AIM : Study and Installation of Weka AI Tool

Name:- Yash Santosh Lale
Roll no. :- 38
Sub :- ET II
section:- 3C
Date:08/03/2025

To install WEKA on your machine, visit WEKA's official website and download the installation file. WEKA supports installation on Windows, Mac OS X and Linux. You just need to follow the instructions on this page to install WEKA for your OS.

The steps for installing on Mac are as follows:

- Download the Mac installation file.

- Double click on the downloaded **weka-3-8-3-corretto-jvm.dmg file**.

You will see the following screen on successful installation.

- Click on the **weak-3-8-3-corretto-jvm** icon to start Weka.

- Optionally you may start it from the command line:

```
java -jar weka.jar
```

The WEKA GUI Chooser application will start and you would see the following screen:


The GUI Chooser application allows you to run five different types of applications as listed here:

- Explorer

- Experimenter

- KnowledgeFlow

- Workbench

- Simple CLI

We will be using **Explorer** in this tutorial.

In this chapter, let us look into various functionalities that the explorer provides for working with big data.

When you click on the **Explorer** button in the **Applications** selector, it opens the following screen:
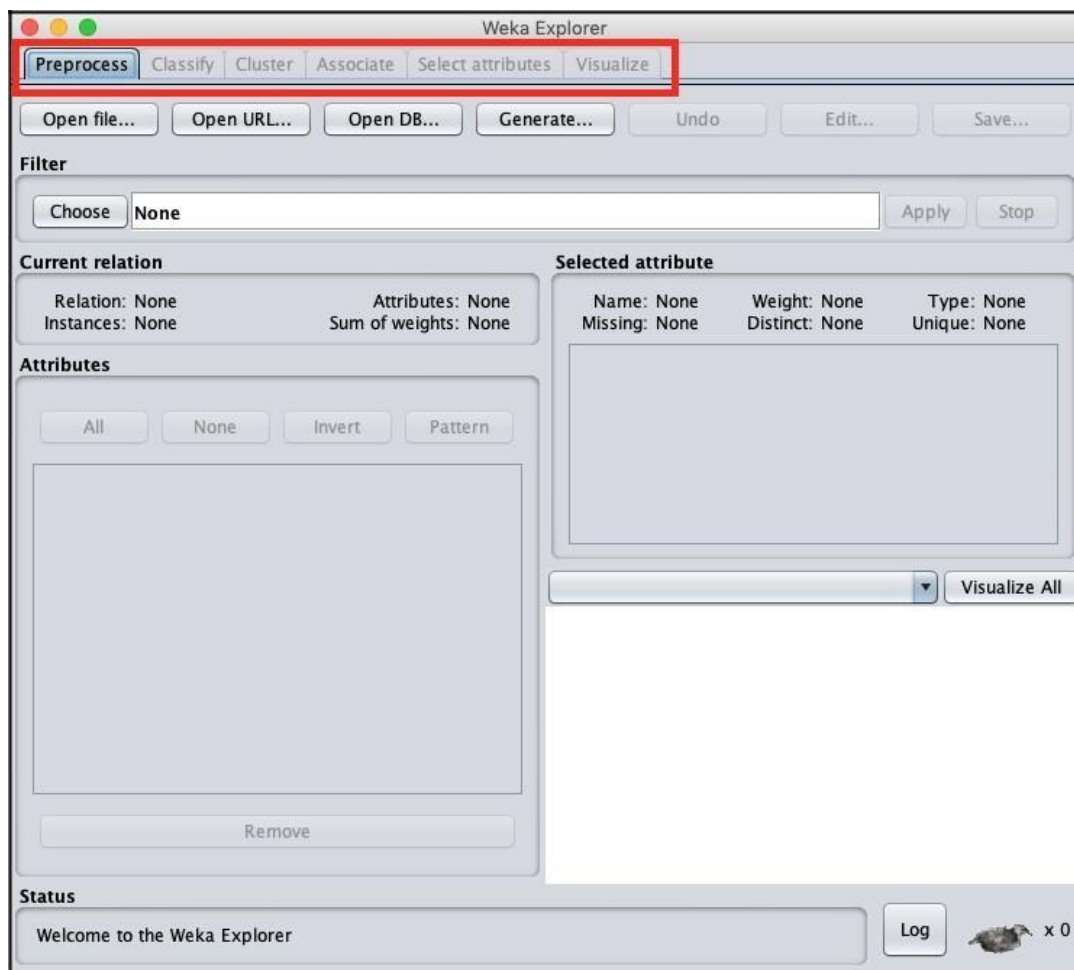
On the top, you will see several tabs as listed here:

- Preprocess

- Classify

- Cluster

- Associate

- Select Attributes

- Visualize

    6 Weka

Under these tabs, there are several pre-implemented machine learning algorithms. Let us look into each of them in detail now.

Preprocess Tab

Initially as you open the explorer, only the **Preprocess** tab is enabled. The first step in machine learning is to preprocess the data. Thus, in the **Preprocess** option, you will select the data file, process it and make it fit for applying the various machine learning algorithms.

Classify Tab

The **Classify** tab provides you several machine learning algorithms for the classification of your data. To list a few, you may apply algorithms such as Linear Regression, Logistic Regression, Support Vector Machines, Decision Trees, RandomTree, RandomForest, NaiveBayes, and so on. The list is very exhaustive and provides both supervised and unsupervised machine learning algorithms.

Cluster Tab

Under the **Cluster** tab, there are several clustering algorithms provided - such as SimpleKMeans, FilteredClusterer, HierarchicalClusterer, and so on.

Associate Tab

Under the **Associate** tab, you would find Apriori, FilteredAssociator and FPGrowth.

Select Attributes Tab

Select Attributes allows you feature selections based on several algorithms such as ClassifierSubsetEval, PrinicipalComponents, etc.

Visualize Tab

Lastly, the **Visualize** option allows you to visualize your processed data for analysis.

As you noticed, WEKA provides several ready-to-use algorithms for testing and building your machine learning applications. To use WEKA effectively, you must have a sound knowledge of these algorithms, how they work, which one to choose under what circumstances, what to look for in their processed output, and so on. In short, you must have a solid foundation in machine learning to use WEKA effectively in building your apps.

In the upcoming chapters, you will study each tab in the explorer in depth.

In this chapter, we start with the first tab that you use to preprocess the data. This is common to all algorithms that you would apply to your data for building the model and is a common step for all subsequent operations in WEKA.

For a machine learning algorithm to give acceptable accuracy, it is important that you must cleanse your data first. This is because the raw data collected from the field may contain null values, irrelevant columns and so on.

In this chapter, you will learn how to preprocess the raw data and create a clean, meaningful dataset for further use.

First, you will learn to load the data file into the WEKA explorer. The data can be loaded from the following sources:

- Local file system

- Web

- Database

In this chapter, we will see all the three options of loading data in detail.

Loading Data from Local File System

Just under the Machine Learning tabs that you studied in the previous lesson, you would find the following three buttons:

- Open file …

- Open URL …

- Open DB …

Click on the **Open file …** button. A directory navigator window opens as shown in the following screen:
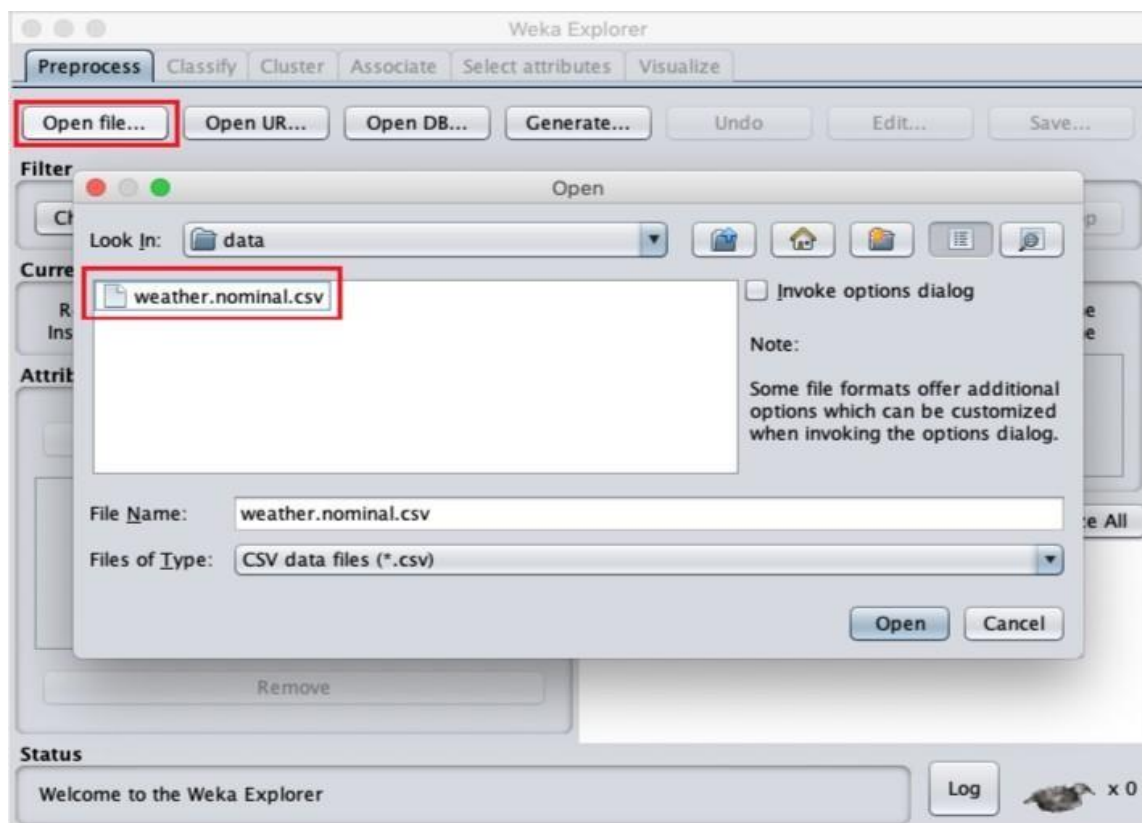
Now, navigate to the folder where your data files are stored. WEKA installation comes up with many sample databases for you to experiment. These are available in the **data** folder of the WEKA installation.

For learning purpose, select any data file from this folder. The contents of the file would be loaded in the WEKA environment. We will very soon learn how to inspect and process this loaded data. Before that, let us look at how to load the data file from the Web.

