



FUNDAMENTALS OF DATA MINING

Fall 2021



DATE OF SUBMISSION: **25/NOVEMBER/2021**_

TEACHER PROFORMA



Name: Ms. Ambreen

Father's Name: Shamsuddin Hamal Khan

Date of Birth:

Date: **04** Month: **August** Year: **1992**

Nationality: Pakistani

Qualification: MS Computer Science

Contact Details:

Tel #: +923004991229

Residence: Gulshan-e-Iqbal Karachi

Email ID: ambreen.khan@bbsul.edu.pk

STUDENT PROFORMA

Name: **Rohit Kumar**

Father's Name: **Motilal**

Section: **5th semester, B**

Batch: **10th**

Roll #: **731**

Date of Birth:

Date: **29** Month: **October** Year: **2000**

Nationality: **Pakistani**

Address: **Abu Mansion Dhobi Gath Flat # 2, Floor # 1, Lea Market Karachi**

Email ID: **rohitramsinghania@gmail.com**



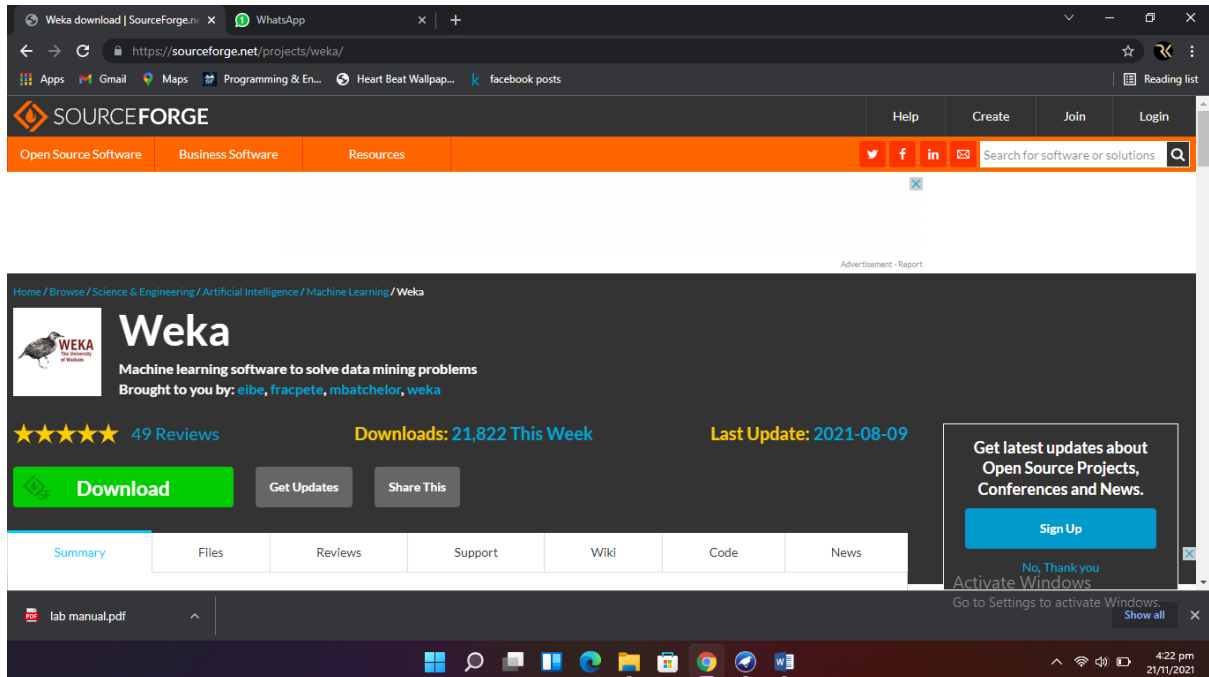
TABLE OF CONTENT:

| S.NO: | TASKS | PAGE # |
|-------|--|--------|
| 1 | Install WEKA and add screenshots of the installation steps. | |
| 2 | Check each feature of WEKA and add screenshots. | |
| 3 | Complete interface of WEKA with screenshots. | |
| 4 | Import dummy dataset from KAGGLE and add screenshots. Also describe some features of KAGGLE briefly, and add screenshots of your work. | |
| 5 | Import dummy dataset from UCI and add screenshots. Also describe some features of UCI briefly, and add screenshots of your work. | |
| 6 | Import downloaded dataset from your device in WEKA and add screenshots of all steps. | |
| 7 | Import any dataset in WEKA and perform preprocessing. Also add screenshots of the output. | |
| 8 | Import any dataset in WEKA and perform Classification. Also add screenshots of the output. | |
| 9 | Import any dataset in WEKA and perform Clustering. Also add screenshots of the output. | |
| 10 | Import any dataset in WEKA and perform Decision Tree and Naïve Bayes Algorithms. Also add screenshots of the outputs. | |

Lab#1

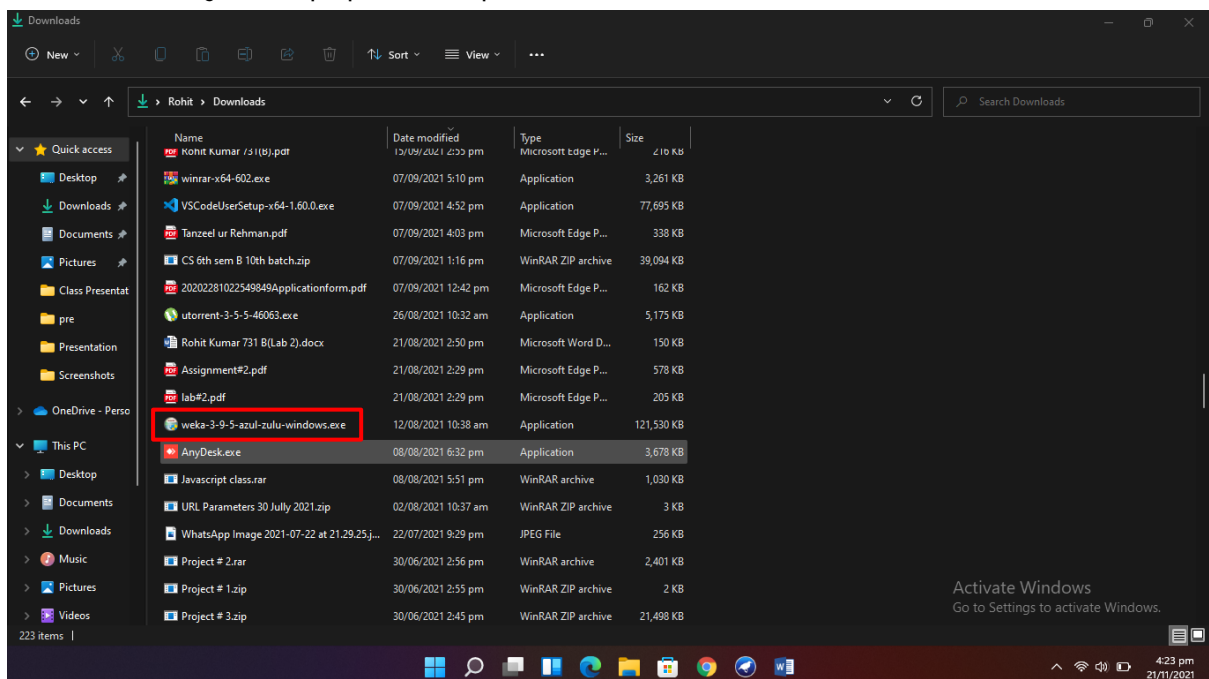
Step#1:

Got to the SourceForge and click on download:



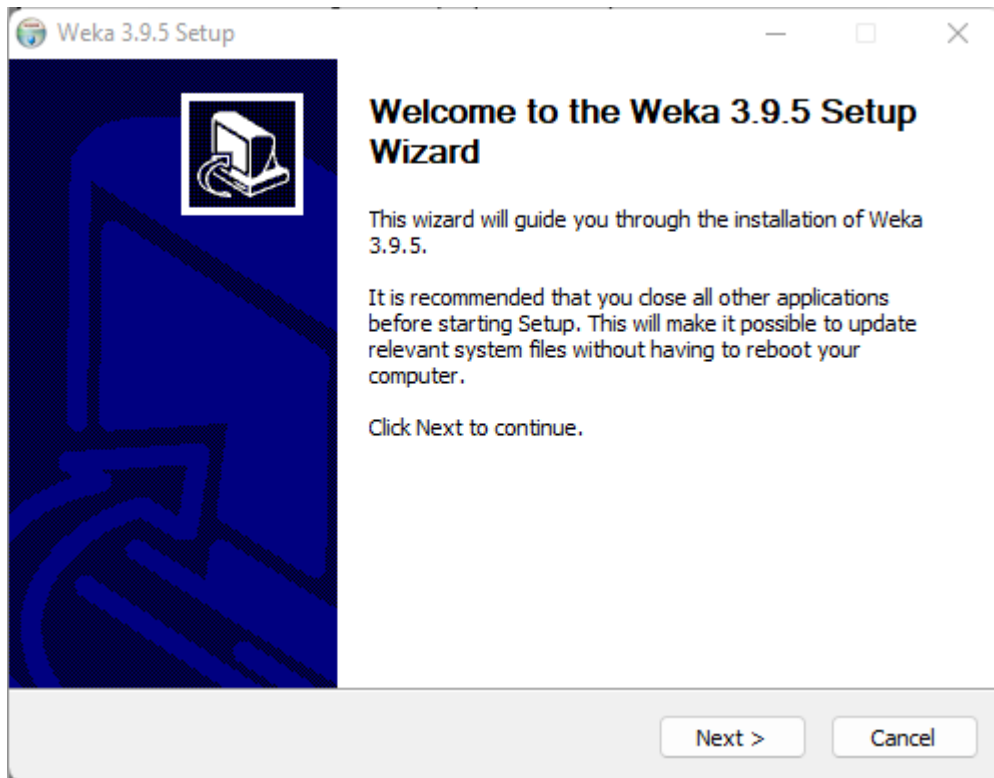
Step#2:

After downloading the setup open the setup from downloads



Step#3:

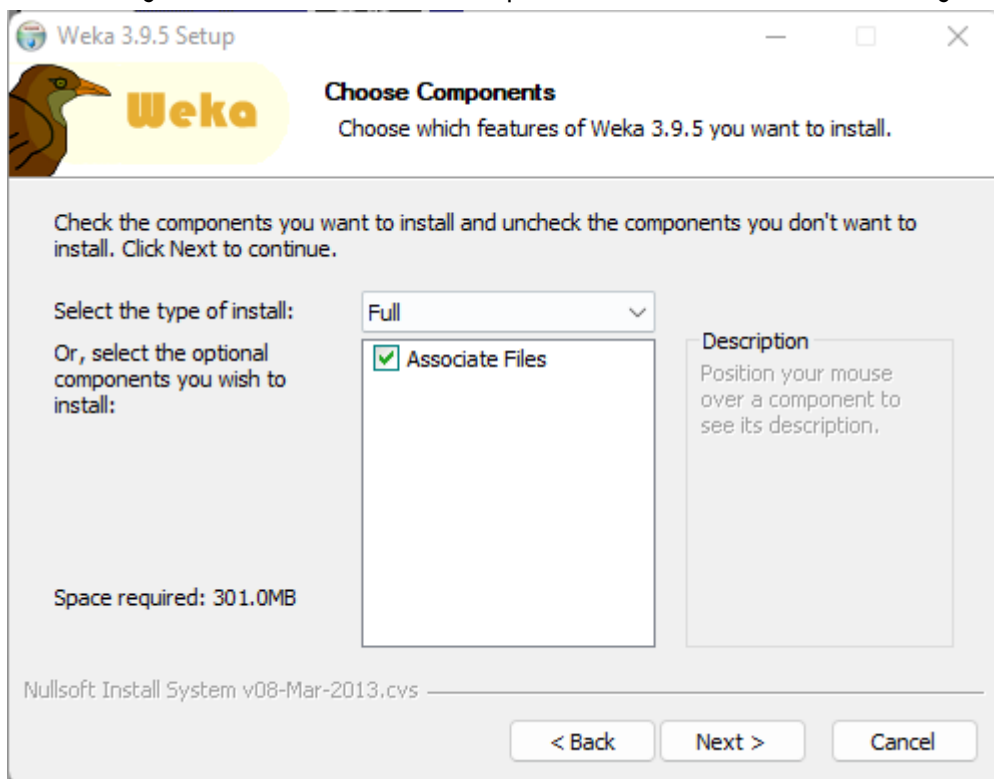
Double click the weka setup to start installing...



click on next when this window appears.

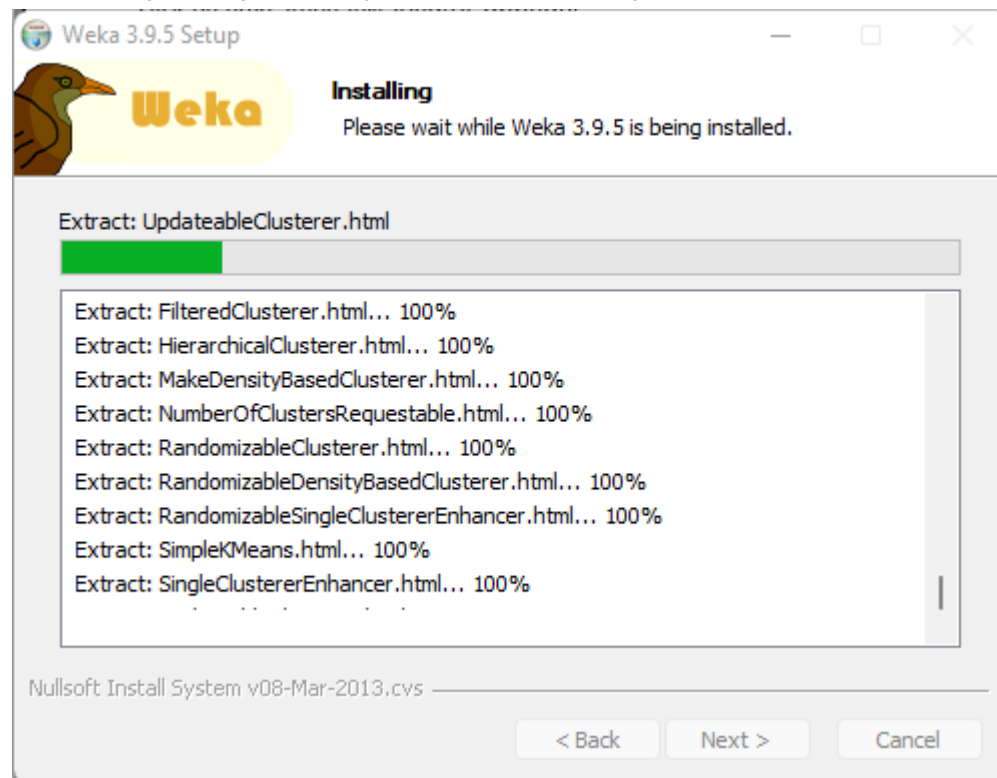
Step#4:

Click on "I agree" when this windows shows up.. and click next next until installation get started.



