attribute→ ReplaceMissingValue.



7. Click on the Apply button and examine the all attribute.

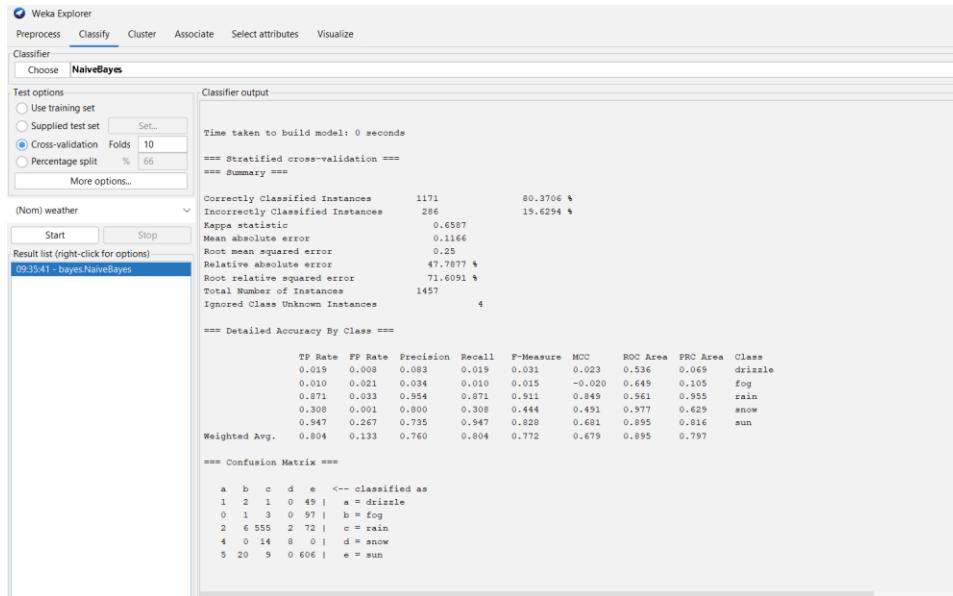


## (2) Implementation of Classification techniques on Weather data set (seattle - weather2014-15) in Data Mining using WEKA tool.

- ✓ Classification is the process of grouping things according to shared characteristics, properties & structure.
- ✓ We will use **J48 classifier & NaiveBayes classifier** for our practical.

### ➤ Steps:

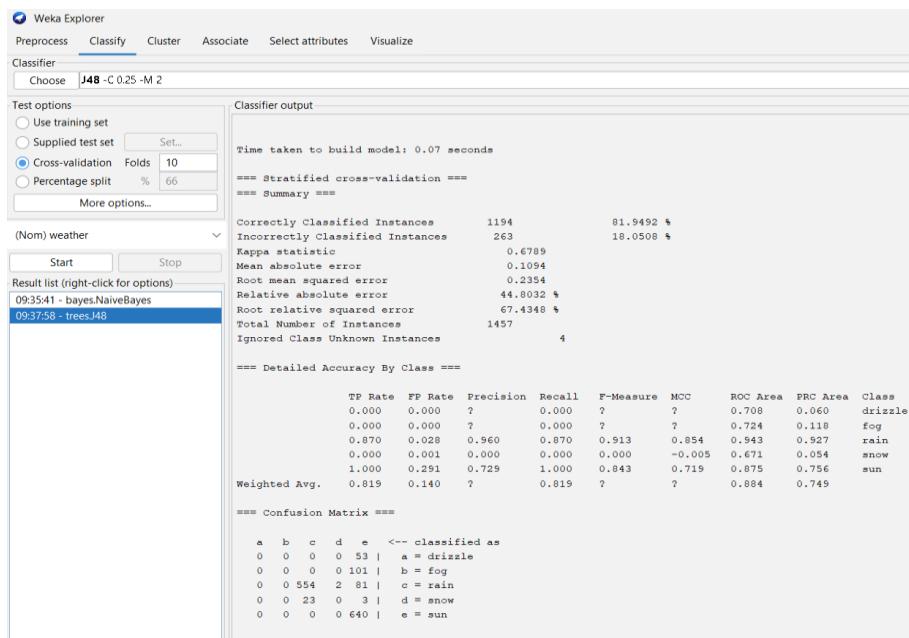
1. Here we use **seattle -weather-2014-15.arff Dataset**.
2. Open WEKA tool.
3. Then click on Explorer Button.
4. Now click on open file button & open seattle -weather-2014-15.arff file.
5. Click on the Choose button in the Classify tab and select the following Classifier Algo. – weka→classifiers→Bayes→NaiveBayes
6. Click on the Start button and examine the Result.



7. Now we will use second classifier that is J48.

8. Click on the Choose button in the Classify tab and select the following Classifier Algo.  
– weka→classifiers→trees→J48

9. Click on the Start button and examine the Result.



## Practical: 11

**Aim: Classify the dataset using decision tree.**

**Solution:**

**of Classification technique on data set (Student Education Data) using Decision Tree in Data Mining using WEKA tool.**

- Classification is the process of grouping things according to shared characteristics, properties & structure.
- There are different classification algorithms used in data mining.
  - Decision Trees
  - Bayesian Classifiers
  - Neural Networks
  - K-Nearest Neighbour
  - Support Vector Machines
  - Linear Regression
  - Logistic Regression
- In this practical, we will use J48 Classifier Algorithm for Decision Tree.

❖ **Decision Trees:**

- Decision tree algorithm falls under the category of supervised learning.
- It is the most powerful & popular tool for classification & regression problems.
- It is very simple & easy to understand.
- Decision Tree is the classification technique that consists of three components:

**(1) Root node :** It represents the test condition for different attributes.

The root node is at the starting of the tree which is also called the top of the tree.

