JSPM'S

Bhivarabai Sawant Institute of Technology & Research Wagholi,

Pune-412207

Department Of Computer Engineering

Academic Year 2020-21

LP-II Mini Project Report

On

"Staff Billing System"

Submitted by: Sanobar Shaikh (BE COMP -52)

Arati Thorat (BE COMP -56)

Under the guidance of

Prof. Vijay Sonawane

Subject : Laboratory Practice –II

DEPARTMENTOFCOMPUTERENGINEERING

BHIVARABAI SAWANT INSTITUTE OF TECHNOLOGY & RESEARCH

WAGHOLI, PUNE – 412 207

CERTIFICATE

This is to certify that the **Sanobar Shaikh** and **Arati Thorat** submitted their Project report on under my guidance and supervision. The work has been done to my satisfaction during the academic year 2020-2021 under SavitribaiPhule Pune University guidelines.

Date: 15-11-2020

Place: BSIOTR, PUNE.

Prof. Vijay Sonawane
Project Guide

Dr.Prof. Gayatri Bhandari
H.O. D.

ACKNOWLEDGEMENT

This is a great pleasure & immense satisfaction to express our deepest sense of gratitude & thanks to everyone who has directly or indirectly helped us in completing our Project work successfully.

We express our gratitude towards guide Prof. Vijay Sonawane and Dr.Prof. G. M. Bhandari Head of Department of Computer Engineering, Bhivarabai Sawant Institute Of Technology and Research, Wagholi, Pune who guided& encouraged us in completing the Project work in scheduled time. We would like to thanks our Principal, for allowing us to pursue our Project in this institute.

Sanobar Shaikh

Arati Thorat

INDEX

TITLE

By using WEKA tool implements algorithms for data preprocessing, classification, regression, clustering, association rules and also visualization. Using WEKA Explorer perform preprocessing, classification, clustering, association, attribute selection, and visualization on your dataset.

PROBLEM DEFINATION

perform preprocessing, classification, clustering, association, attribute selection, and visualization on your dataset.

PREREQUISITE

Knowledge of data processing and WEKA tool.

INTRODUCTION

In this project we implements algorithms for data preprocessing, classification, regression, clustering, association rules and also visualization. By using WEKA tool.

We perform different function like data pre-processing, classification, clustering, association attribute selection, and visualization by using WEKA tool. At the end of each problem there is a representation of the results

SYSTEM REQUIREMENTS

Software Requirements:

- Operating System: Windows 10
- Front end: Eclipse-jee-mars-2-win32-x86_64
- Database : MySQL Server 5.5

Hardware Requirements:

- System type: 64-bit Operating System
- Processor :Intel(R) Core(TM)i3-5005U CPU 2.00GHz
- Installed Memory(RAM):8.00GB

LEARNING OBJECTIVES AND OUTCOMES

Learning Objectives:

We are going to learn how implements algorithms for data preprocessing, classification, regression, clustering, association rules and also visualization. By using WEKA tool.

Outcomes:

You are able to perform preprocessing, classification, clustering, association, attribute selection, and visualization by using WEKA tool.

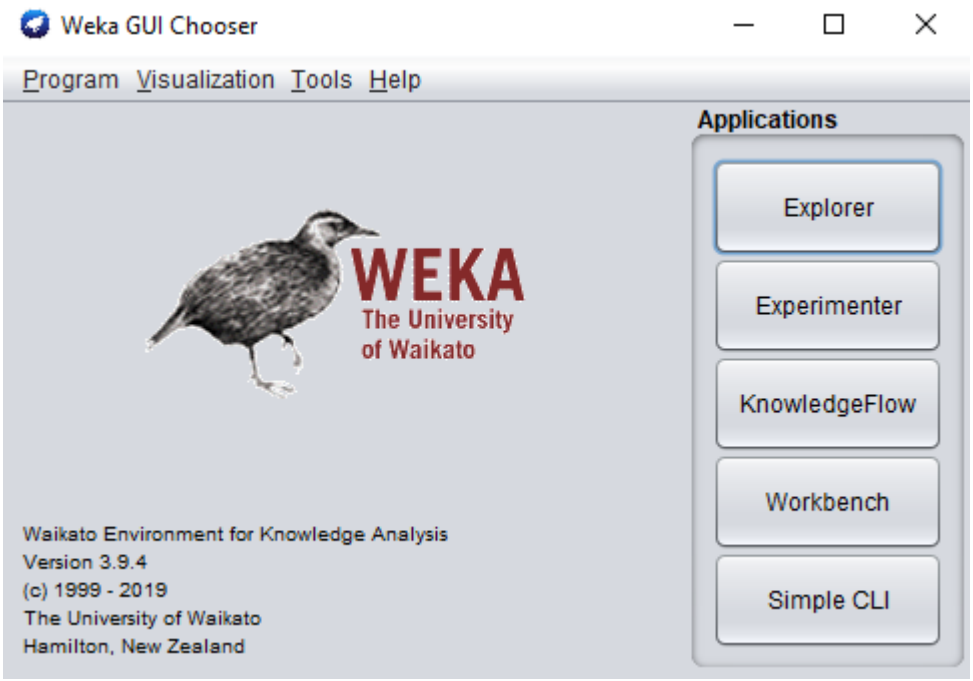
THEORY CONCEPTS

1) What is Data Pre-Processing?

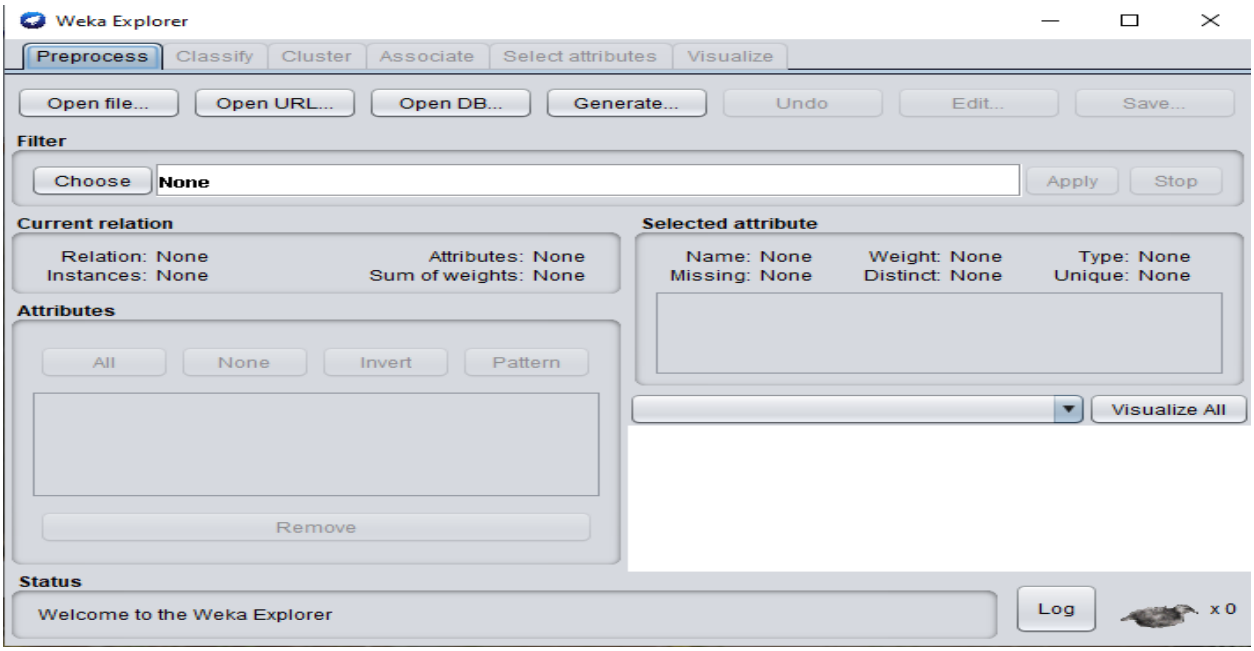
Data preprocessing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, and missing values, etc.