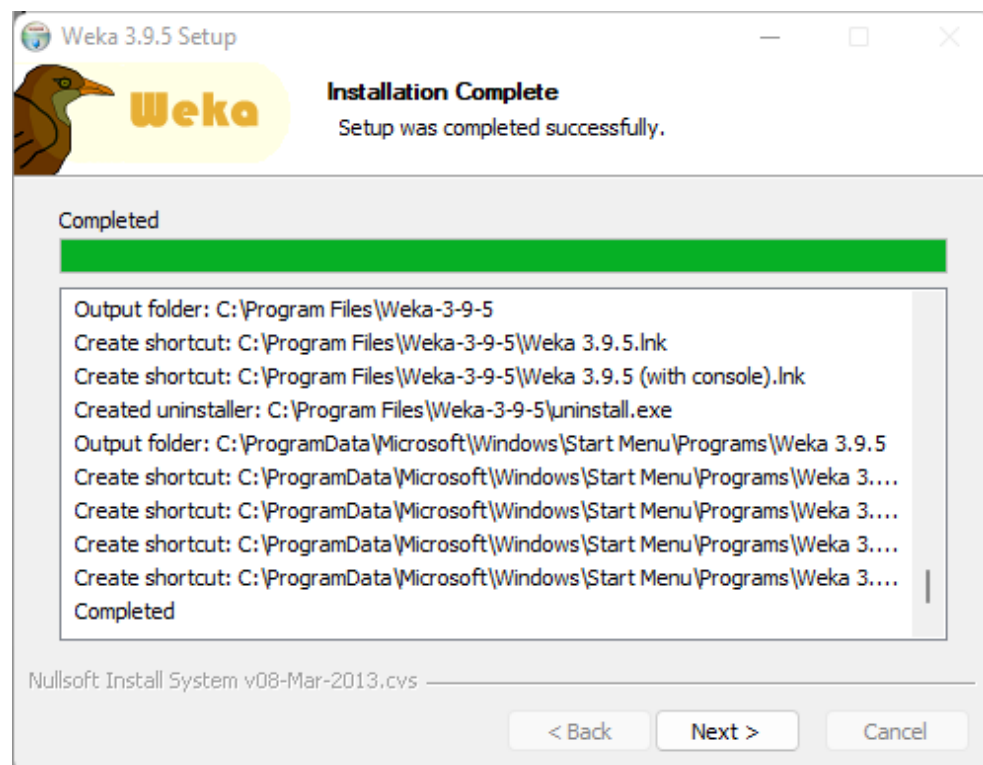
Step#5:

After clicking next again and again, it will start installing



Step#6:

After all steps it will be installed successfully...

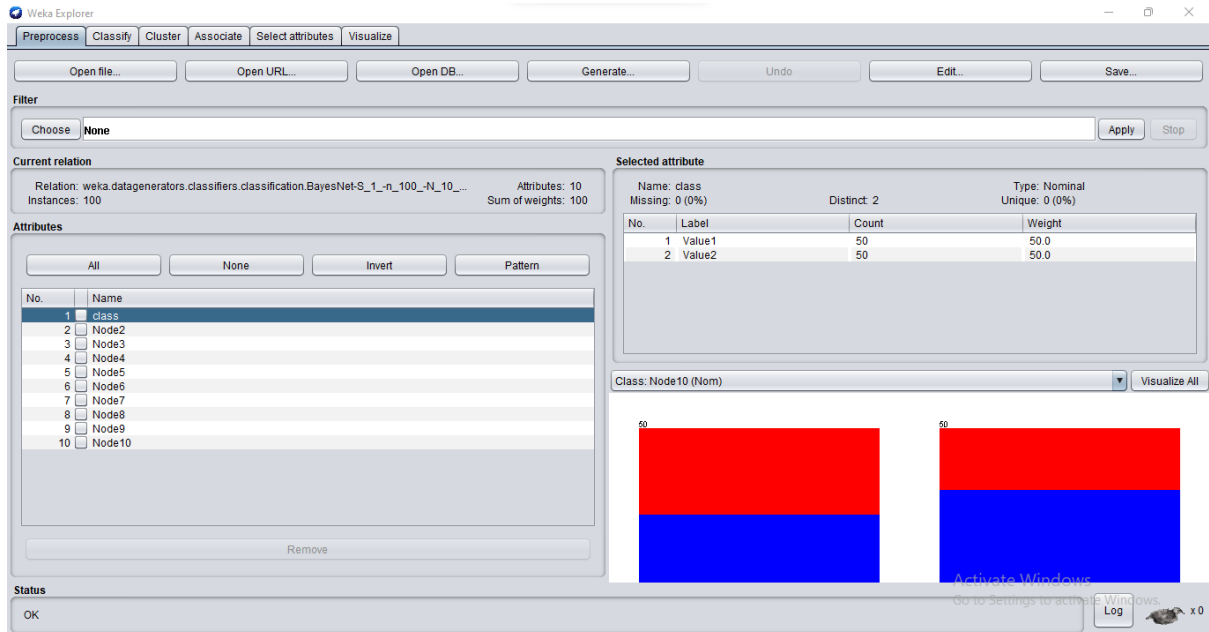


Lab#2

Step#1:

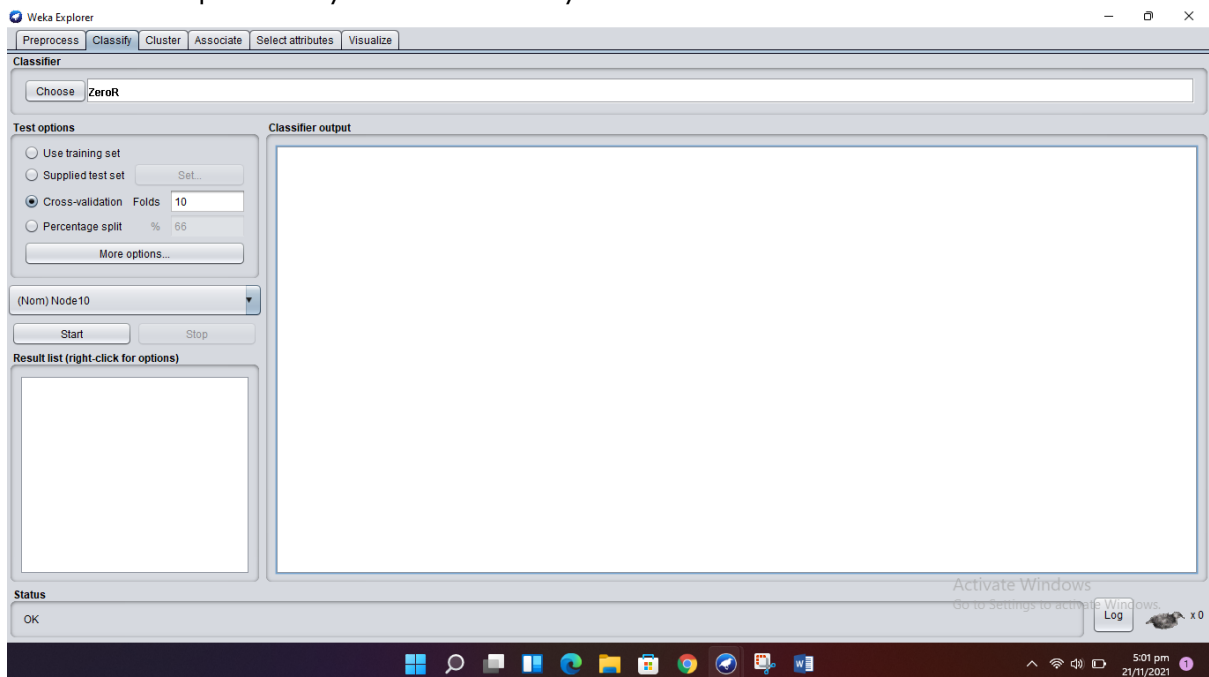
In this lab we will explore all the interface of weka...Open explorer...

The first option is preprocessor... it has all the preprocessor algorithm for the given data:



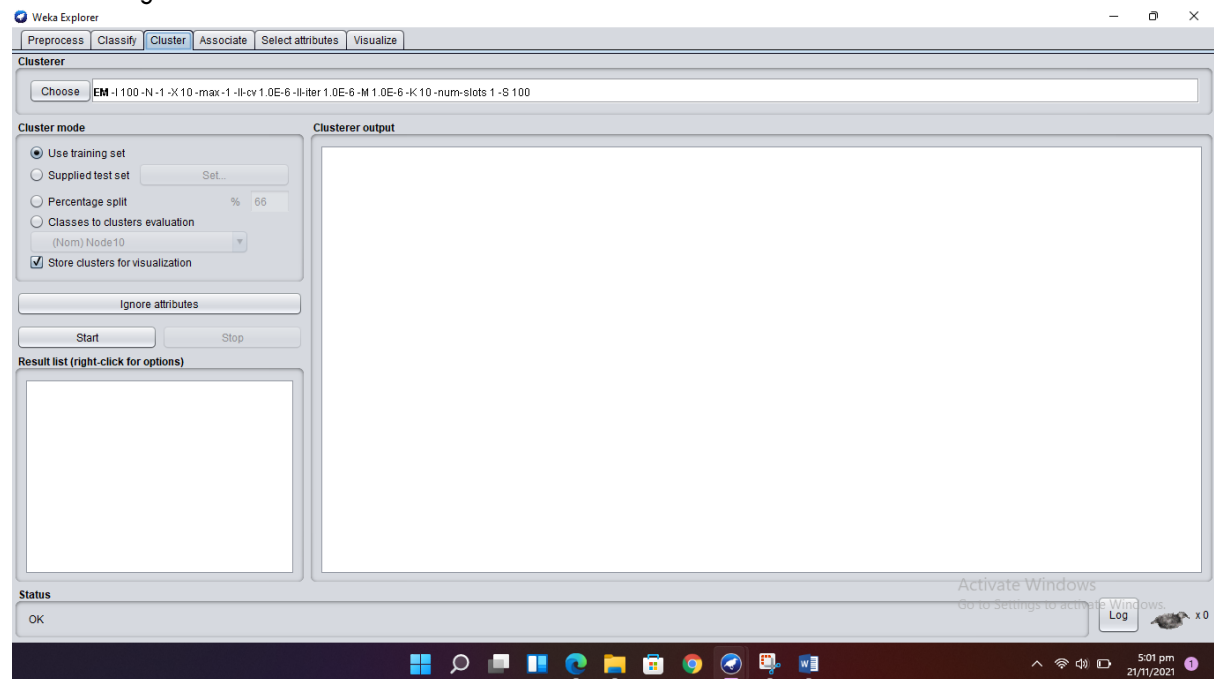
Step#2:

Click on second option classify.. in which we classify the data...



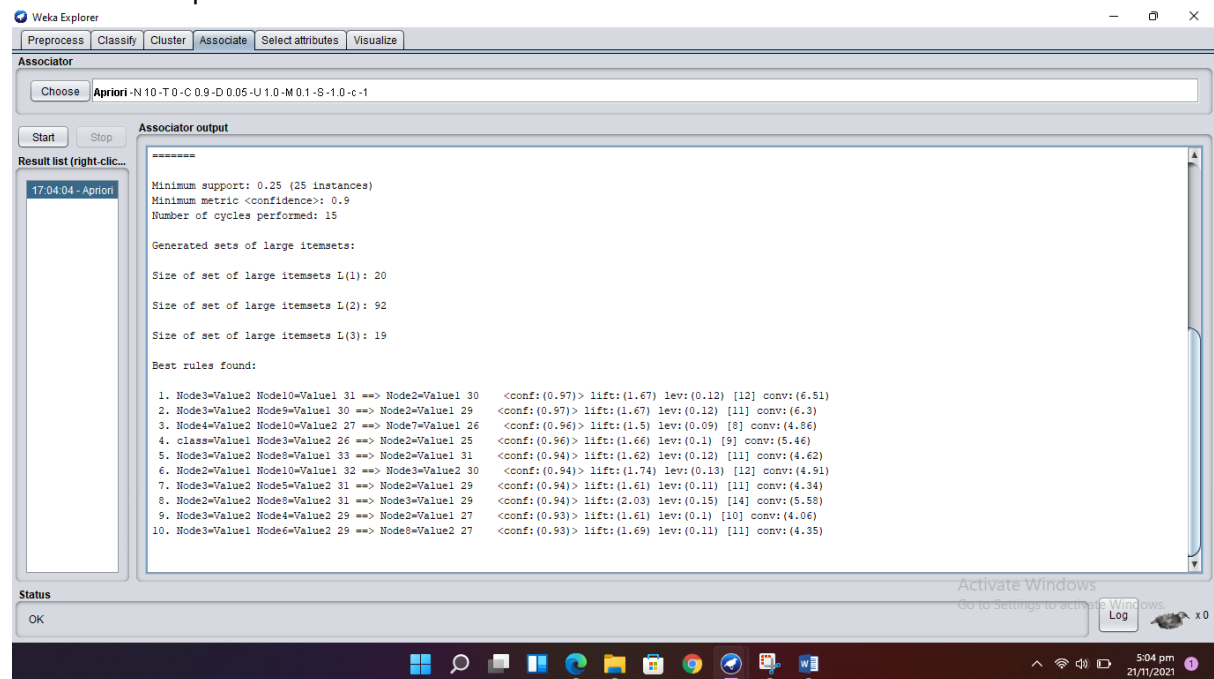
Step#3:

Click on third option which is clustering... in which we make cluster for the given dataset using different algorithm.