**(2) Branch (edge or link) :** It represents all possible outcomes that can be there in the test

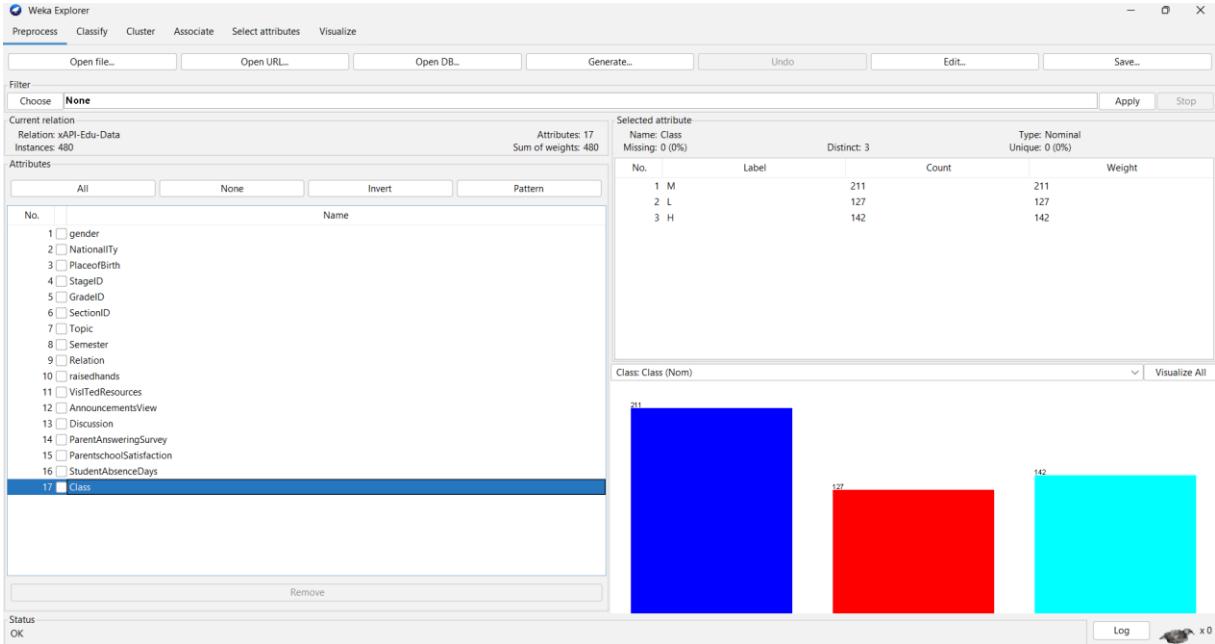
**(3) Leaf node :** This node contain the label of the class to which it belongs.

**Steps:**

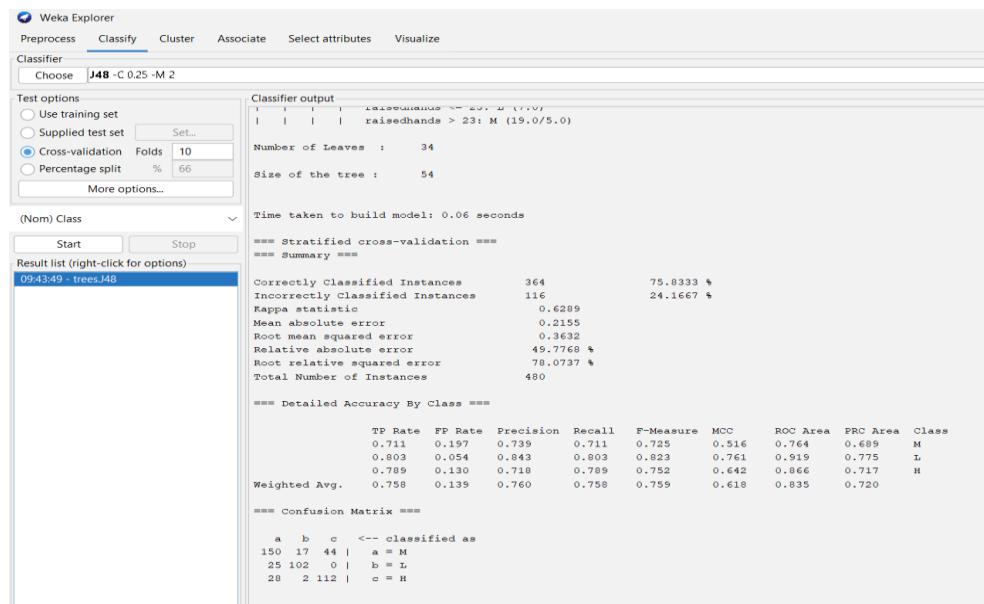
1. Here we use **xAPI-Edu-Data.arff Dataset**.

