

Making predictions on new data using Weka

Daniel Rodríguez
daniel.rodriiguezg@uah.es
University of Alcala

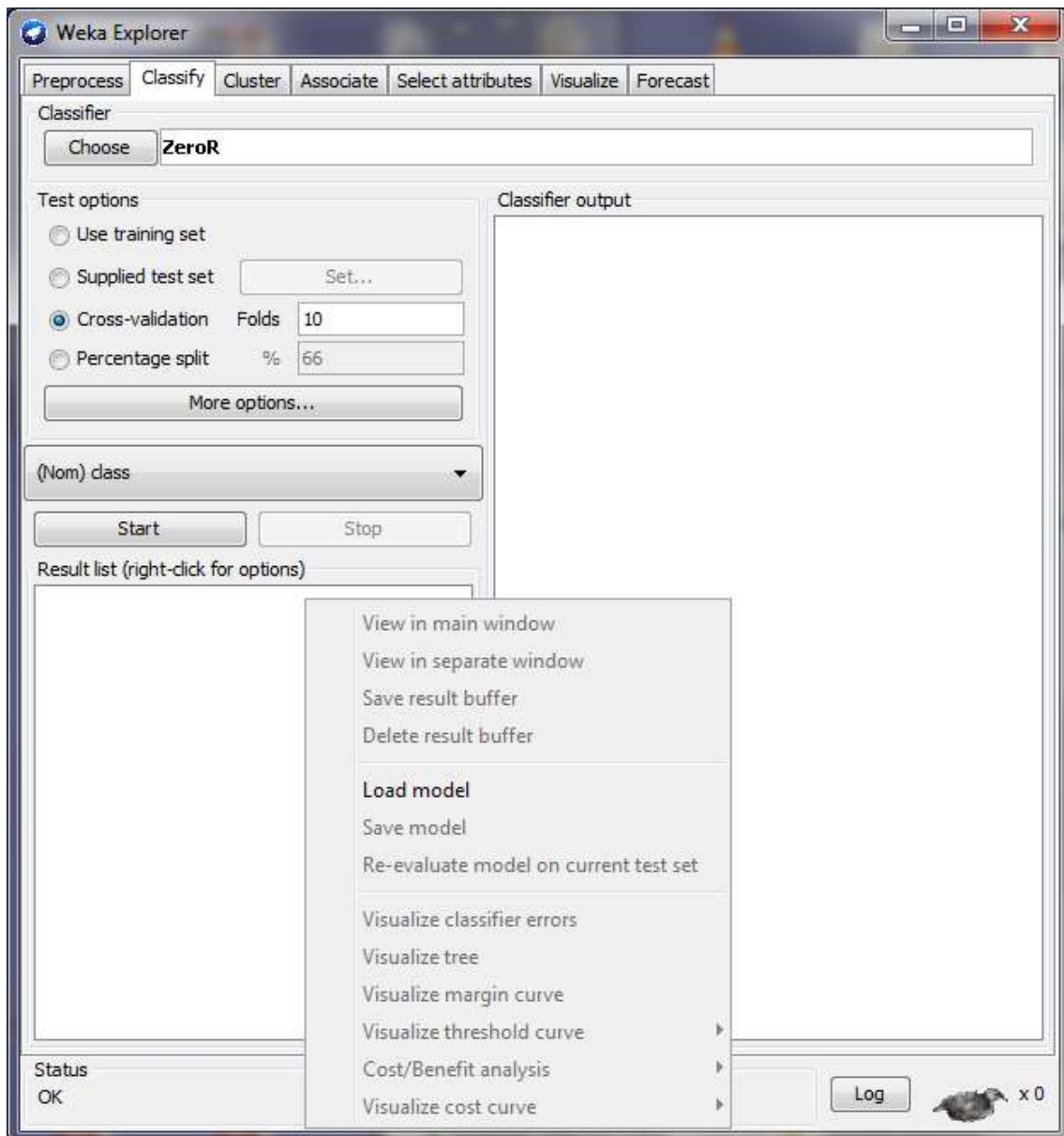
Once we have learned a model, it can be used to classify new unseen data. These notes describe the process of doing some both graphically and from the command line.

