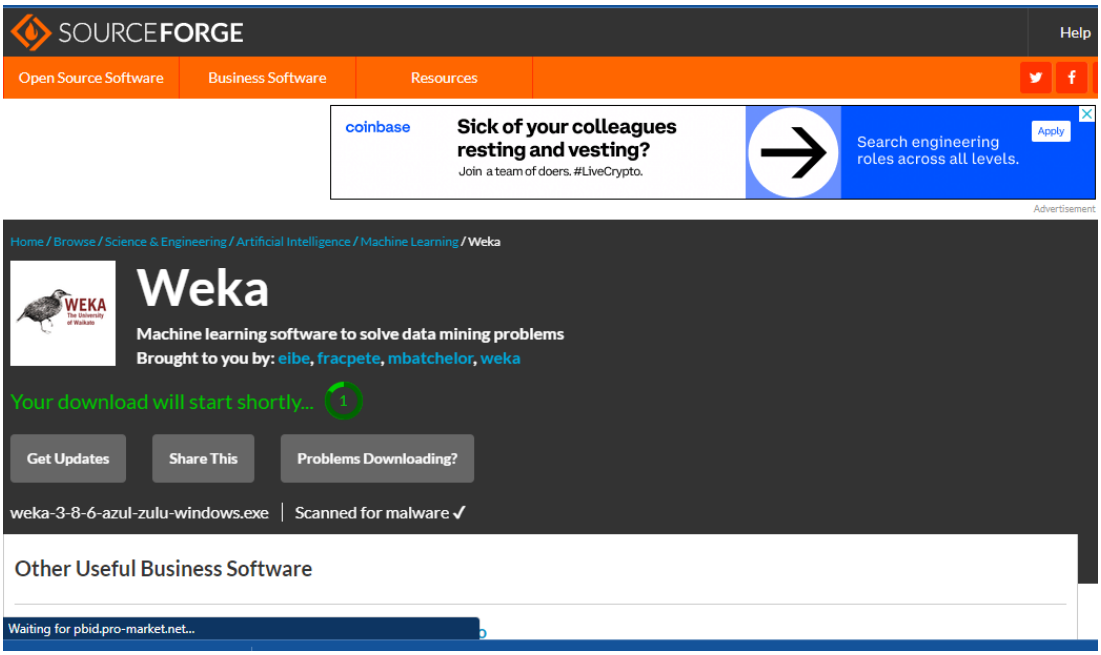
Prajakta Deokule (3330)