2. Open WEKA tool.
3. Then click on Explorer Button.

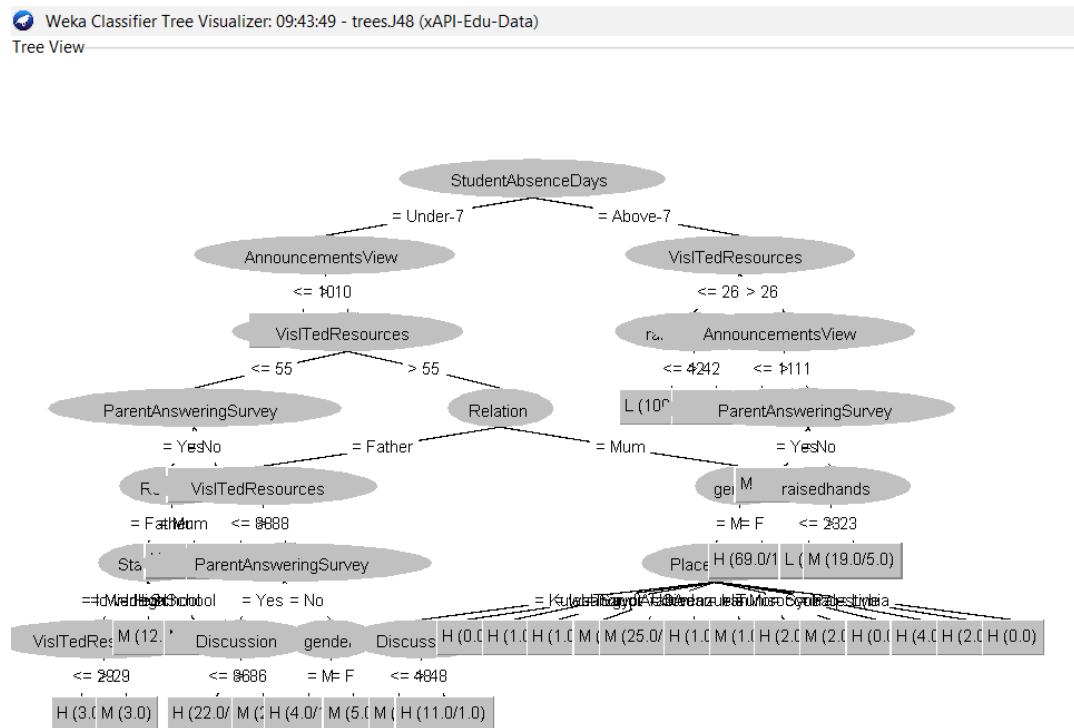
Now click on open file button & open xAPI-Edu-Data.arff file.



4. Click on the Choose button in the Classify tab and select the following Classifier Algo. - weka→classifiers→trees→J48
5. Click on the Start button and examine the Result.



6. Now, Right click on trees.J48 & click on Visualize tree option.



## **Practical: 12**

**Aim: To perform Classification on various datasets and compare the accuracy of various Algorithms.**

**Solution:**

In this practical, we will use **Naïve Bayes & J48 classifier algorithm**. First of all let us try to understand these 3 algorithms.

### **Naïve Bayes Algorithm:**

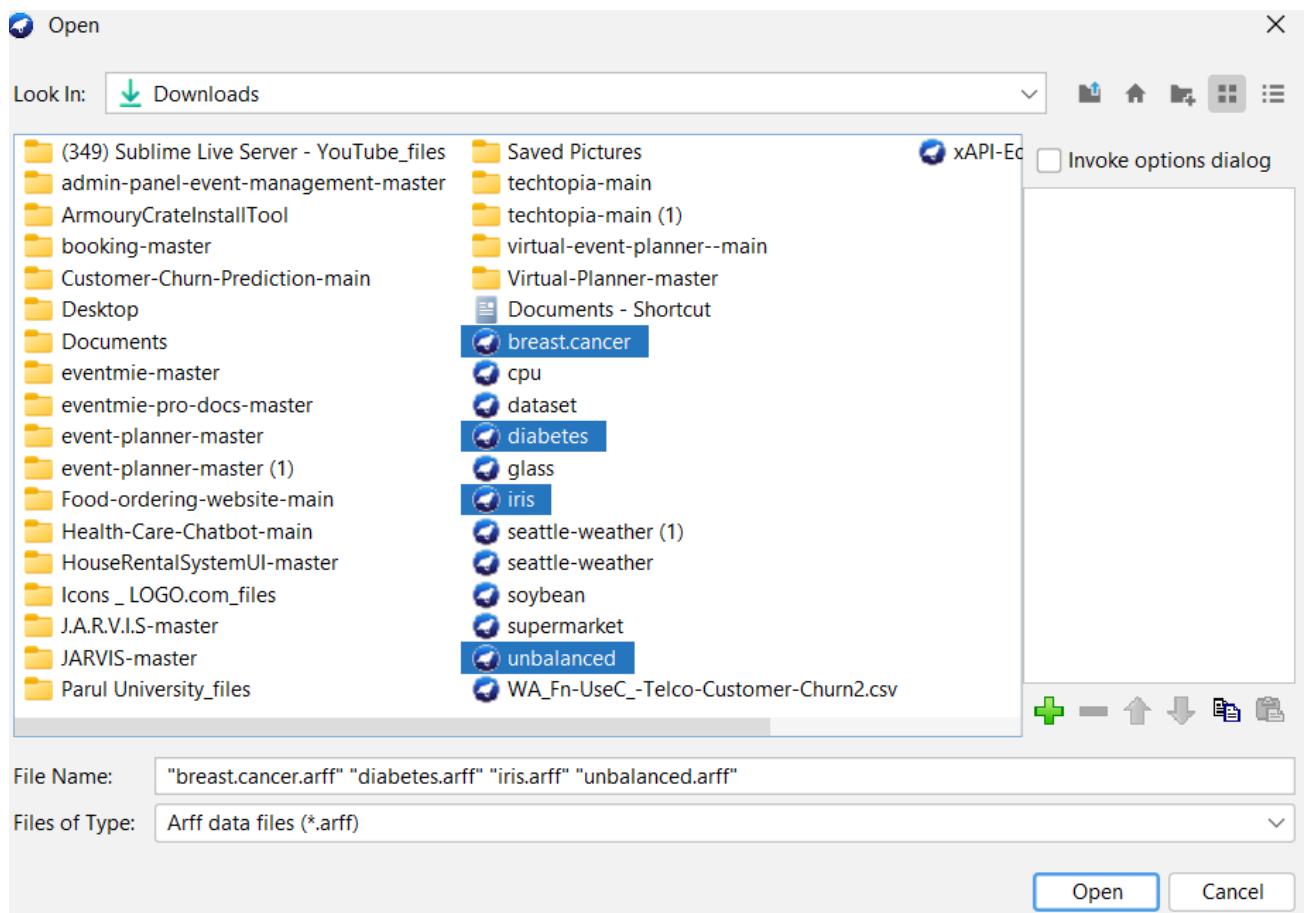
- Naïve Bayes algorithm is a supervised learning algorithm & used for solving classification problems.
- It is based on Bayes' Theorem.
- It is one of the simple & most effective Classification algorithms.
- It is extremely fast compared to other algorithms.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.
- 