First, the file with cases to predict needs to have the same structure that the file used to learn the model. The difference is that the value of the class attribute is “?” for all instances (question marks represent missing values in Weka). For example assuming that we have learnt a decision tree using the diabetes datasets included in Weka, the following file will be used to predict the 5 cases included in the ar-ff file:

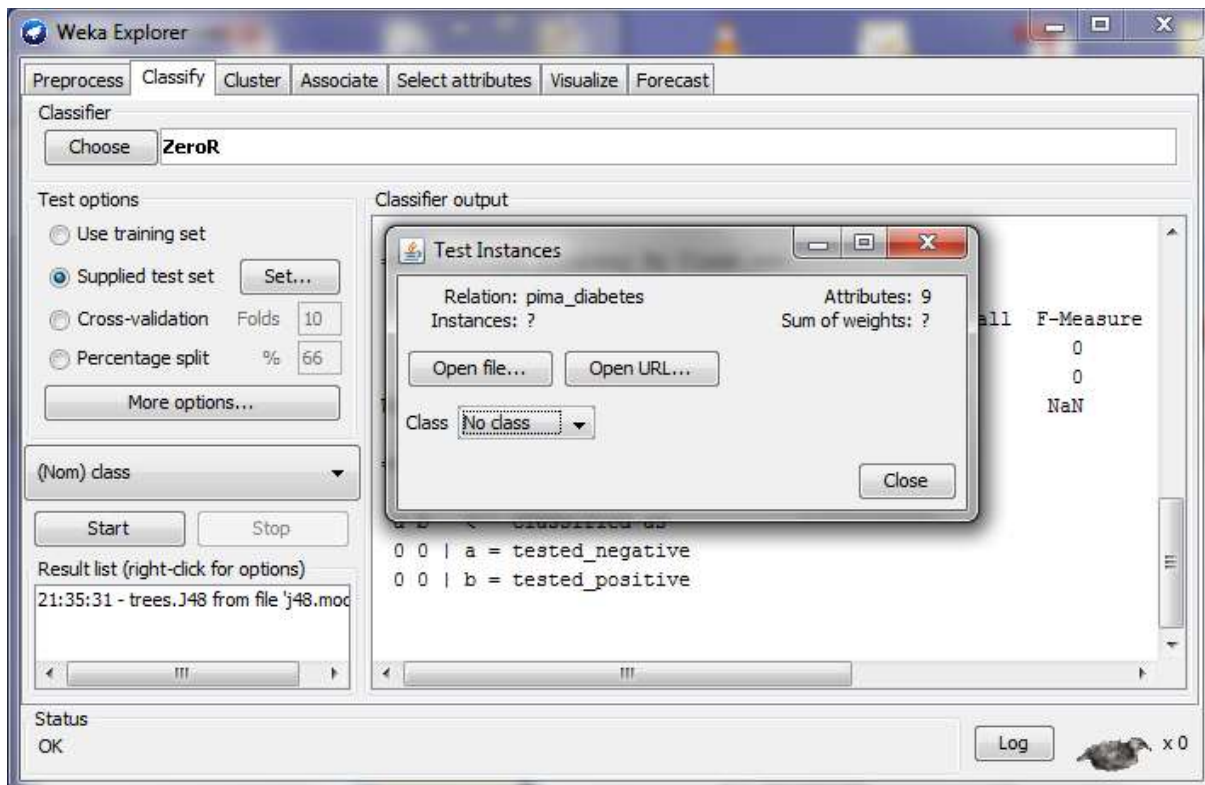
```
@relation pima_diabetes
@attribute 'preg' real
@attribute 'plas' real
@attribute 'pres' real
@attribute 'skin' real
@attribute 'insu' real
@attribute 'mass' real
@attribute 'pedi' real
@attribute 'age' real
@attribute 'class' { tested_negative, tested_positive}
@data
6,148,72,35,0,33.6,0.627,50,?
1,85,66,29,0,26.6,0.351,31,?
8,183,64,0,0,23.3,0.672,32,?
1,89,66,23,94,28.1,0.167,21,?
0,137,40,35,168,43.1,2.288,33,?
```

Using Weka's Explorer

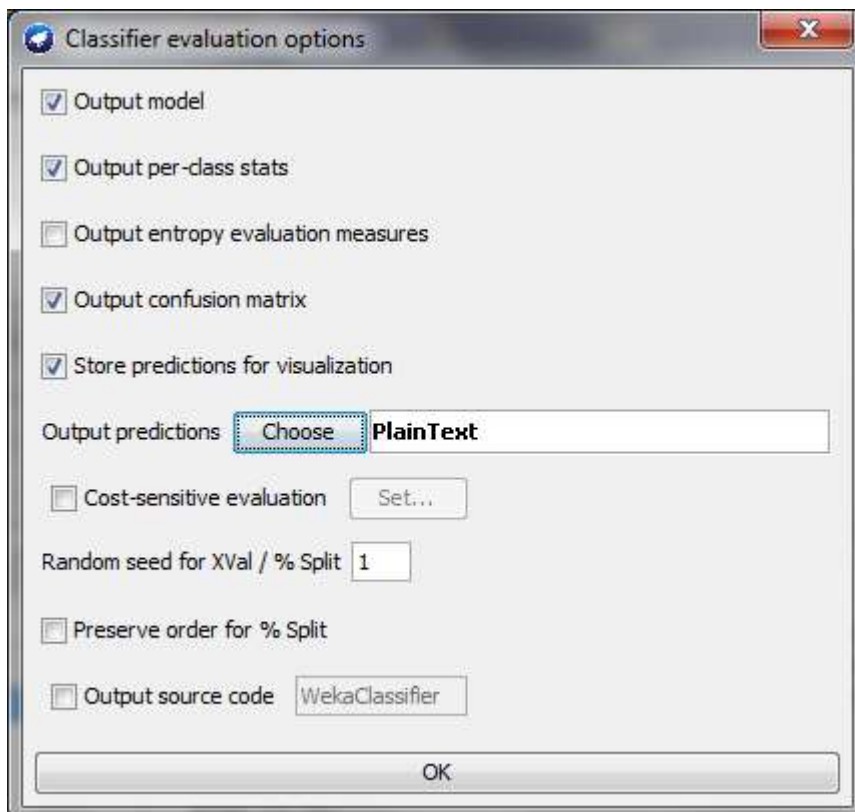
First, we load the saved model with the right click menu on the "Result list" panel:



