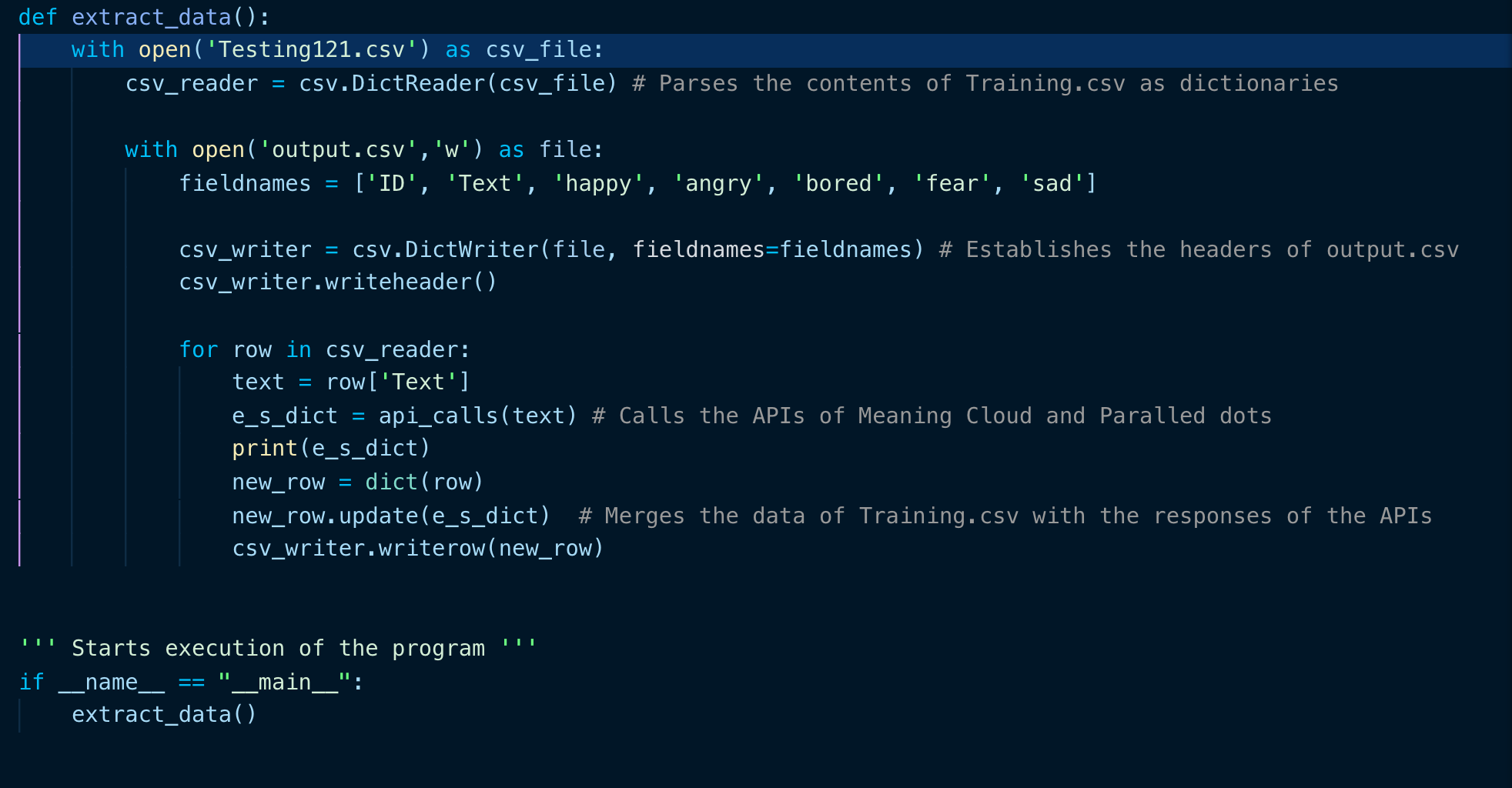
# H3 – Testing k-NN using emotion features