### **J48 Algorithm:**

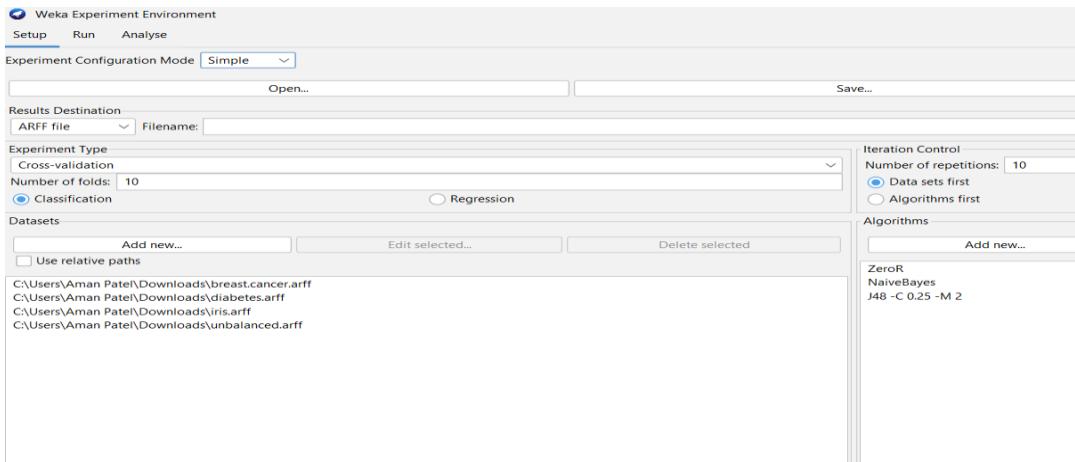
- J48 is a **machine learning decision tree classification algorithm based on Iterative Dichotomiser 3**.
- It is very helpful in examine the data categorically and continuously.
- It can help not only to make accurate predictions from the data but also to explain the patterns in it. **Steps:**

1. Open WEKA tool.
2. Then click on Experimenter Application.
3. Weka Experiment Environment window will be opened.
4. There are 3 sections in this Weka Experiment Environment window.
  - (1) Setup
  - (2) Run
  - (3) Analyse
  - (4) Now, click on New Button. So, all inactive tabs will be activated.

- (5) Select classification radio button. By default, it is selected.
- (6) Now, select different datasets which you like to work upon. For that, click on Add New button under Datasets window and select datasets.
- (7) Here we use 4 different datasets like **iris.arff**, **breast-cancer.arff**, **diabetes.arff** &**unbalanced.arff**.



- (8) Now, we will select different algorithm. For that, click on Add New button under Algorithms window.
- (9) A small window will be opened. By default, Zero-R algorithm will be opened.
- (10) We will add other two algorithms that is Naïve Bayes & J48 classifier algorithms like this.
- (11) Now, we will see 2nd option that is Run Option. For that click on Run option.



(12) Then click on Strat button. We could see the result immediately.



- (13) Now, we will see 3rd option that is Analyse Option. For that click on Analyse option.  
 (14) Then click on Experiment button.  
 (15) Now, click on Performance tab button under Action window.  
 (16) You can see the comparison table of all datasets.  
 (17) So, according to accuracy of algorithms, you can identify that which algorithm is best.

```
Tester:      weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -result-matrix "weka.experiment.ResultMatrixPlainText -mean-prec 2 -stddev-prec 2 -c  
Analysing:  Percent_correct  
Datasets:   4  
Resultsets: 3  
Confidence: 0.05 (two tailed)  
Sorted by:  -  
Date:       9/26/24, 10:07 AM
```

Dataset	(1) rules.Ze		(2) bayes	(3) trees		
breast-cancer	(100)	70.30		72.70	74.28	
pima_diabetes	(100)	65.11		75.75	v 74.49	
iris	(100)	33.33		95.53	v 94.73	
unbalanced	(100)	98.60		90.83	*	98.60

-----  
(v/ /\*) | (2/1/1) (2/2/0)

Key:  
(1) rules.ZeroR '' 48055541465867954  
(2) bayes.NaiveBayes '' 5995231201785697655  
(3) trees.J48 '-C 0.25 -M 2' -217733168393644444

## **Practical: 13**

**Aim: Perform Association technique on Weather & Agriculture dataset.**

**Solution:**

**(1) Perform Association technique using Apriori Algorithm on Agriculture dataset (Soyabean) in Data mining using WEKA tool.**

➤ What is Association in data mining?

Association is a data mining function that discovers the probability of the co- occurrence of items in a collection. The relationships between co-occurring items are expressed as Association Rules.

➤ How do you do association rule mining?

**Steps involved in Association Rule Mining:**

Step 1: Find all frequent item sets. An itemset is a set of items that occurs in a shopping basket. ... Step 2: Generate strong association rules from the frequent item sets. Association rules are generated by building associations from frequent item sets generated in step 1.

