PREPROCESSING:

It is a process of identifying the unwanted data (data cleaning) before loading the data from the data base.

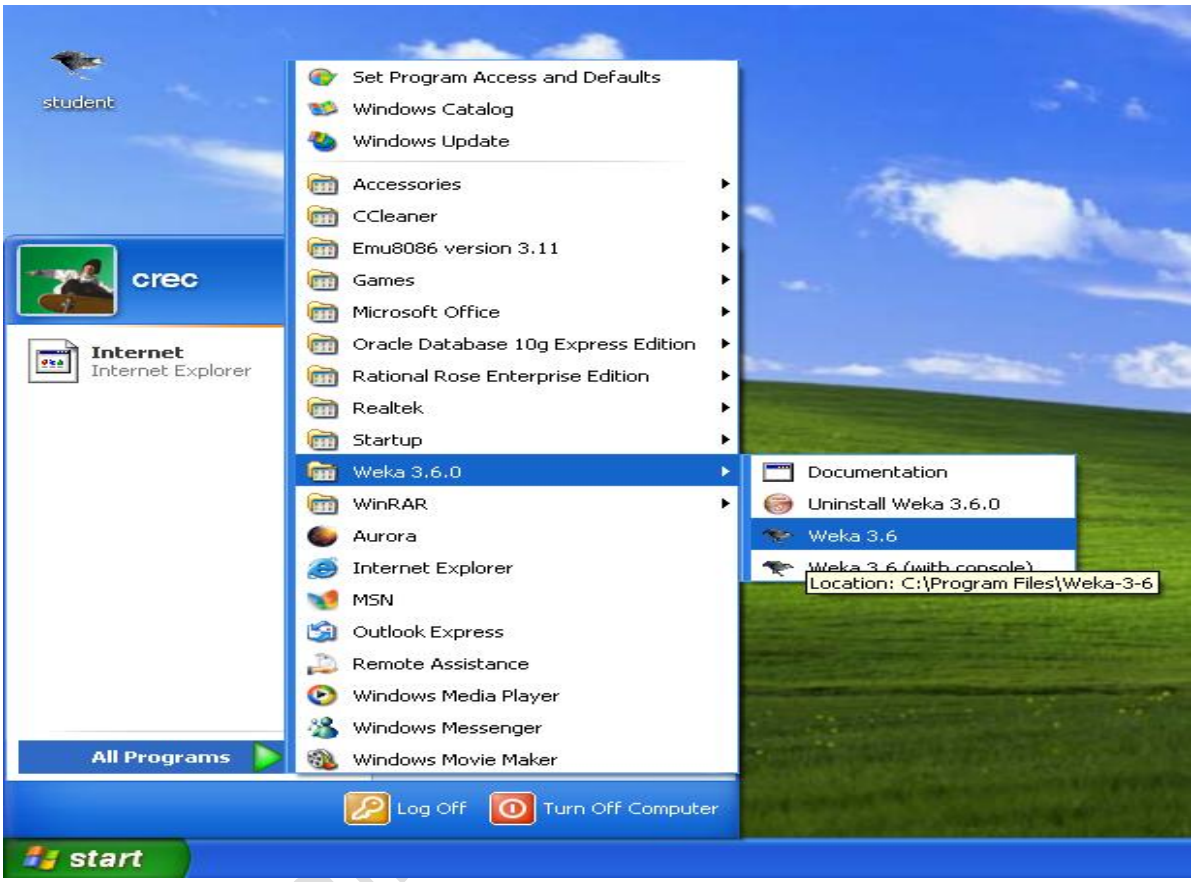

Figure-1

1. Open the WEKA application as shown in the above figure-1

Figure-2

2. Now Click on Explorer as shown in the above figure-2.

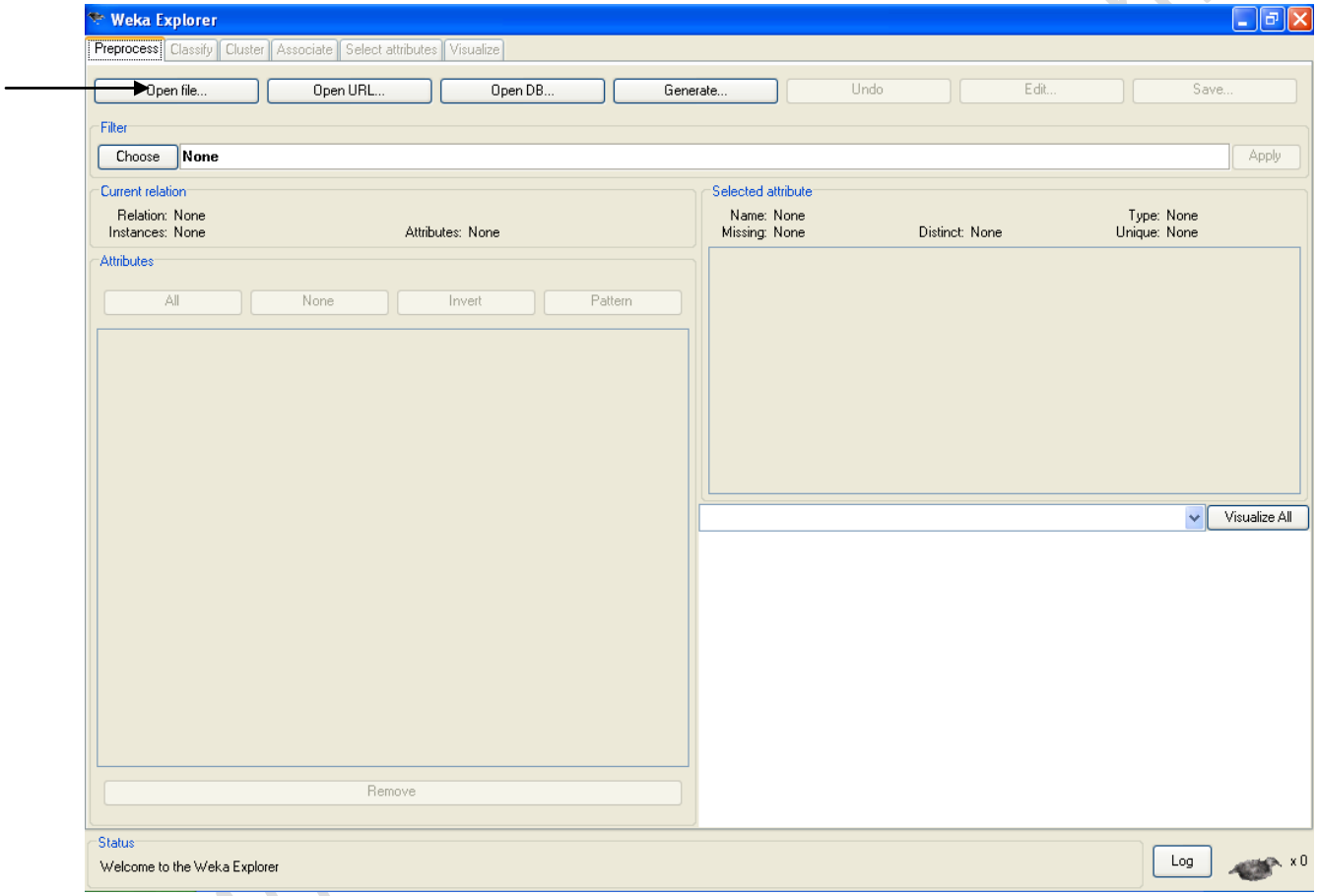

Figure-3

3. Now open file by choosing the “open file” button as shown in the above figure-3.

Relation specifies the name of the database used, instances specify the objects involved, and attributes specify the number of attributes used in the data base or relation.