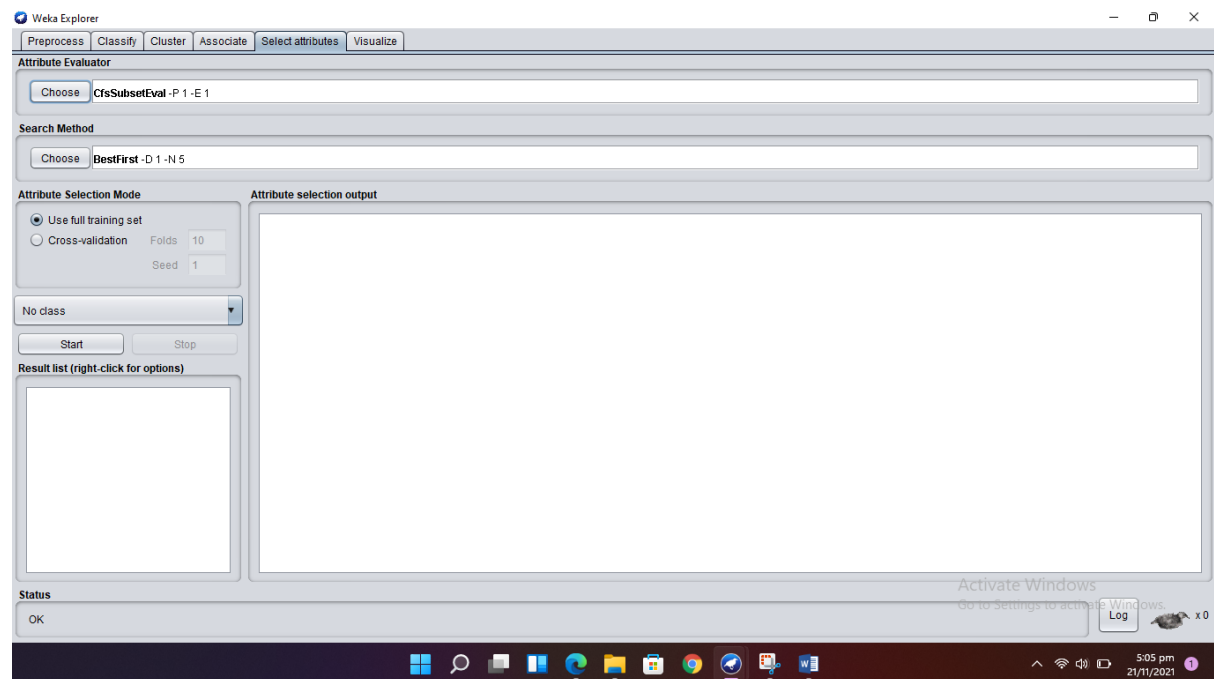
Step#4:

Click on fourth option which is associate...which create association with the dataset.



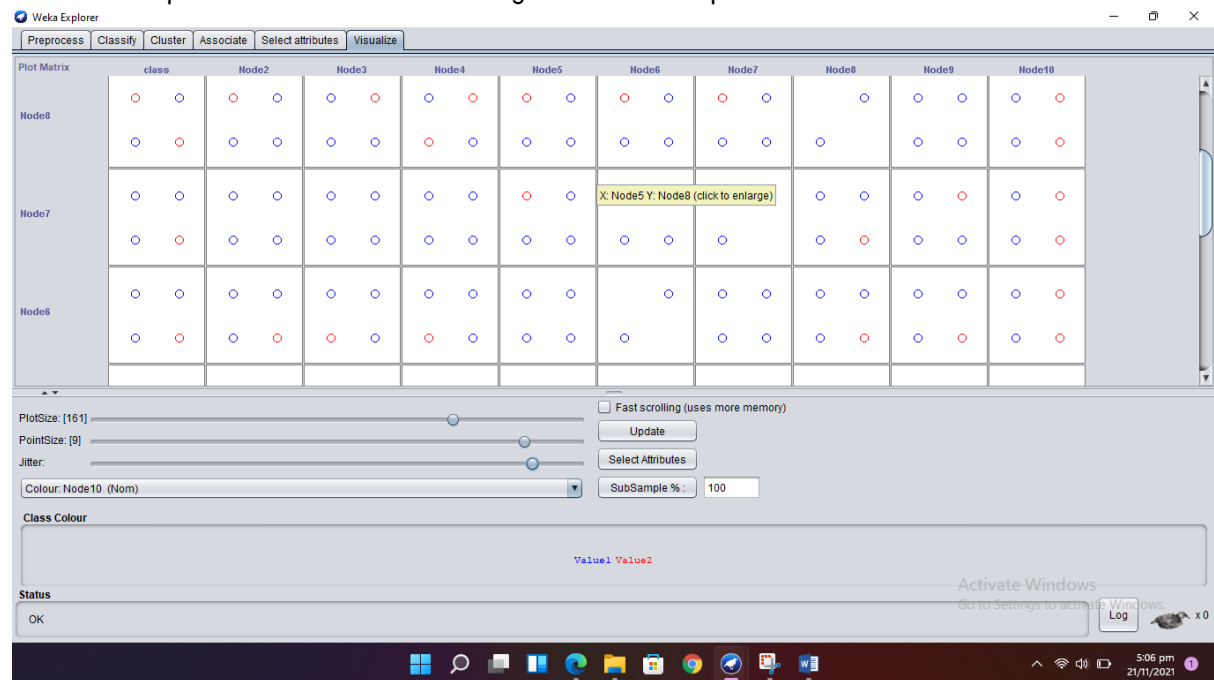
Step#5:

Click on fifth option which is Select Attributes... which is used to select specific attributes.



Step#6:

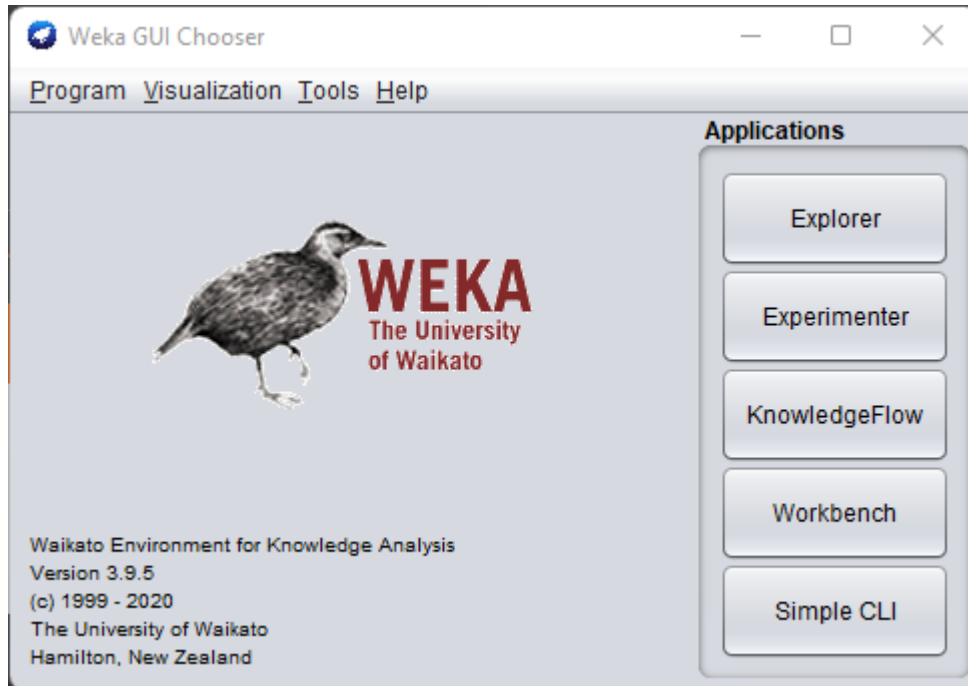
Click on last option which is Visualize. Which gives the visual representation of the data.



Lab#3

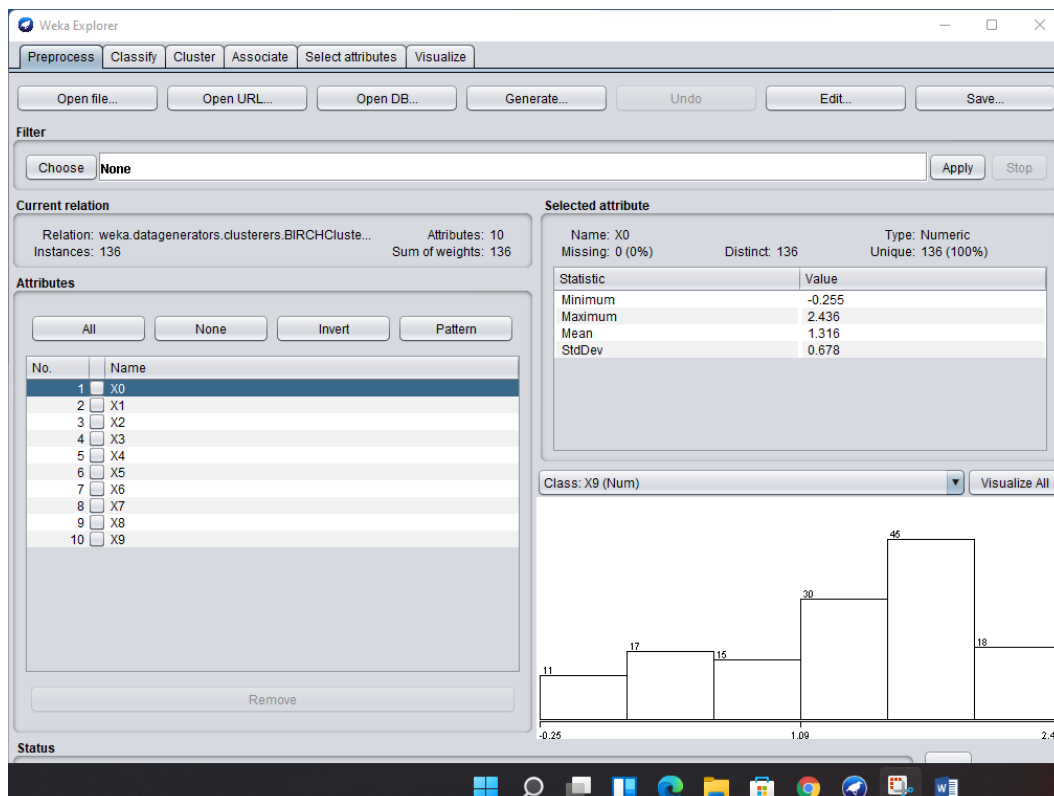
Step#1:

Open weka tool and go to explorer:



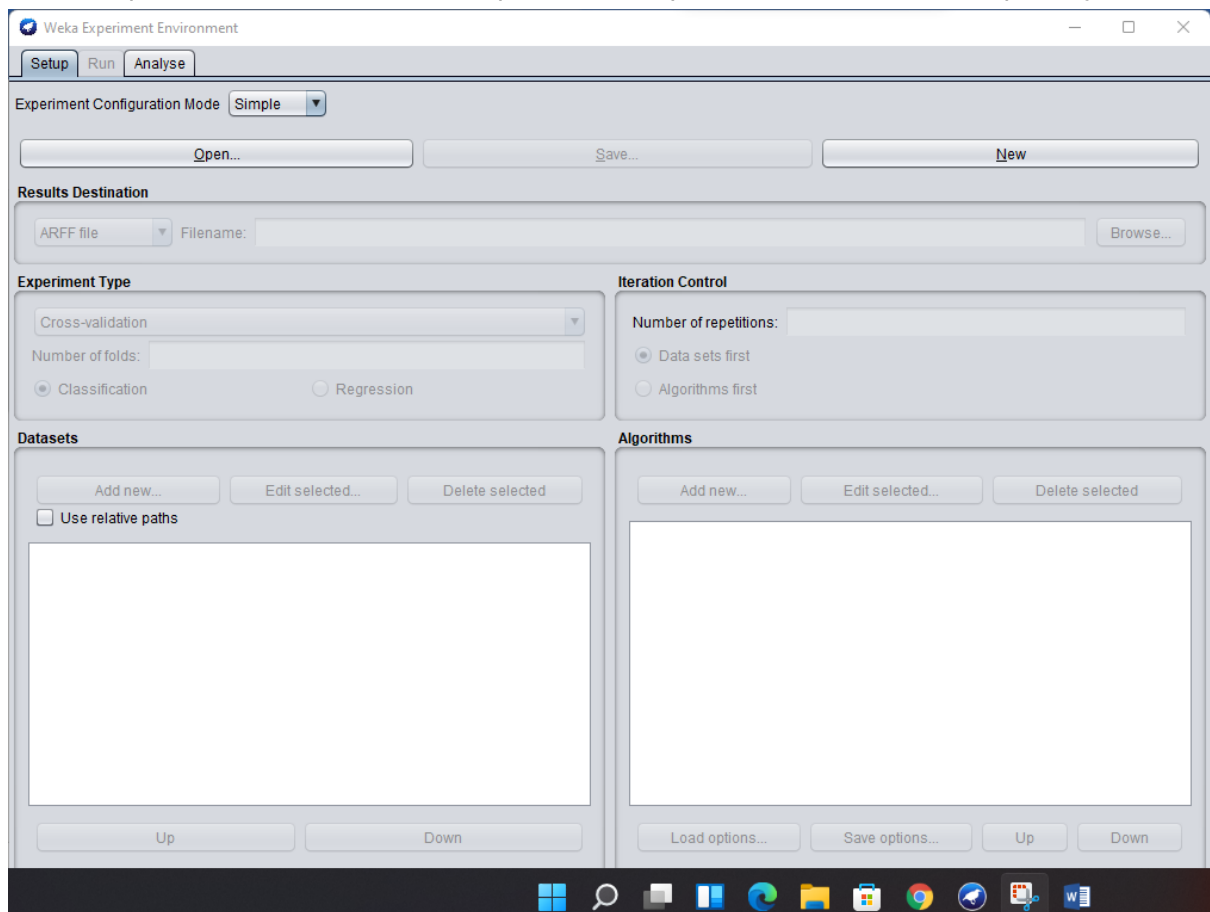
Step#2:

Open explorer from weka tool...it will show all the possible tools for datamining as we have studied in KDD.