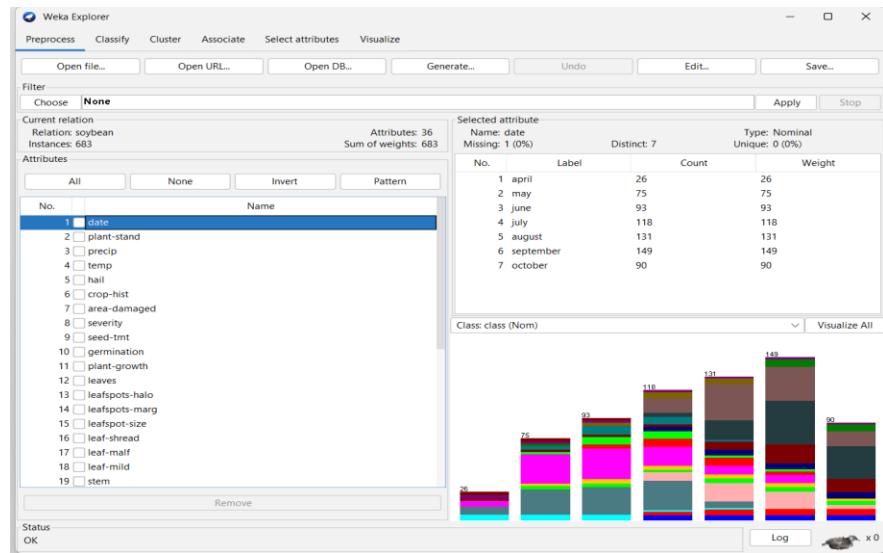
➤ **Introduction to Apriori Algorithm:**

There are many algorithms used to mining data.

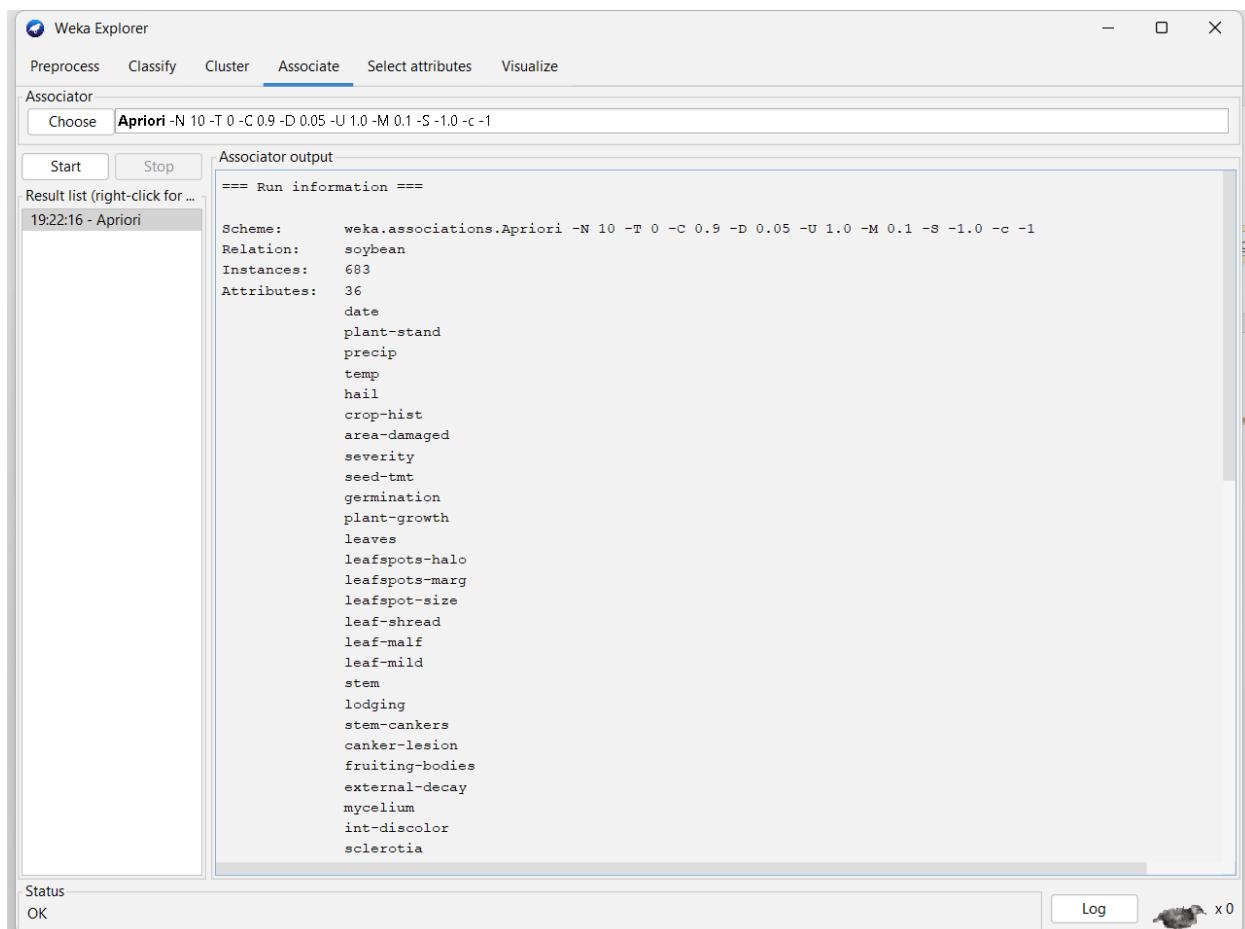
- Apriori & Cluster are the first-rate & most famed algorithms.
- Apriori is the simple algorithm, which applied for mining of repeated the patterns from the transaction dataset to find frequent item sets & association between various item sets.