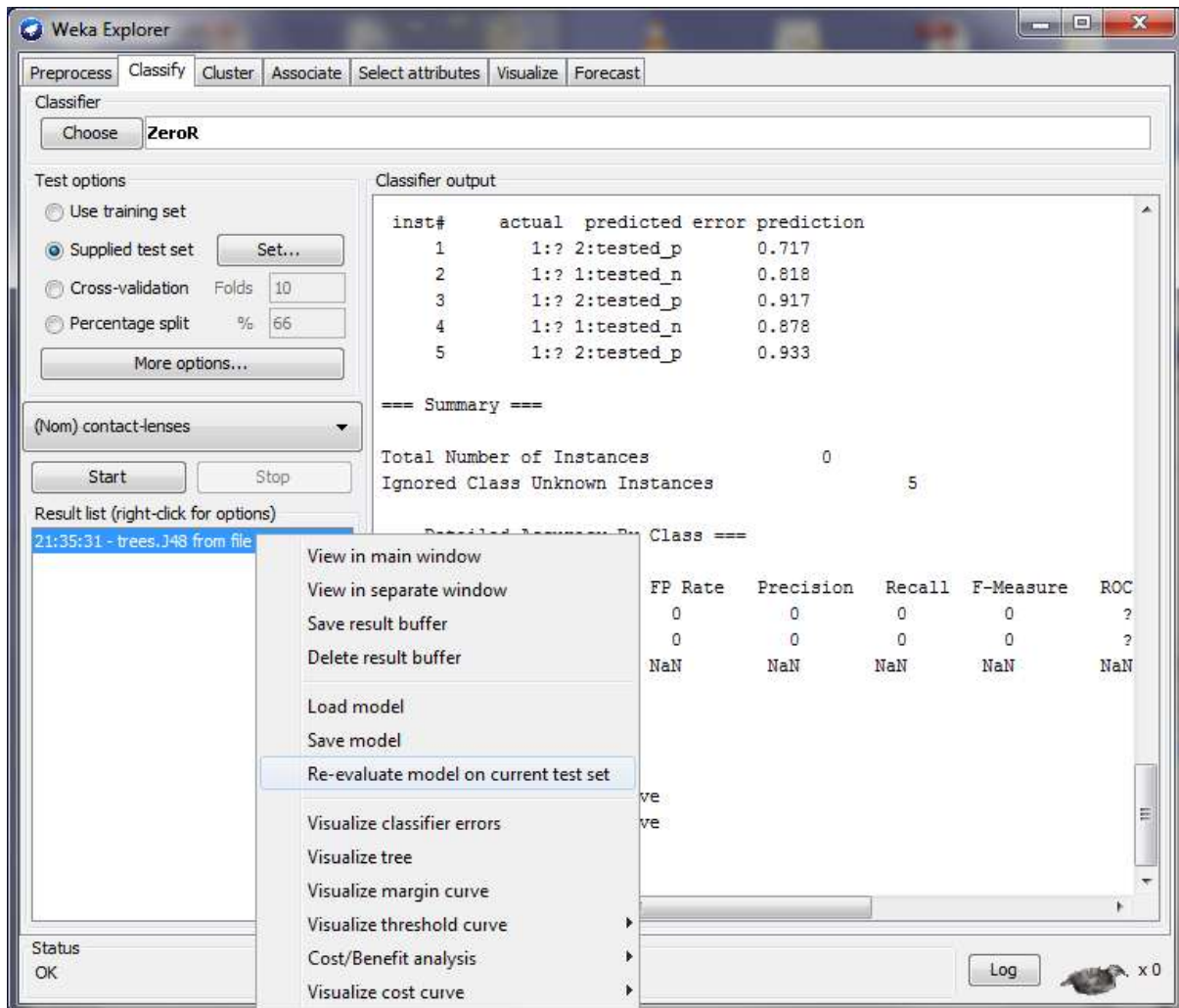
In the “Test Options”, we have to select “Supplied test set”, and once the file is loaded we select “No class” from the list of attributes.