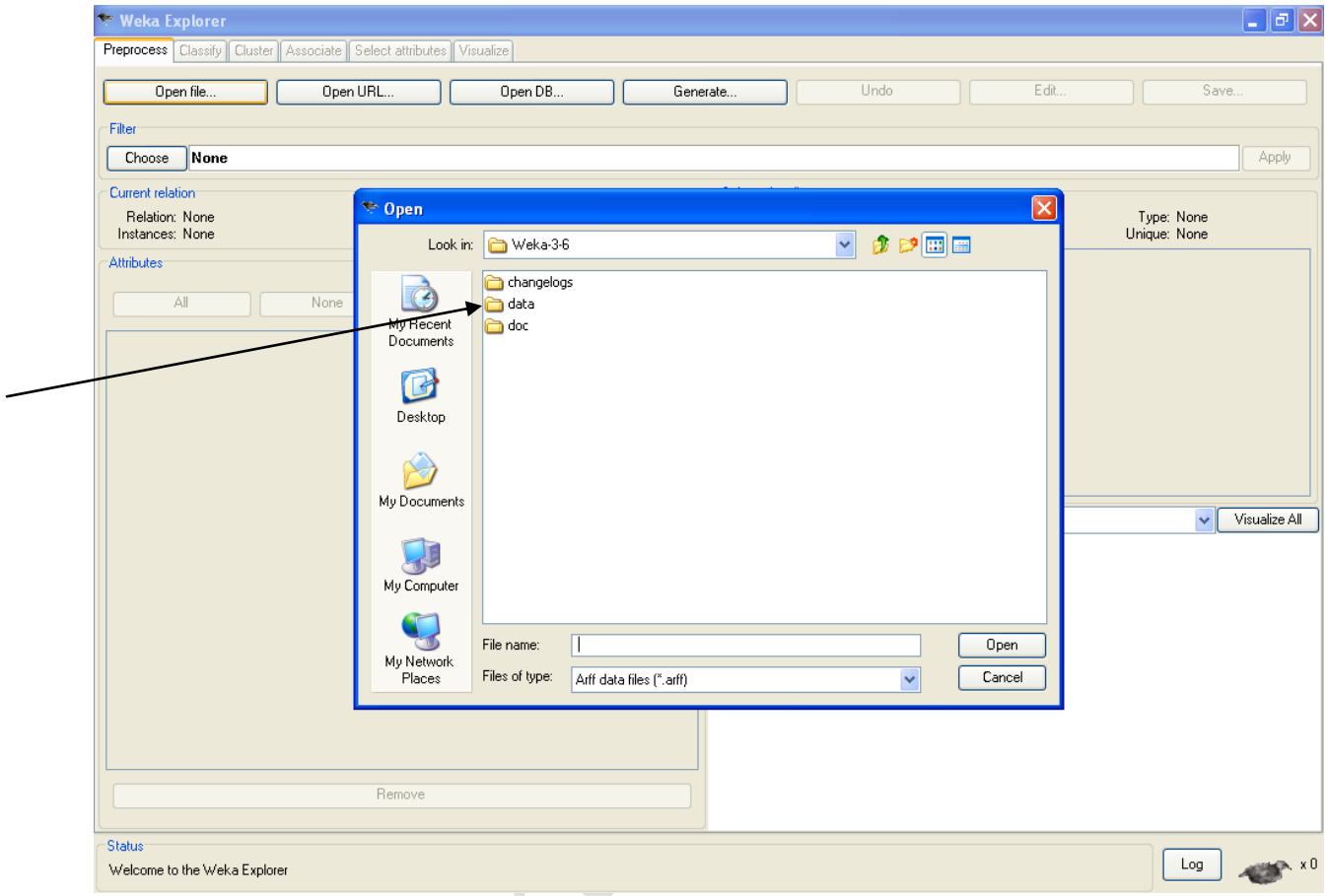

Figure-4

4. Now choose the data folder in the open dialogue box as in figure-4.

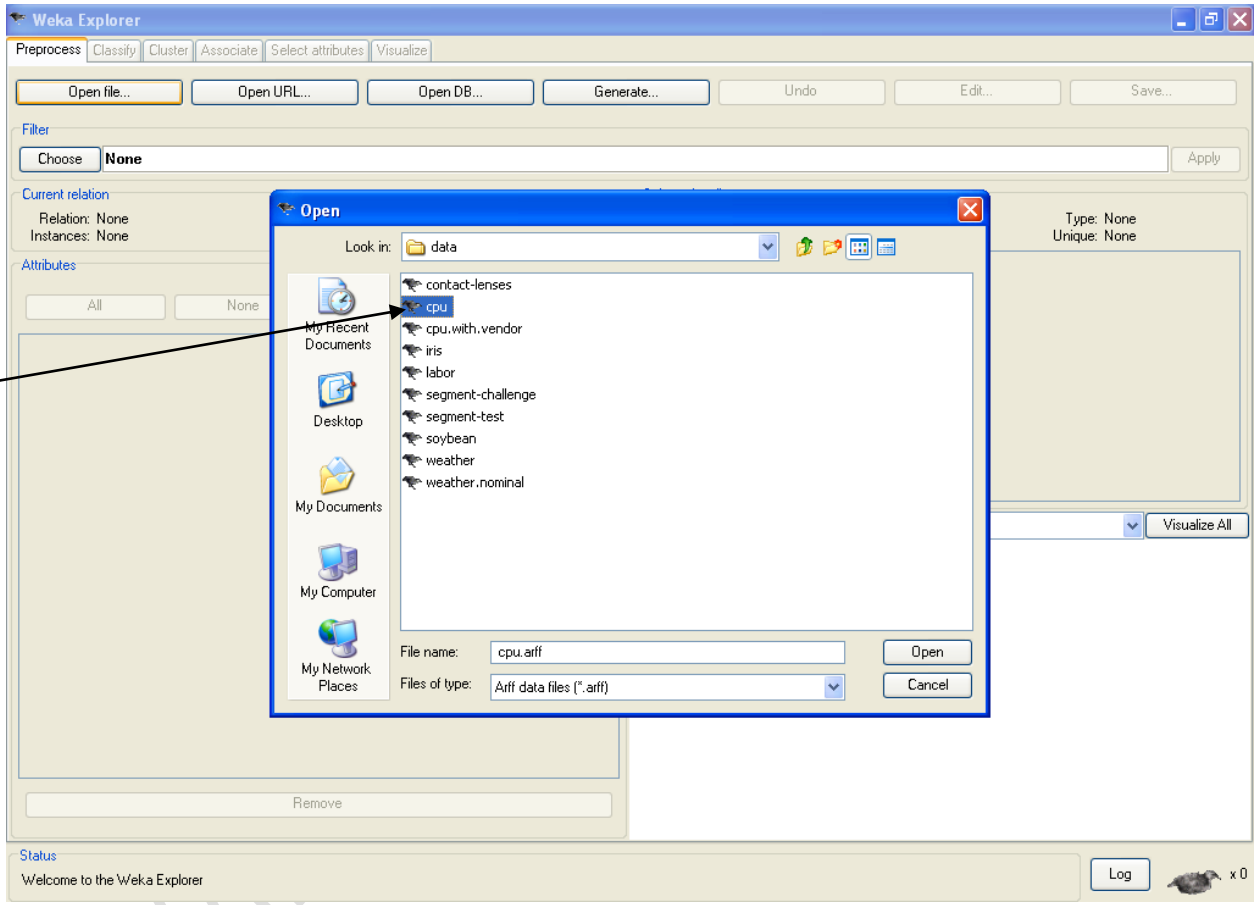

Figure-5

5. Now choose the “cpu.arff” file in the above figure-5.

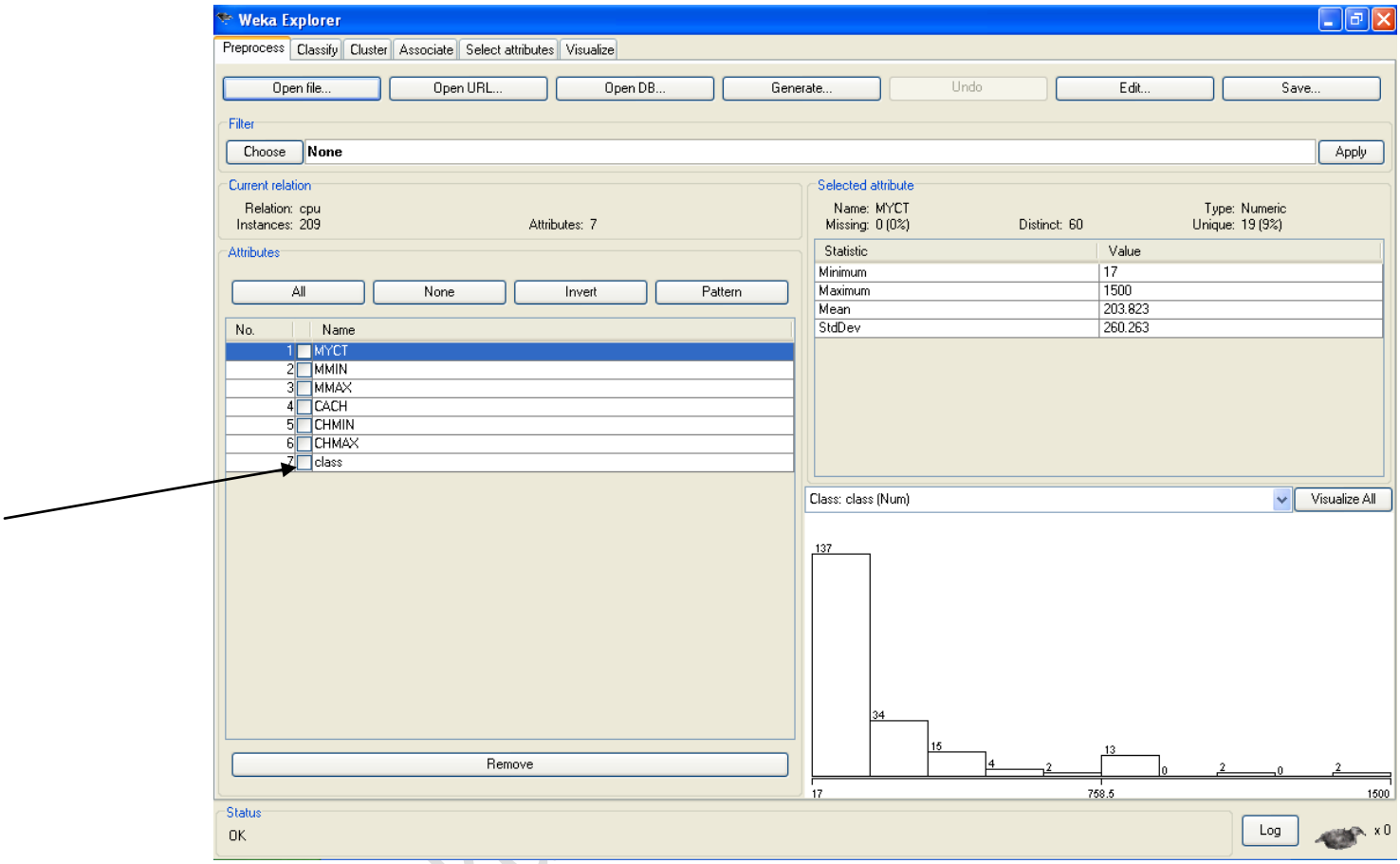

Figure-6

6. Select any attribute in the attributes section (for e.g. class as shown in Figure-6) and click on remove button as shown in the figure-7 below.

   On right Side of the window we have the selected attribute field which specifies the following options like:

   Name: - It specifies the name of the attribute.

   Missing: - It specifies the missing values in the attribute.

   Type: - It specifies the type of attribute used in the database like nominal or numeric or string etc…

   Unique: - It specifies unique attributes available in the relation.