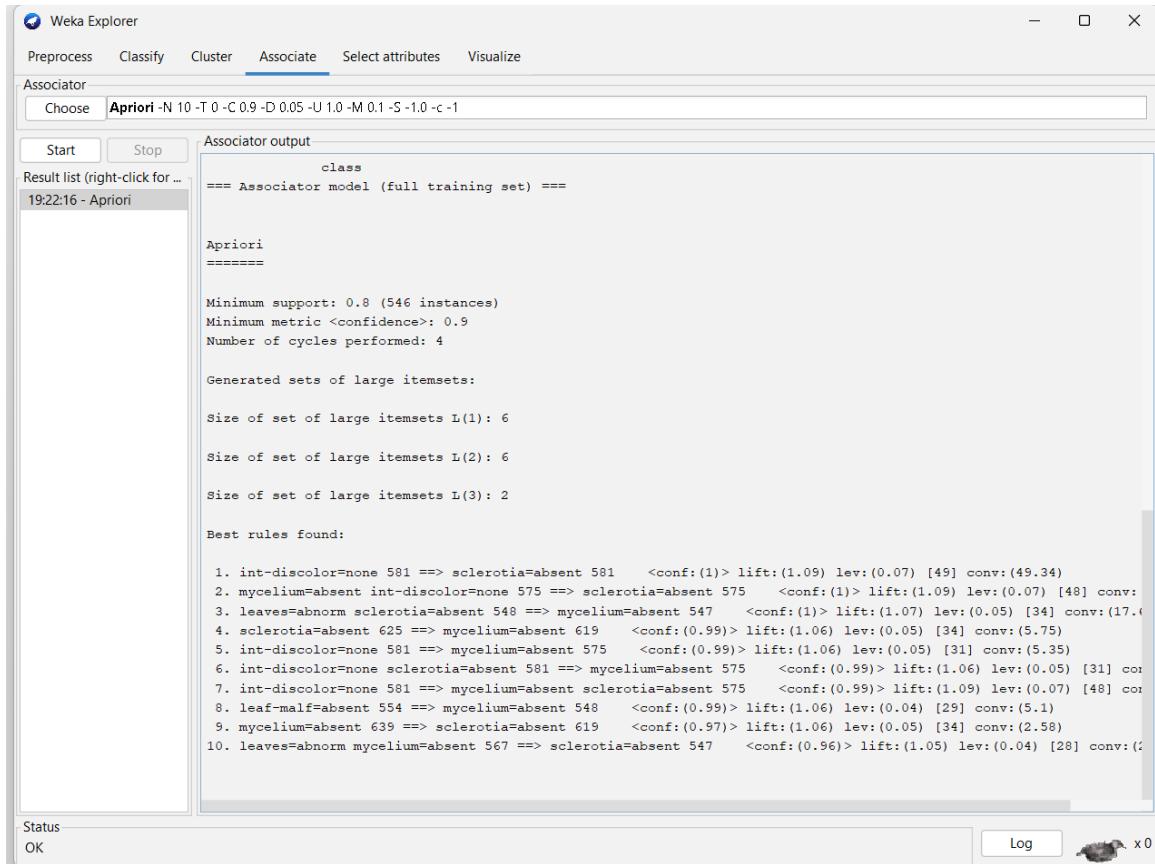
➤ **Steps:**

1. Here we will use soybean.arff Dataset.
2. Open WEKA tool.
3. Then click on Explorer Button.
4. Now click on open file button & open soyabean.arff file



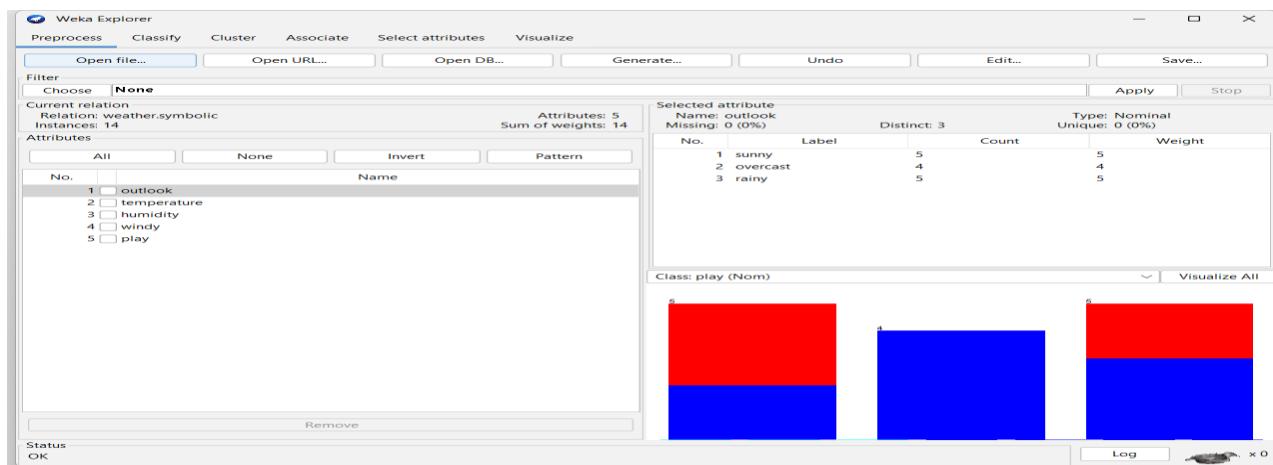
5. on the Associate tab and select the **Apriori Algorithm**.
6. Click on the Start button and examine the Result.