A1

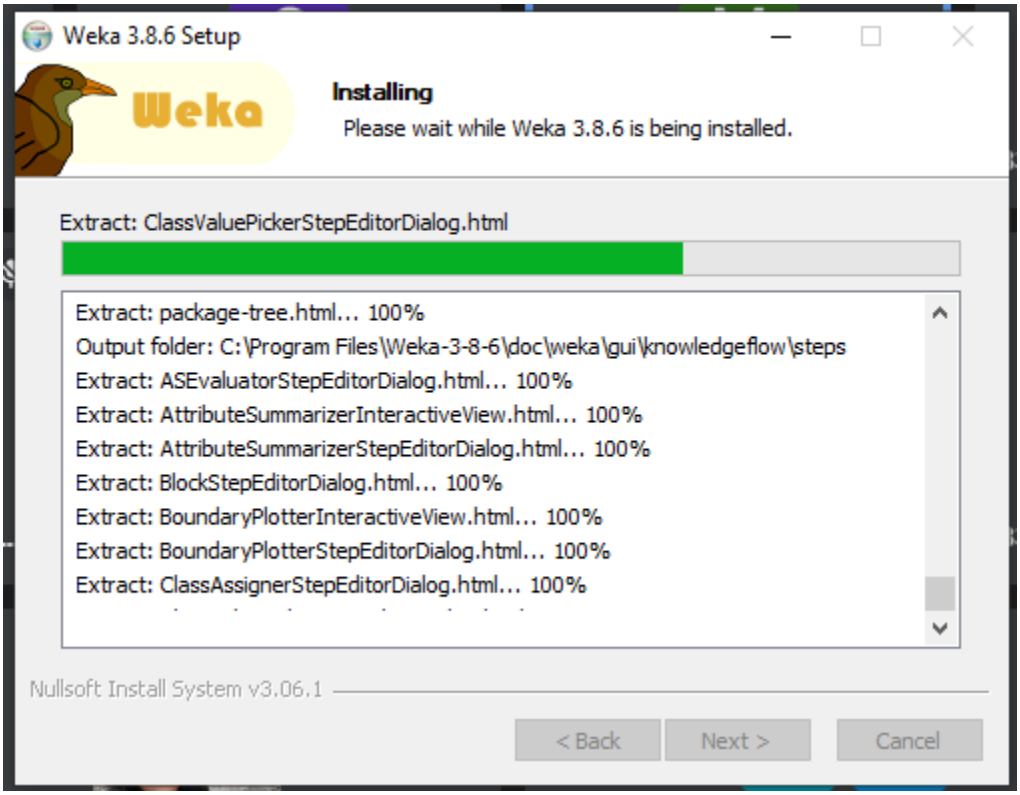
Download and Install WEKA

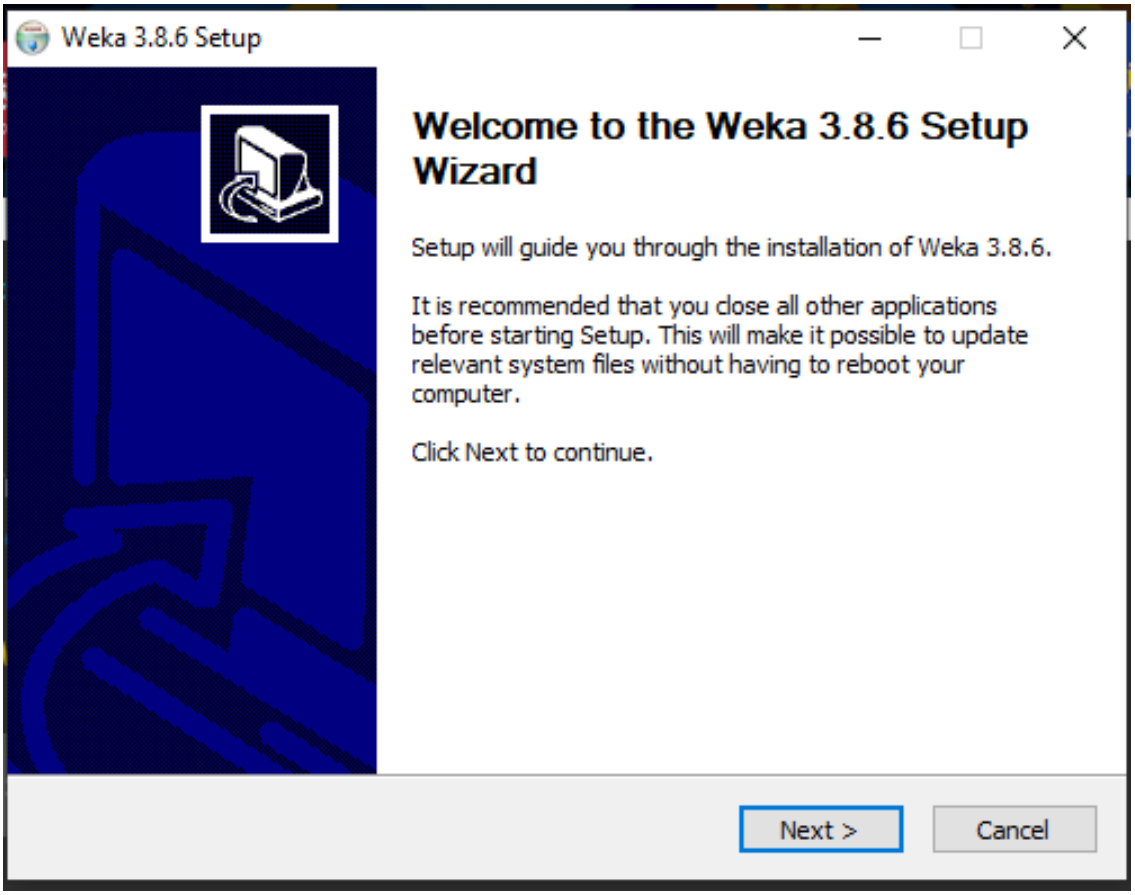
Aim - Explore WEKA Data Mining / Machine Learning Toolkit and perform the following operations: Understand the features of WEKA toolkit, Study the arff file format, explore the available datasets in WEKA