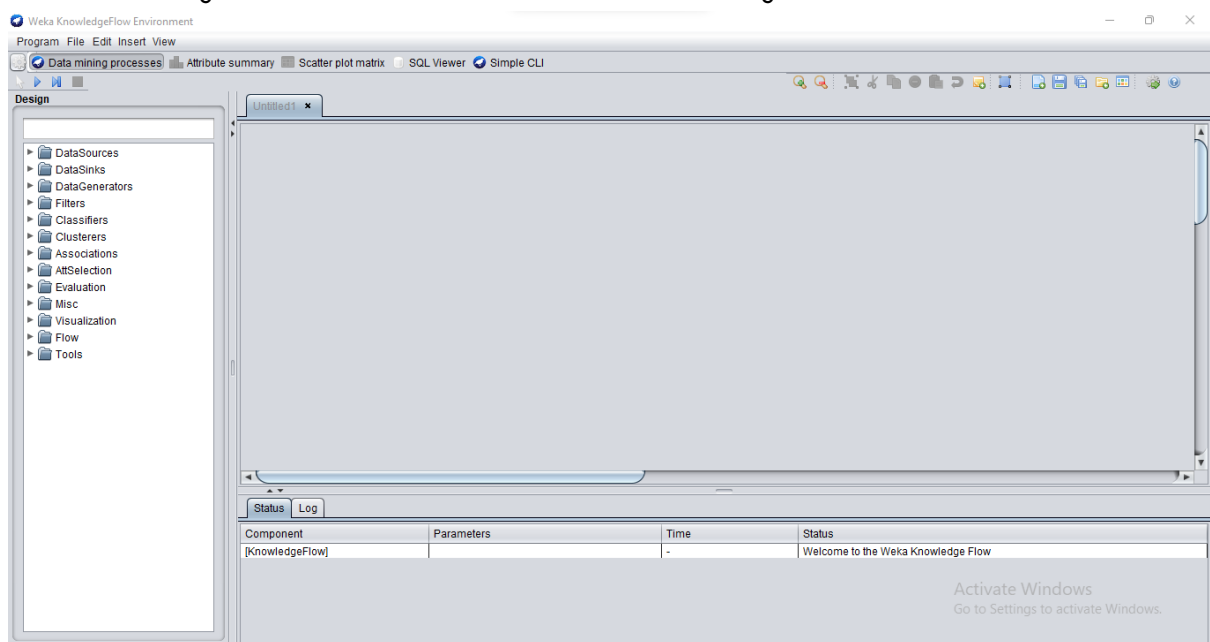
Step#3:

Click on Experimenter... it will show the setup, run and analyze tabs each tab work on depending dataset.



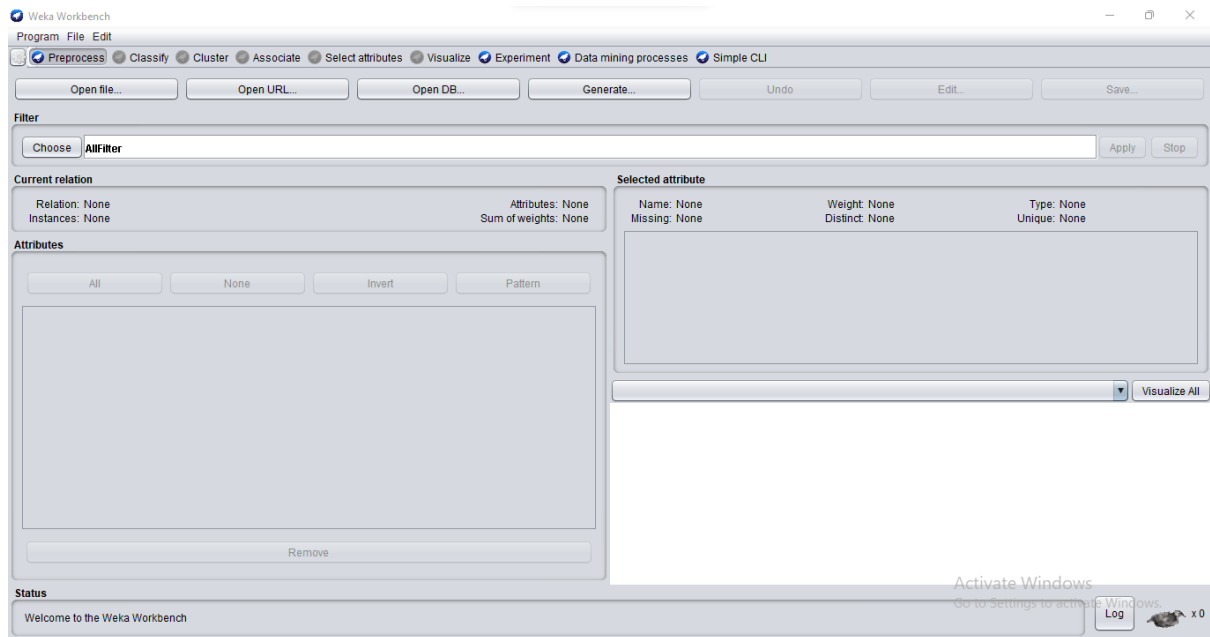
Step#4:

Click on KnowledgeFlow... It will have all the features for datamining with their attributes and instances..



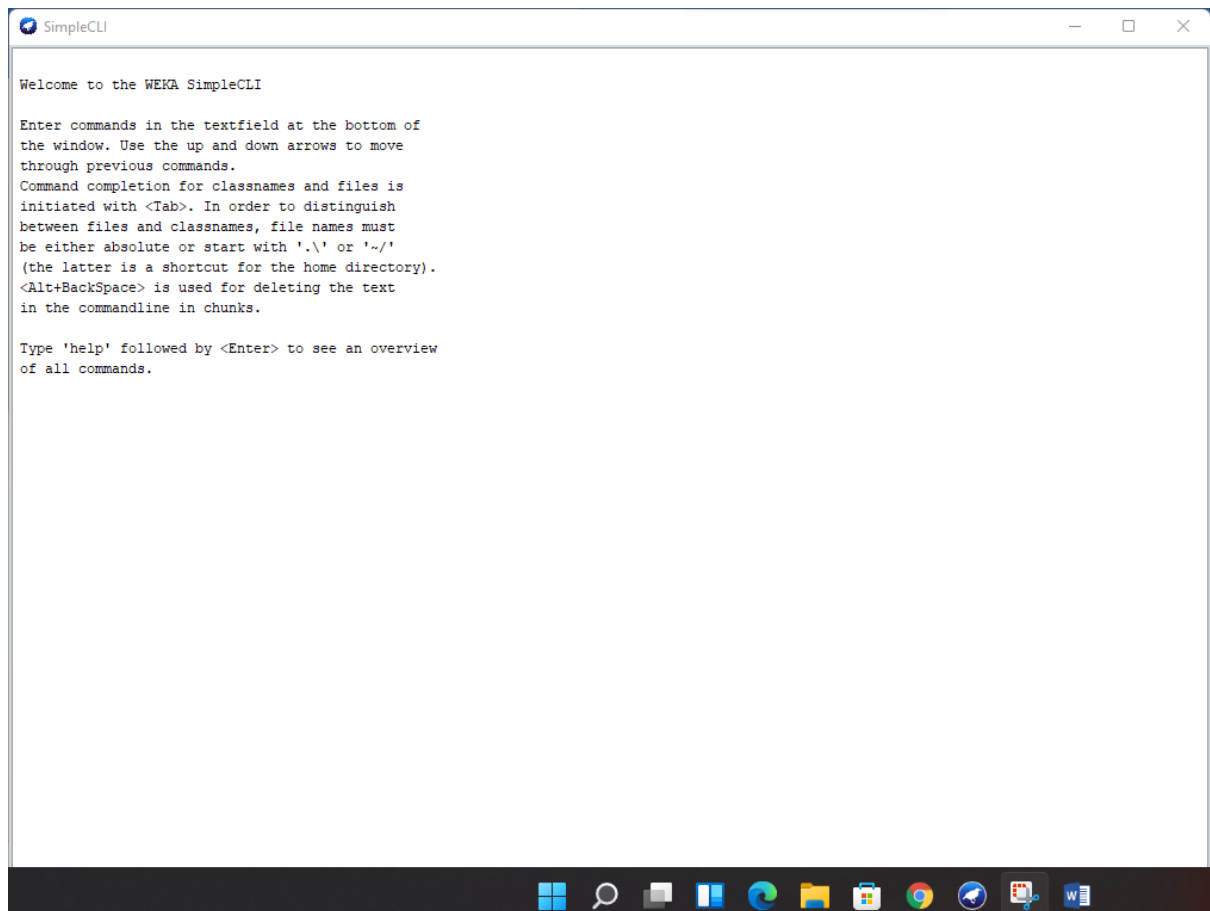
Step#5:

Click on WorkBench... It will have all the features of explorer with different UI.



Step#6:

Click on Simple CLI... it is a CLI or terminal used for data mining.



Lab#4

Step#1:

Got to the Kaggle.com and search for any desired dataset you want:

The screenshot shows the Kaggle website interface. On the left is a sidebar with navigation links: Create, Home, Competitions, Datasets (selected), Code, Discussions, Courses, and More. The main content area displays the 'Breast Cancer Wisconsin (Diagnostic) Data Set' by UCI Machine Learning, updated 5 years ago (Version 2). It includes a search bar at the top, a 'Download (125 kB)' button, and a 'New Notebook' button. Below the dataset title, there's a section for 'Your Dataset download has started' with a '2452' upvote count and a row of user avatars. Further down, it shows 'Usability 8.5', 'License CC BY-NC-SA 4.0', and 'Tags cancer, healthcare'. The description mentions that features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. The Windows taskbar at the bottom shows the time as 5:14 pm on 21/11/2021.

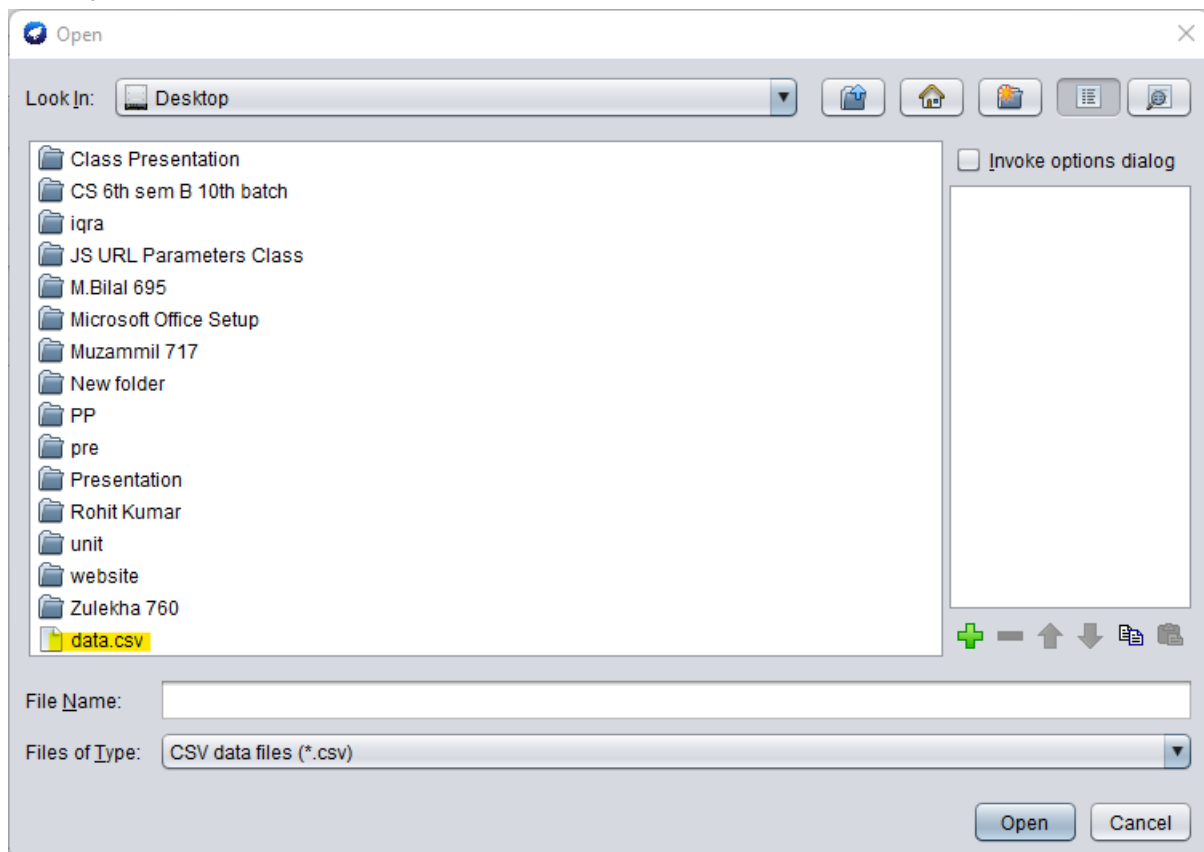
Step#2:

Download the dataset from the website then open the weka tool and click on open file in explorer.

The screenshot shows the Weka Explorer application window. The 'Preprocess' tab is selected. At the top, there are buttons for 'Open file...', 'Open URL...', 'Open DB...', 'Generate...', 'Undo', 'Edit...', and 'Save...'. Below these is a 'Filter' section with a 'Choose' button and a text field set to 'None', with 'Apply' and 'Stop' buttons. The main area is divided into 'Current relation' and 'Selected attribute' sections. 'Current relation' shows 'Relation: None' and 'Instances: None'. 'Selected attribute' shows 'Name: None', 'Missing: None', 'Weight: None', 'Distinct: None', 'Type: None', and 'Unique: None'. There is also an 'Attributes' section with buttons for 'All', 'None', 'Invert', and 'Pattern', and a 'Remove' button. The Windows taskbar at the bottom shows the time as 5:14 pm on 21/11/2021.