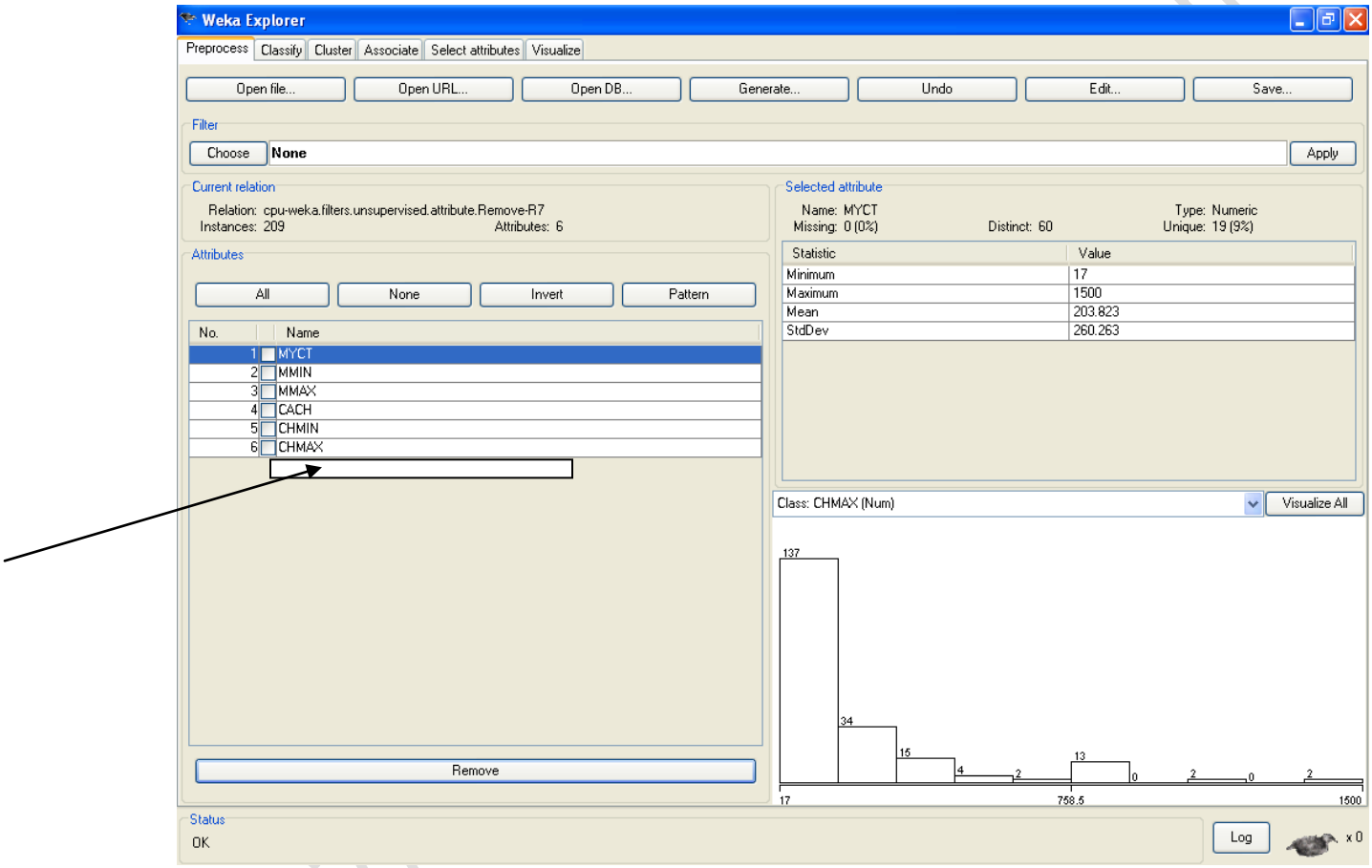

Figure-7

7. The above diagram shows the cpu.arff after removing the attribute “class”.

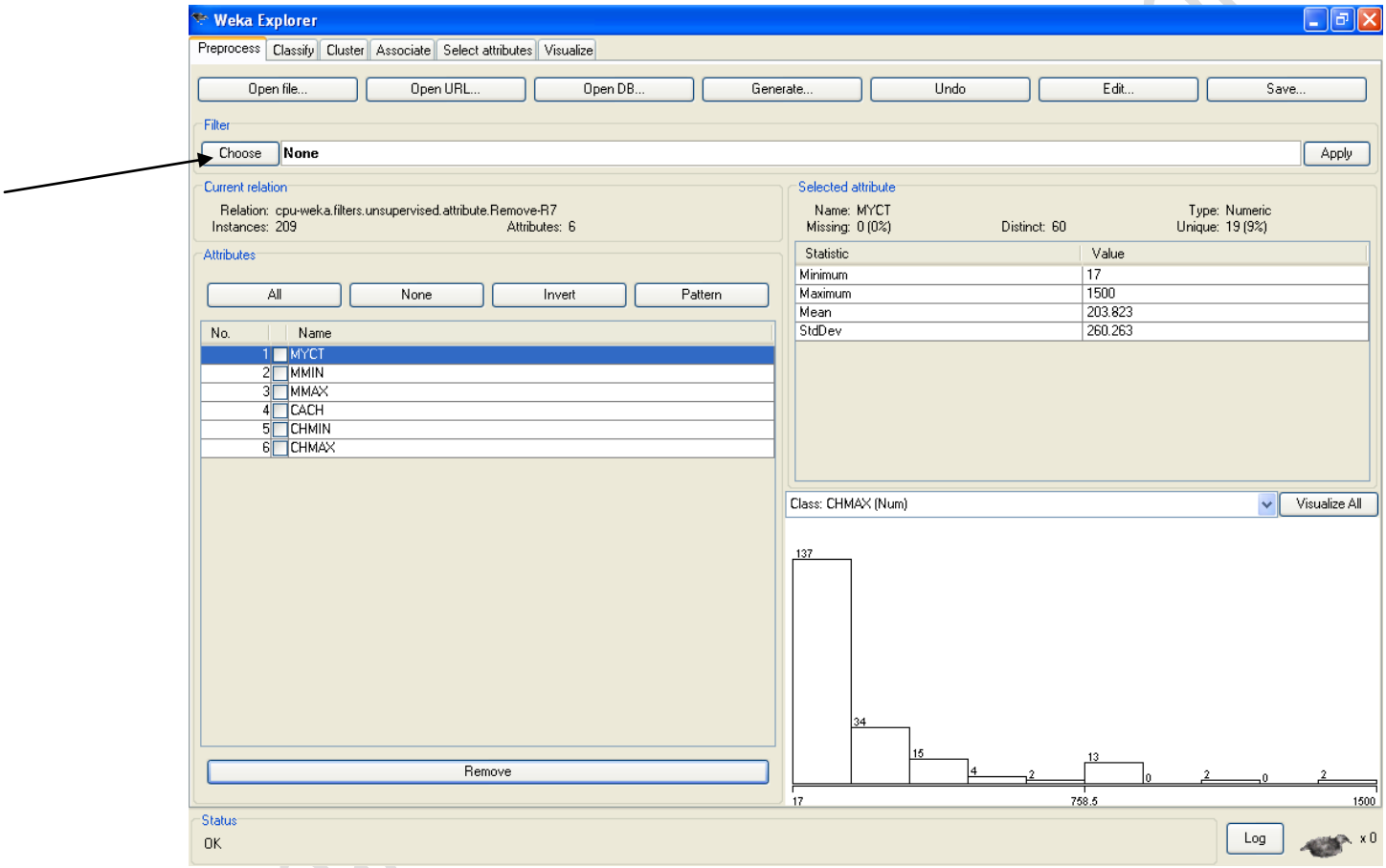

Figure-8

8. Now click on the “choose” button from filter as in figure-8 and expand the “unsupervised” option and select the “Discretize” option, which is as shown in figure-9.

Discretize is used to discretize or convert numeric attributes into nominal ones, based on the class information.