Steps

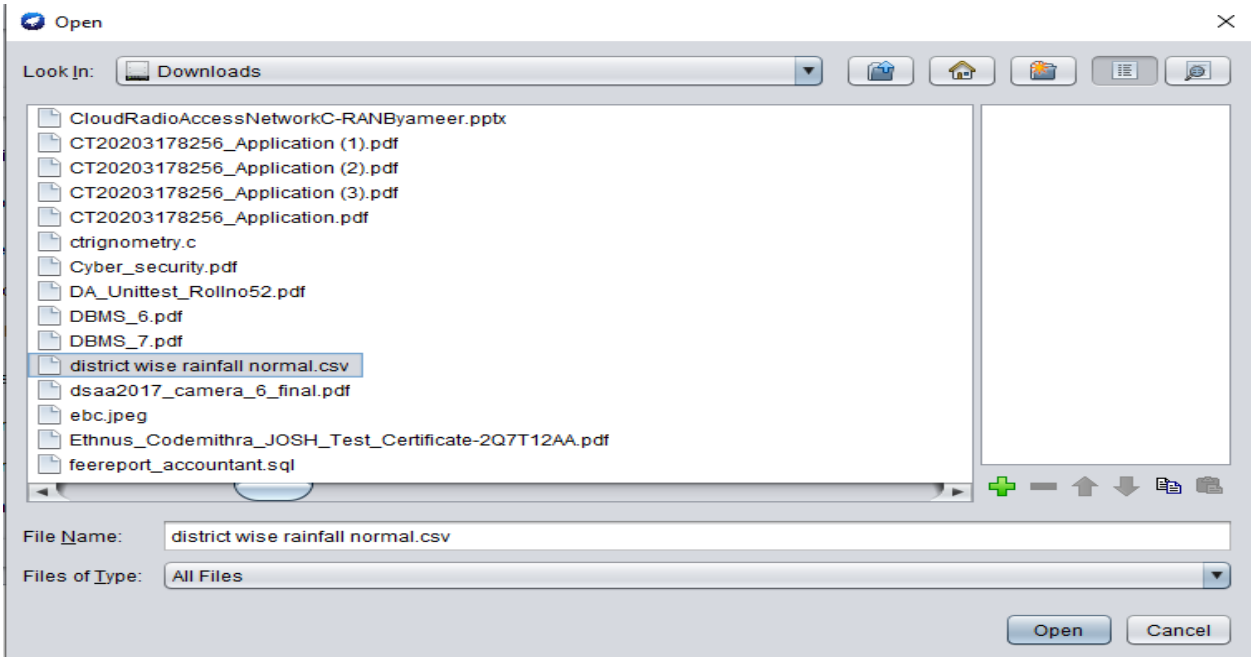
1. First Download Latest Weka tool version.
2. After installiation open the weka tool.