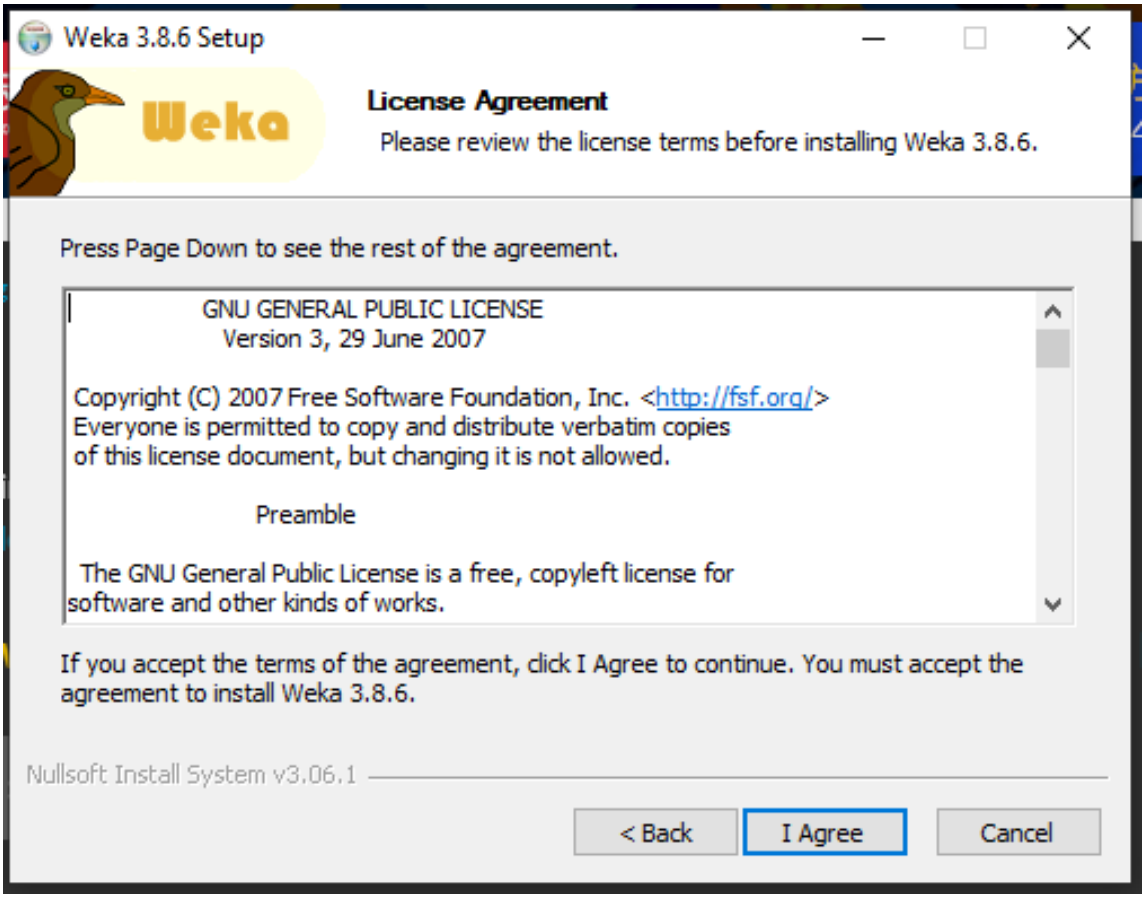
Downloading WEKA-



Installing WEKA –

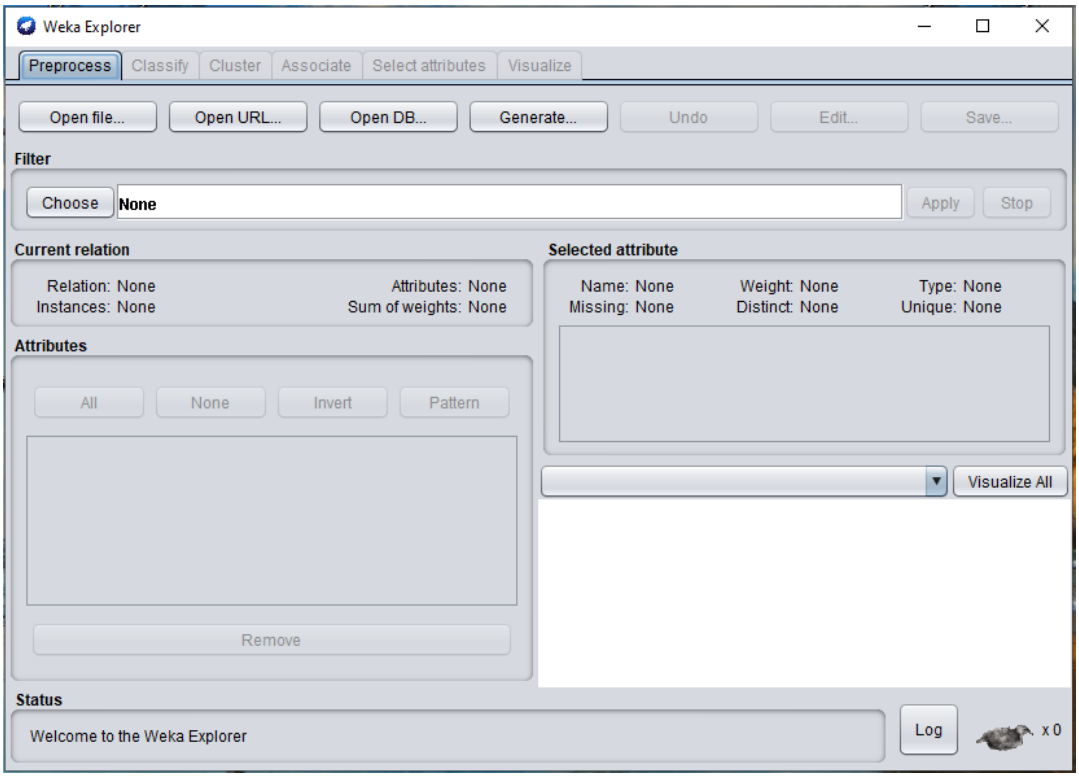



Weka 3.8.6 Setup
Welcome to the Weka 3.8.6 Setup Wizard
Setup will guide you through the installation of Weka 3.8.6.
It is recommended that you close all other applications before starting Setup. This will make it possible to update relevant system files without having to reboot your computer.
Click Next to continue.
Next >
Cancel

