Tecnológico de Monterrey

Campus Querétaro

Cinthia Valdéz Guillen A01203141
Cristian Alexsis González Jaime A01204478
Luis Pablo Morales Cruz A01205724

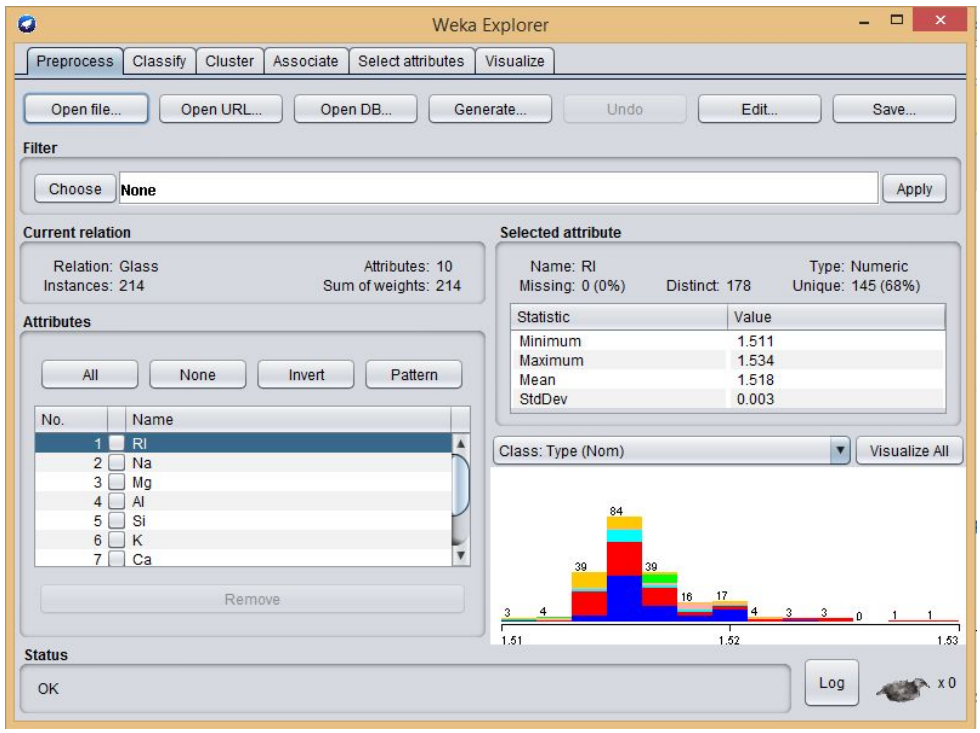
Report Lab05 Using Data Mining Tools WEKA

Delivery date: 09/11/2016

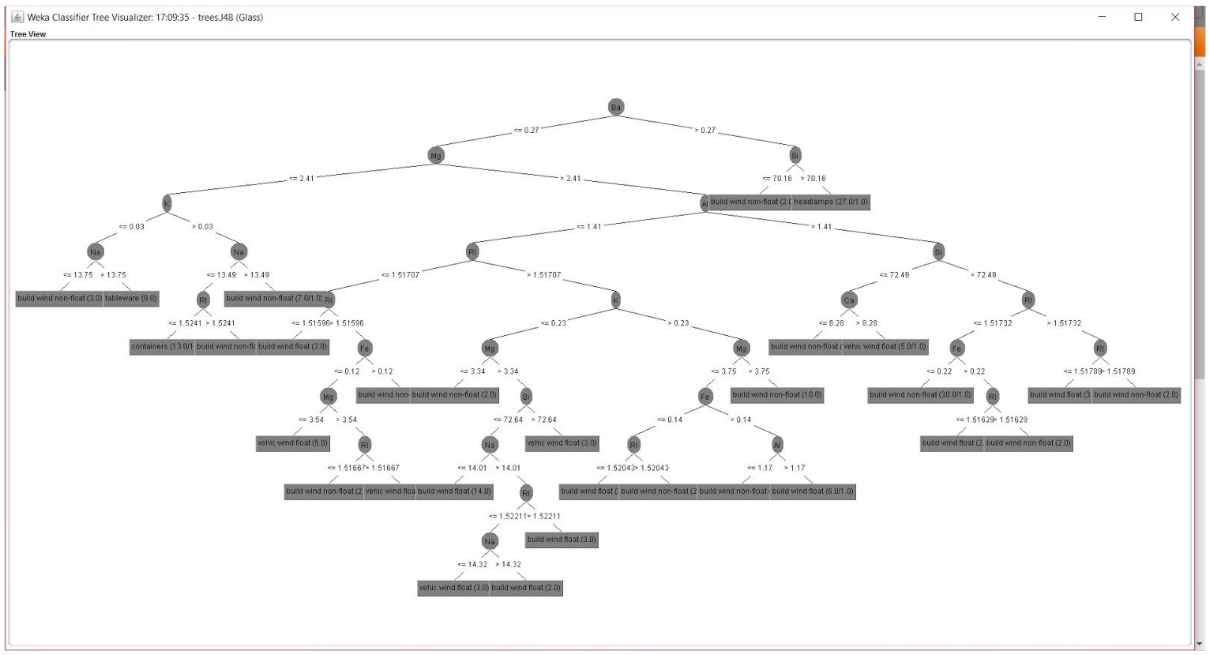
So on this lab we have to develop a ID3, so we chose C++ as a programing language the big challenge here was the generation of the tree, comparing to weka, the challenge was to know the correct format of the data in order to be recognized by weka software, we always say that based on the problem choose a solution, but for this time we have to say that WEKA is better, why? First of all with a tutorial of 1hr of reading you can achieve the same and even more than 13 hrs of implementation, weka have a graphical display, our implementation don't have that feature, at the beginning weka could look very hard to understand but after you read the tutorial everything get clear and weka allow us to do several standard data mining tasks like clustering, data processing, classification, regression, visualization ,etc.