**Introduction**

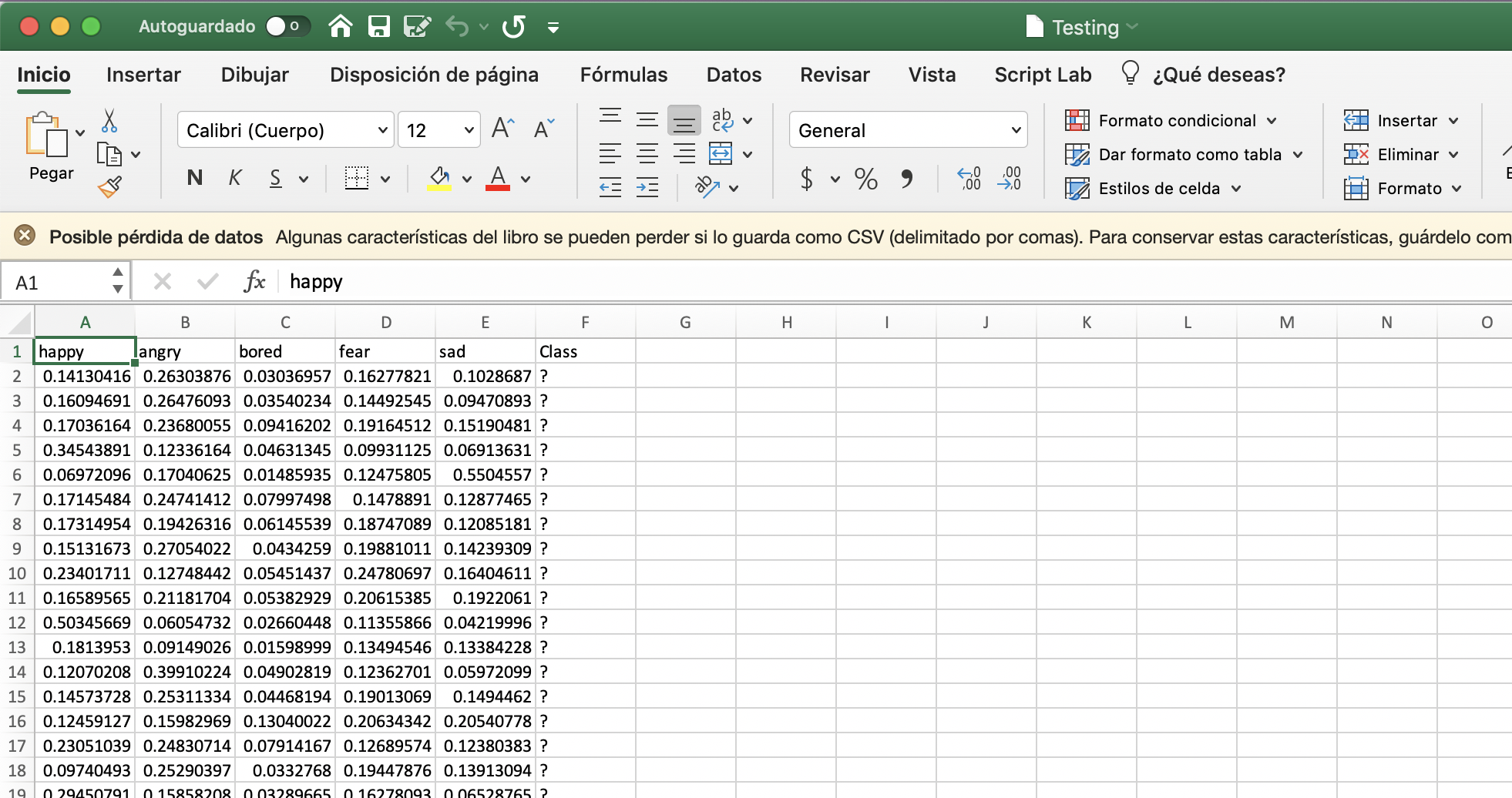
A k-NN algorithm (k-Nearest Neighbor) is a widely used algorithm classifier in machine learning. It works by selecting the nearest class to our data according to the number of neighbors that surround it. The number of neighbors is selected by the user and different types of distances like Euclidean or Manhattan can be chosen.

**Development**

Using the python script from HW1, with some modifications that excluded sentiment features, emotions from Testing.csv were extracted. The columns from output1-60.csv, output61-120.csv and output121-150 were copied to the original Testing.csv, ID and Text columns were removed. Also, a new Column named “Class” was added and filled with the symbol “?”.