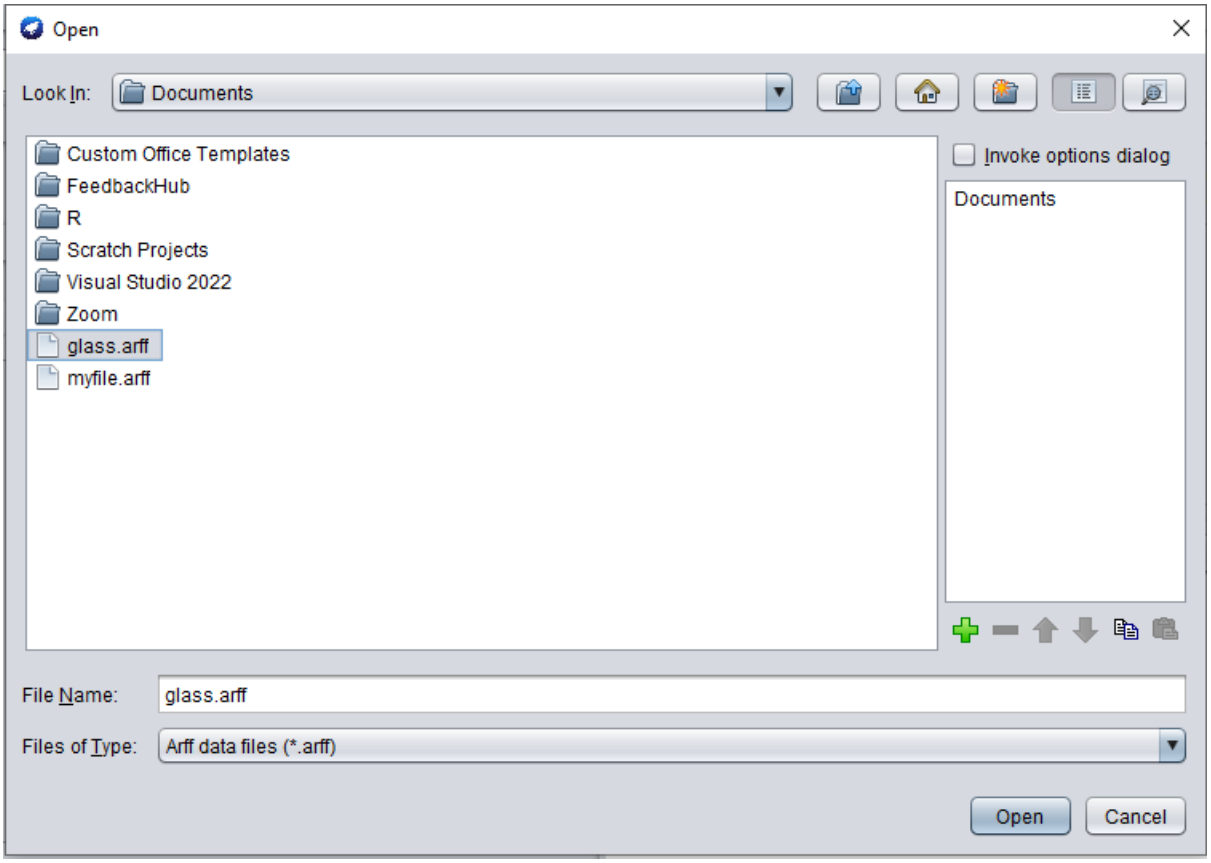

Weka 3.8.6 Setup
Weka
License Agreement
Please review the license terms before installing Weka 3.8.6.
Press Page Down to see the rest of the agreement.
GNU GENERAL PUBLIC LICENSE
Version 3, 29 June 2007
Copyright (C) 2007 Free Software Foundation, Inc. <http://fsf.org/>
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.
Preamble
The GNU General Public License is a free, copyleft license for
software and other kinds of works.
If you accept the terms of the agreement, click I Agree to continue. You must accept the agreement to install Weka 3.8.6.
Nullsoft Install System v3.06.1
< Back
I Agree
Cancel

Exploring WEKA

The Preprocess panel is the start point for knowledge exploration. From this panel you can load datasets and browse the characteristics of attributes.