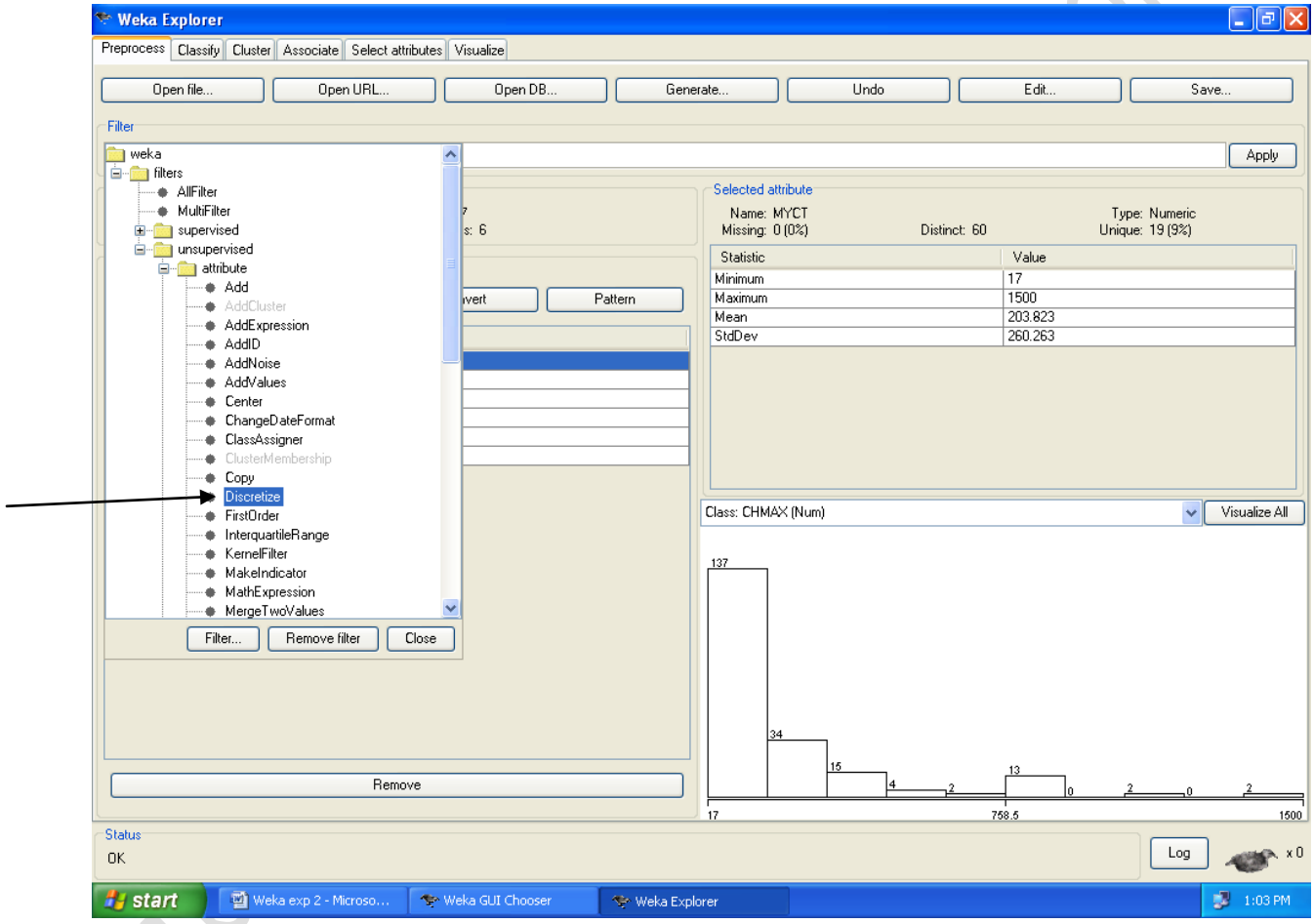

Figure-9

9. Now left click the object to edit the properties or right click and select show properties to edit the properties which is as shown in the figure-10 as follows.