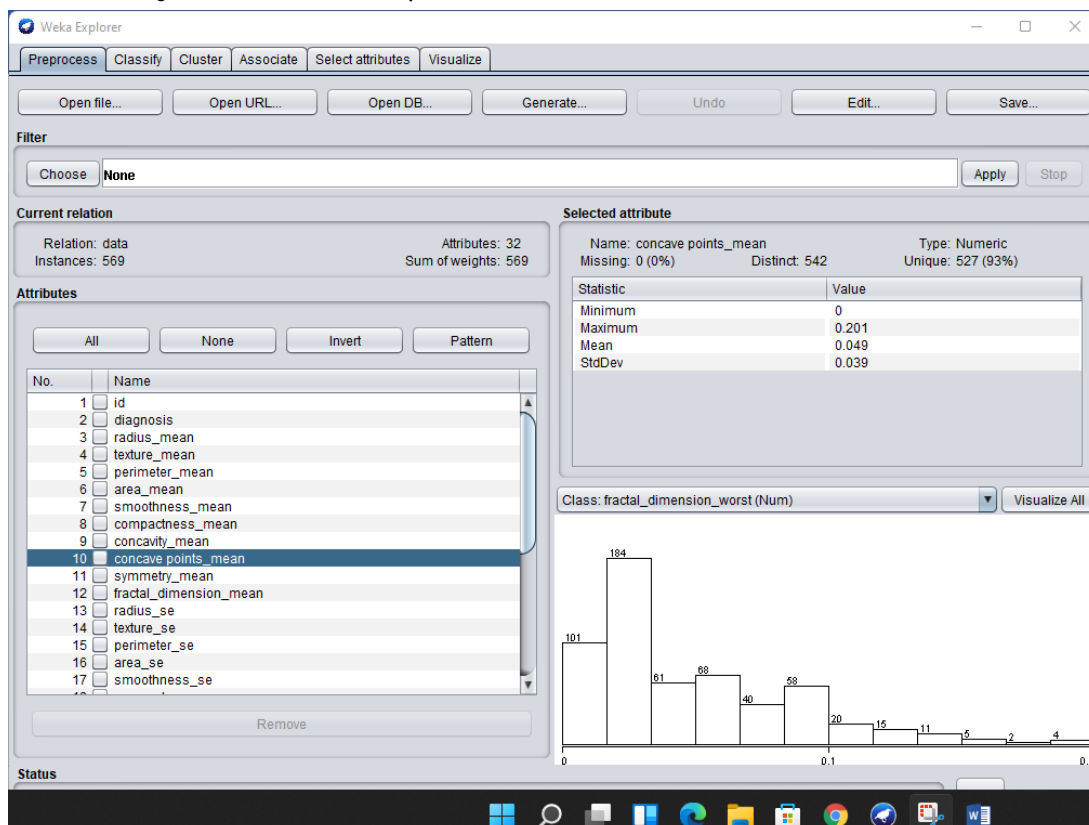
Step#3:

Click Open file and select the downloaded data...



Step#4:

After selecting the data it will be imported in weka.



Lab#5

Step#1:

Got to the www.archive.ics.uci.edu and search for any desired dataset you want:

The screenshot shows the UCI Machine Learning Repository website. The browser tabs include 'lab manual.pdf', 'uci dataset - Bing', and 'UCI Machine Learning Repository'. The address bar shows 'https://archive.ics.uci.edu/ml/datasets/Breast+Cancer'. The page features the UCI logo, a search bar, and a navigation menu. A banner at the top promotes a beta version of the repository. The main content area is titled 'Breast Cancer Data Set' and includes links for 'Data Folder' and 'Data Set Description'. An abstract mentions 'Breast Cancer Data (Restricted Access)'. A diagram of a neural network is displayed. A table provides data set characteristics, attribute characteristics, and associated tasks. The source is listed as 'Matjaz Zwitter & Milan Soklic (physicians)'.

| | | | | | |
|----------------------------|----------------|-----------------------|-----|---------------------|------------|
| Data Set Characteristics: | Multivariate | Number of Instances: | 286 | Area: | Life |
| Attribute Characteristics: | Categorical | Number of Attributes: | 9 | Date Donated | 1988-07-11 |
| Associated Tasks: | Classification | Missing Values? | Yes | Number of Web Hits: | 591128 |

Step#2:

Click on Data Folder... and download the data..

The screenshot shows the 'Index of /ml/machine-learning-databases/breast-cancer' page. The browser tabs include 'lab manual.pdf', 'uci dataset - Bing', and 'Index of /ml/machine-learning-databases/breast-cancer/'. The address bar shows 'https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer/'. The page lists a 'Parent Directory' and links to 'Index', 'breast-cancer.data', and 'breast-cancer.names'. A footer note mentions the server configuration: 'Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips SVN/1.7.14 Phusion_Passenger/4.0.53 mod_perl/2.0.11 Perl/v5.16.3 Server at archive.ics.uci.edu Port 443'.

Step#3:

After downloading the data Import this data in weka tool.

The screenshot shows the Weka Explorer application window. The 'Preprocess' tab is active. The 'Current relation' is 'anneal' with 898 instances and 39 attributes. The 'Selected attribute' panel shows a list of attributes with their counts and weights. A bar chart at the bottom visualizes the distribution of the selected attribute 'family'.

Current relation

Relation: anneal
Instances: 898
Attributes: 39
Sum of weights: 898

Attributes

All None Invert Pattern

| No. | Name |
|-----|-----------------|
| 1 | family |
| 2 | product-type |
| 3 | steel |
| 4 | carbon |
| 5 | hardness |
| 6 | temper_rolling |
| 7 | condition |
| 8 | formability |
| 9 | strength |
| 10 | non-ageing |
| 11 | surface-finish |
| 12 | surface-quality |
| 13 | enamelability |
| 14 | bc |
| 15 | bf |
| 16 | bt |
| 17 | bw/me |

Selected attribute

Name: family
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

| No. | Label | Count | Weight |
|-----|-------|-------|--------|
| 1 | ? | 772 | 772.0 |
| 2 | GB | 0 | 0.0 |
| 3 | GK | 0 | 0.0 |
| 4 | GS | 0 | 0.0 |
| 5 | TN | 67 | 67.0 |
| 6 | ZA | 0 | 0.0 |
| 7 | ZF | 0 | 0.0 |
| 8 | ZH | 0 | 0.0 |
| 9 | ZM | 0 | 0.0 |
| 10 | ZC | 59 | 59.0 |

Visualize All

Class: class (Nom)

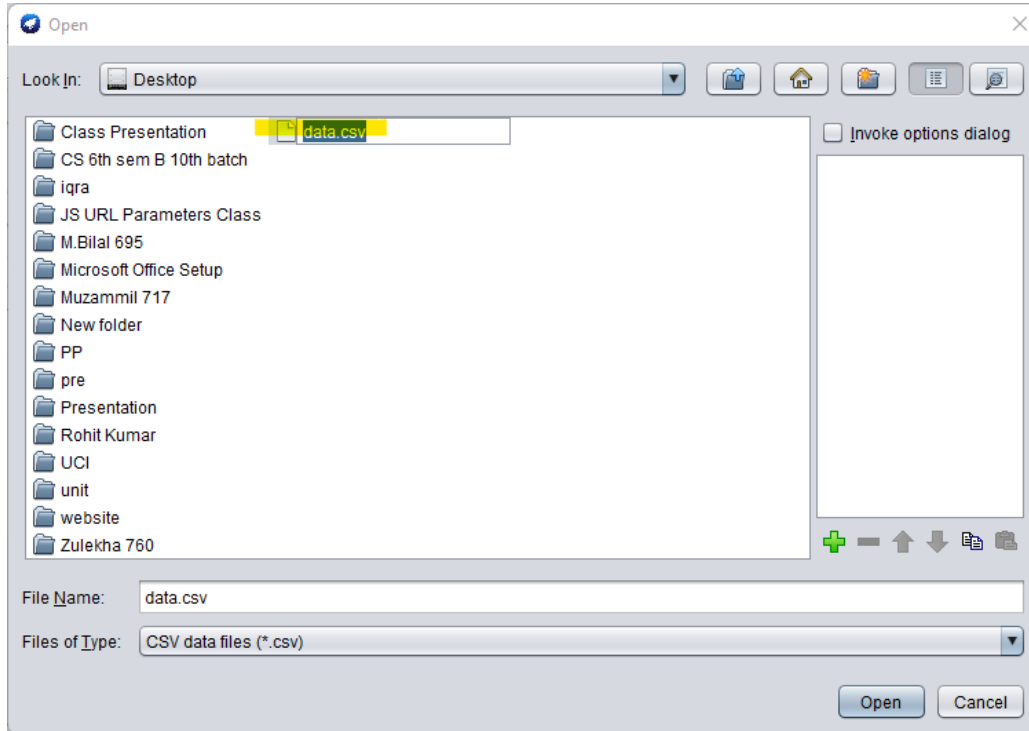
Bar chart visualization showing the distribution of the selected attribute 'family'.

| Label | Count |
|-------|-------|
| ? | 772 |
| GB | 0 |
| GK | 0 |
| GS | 0 |
| TN | 67 |
| ZA | 0 |
| ZF | 0 |
| ZH | 0 |
| ZM | 0 |
| ZC | 59 |

Lab#6

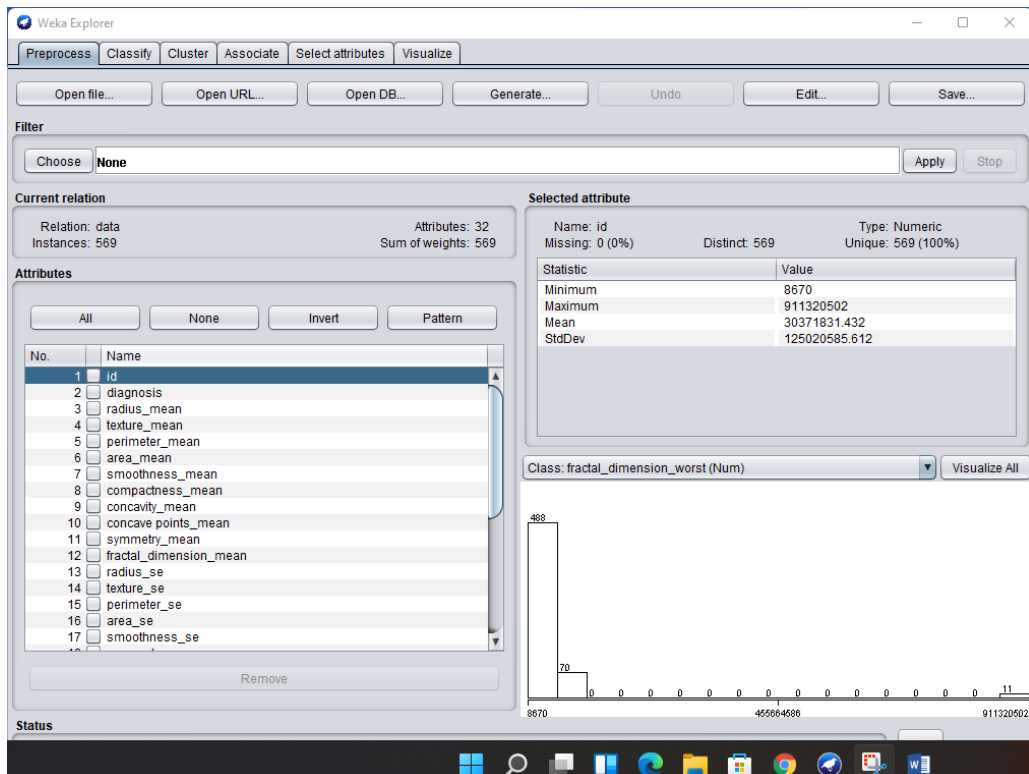
Step#1:

In this lab we will import the data from our local device to weka tool...open explorer and click on open file:



Step#2:

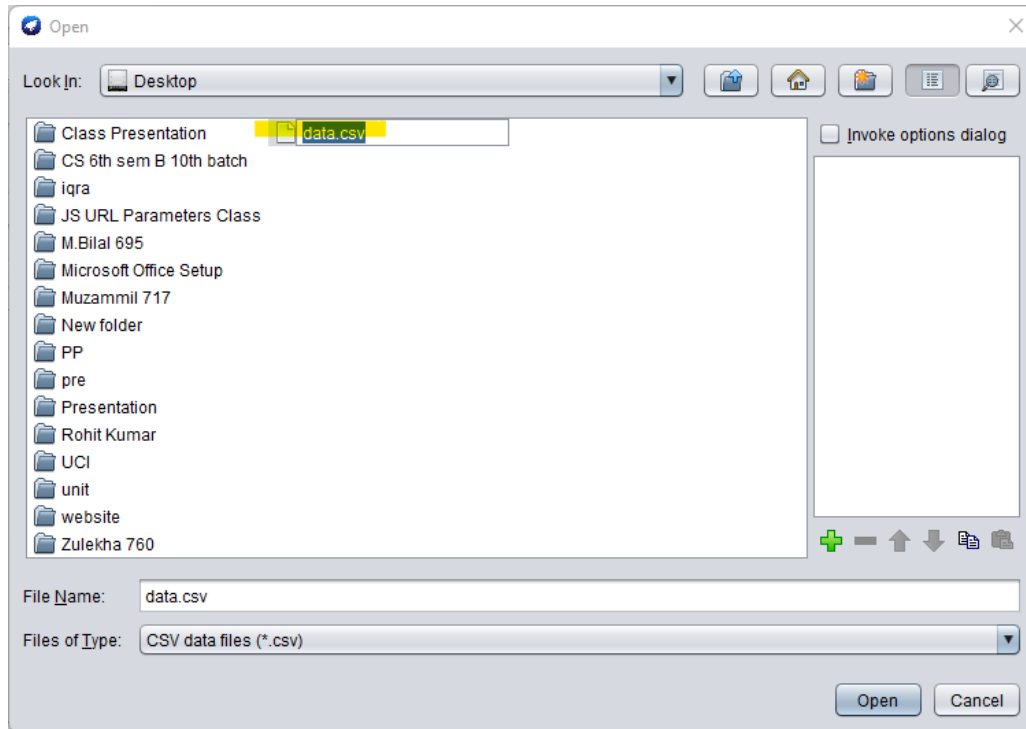
Click on the dataset file then it will be imported into weka..