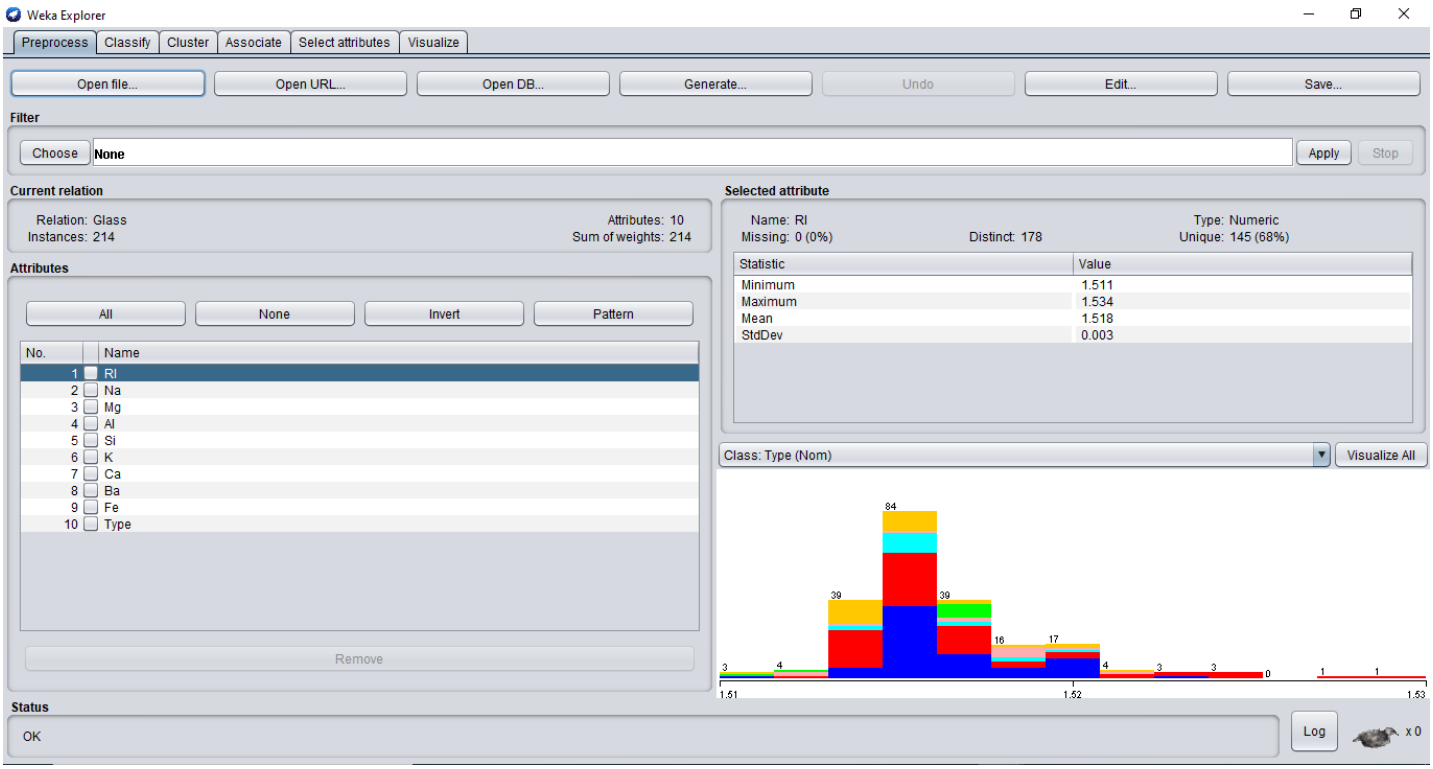


Loading glass.arff



Glass Dataset – arff file

10 Attributes and 214 Instances



Visualize all the attributes