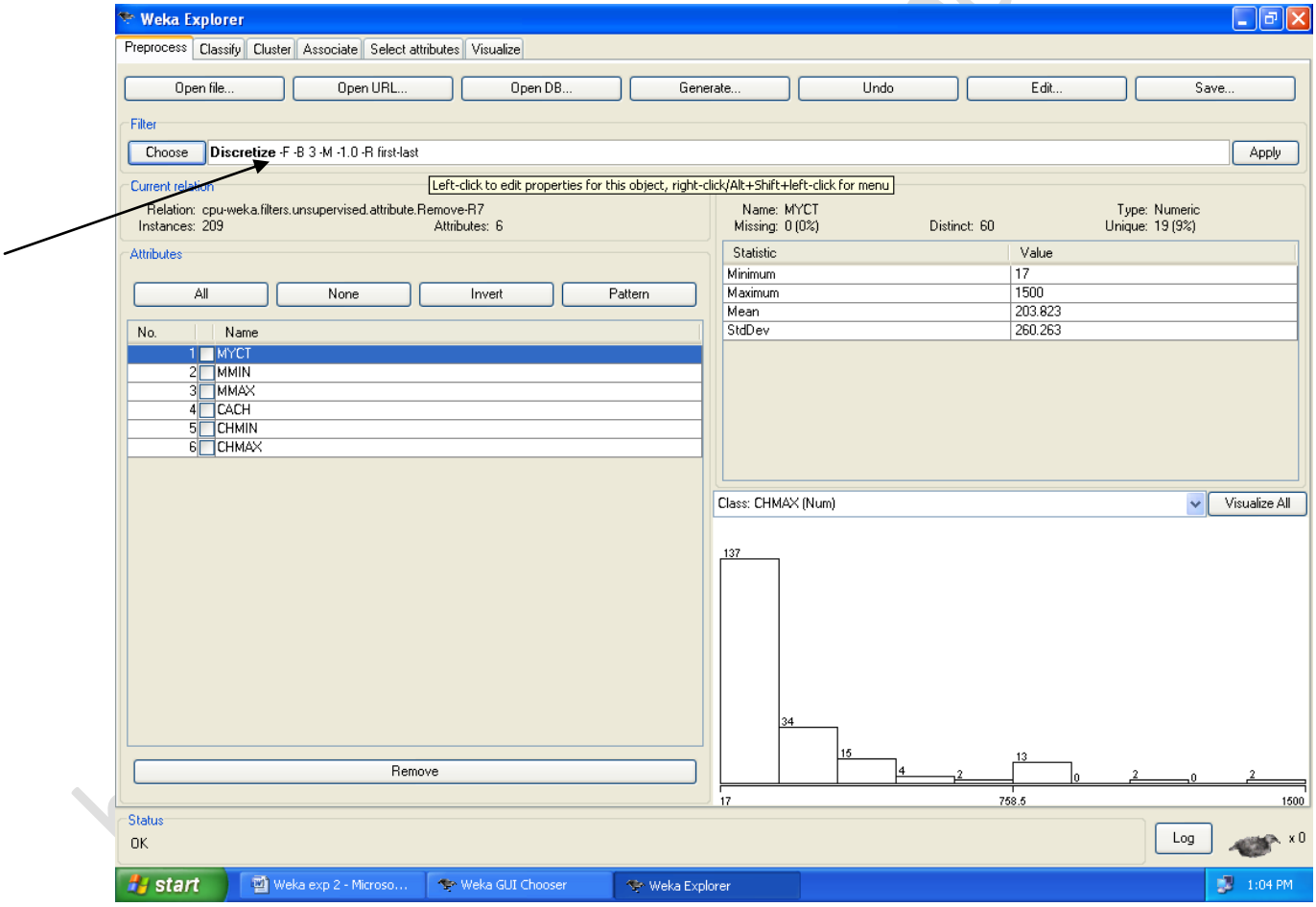

Figure-10

10. In the "GenericObjectEditor" change the bins value to either 2 or 3 or as our desire and make the "useEqualFrequency" option as "TRUE" and click on ok, which is as shown in the figure-11 as follows.