Loading Data from Web

Once you click on the **Open URL ...** button, you can see a window as follows:
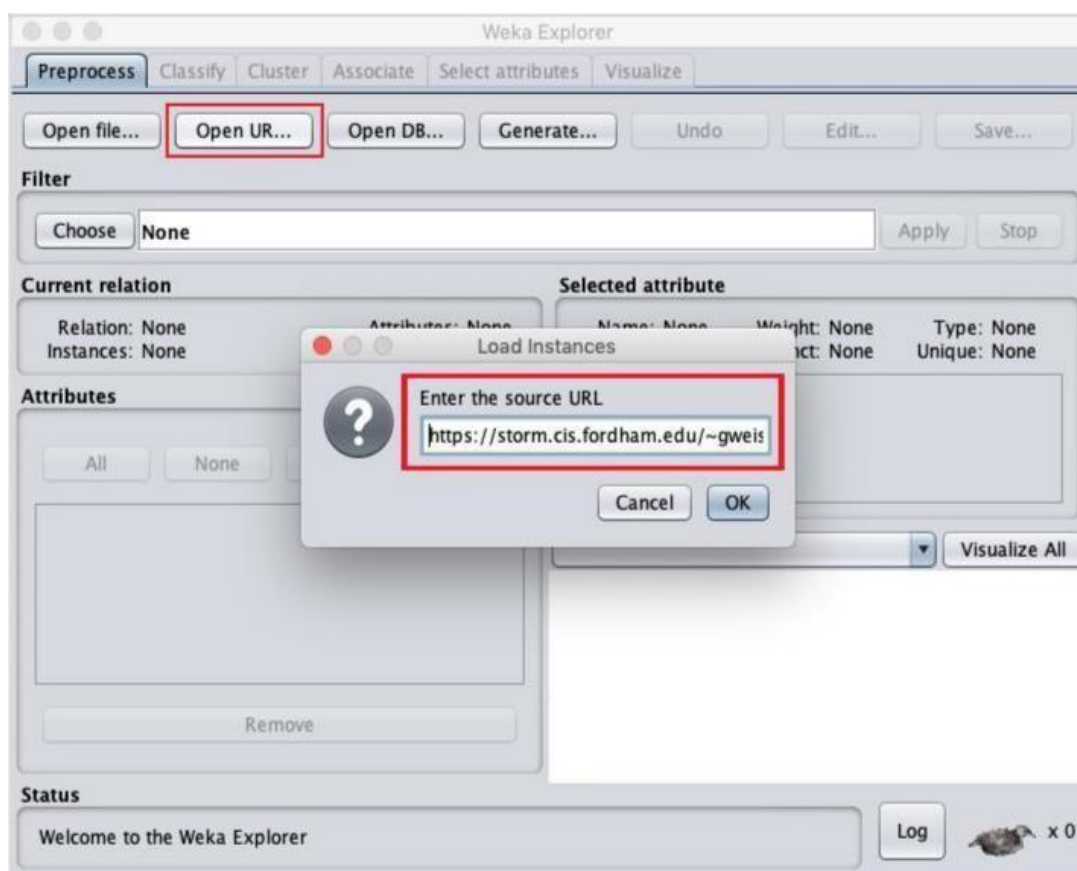
We will open the file from a public URL Type the following URL in the popup box:

https://storm.cis.fordham.edu/~gweiss/data-mining/weka-data/weather.nominal.arff

You may specify any other URL where your data is stored. The **Explorer** will load the data from the remote site into its environment.