Classifier Panel

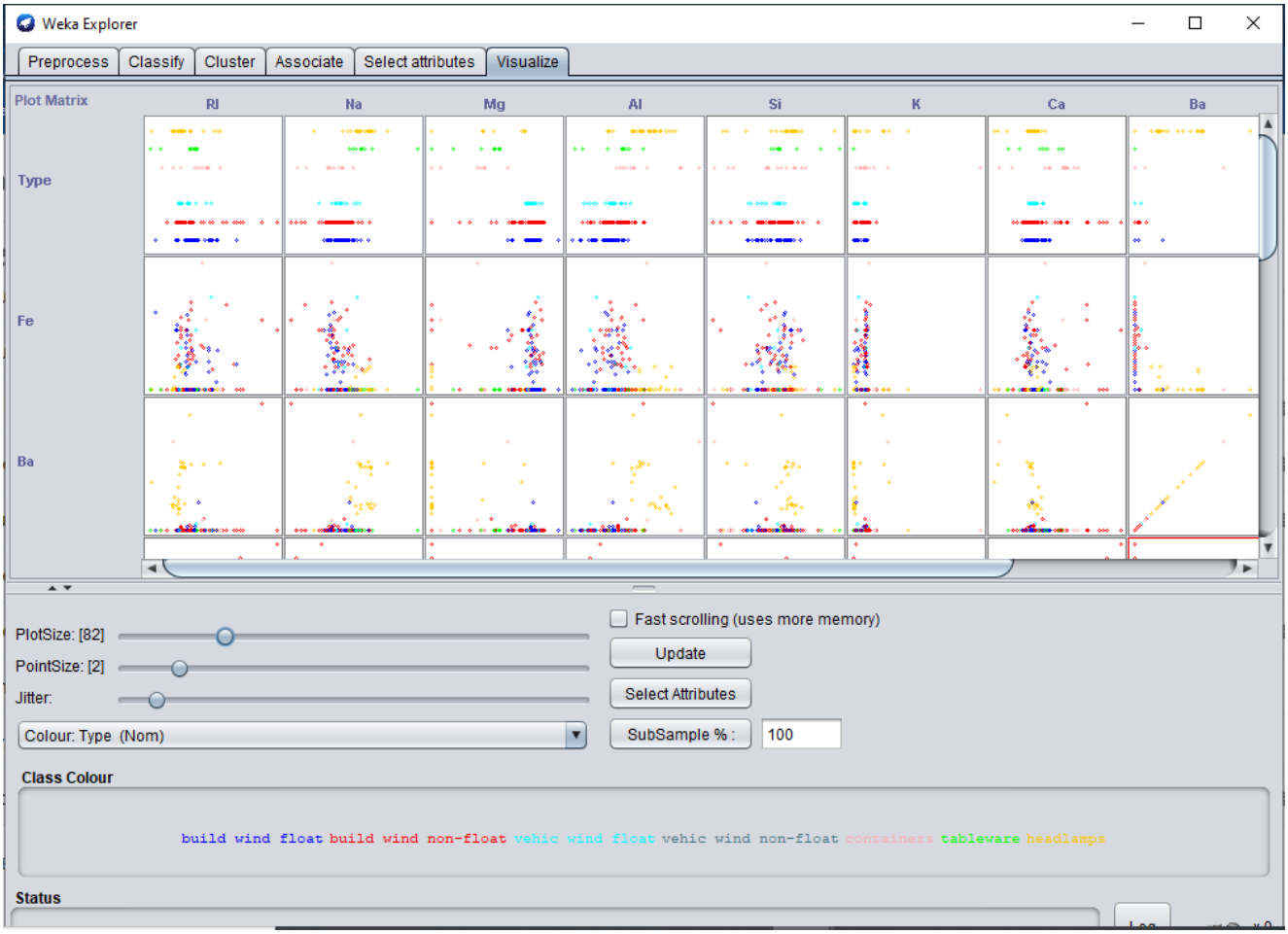
The classifier panel allows you to configure and execute any of the weka classifiers on the current dataset.