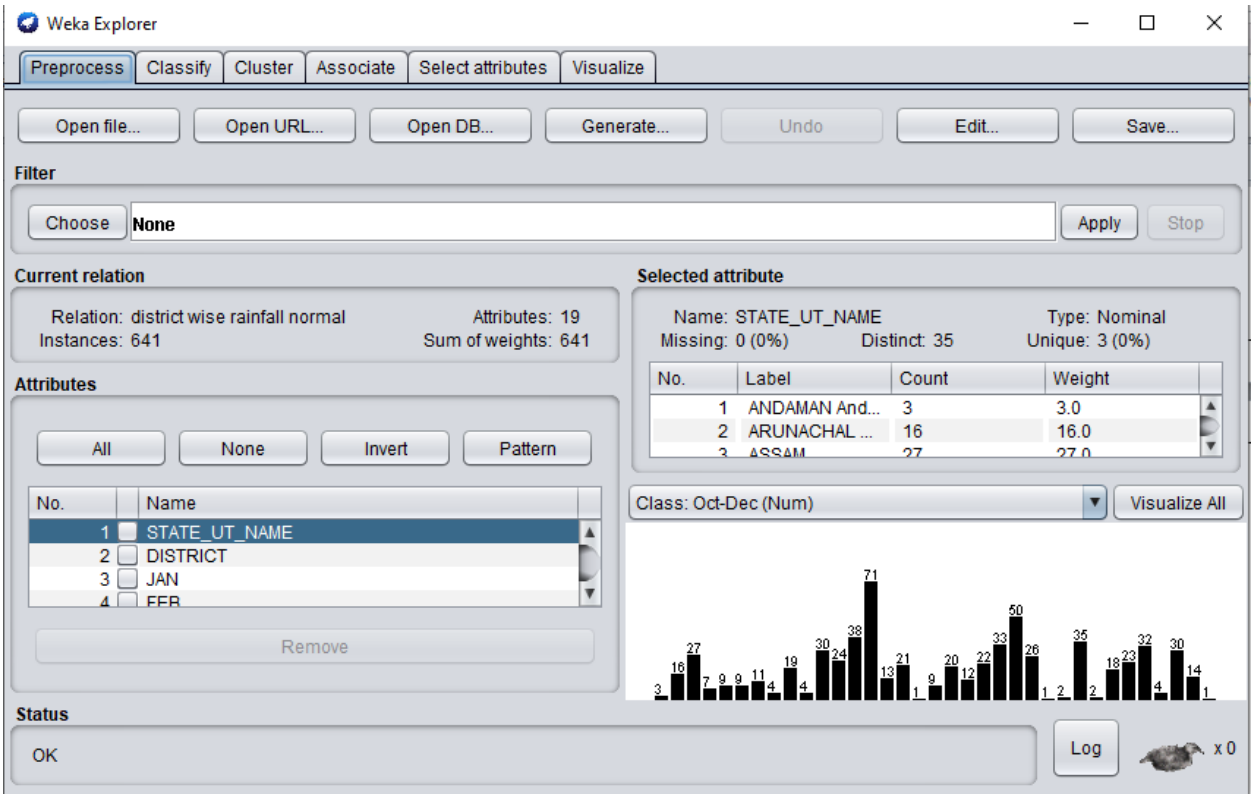
3. Click on Explorer



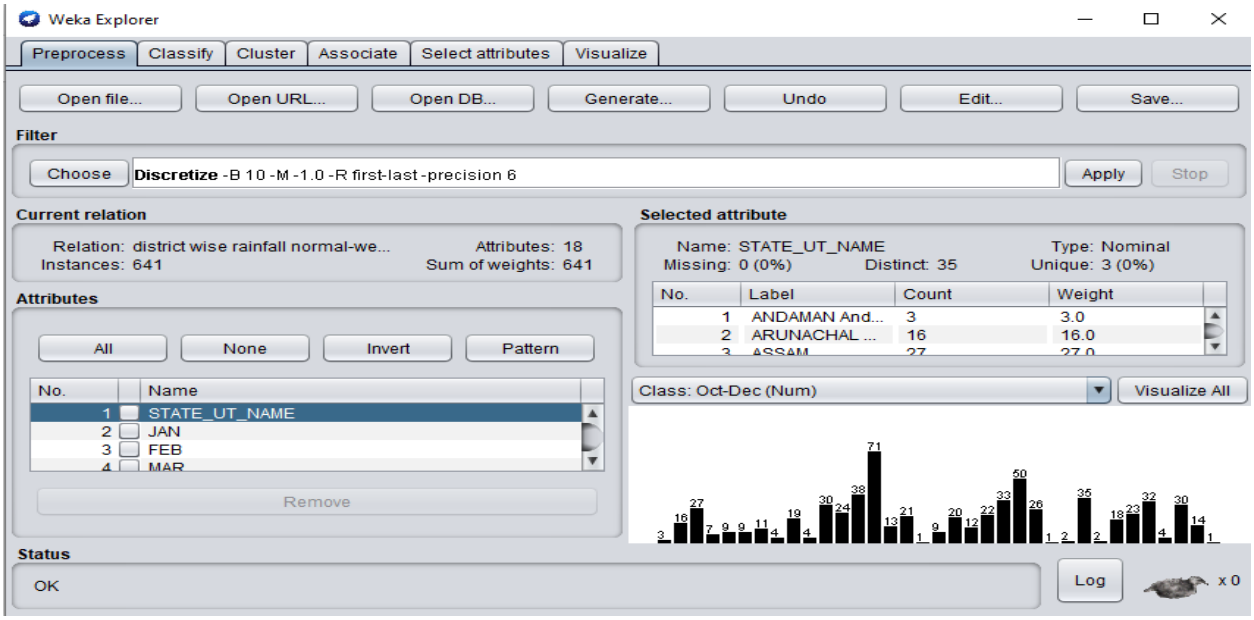
4. After opening the explorer you will see different options but these options are disable after loading dataset these options will be enable so load the dataset Click -> preprocess -> open file ->select file -> district wise rainfall normal.cvs



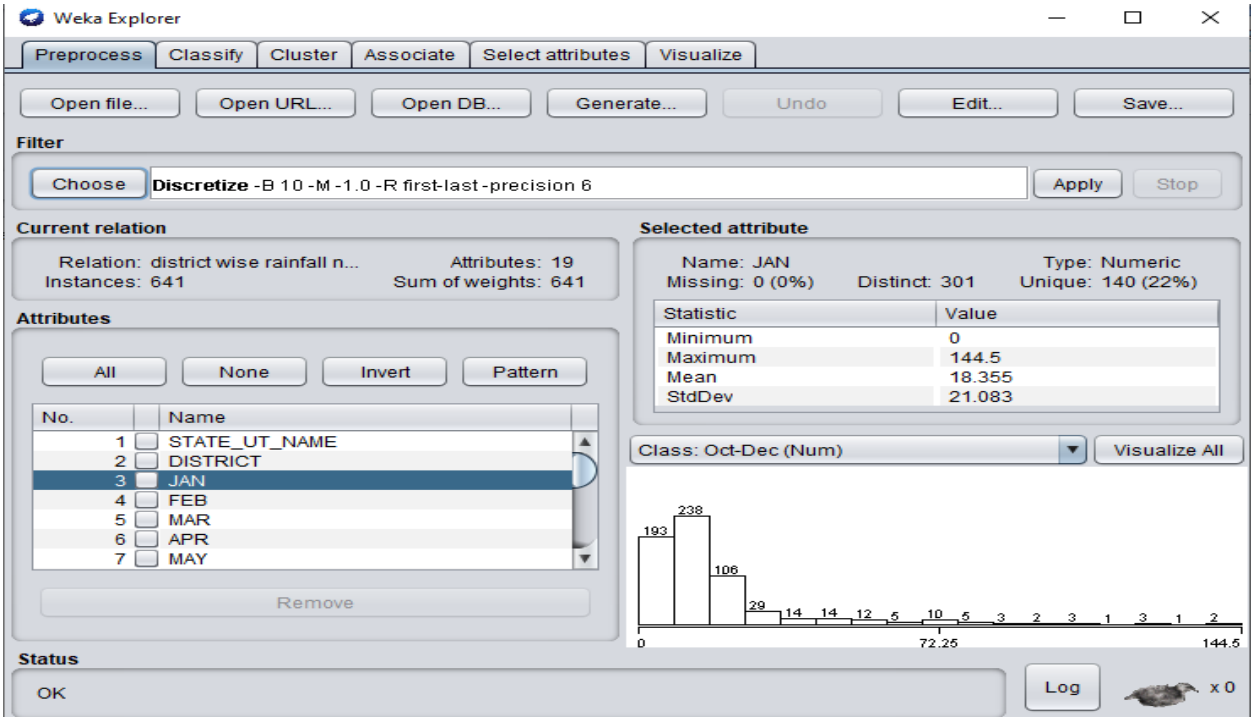
5. After opening file all the options are enable and scree looks like this here we have to perform different functions.