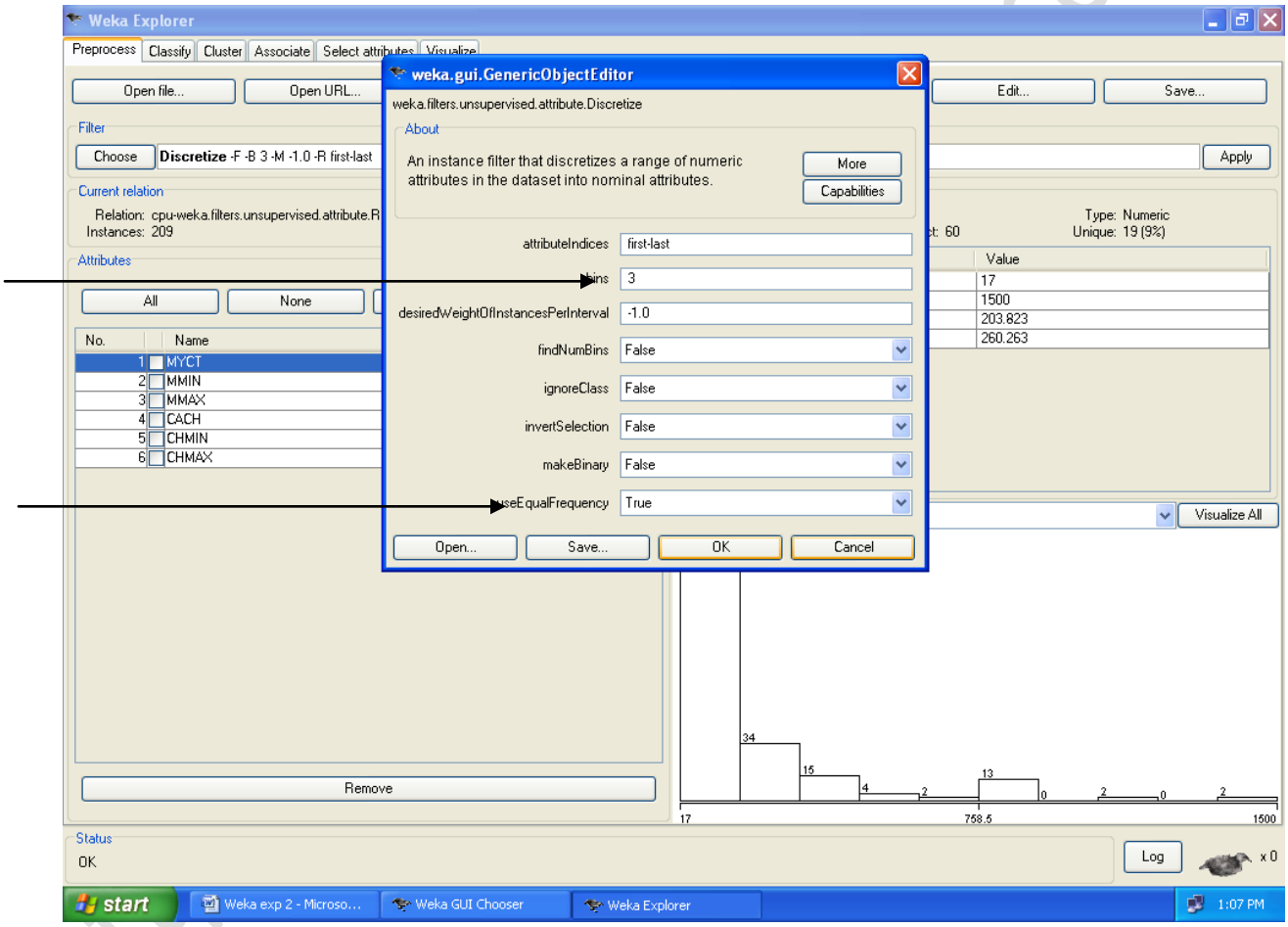

Figure-11

11. Now apply the properties by clicking on “apply” button in the filter where the Discretize object contains 3 bins and useEqualFrequency is set to TRUE. After applying the properties to the object the result obtained is as shown in the following figure-12.