Cluster Panel

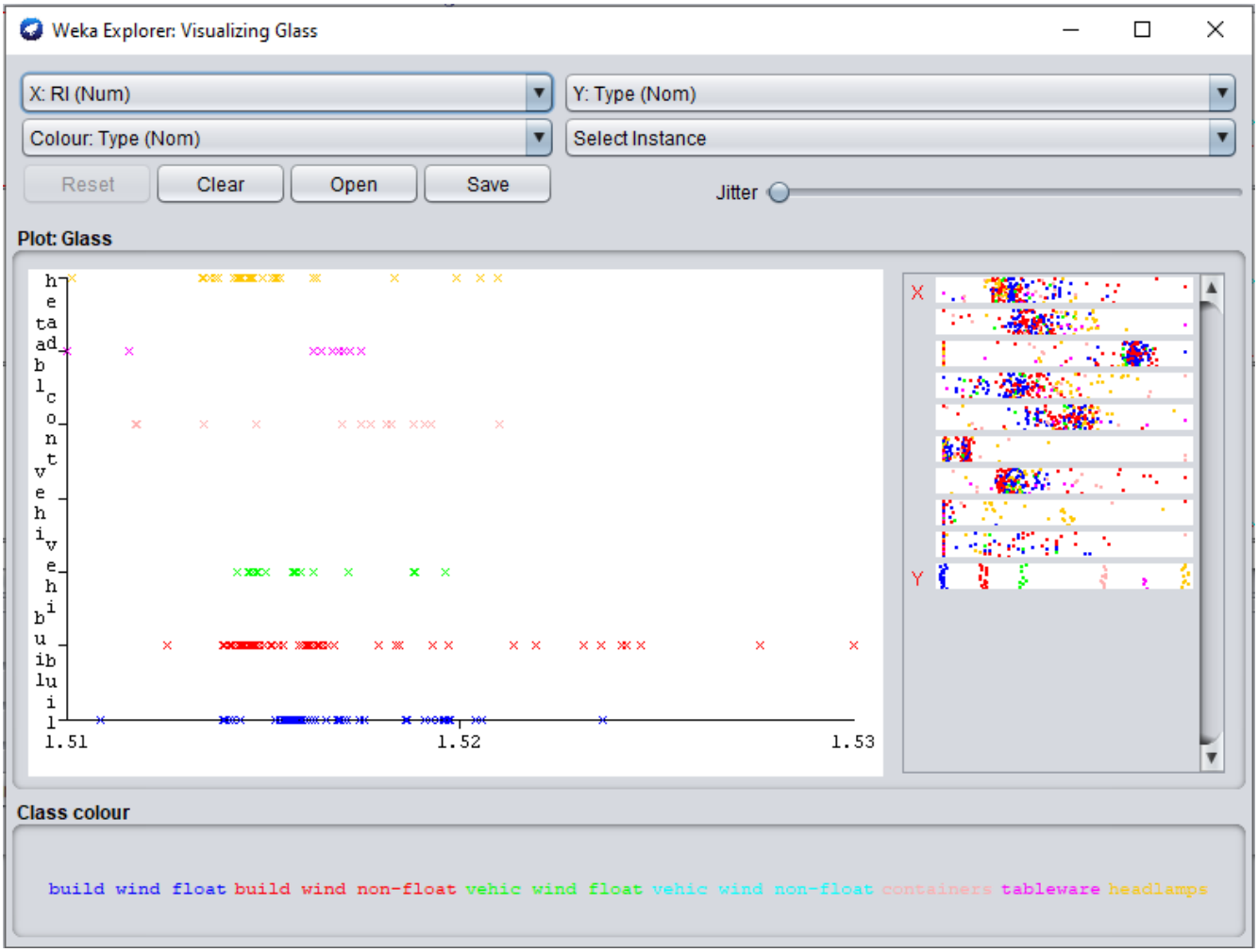
From the cluster panel you can configure and execute any of the weka clusterers on the current dataset.