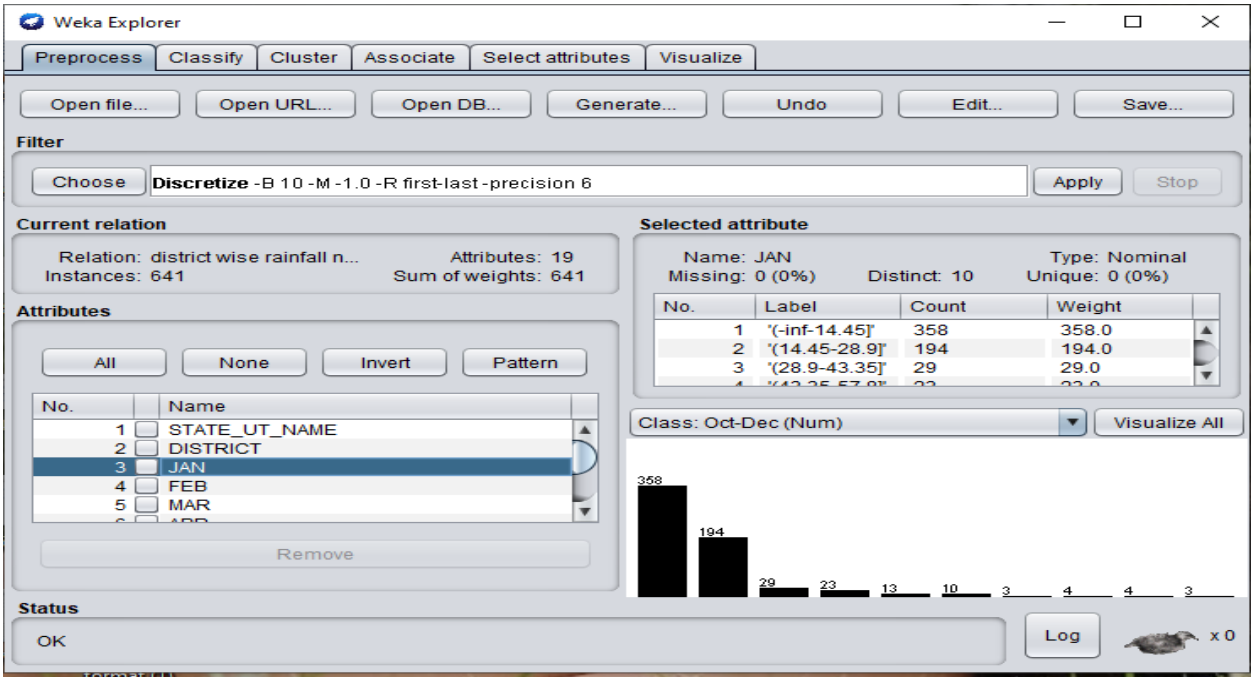
6. In preprocessing we can remove the attributes which we doesn't need
   Ex. Here I remove the attribute STATE_UT_NAME click on select box of -> click-> Remove



7. Next step is Discretize. The filter will convert Numeric values into Nominal Click-> Choose ->filters -> Supervised -> Attribute -> Discretize and click on ‘Apply’ button. Here the attribute JAN has numeric value .



Here the attribute has a nominal value.

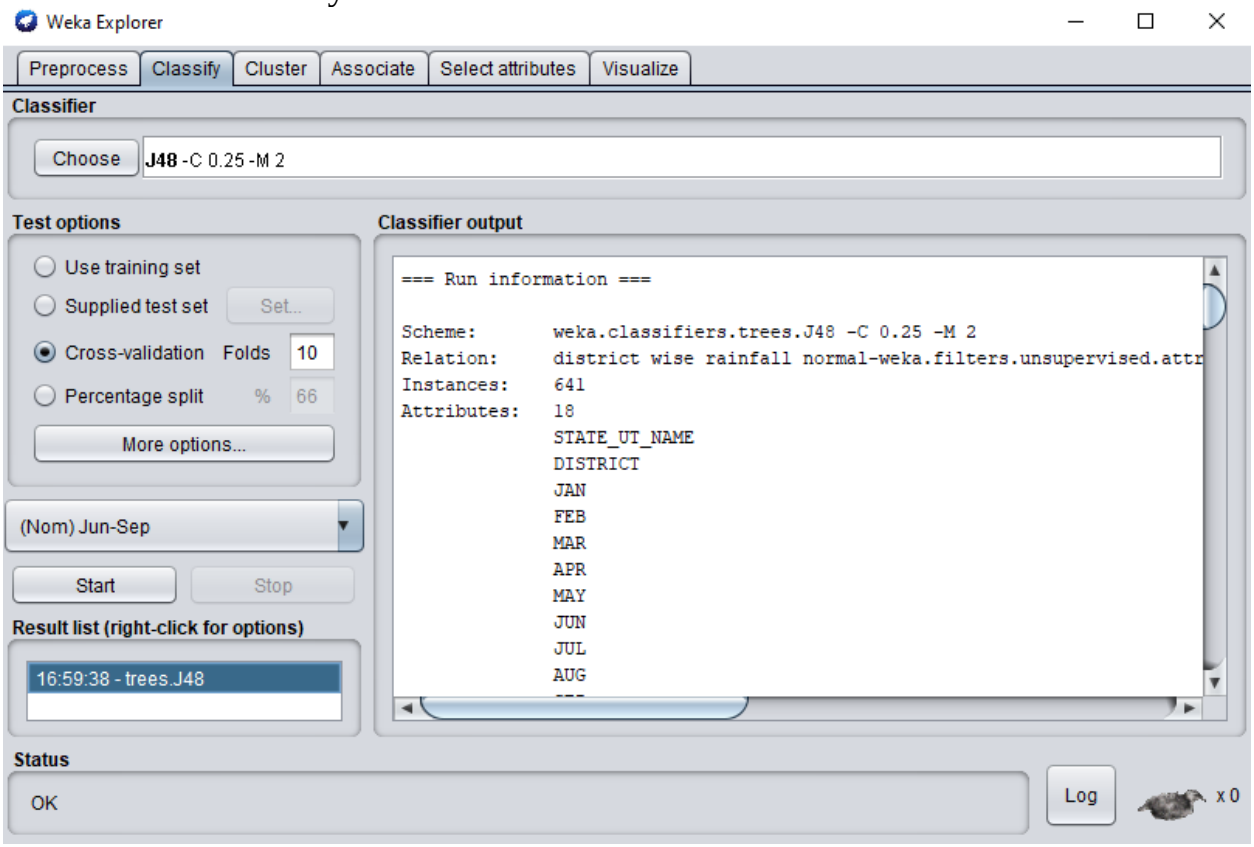


2. What is classification?

Classification is a data mining function that assigns items in collection to target categories or classes. The goal of classification is to accurately predicat the target class for each case in the data Foe example a classification model could be used to identify loan applications as low ,medium,or high credit risks.