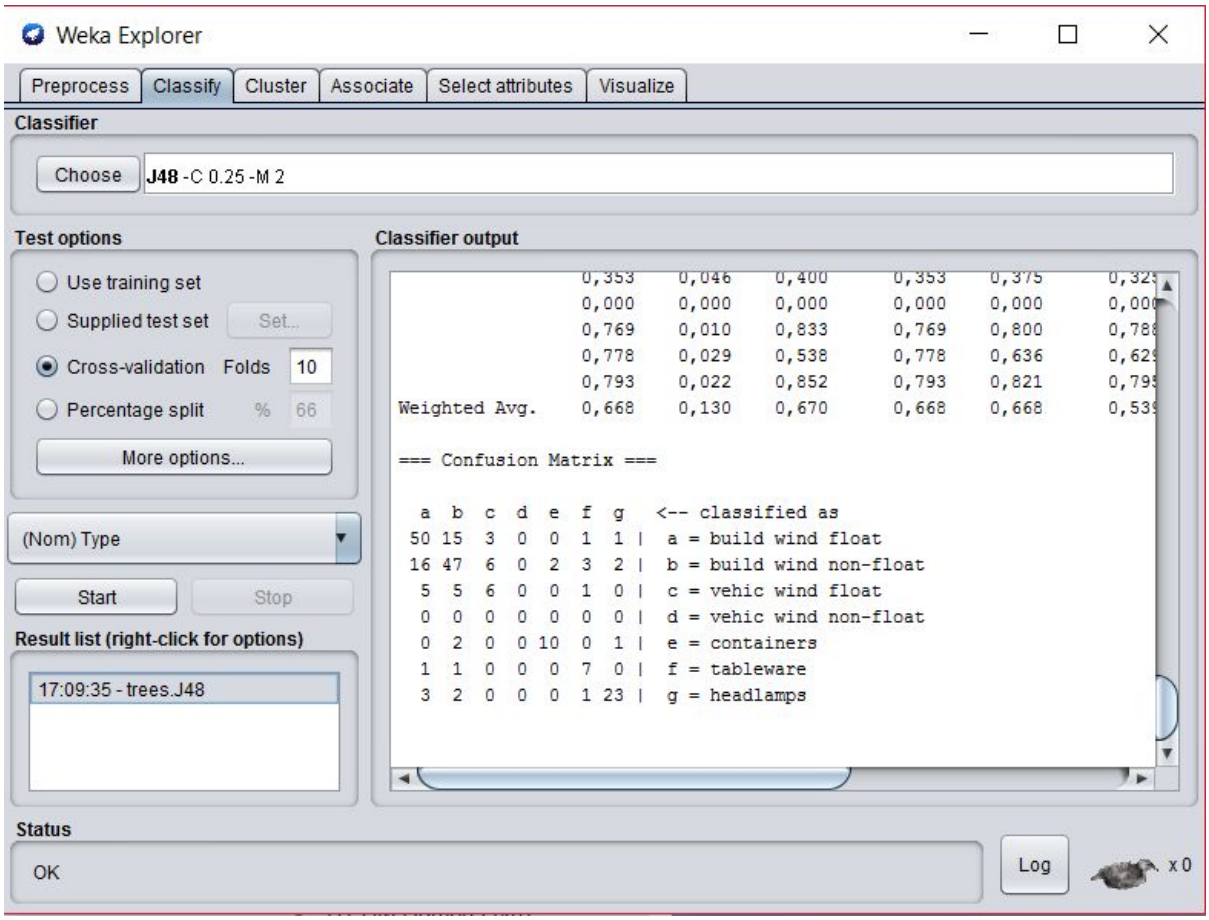
Another advantage for the use of weka is that you will have the basic functions for data mining and you will be sure that the data analysis will be ok, but a disadvantage could be that may need an specific function that weka doesn't support.

If we could have more time for the develop of our program there could be an infinite of advantages because we could set that functions that we specific need and this takes time and skills that maybe we don't have and if it is not tested the program could have some errors.



```
c100tias@c100tias-VirtualBox:~/Documentos/Lab_05$ ./id3 < testcases/2.in
A: TRUE
  ANSWER: TRUE
A: FALSE
  B: TRUE
    ANSWER: TRUE
  B: FALSE
    ANSWER: FALSE
```





For the data set for testing our program we use the ones that were given by the teacher Ruben in the following link https://github.com/rhomeister/TC2011_lab_id3/tree/master/testcases , and for weka we have to pick a data set from http://archive.ics.uci.edu/ml/datasets.html and we adapt it with the format that weka need (file will be atached on file), let be clear that most of the data sets are so big and in most of the times the data is not clear at all, so we decided to choose a data set that was more simple and of course a data that meets the specifications of weka for ID3 algorithm (boolean values and ordered values).

So for final both softwares do the same the only difference is that weka shows you in a graphical way the result and have other options that the program we develop like edit the data set and it have filters for the data,etc, of course weka is better so as decision trees allows for forward and backward calculation paths to happen and hence the choice of the correct decision to take, it is obvious that decision trees are very helpful in every place where making a decision is critical, like a company or the stock exchange.