Visualize Panel

This panel displays a scatter plot matrix for the current dataset. The size of the individual cells and the size of the points they display can be adjusted using the slider controls at the bottom of the panel. The number of cells in the matrix can be changed by pressing the "Select Attributes" button and then choosing those attributes to displayed.



Visualizing the scatter plot for RI vs Type of glass



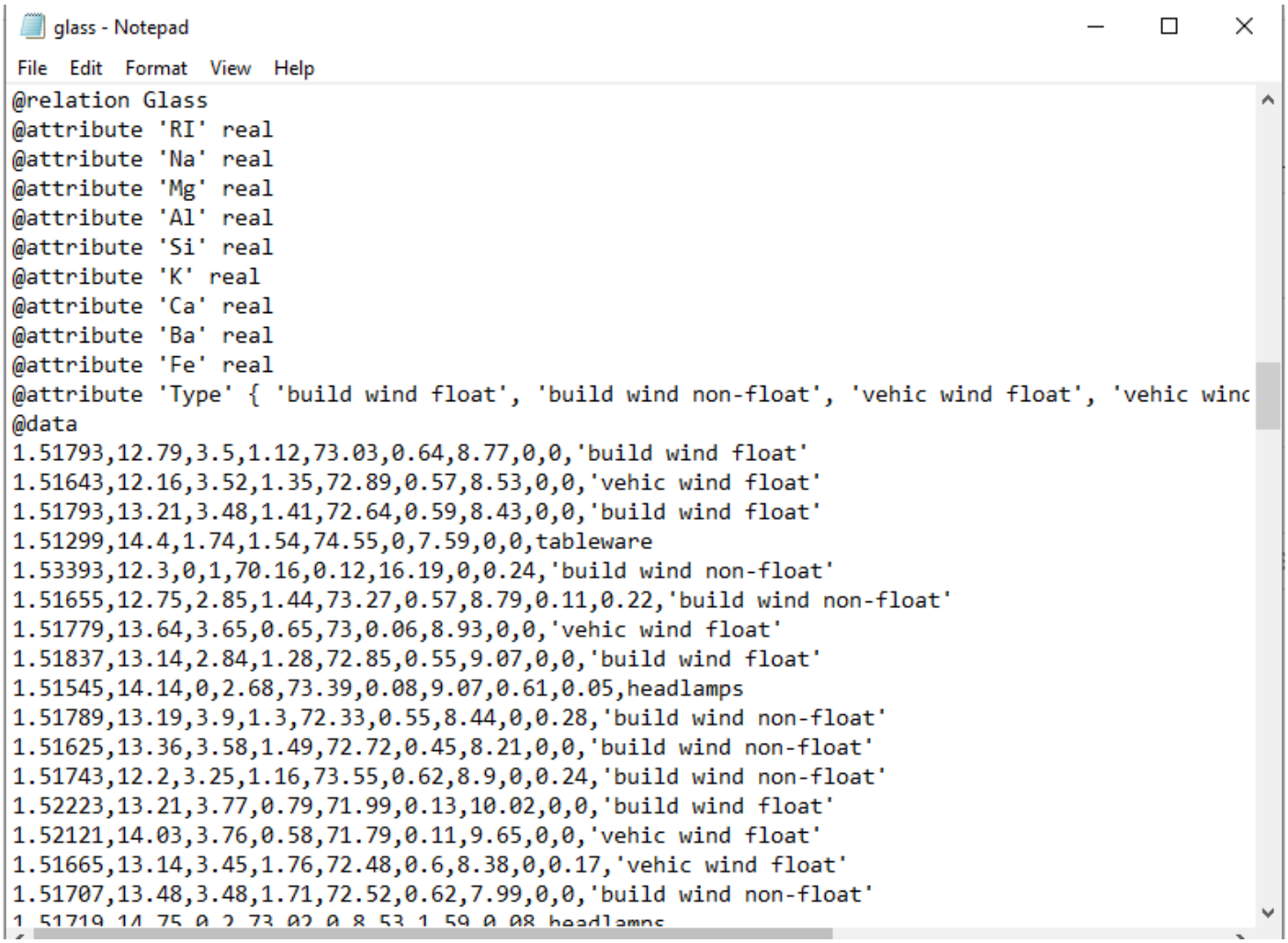
What is arff file format?

ARFF stands for Attribute-Relation File Format. It 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 describes the attribute types.
- The data section contains a comma separated list of data.



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

   Here name=Glass

2. The @attribute tag defines the attributes.
   (RI,Na,Mg,Al,Si,K,Ca,Ba,Fe)
   Target or Class variable is Type

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