Loading Data from DB

Once you click on the **Open DB ...** button, you can see a window as follows:
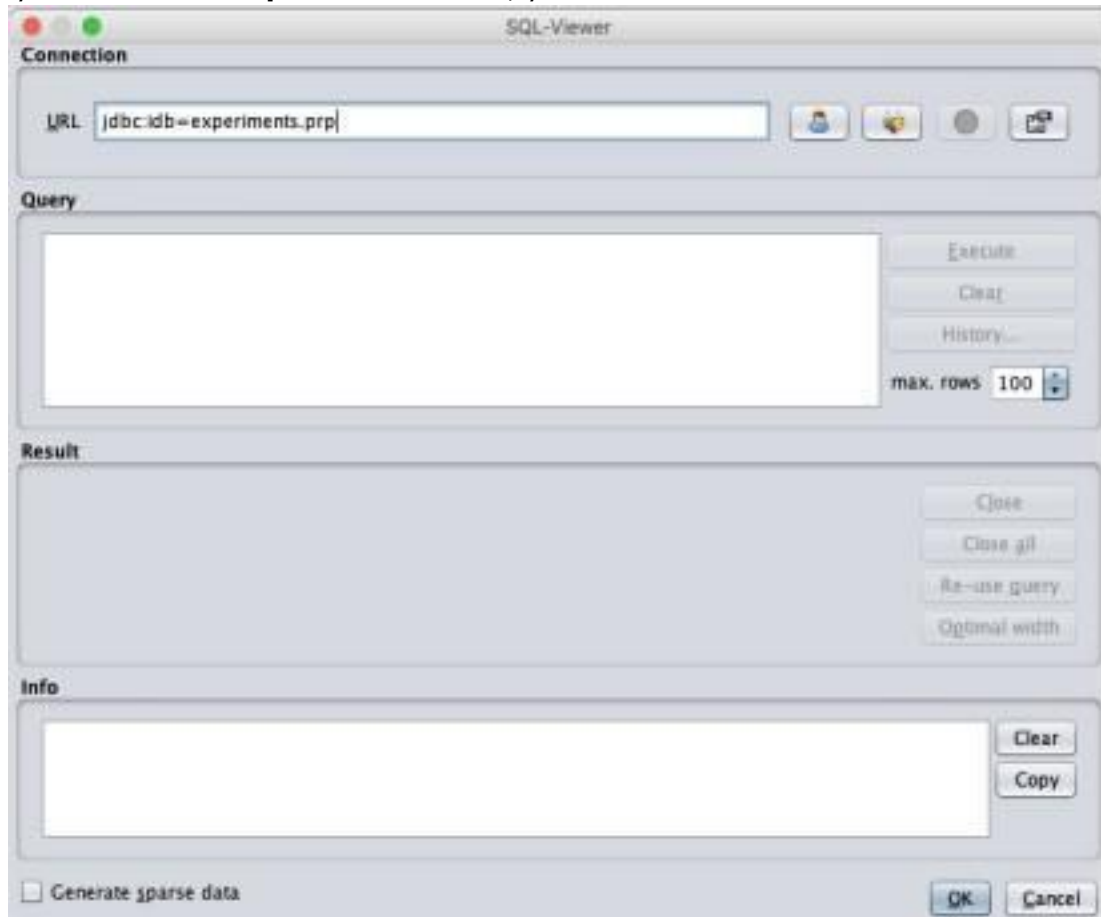
Set the connection string to your database, set up the query for data selection, process the query and load the selected records in WEKA.
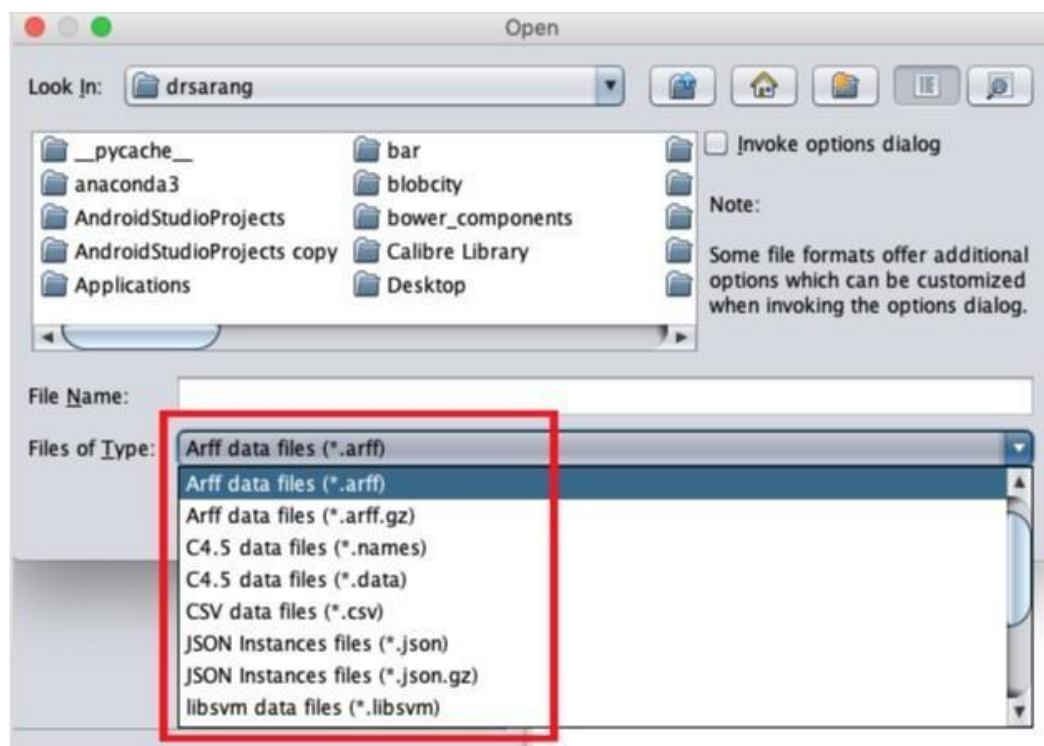
WEKA supports a large number of file formats for the data. Here is the complete list:

- arff

- arff.gz

- bsi

- csv

- dat

- data

- json

- json.gz

- libsvm

- m

- names

- xrff

- xrff.gz

The types of files that it supports are listed in the drop-down list box at the bottom of the screen. This is shown in the screenshot given below.

As you would notice it supports several formats including CSV and JSON. The default file type is Arff.

Arff Format

An **Arff** file contains two sections - header and data.

- The header describes the attribute types.

- The data section contains a comma separated list of data.

As an example for Arff format, the **Weather** data file loaded from the WEKA sample databases is shown below:
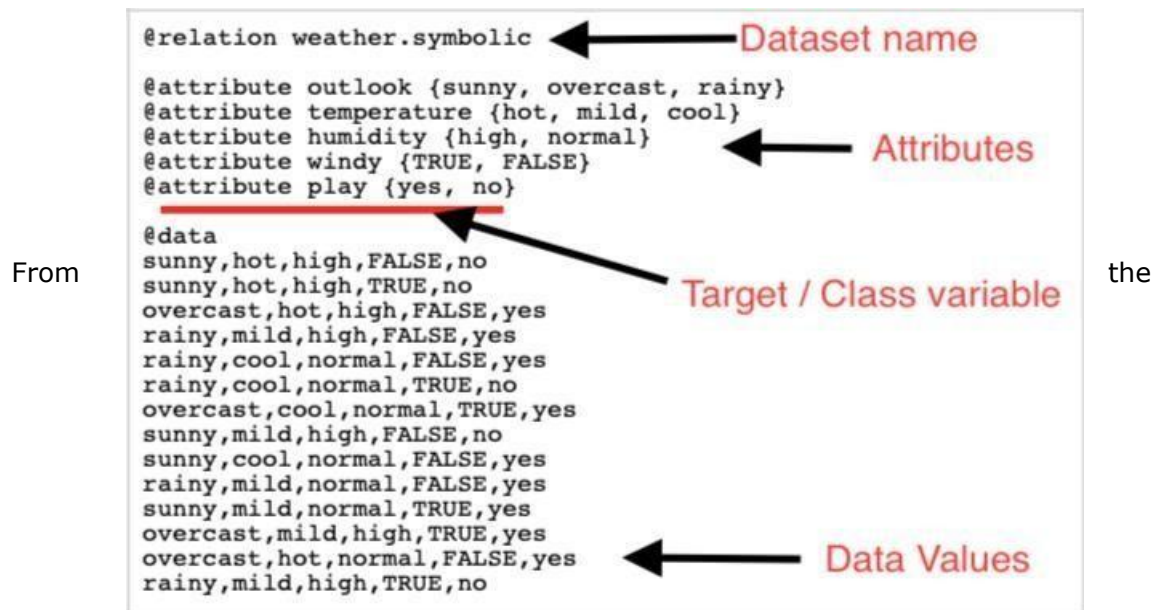
From screenshot, you can infer the following points:

- The @relation tag defines the name of the database.

- The @attribute tag defines the attributes.

- The @data tag starts the list of data rows each containing the comma separated fields.

- The attributes can take nominal values as in the case of outlook shown here:

```
@attribute outlook (sunny, overcast, rainy)
```

- The attributes can take real values as in this case:

```
@attribute temperature real
```

- You can also set a Target or a Class variable called play as shown here:

```
@attribute play (yes, no)
```

- The Target assumes two nominal values yes or no.