Lab#7

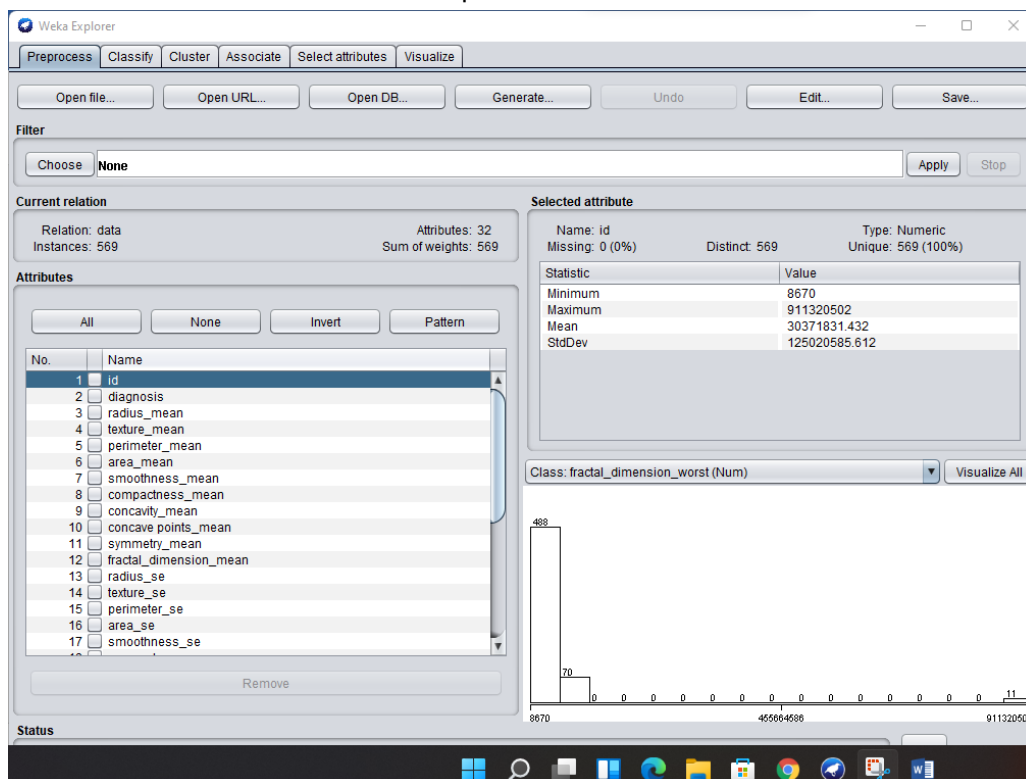
Step#1:

In this lab we will import the data from our local device to weka tool...open explorer and click on open file:



Step#2:

Click on the dataset file then it will be imported into weka..



Step#3:

Now choose any preprocessing algorithm(Normalize) of your choice and apply that algorithm.

The screenshot displays the Weka Explorer interface with the 'Preprocess' tab selected. The 'Filter' section shows 'Choose' and 'Normalize -S 1.0 -T 0.0' with 'Apply' and 'Stop' buttons. The 'Current relation' section indicates 'Relation: data-weka.filters.unsupervised.attribute.Norm...' with 32 attributes and 569 instances. The 'Attributes' list on the left includes 'id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean', 'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean', 'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean', 'radius_se', 'texture_se', 'perimeter_se', 'area_se', and 'smoothness_se'. The 'Selected attribute' section shows statistics for 'id': Minimum 0, Maximum 1, Mean 0.033, and StdDev 0.137. A histogram at the bottom right shows the distribution of 'fractal_dimension_worst' with a peak at 488 and a tail extending to 11.

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose | Normalize -S 1.0 -T 0.0 | Apply | Stop

Current relation

Relation: data-weka.filters.unsupervised.attribute.Norm... | Attributes: 32 | Sum of weights: 569
Instances: 569

Attributes

All | None | Invert | Pattern

| No. | Name |
|-----|------------------------|
| 1 | id |
| 2 | diagnosis |
| 3 | radius_mean |
| 4 | texture_mean |
| 5 | perimeter_mean |
| 6 | area_mean |
| 7 | smoothness_mean |
| 8 | compactness_mean |
| 9 | concavity_mean |
| 10 | concave points_mean |
| 11 | symmetry_mean |
| 12 | fractal_dimension_mean |
| 13 | radius_se |
| 14 | texture_se |
| 15 | perimeter_se |
| 16 | area_se |
| 17 | smoothness_se |

Remove

Selected attribute

Name: id | Missing: 0 (0%) | Distinct: 569 | Type: Numeric | Unique: 569 (100%)

| Statistic | Value |
|-----------|-------|
| Minimum | 0 |
| Maximum | 1 |
| Mean | 0.033 |
| StdDev | 0.137 |

Class: fractal_dimension_worst (Num) | Visualize All

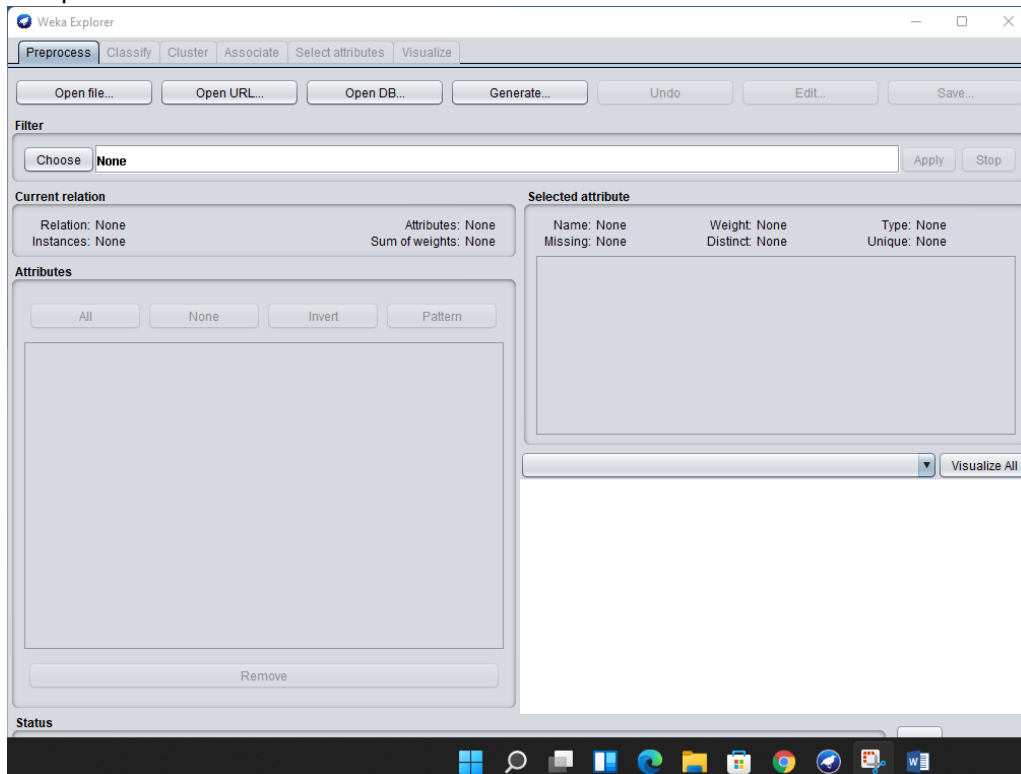
488 | 70 | 0 | 0.5 | 1 | 11

Status

Lab#8

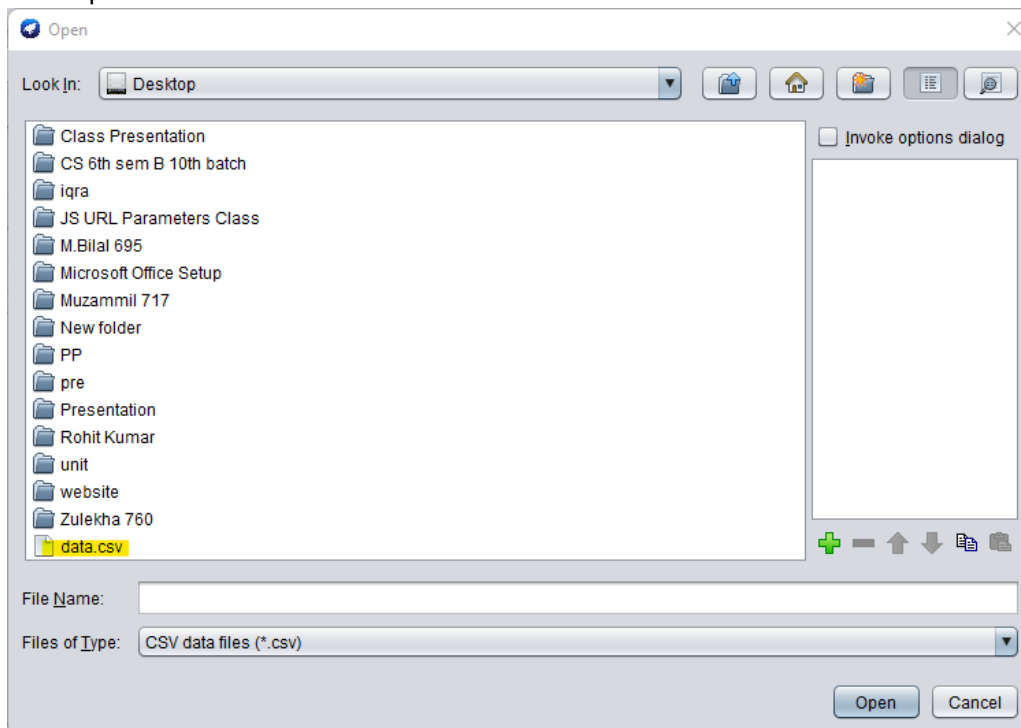
Step#1:

Import the Downloaded the dataset from the website then open the weka tool and click on open file in explorer.



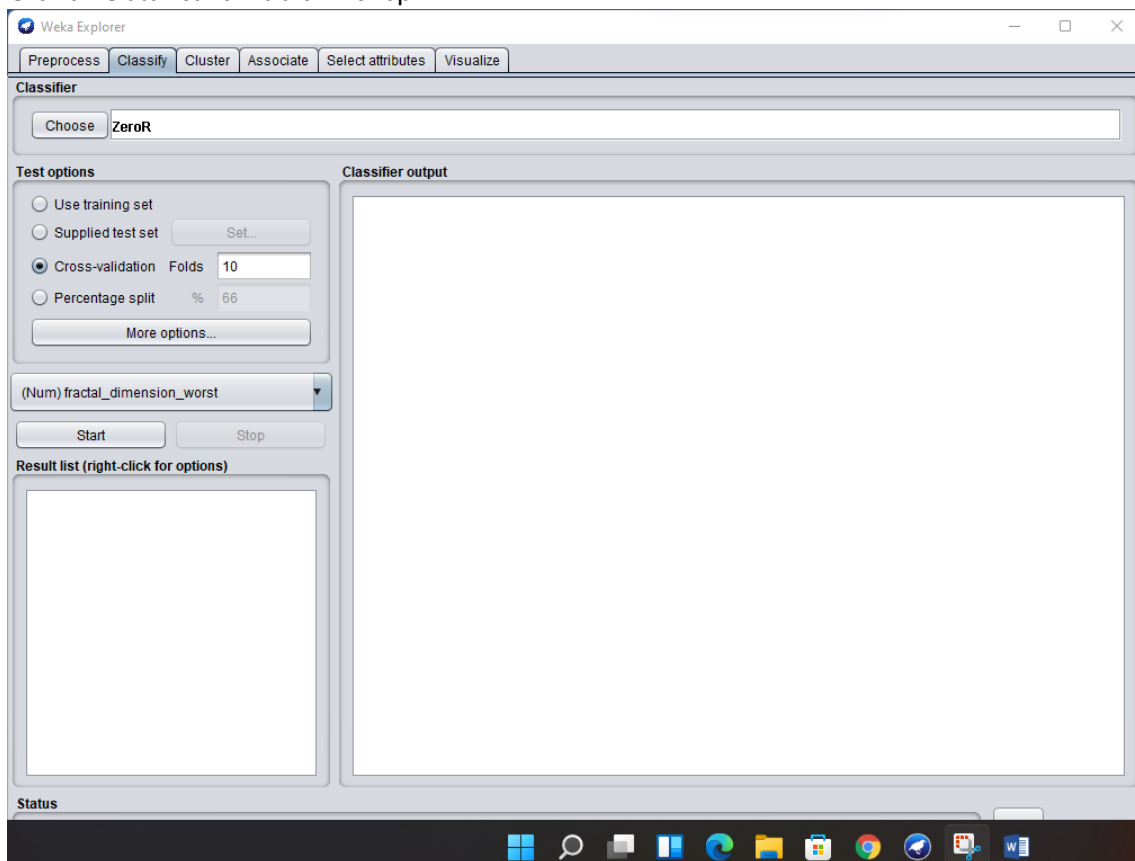
Step#2:

Click Open file and select the downloaded data...



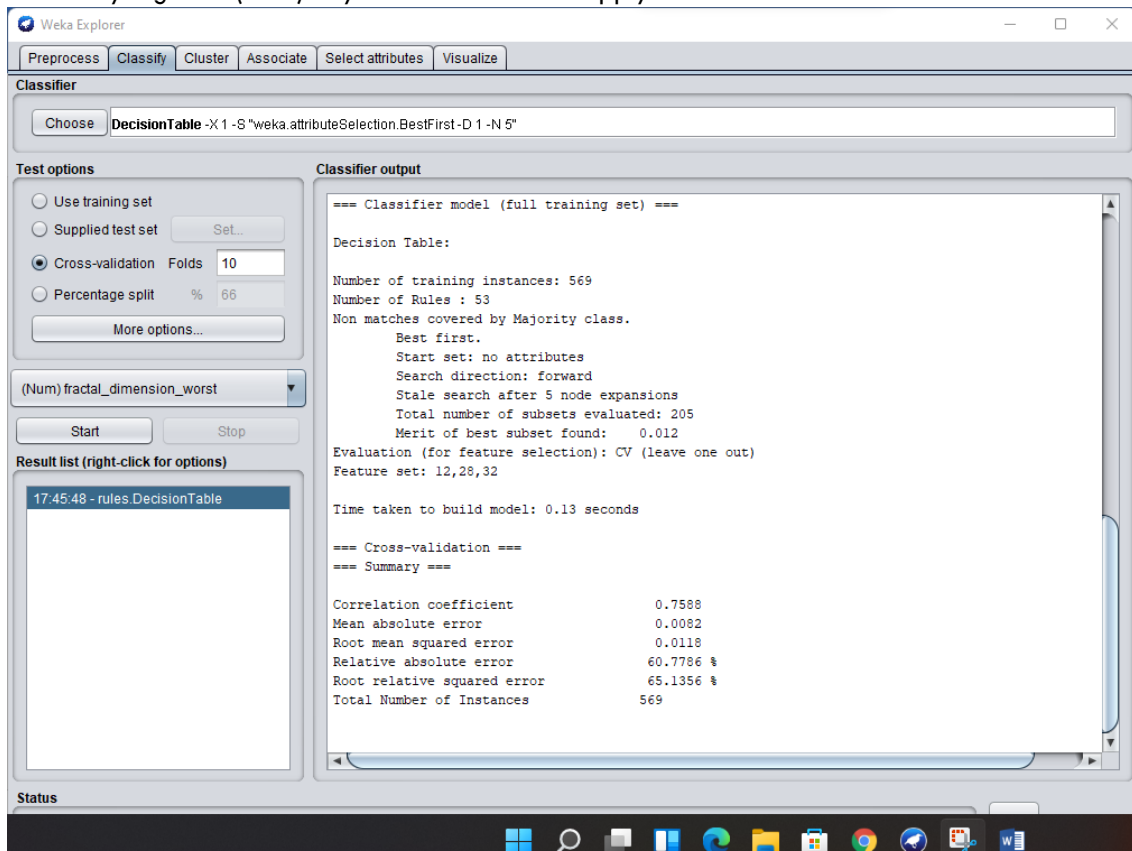
Step#3:

Click on Classification tab on the top...