Then, clicking “More Options”, a new window opens and we choose **PlainText** from ‘Output predictions’



Finally, we need to right click in the model and run “Re-evaluate model on current test set”.



The results are shown in the “Classifier output” panel, under “Predictions on test data”. The “predicted” column contains tested_p or tested_n for each of the lines in the test file.

Using the command line

It is explained in the following link:

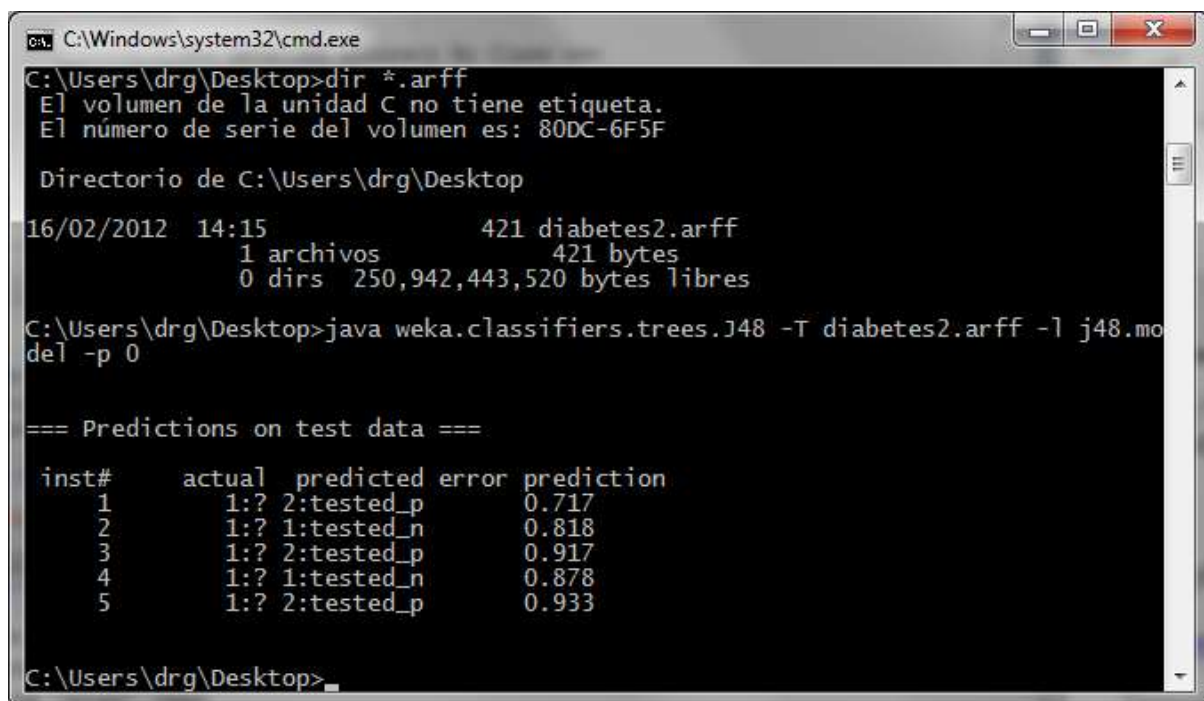
<http://weka.wikispaces.com/Making+predictions>

An example using our data:

```
java weka.classifiers.trees.J48 -T diabetes2.arff -l j48.model -p 0
```

You need to add the `weka.jar` file into the `CLASSPATH` environment variable (or using `-cp`) and the `'bin'` directory of your java installation in the `PATH` variable)

And the output should look like this:



```
C:\Windows\system32\cmd.exe
C:\Users\drg\Desktop>dir *.arff
El volumen de la unidad C no tiene etiqueta.
El número de serie del volumen es: 80DC-6F5F

Directorio de C:\Users\drg\Desktop

16/02/2012  14:15                421 diabetes2.arff
              1 archivos                421 bytes
              0 dirs 250,942,443,520 bytes libres

C:\Users\drg\Desktop>java weka.classifiers.trees.J48 -T diabetes2.arff -l j48.mo
del -p 0

=== Predictions on test data ===

inst#    actual  predicted error prediction
   1     1:?  2:tested_p  0.717
   2     1:?  1:tested_n  0.818
   3     1:?  2:tested_p  0.917
   4     1:?  1:tested_n  0.878
   5     1:?  2:tested_p  0.933

C:\Users\drg\Desktop>
```