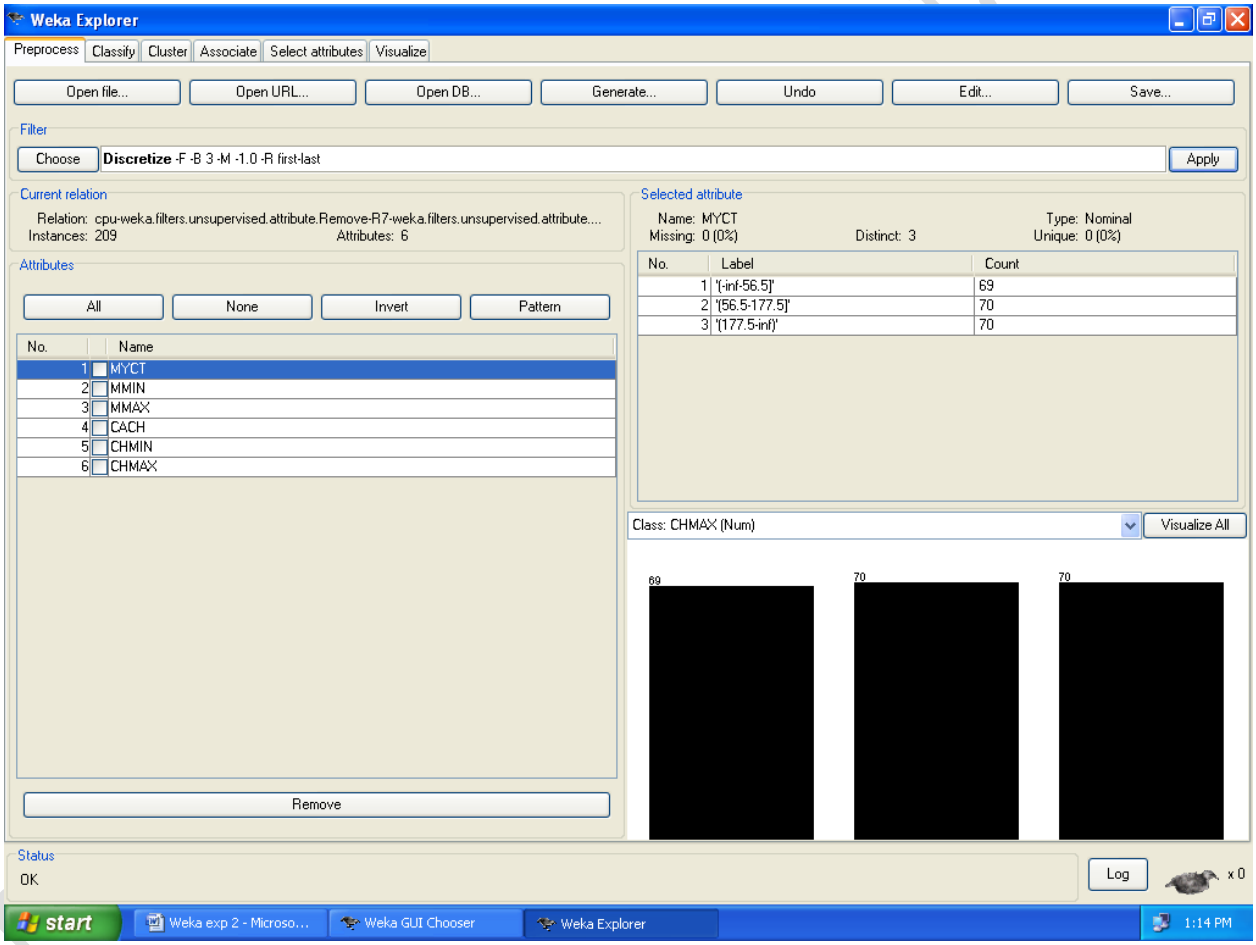

Figure-12

12. We can observe the change in the result in the visualize which is as shown in the figure-12.

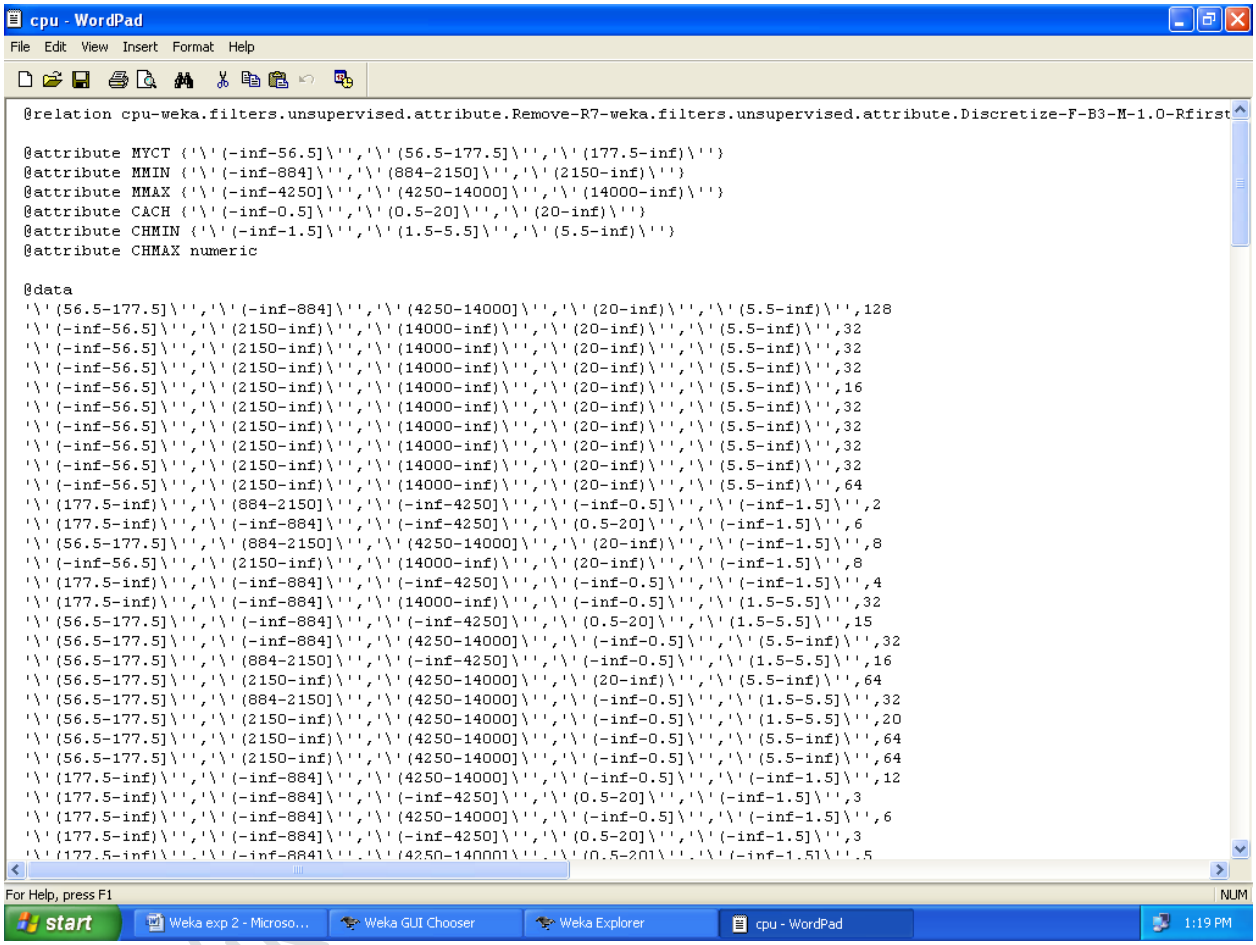

Figure-13

13. Finally the result is obtained as shown in the figure-13.