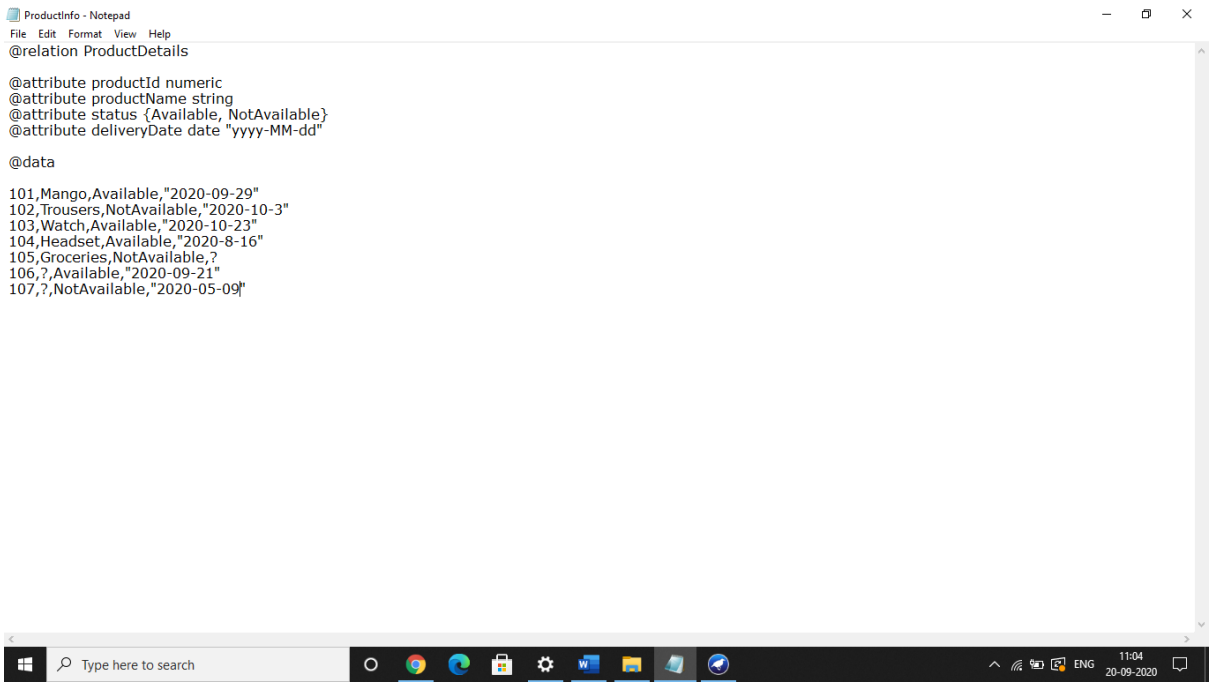
Steps to create ARFF File

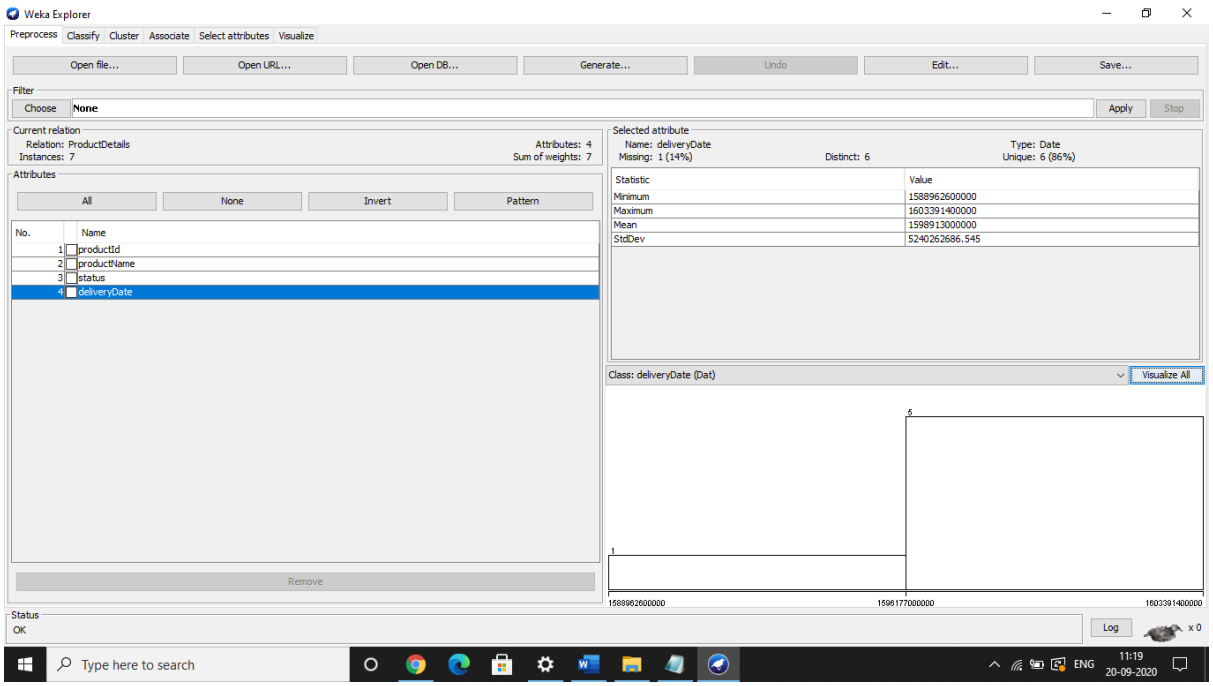
Step1: Open a notepad file.