Step#4:

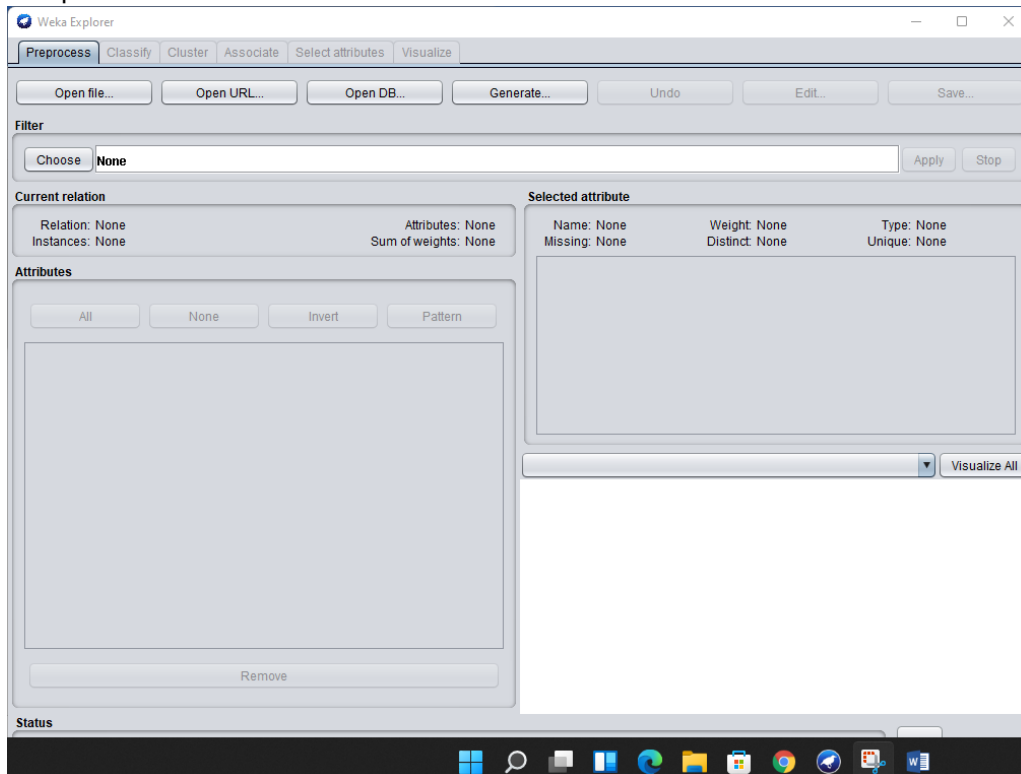
Select any algorithm(Vote) of your choice then click apply.



Lab#9

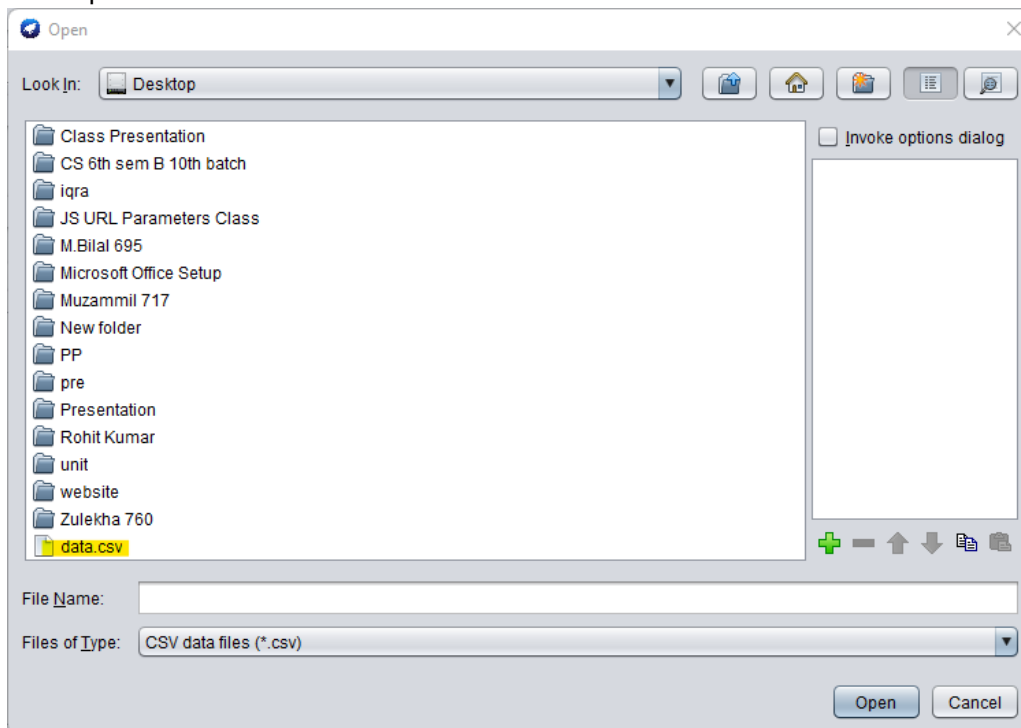
Step#1:

Import the Downloaded the dataset from the website then open the weka tool and click on open file in explorer.



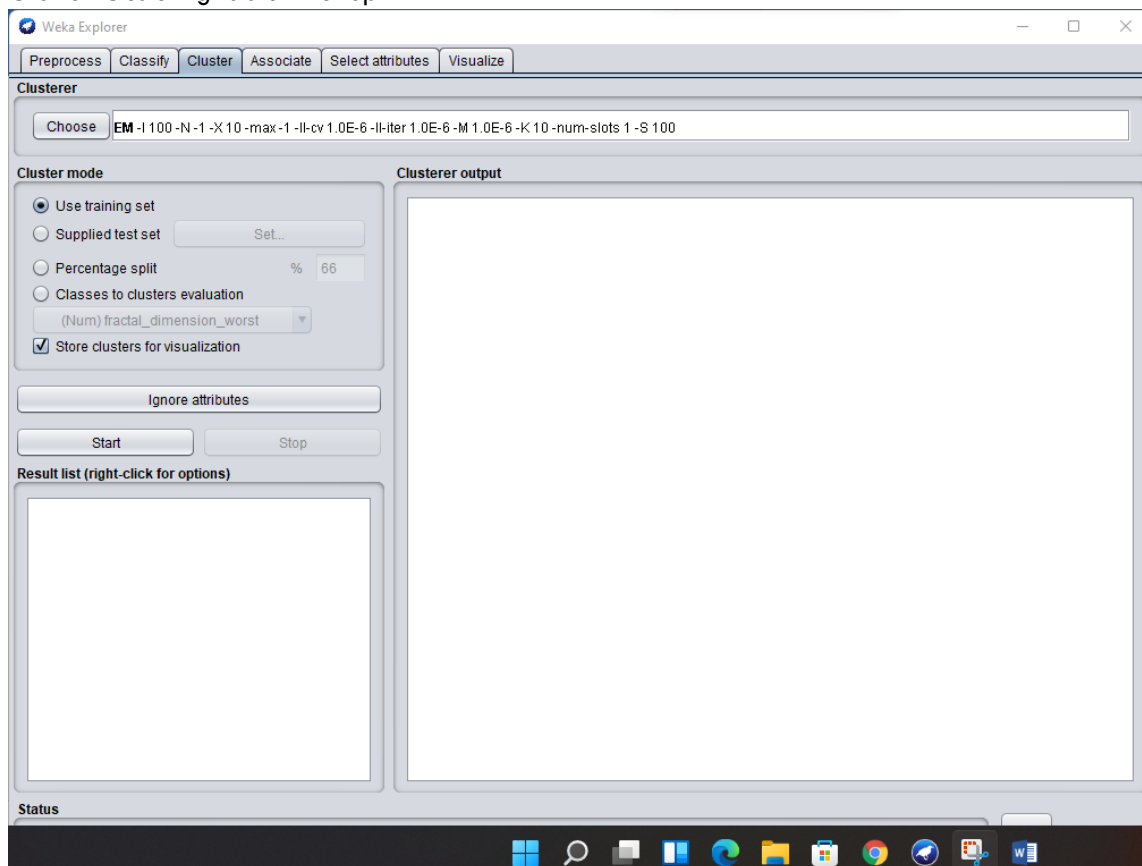
Step#2:

Click Open file and select the downloaded data...



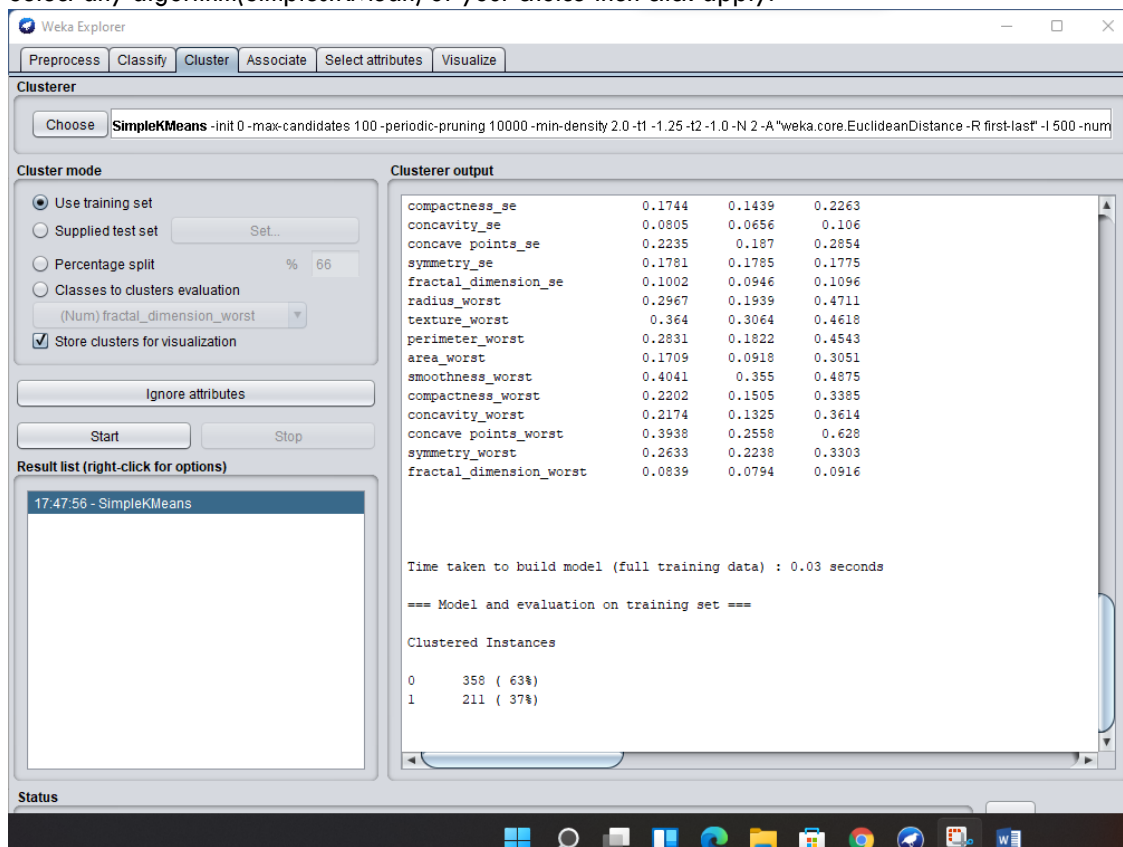
Step#3:

Click on Clustering tab on the top...



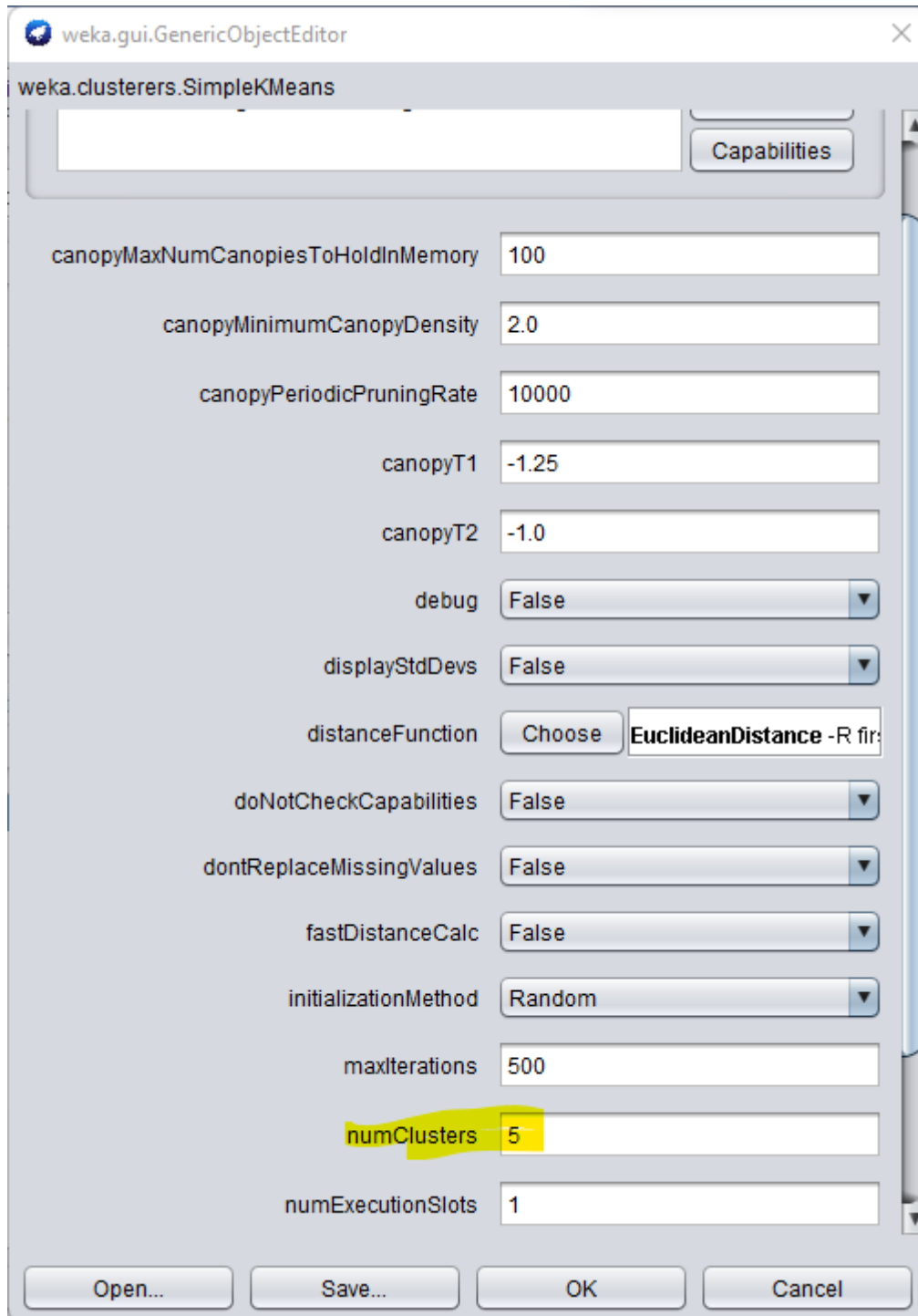
Step#4:

Select any algorithm(SimplestKMean) of your choice then click apply.