Steps:

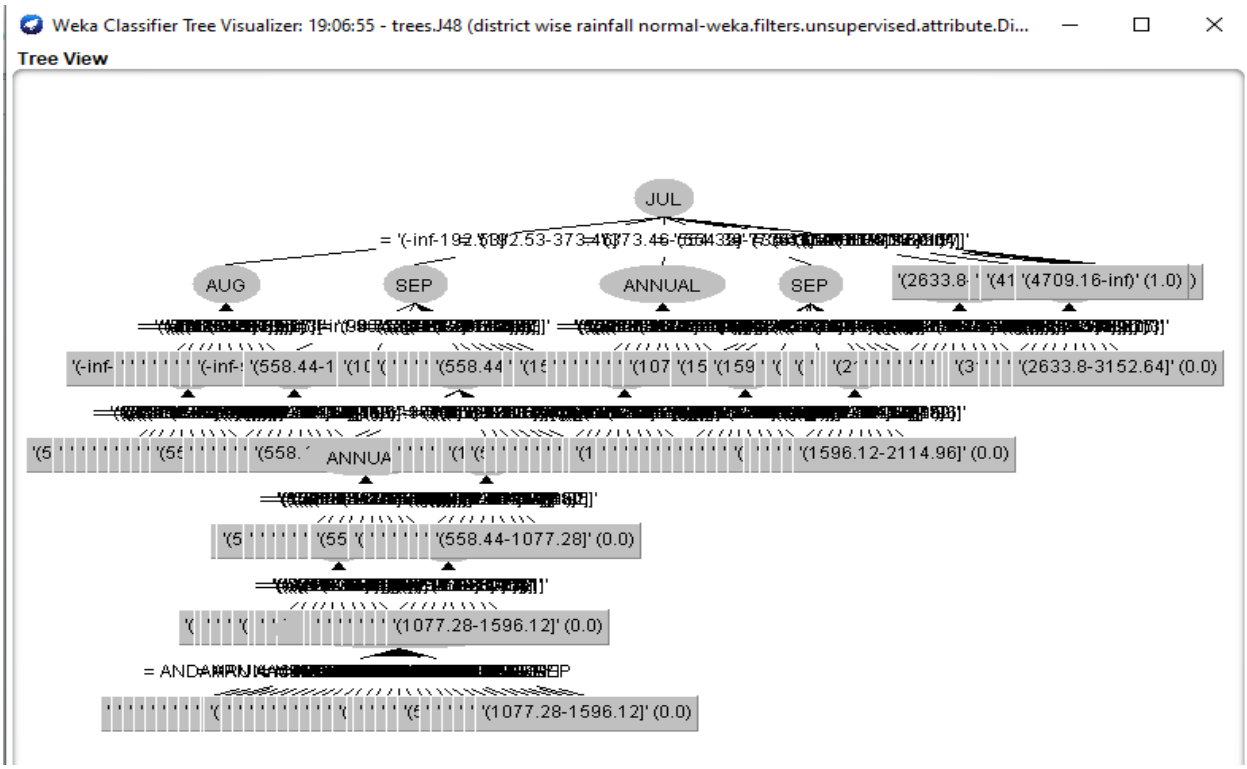
1. Click -> classify -> Choose ->Weka ->Classifiers -> trees -> J48



2. Visualization of Results

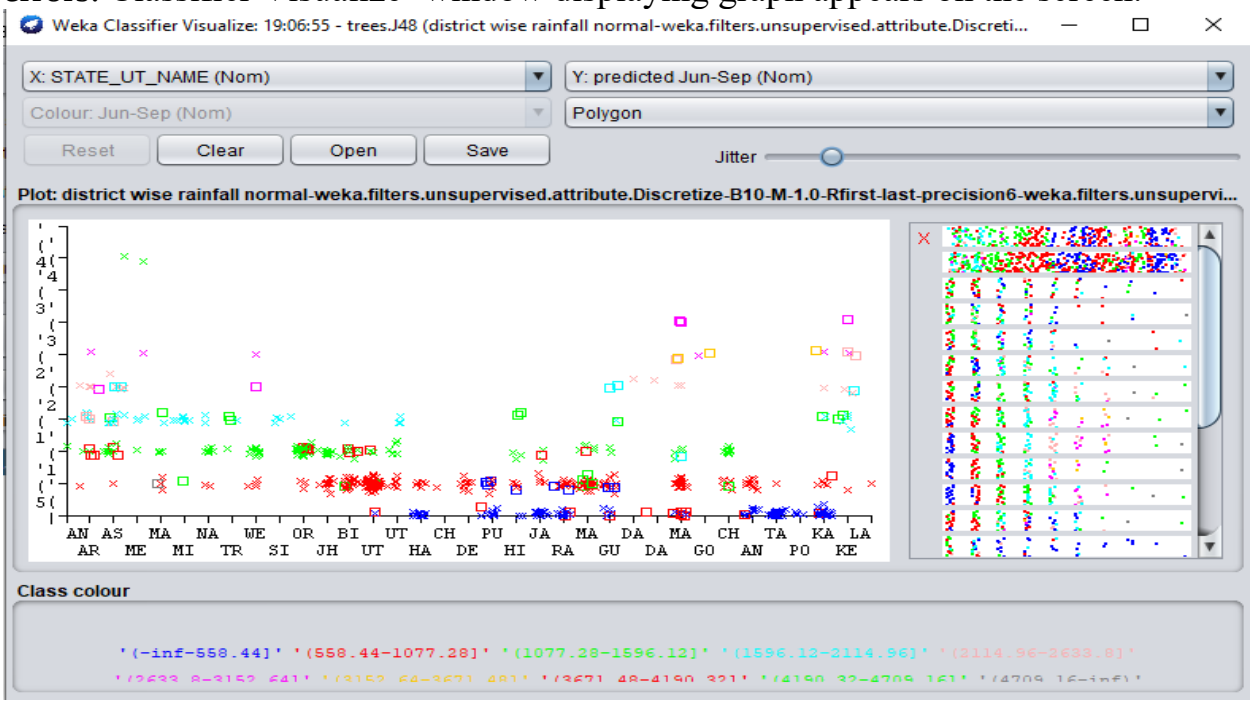
1. Visualize Tree

Classify -> in Result list ->right click on result->click on visualize tree



2. Visualize classifier errors

Classify -> in Result list ->right click on result->click on visualize classifier errors. Classifier Visualize' window displaying graph appears on the screen.