Step2: Enter the relation, attribute details along with data and save it with .arff extension.



- Table name or relation name is declared as @relation, attributes with @attribute, data with @data.
- Data can be of 4 types **numeric(integers), string(characters), date(YYYY-MM-DD) and nominal(class type).**
- Missing values in the data can be represented with "?" otherwise leads to **premature EOL error.**
- If the date format is violated, an error unparseable date error occurs.

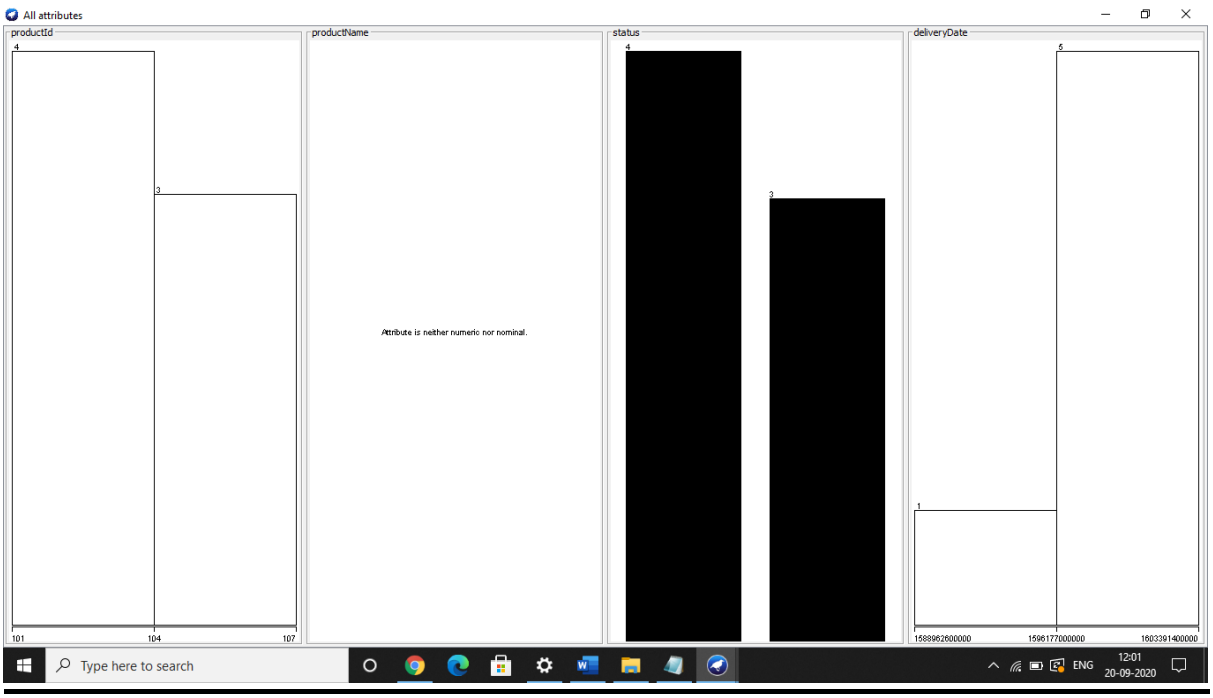
Step3: Open the arff file in weka.