Step#5:

you can also change the number of cluster by pressing on the algorithm stripe on weka tool...It will change the number of cluster in the algorithm...



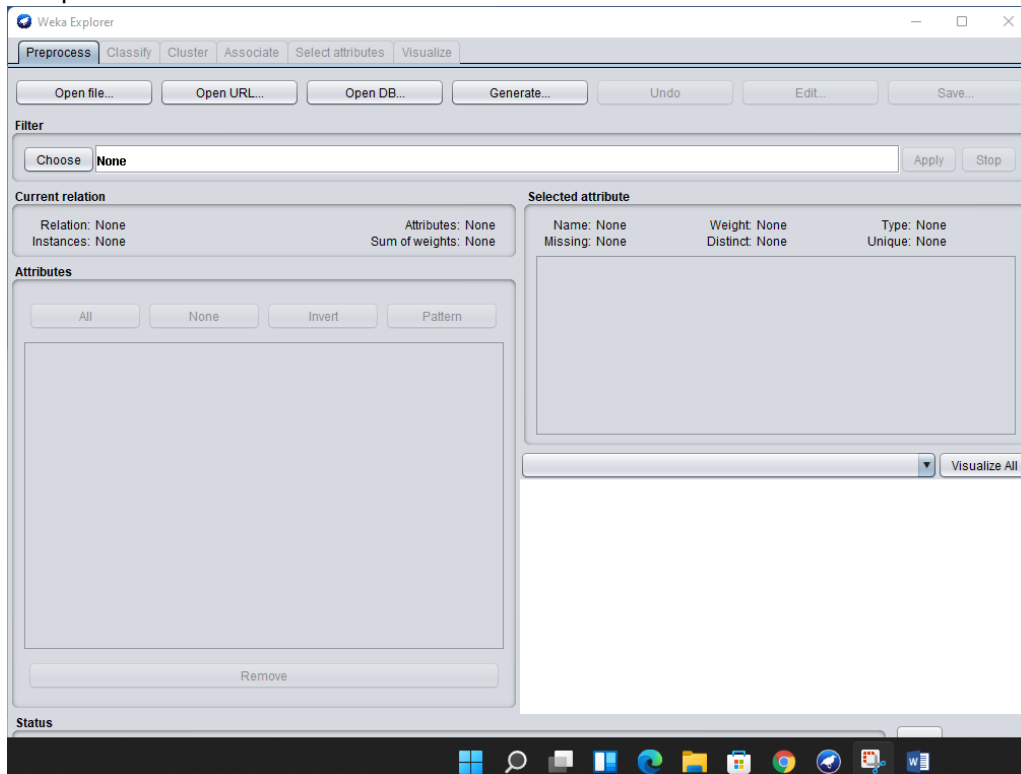
The screenshot shows the 'weka.gui.GenericObjectEditor' window for the 'weka.clusterers.SimpleKMeans' algorithm. The window contains various configuration fields and buttons. The 'numClusters' field is highlighted with a yellow background and contains the value 5. The 'distanceFunction' field is set to 'EuclideanDistance -R first'. The 'Capabilities' button is visible in the top right corner. The 'Open...', 'Save...', 'OK', and 'Cancel' buttons are at the bottom.

| Property | Value |
|------------------------------------|-----------------------------------|
| canopyMaxNumCanopiesToHoldInMemory | 100 |
| canopyMinimumCanopyDensity | 2.0 |
| canopyPeriodicPruningRate | 10000 |
| canopyT1 | -1.25 |
| canopyT2 | -1.0 |
| debug | False |
| displayStdDevs | False |
| distanceFunction | Choose EuclideanDistance -R first |
| doNotCheckCapabilities | False |
| dontReplaceMissingValues | False |
| fastDistanceCalc | False |
| initializationMethod | Random |
| maxIterations | 500 |
| numClusters | 5 |
| numExecutionSlots | 1 |

Lab#10

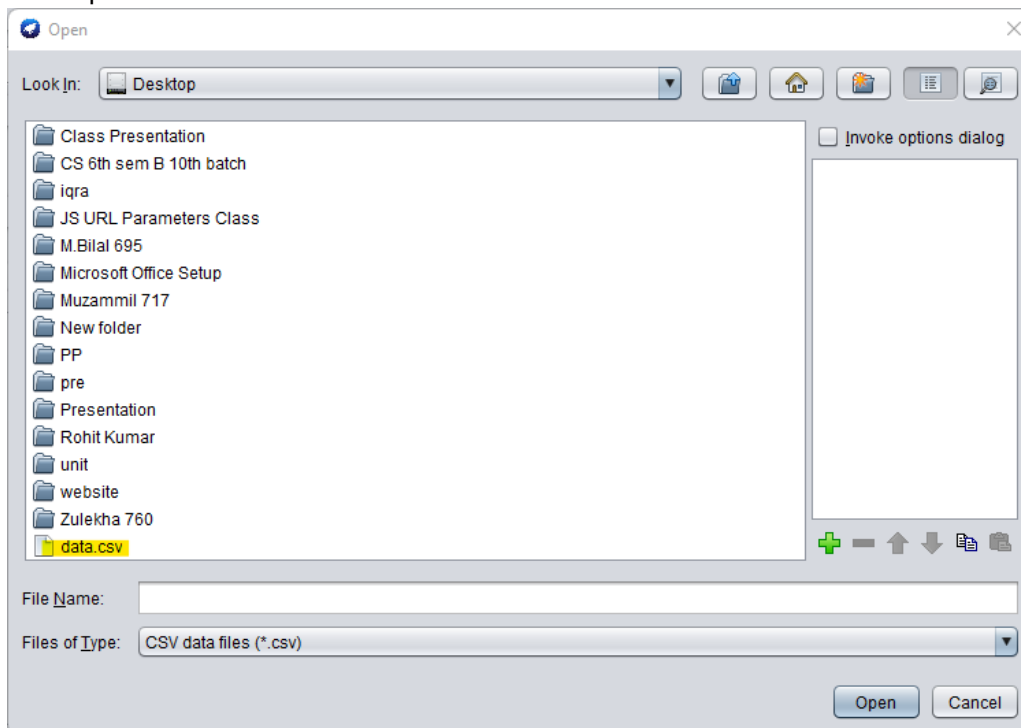
Step#1:

Import the Downloaded the dataset from the website then open the weka tool and click on open file in explorer.



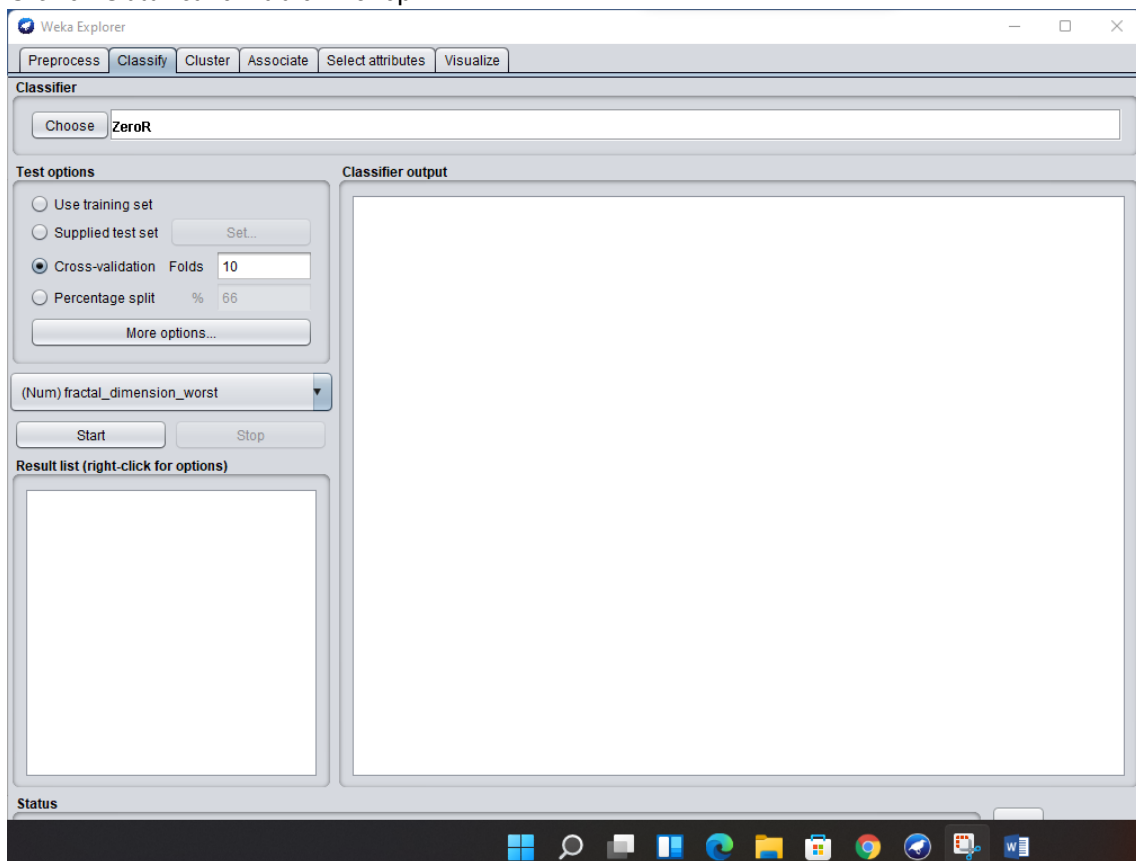
Step#2:

Click Open file and select the downloaded data...



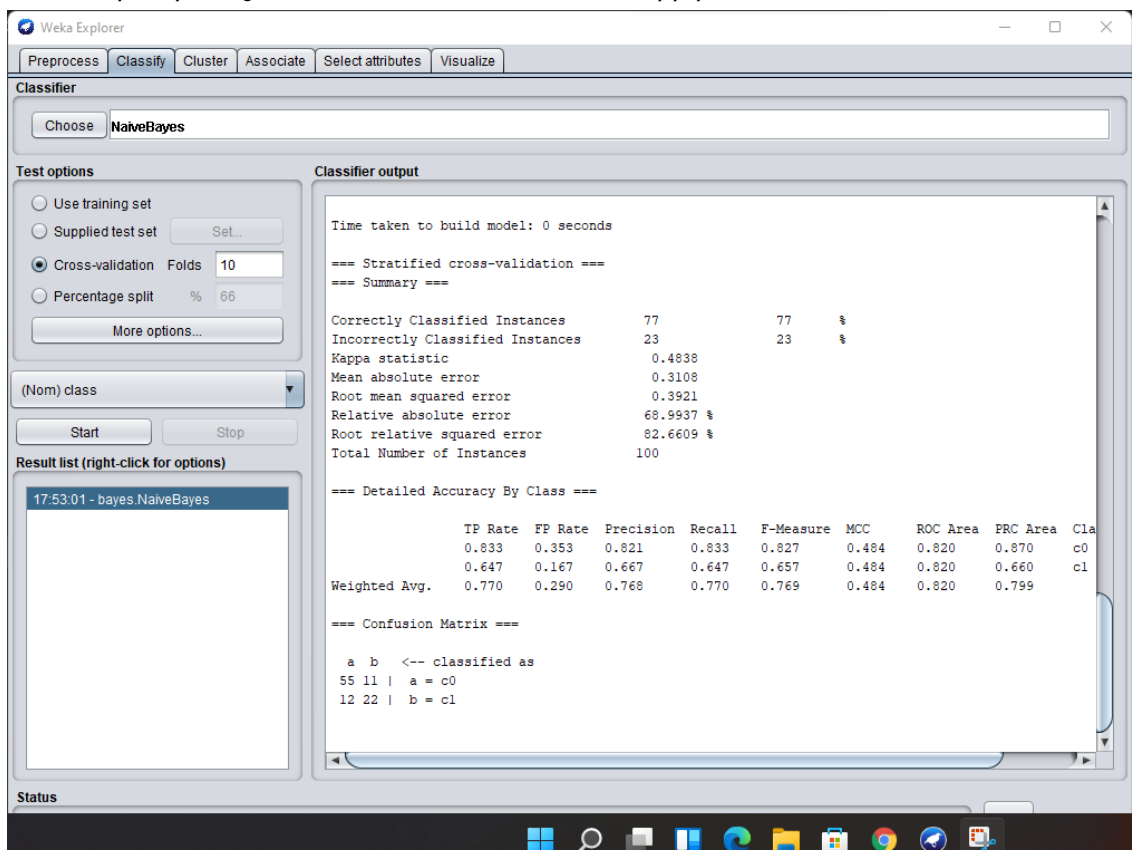
Step#3:

Click on Classification tab on the top...



Step#4:

Select NaiveBayes algorithm from the choose button and apply.



Step#5:

Now select J48 tree algorithm from classification and apply and right click on the algorithm then click on visualize tree it will open the window.