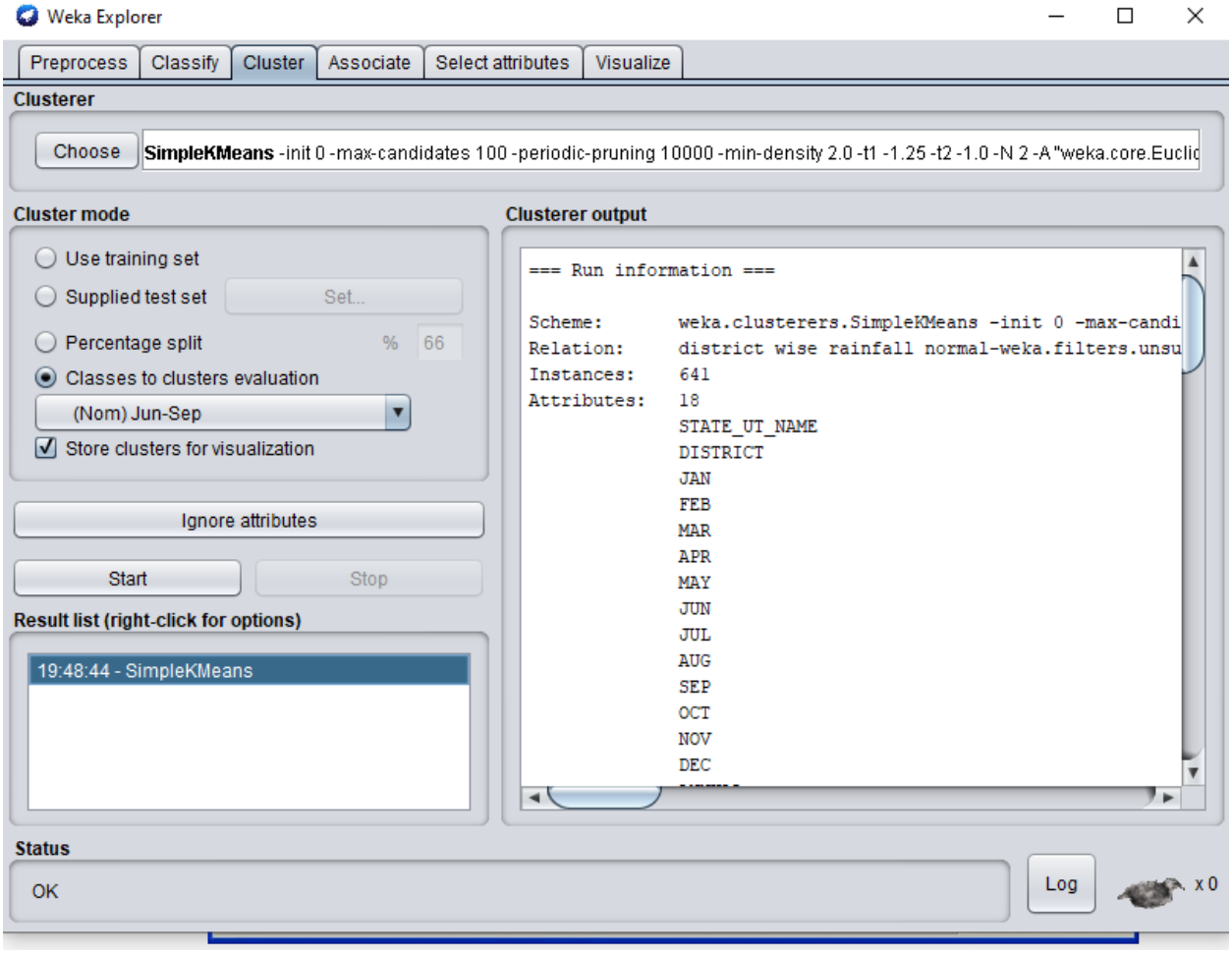


3. What is Clustering Data ?

In a clustering, a group of different data objects is classified as similar objects. One group means a cluster of data. Data sets are divided into different groups in the cluster analysis which is based on the similar of the data.

Steps:

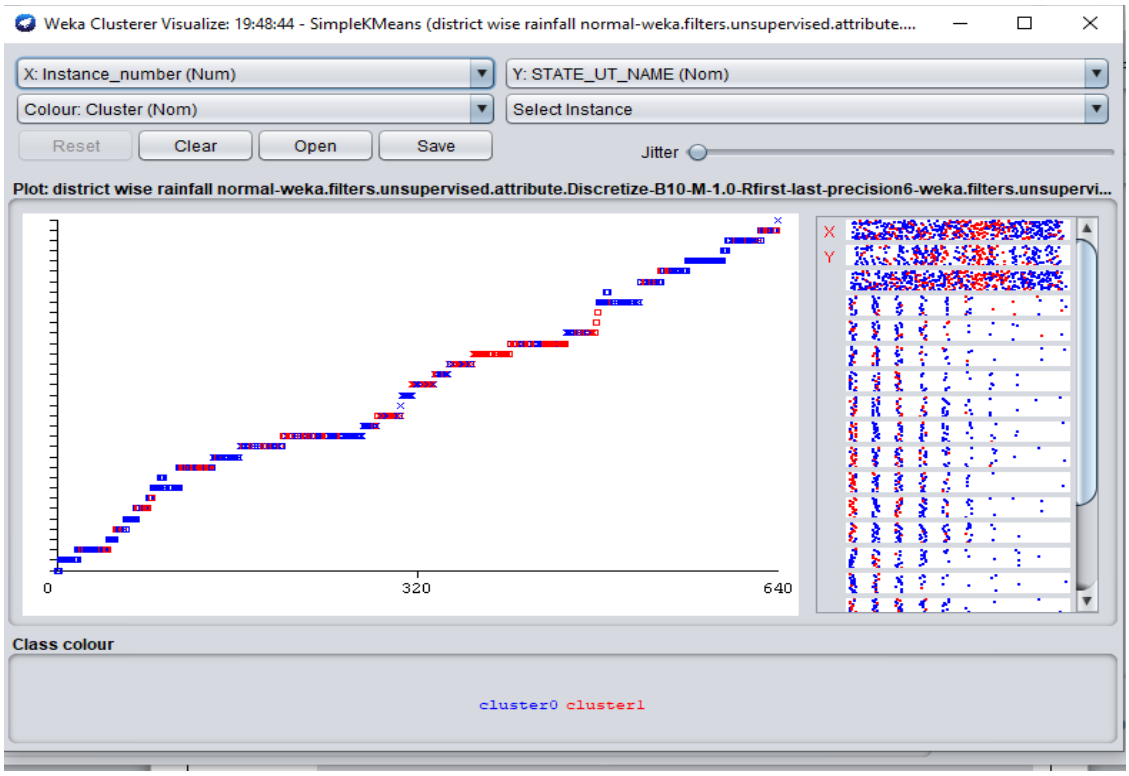
1. Click -> cluster -> Choose ->Weka ->Clusters-> SimpleKMeans