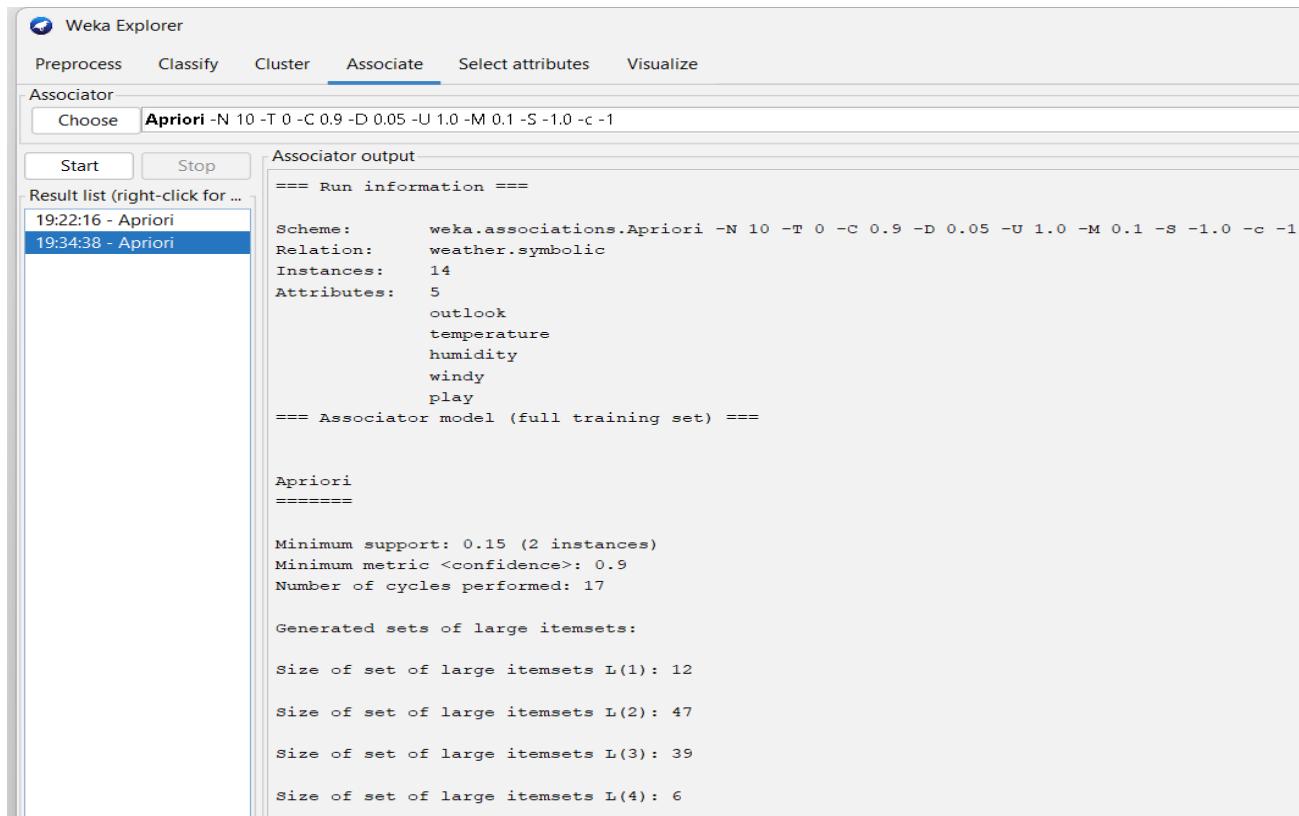
## (2) Perform Association technique using Apriori Algorithm on Weather dataset in Data mining using WEKA tool. ➤ Steps:

- (1) Here we will use weather.nominal.arff Dataset.
- (2) Open WEKA tool. Then click on Explorer Button.
- (3) Now click on open file button & open weather.nominal.arff file.



(4) Click on the Associate tab and select the **Apriori Algorithm**.

(5) Click on the Start button and examine the Result.