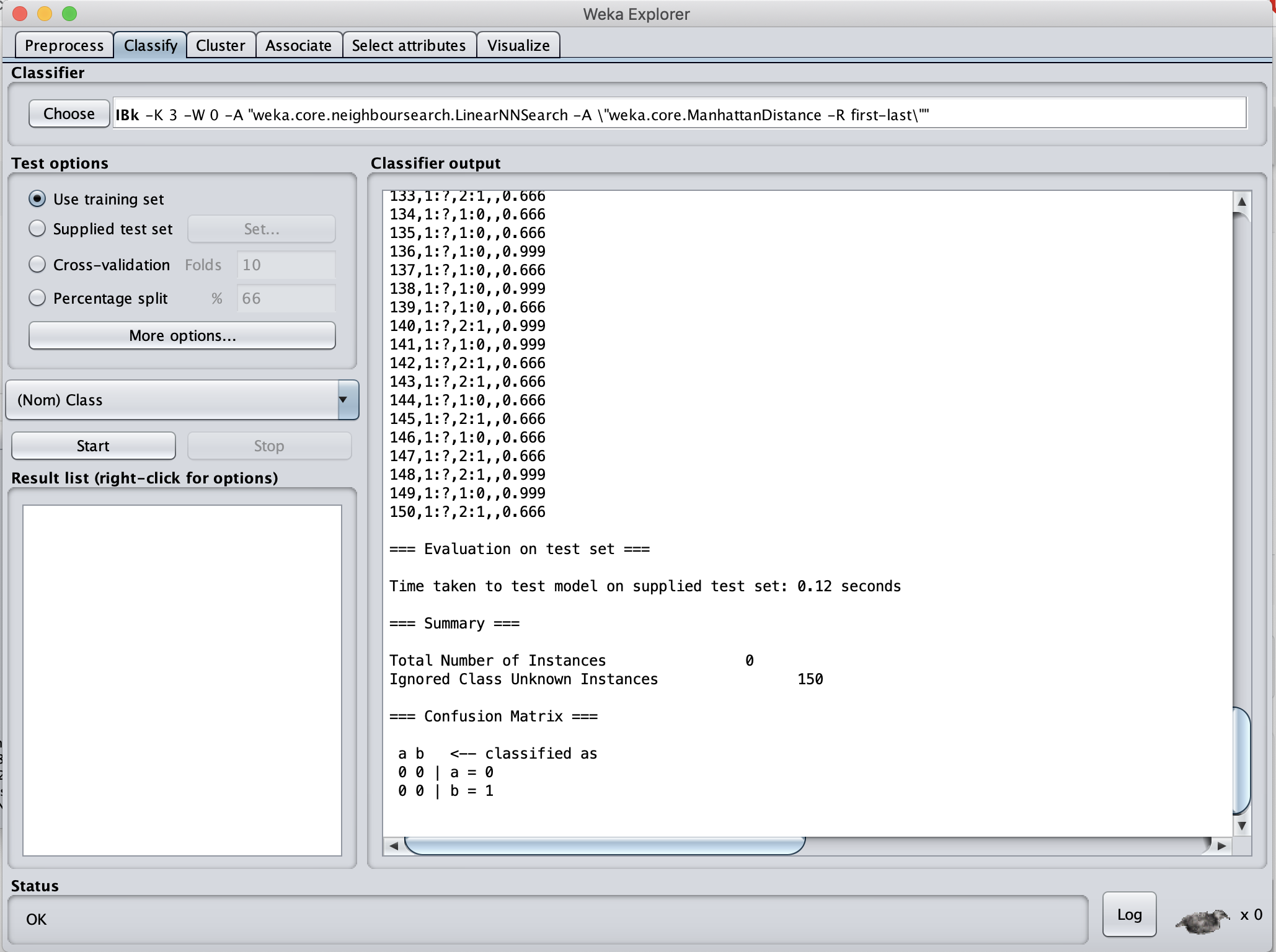
*Figure 1. output.csv features are shown.*

**

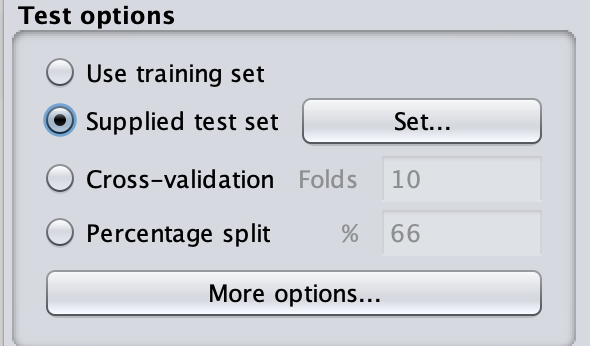
*Figure 2. Modified Testing.csv showing the new emotion’s features*

After all the emotion features were extracted, both Training.csv and Testing.csv were imported to Weka and then converted to .arff files for further manipulation. It is important to mention that in both files, Class attribute had to be changed to Nominal.