2. Visualization of Results

1. Visualize Cluster Assignments

2. Cluster -> in Result list ->right click on result->click on visualize Cluster Assignments

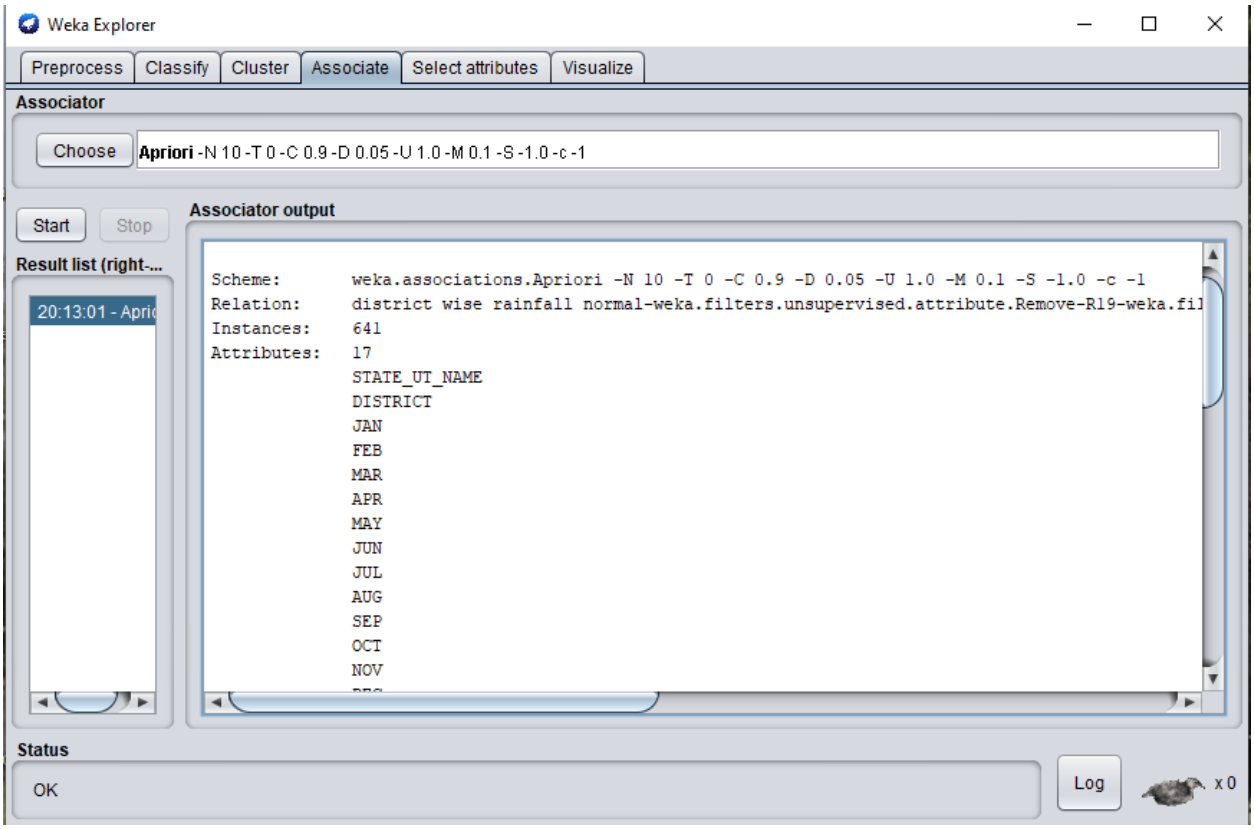


4. What is Associations ?

Association is a data mining functions that discovers the probability of the co-occurrence of items in a collection. The relationship between co-occurring items are expressed as association rules. Association rules are often used to analyze sales transations.

Steps:

1. Click -> Association -> Choose -> Apriori

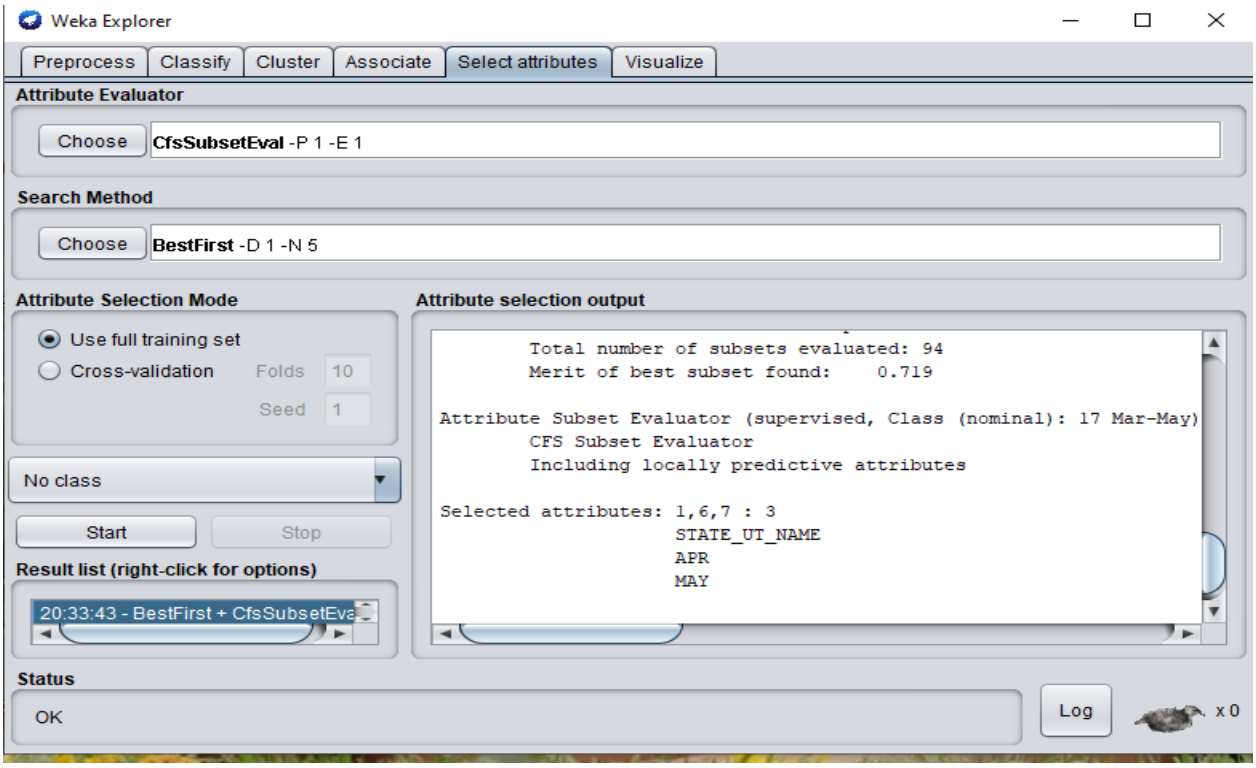


5 . What is Select Attribute?

The attribute selection task essentially consists in selecting a subset of originally available to be subsequently used for model creation. General purpose attribute selection algorithms can be applied to select attribute for arbitrary target algorithms, and sometimes also for different target tasks.