Observations:

- Attributes are productid – numeric, productName – string, status – nominal, deliveryDate – date.
- Number of records – 7
- Missing values deliveryDate - 1(14%), productName – 2(29%)

Histogram:

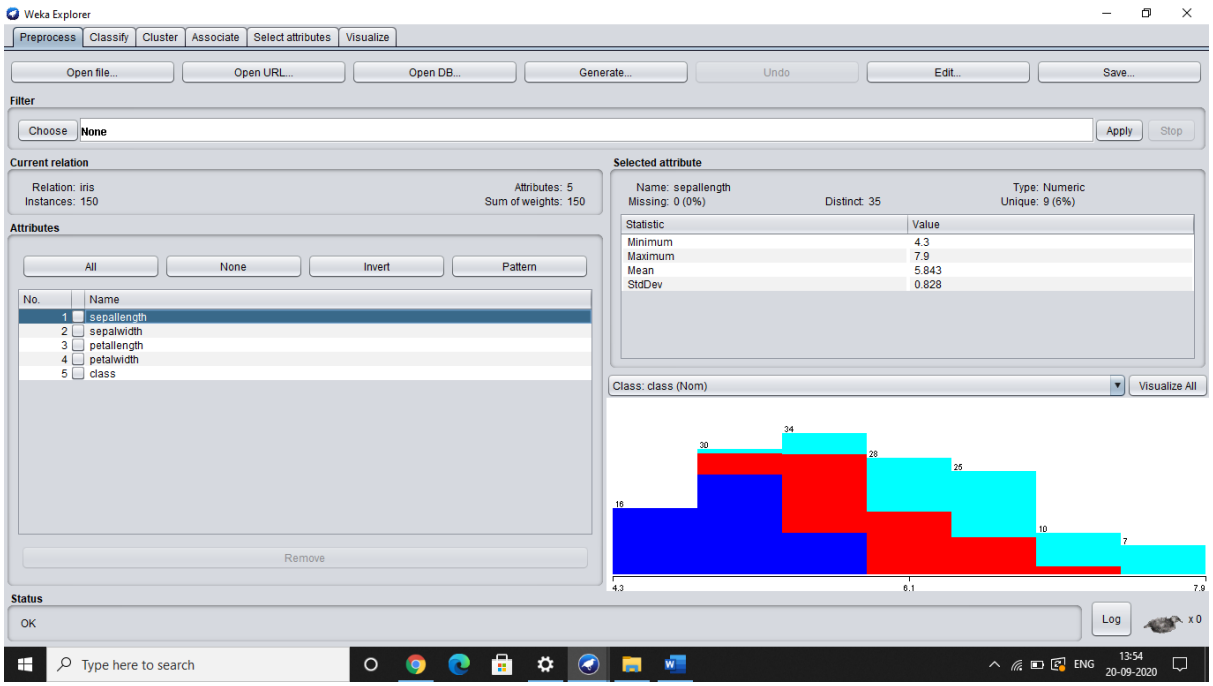


Analyzing the sample datasets

Dataset1: Iris

Steps to open sample dataset:

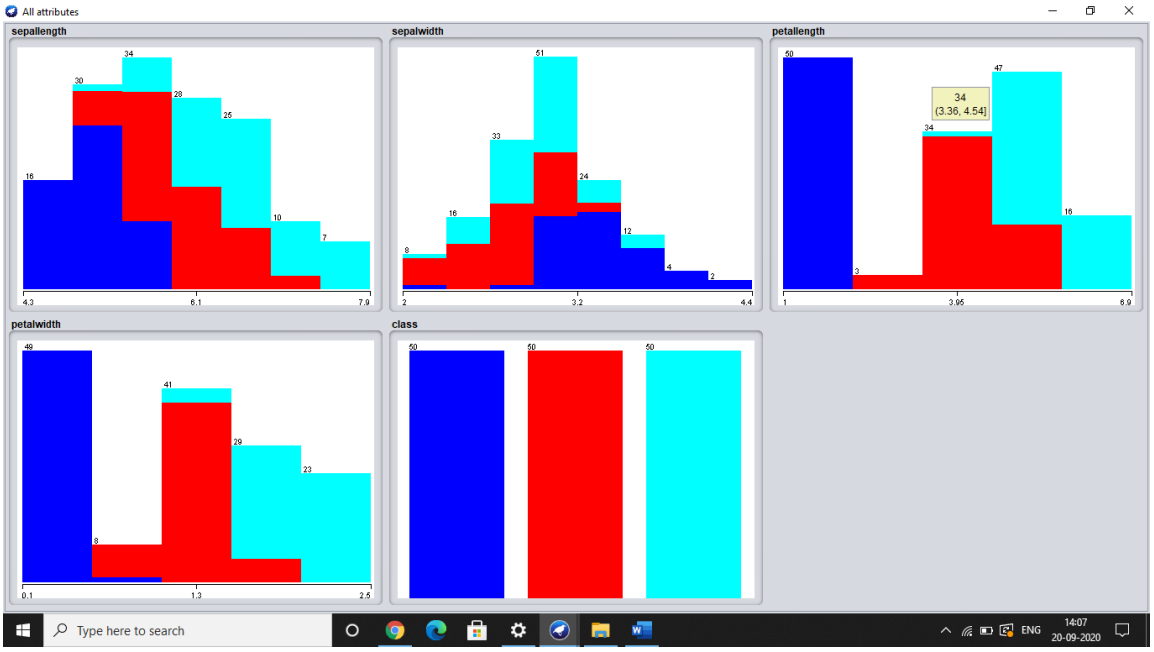
Open weka -> Explorer-> Open file ->Program files->weka -> data-> select the required dataset.



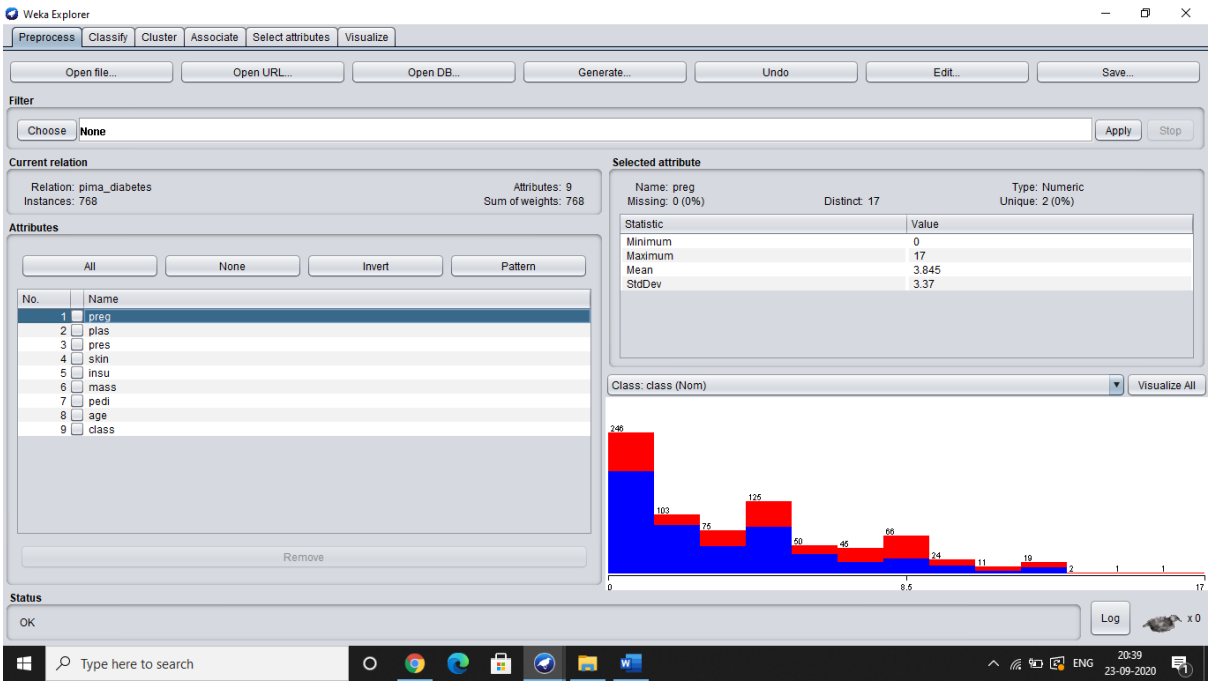
Observations:

- Attributes: Name – type – unique percentage
  - **a)** Sepallength – Numeric – 9
  - **b)** Sepalwidth – Numeric – 5
  - **c)** Petallength – Numeric - 10
  - **d)** Petalwidth – Numeric - 2
  - **e)** Class – Nominal
- Records: 150
- There are no missing values in the dataset.
- It is a data set through which we classify flowers into 3 categories using given information like sepal and petal measurements.

Histogram:


All attributes
sepallength
sepalwidth
petallength
petalwidth
class
34
(3.36, 4.54]
Type here to search
ENG
14:07
20-09-2020

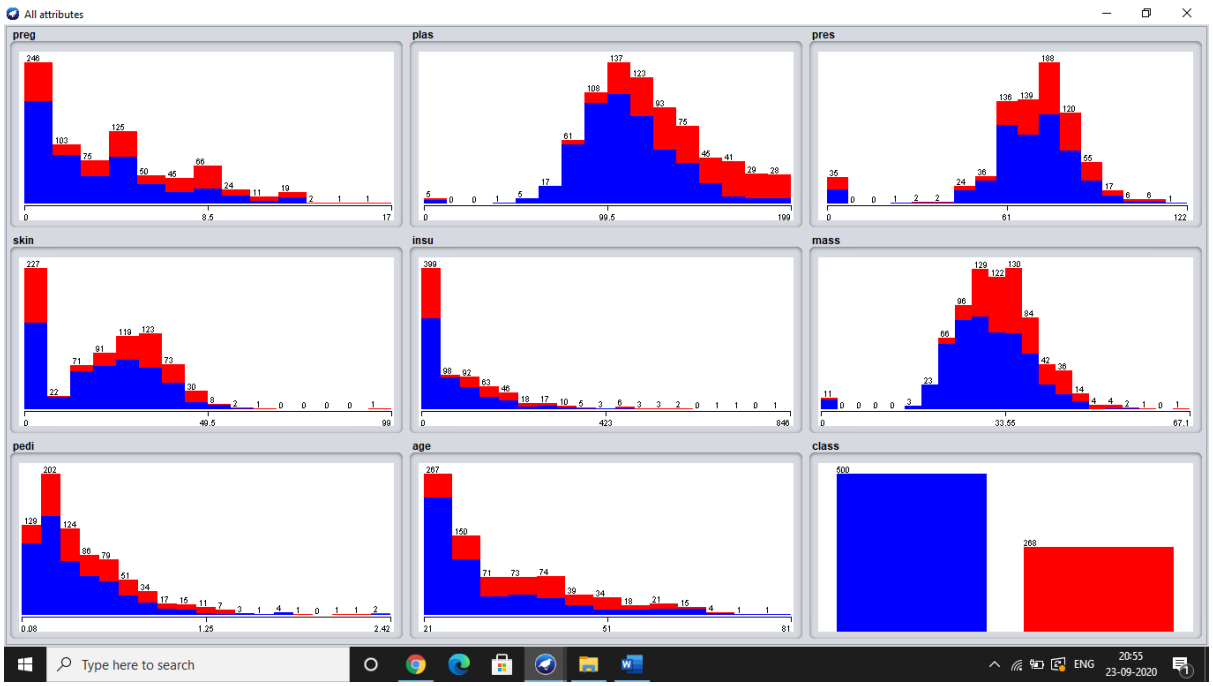
Dataset2: pima_diabetes



Observations:

- **Attributes:** Name – type – unique percentage
  - **a)** Preg – numeric – 2
  - **b)** Plas – numeric - 19
  - **c)** Pres – numeric - 8
  - **d)** Skin – numeric - 5
  - **e)** Insu – numeric - 93
  - **f)** Mass- numeric - 76
  - **g)** Pedi – numeric- 346
  - **h)** Age - numeric – 5
  - **i)** Class – nominal – 0
- **Records:** 768
- There are no missing values in the dataset

Histogram:



Using the dataset we can classify the person into classes of diabetic and non-diabetic