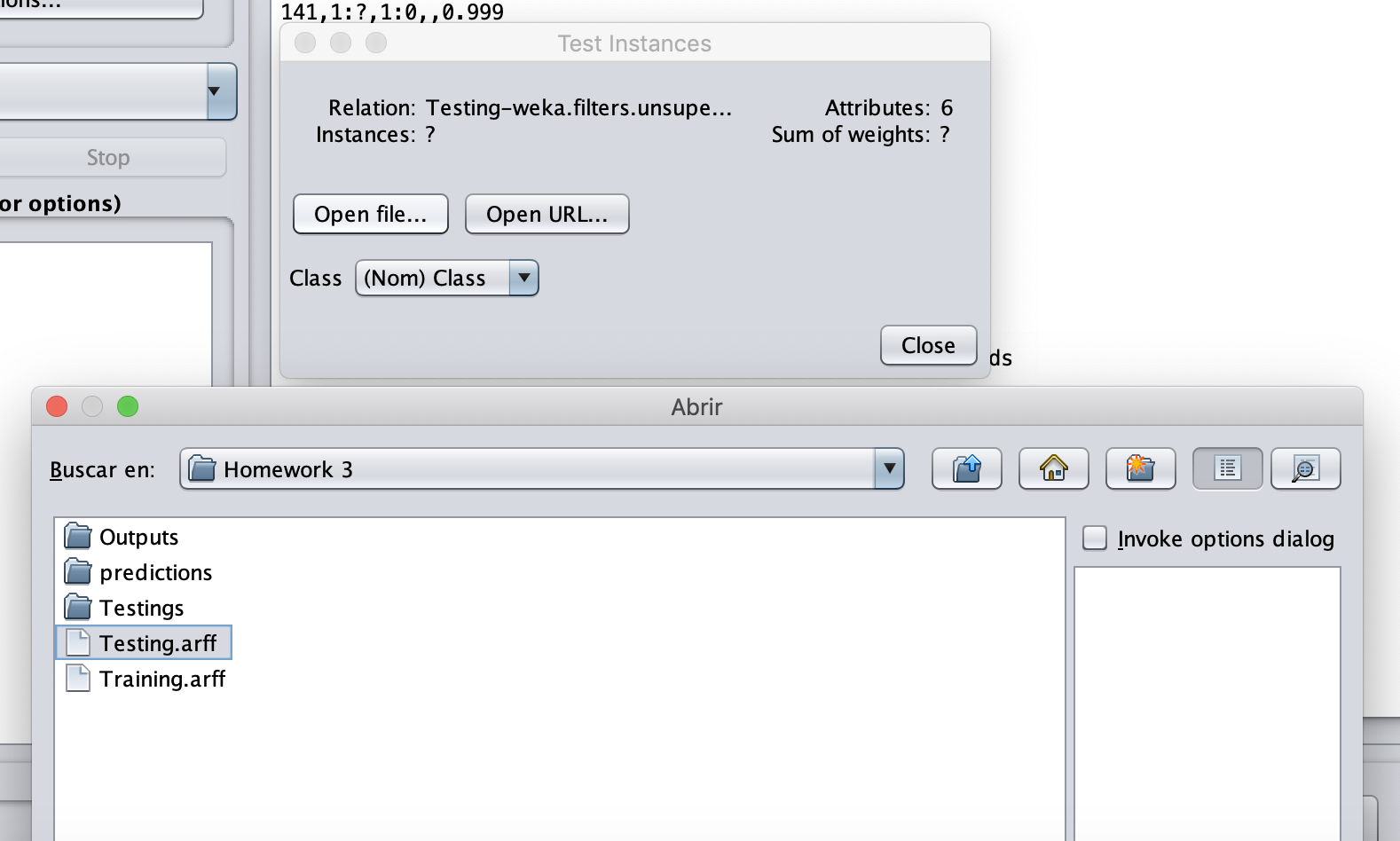
In the Classify tab of Weka, inside of the Lazy folder, IBk was picked, that is the k-NN classifier that was used for this homework. Considering that we have already uploaded the Training.arff file, The “Supplied test set” is chosen.