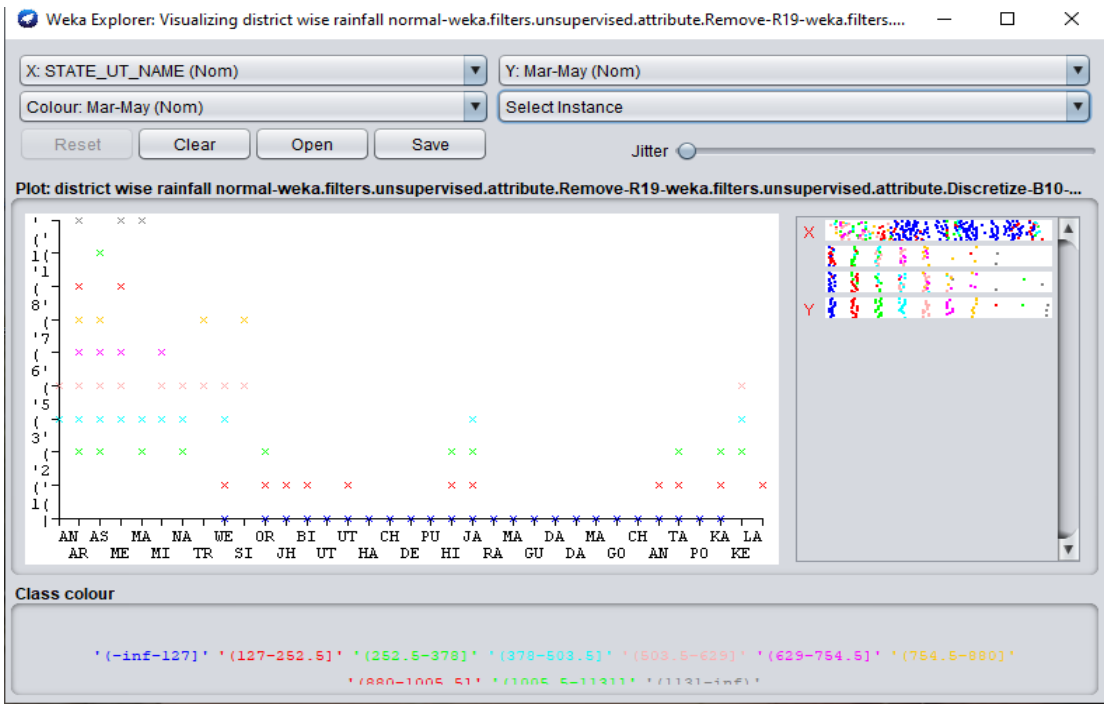
Steps:

1. Select Attribute -> in Attribute Evaluator -> choose -> CfsSubsetEval
   In Search Method -> Choose -> Best First -> click on start



2. Visualizing Results

1. Select attribute> in Result list ->right click on result->click on visualize reduced data

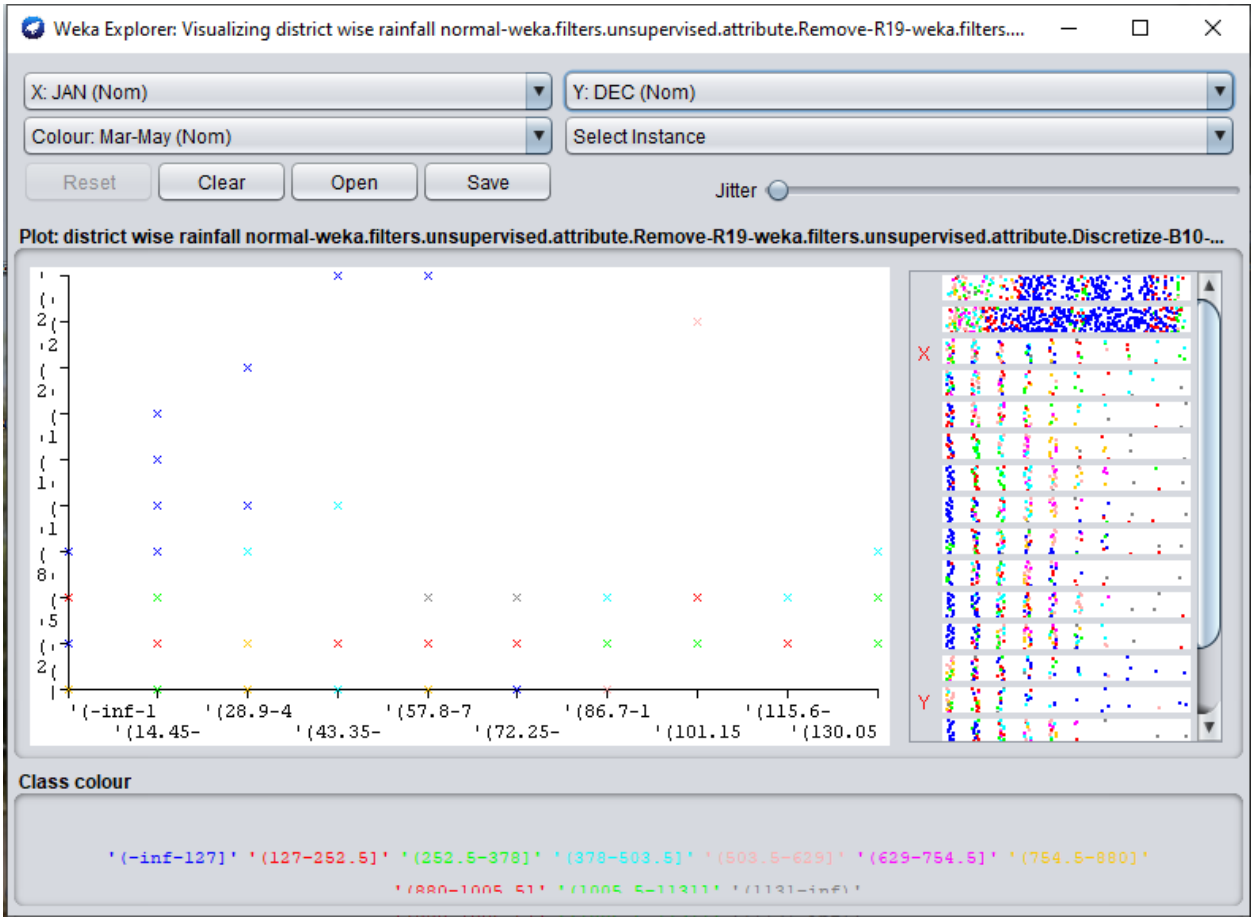


6. What is Data Visualization ?

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization toolsprovide an accessible way to see and understand trends, outliers, and patterns in data.

Steps:

1. Visualize -> Click on JAN you will see the visualization window



FUTURE SCOPE AND CONCLUSION

Future Scope:

In this project we perform multiple functions like pre-processing, clustering, visualization etc. we can also perform different function on dataset. Also we can design different algorithms for data pre processing

Conclusion:

By this project we are able to perform preprocessing, classification, clustering, association, attribute selection, and visualization by using WEKA tool.