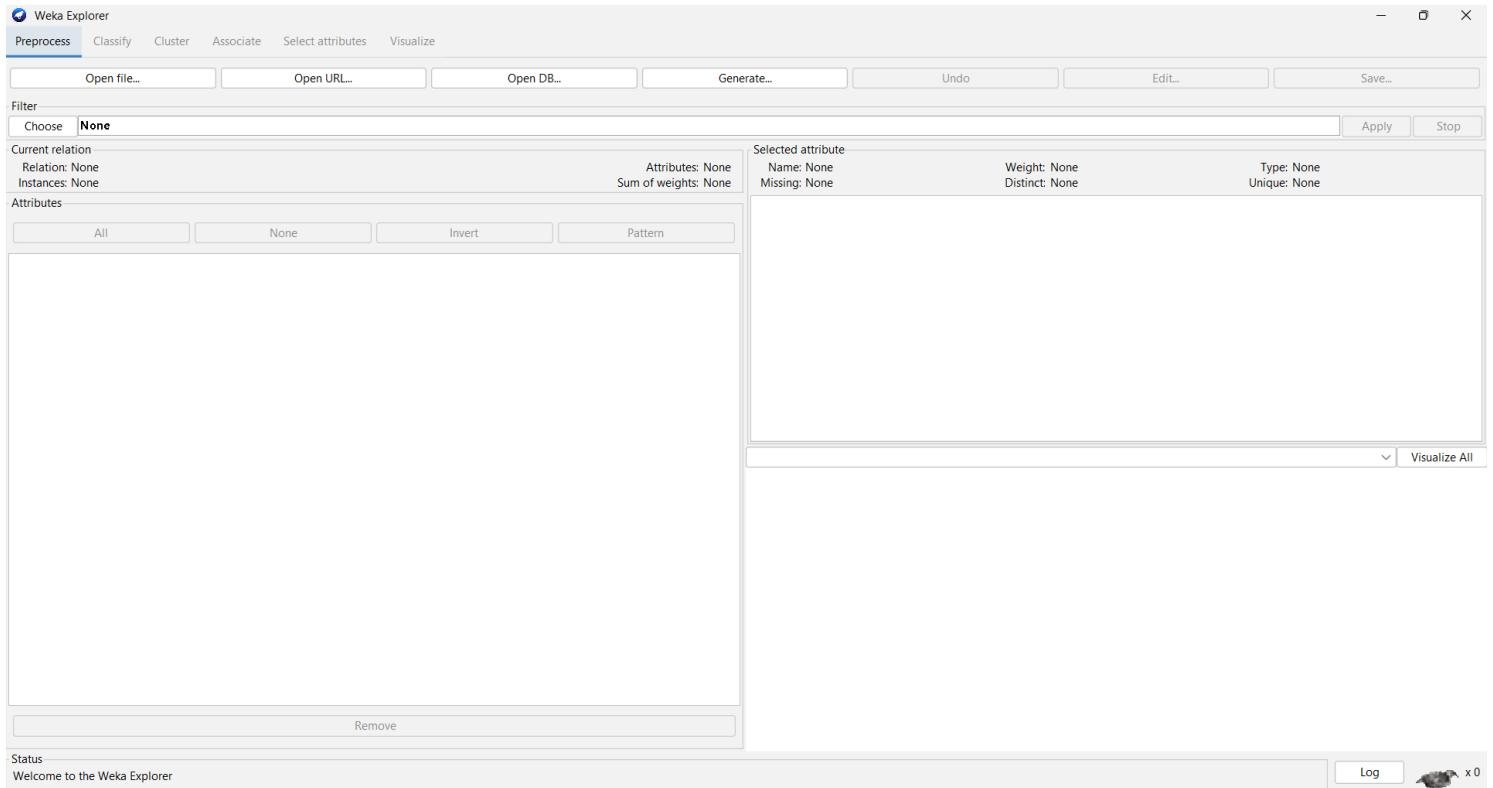
## Practical: 14

**Aim: Demonstration of Weka Explorer, Mining techniques and Attribute Relation File Format (ARFF).**

**Solution:**

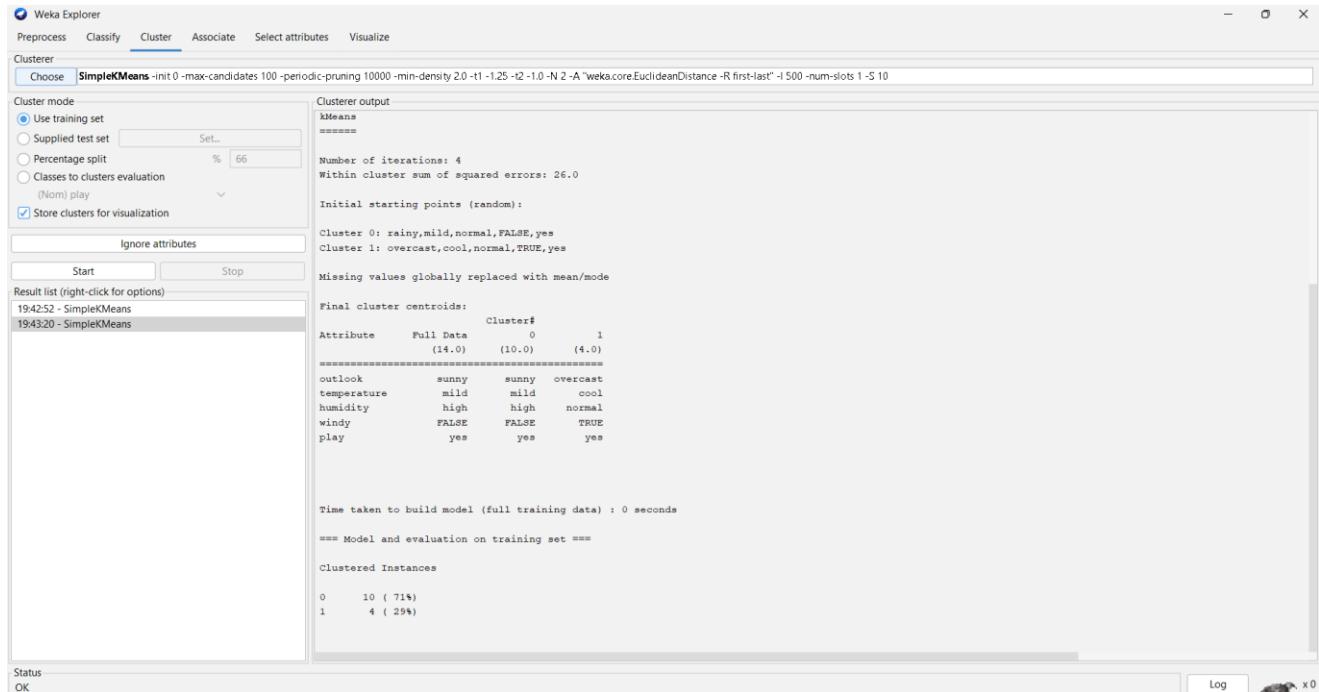
### **[1] WEKA Explorer:**

- The Weka Explorer is an easy-to-use graphical user interface that harnesses the power of the weka software.
- Weka tool can be opened using three methods:
  - 1) From C:\Program Files directory
  - 2) From your desktop by selecting WEKA tool icon
  - 3) From the Windows task bar -> ‘Start’ -> ‘Programs’ -> ‘Weka 3.8.6’
- After that, the main window of WEKA tool (Weka GUI chooser) will be opened like this on your screen. Explorer is an environment for exploring data.



## [2] Mining Techniques:

1. Preprocessing
2. Classification
3. Decision Tree
4. Clustering
5. Association
6. Visualization



### [3] Attribute Relation File Format (ARFF):

- An ARFF (Attribute-Relation File Format) file is **an ASCII text file that describes a list of instances sharing a set of attributes.**
- ARFF files were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software
- ARFF files have two distinct sections.
- The first section is the **Header** information, which is followed by the **Data** information.
- The Header of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types.



```
menubar.txt | email i | $dbc = | a1.html
File Edit View
@relation weather.symbolic
@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```