*Figure 3. Classify tab of Weka.*



*Figure 4. Supplied test set option is picked in Test options*

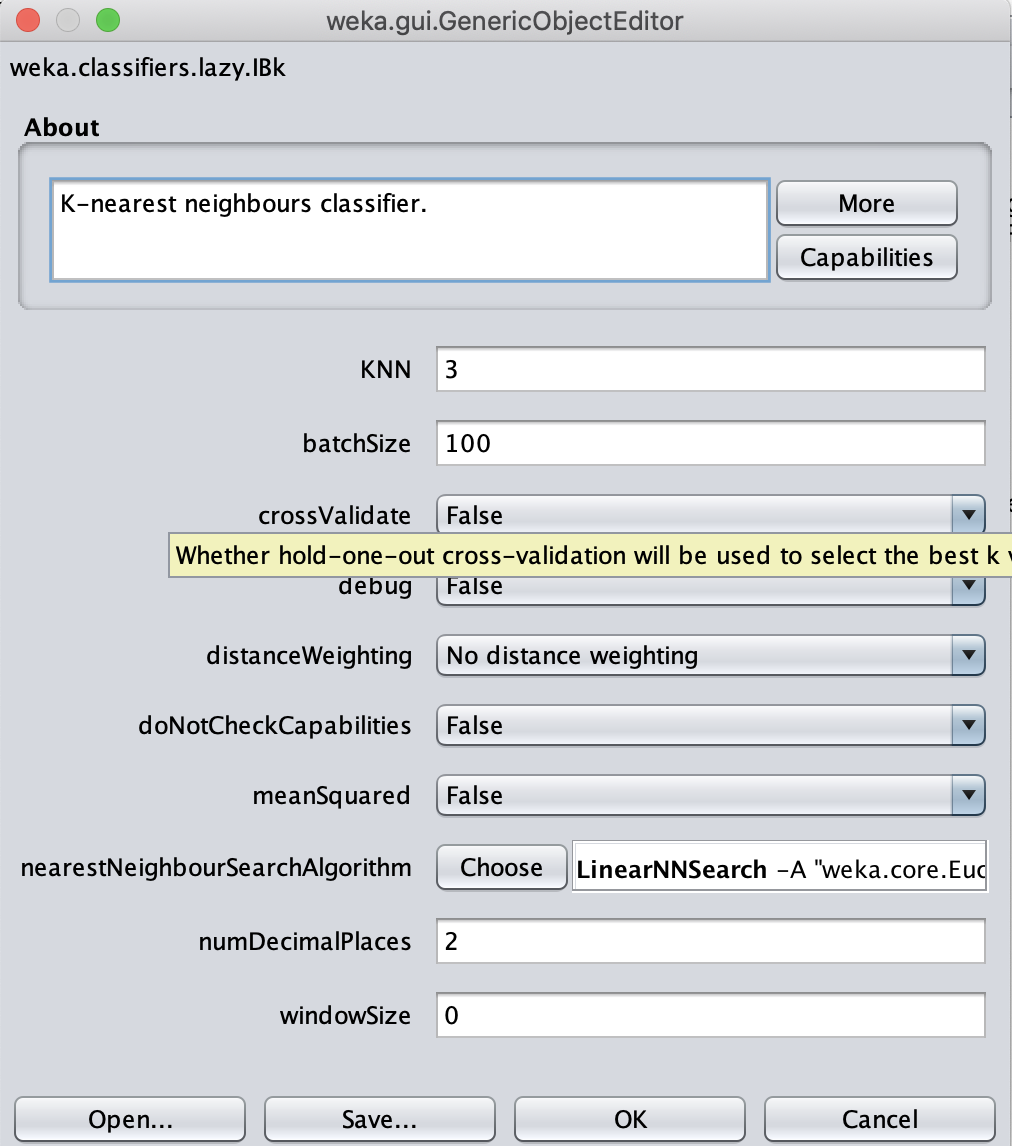
**

*Figure 5. Testing.arff is chosen as the test set.*

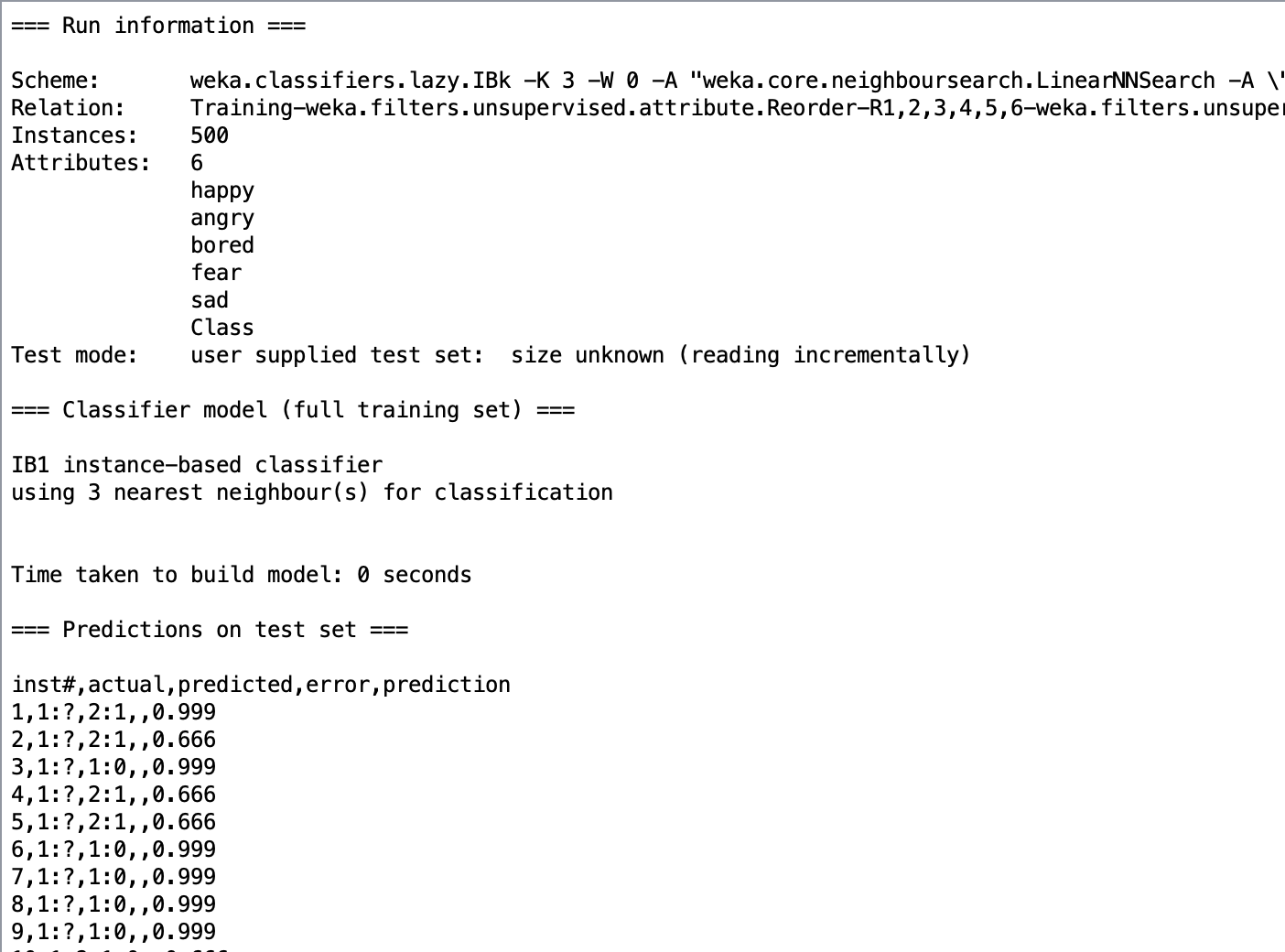
After all the above steps are made, now we can customize our options for our k-NN classifier.

**Results**

*k-NN with N=3 and Euclidean distance*

**

*Figure 6. k-NN configuration panel.*

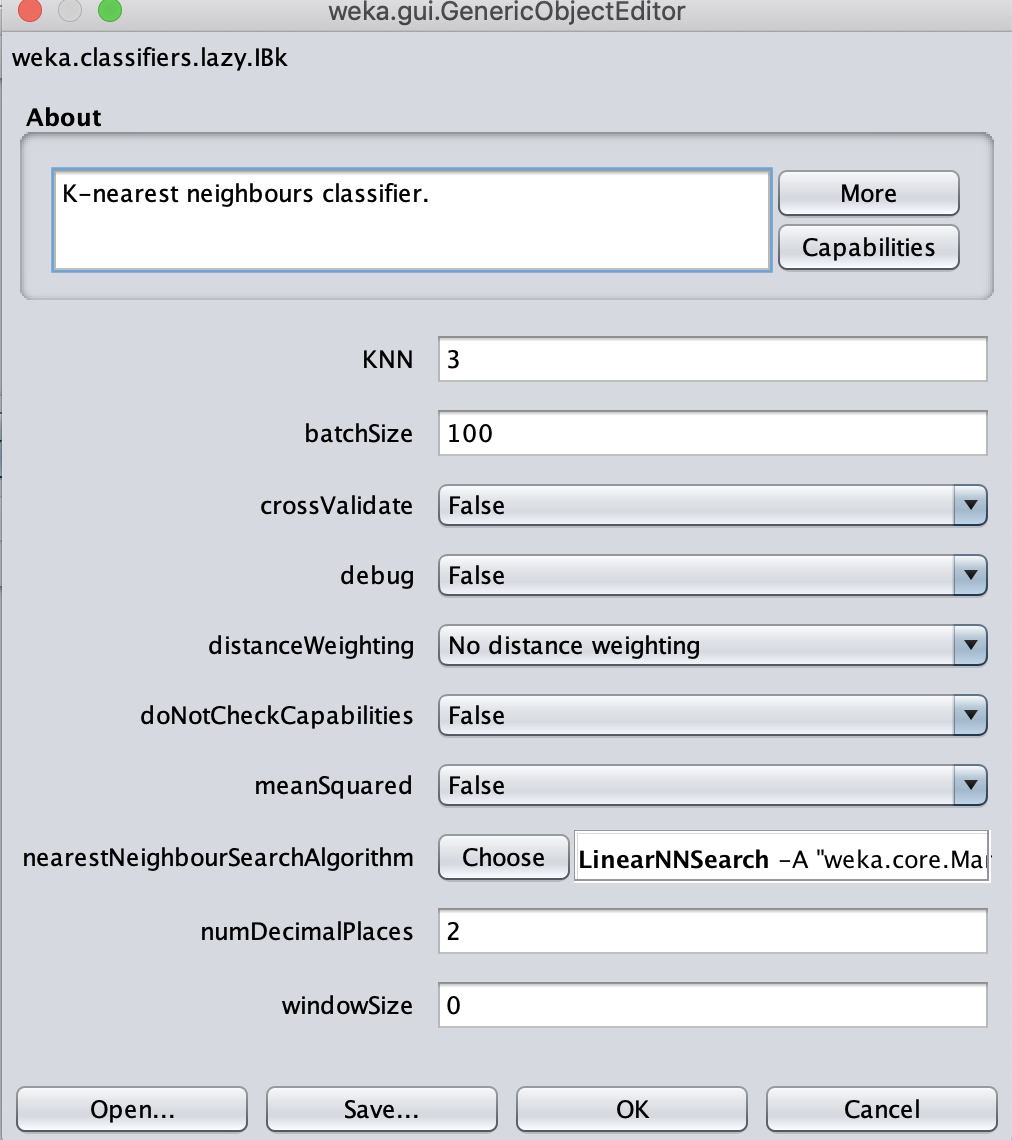
**

*Figure 7. CSV output of the Testing dataset.*

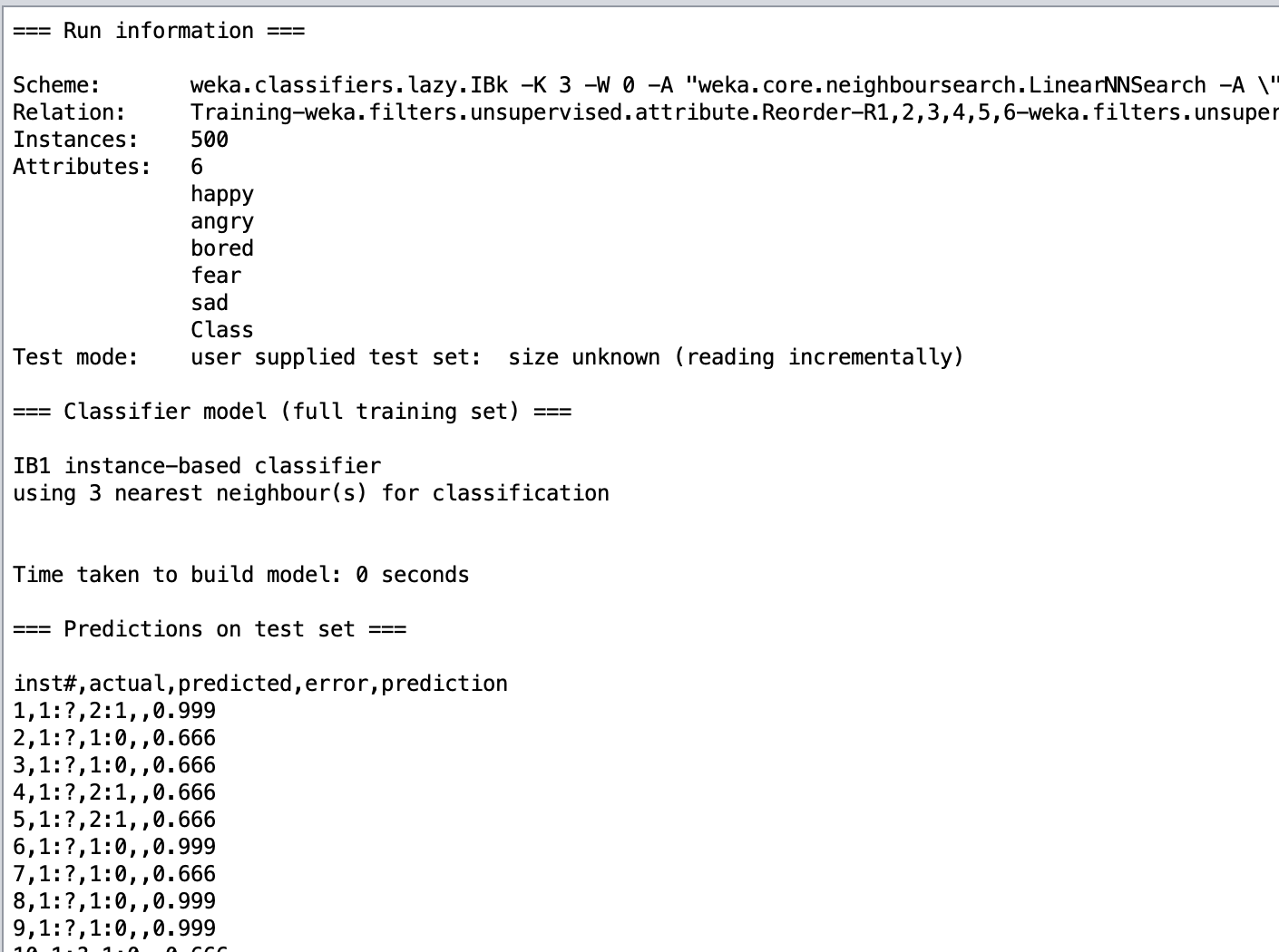
**

*Figure 8. Result obtained in Kaggle.*

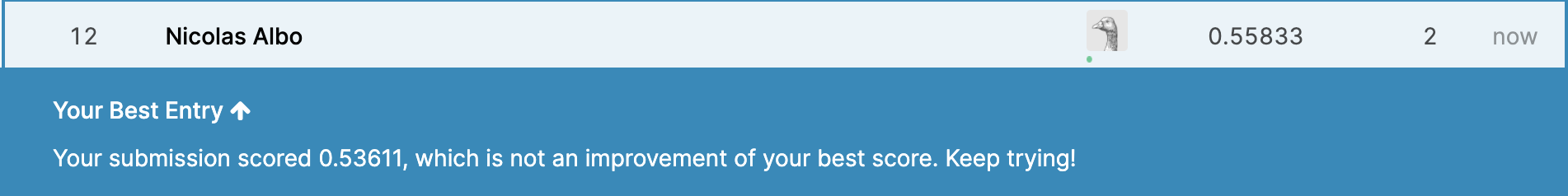
*k-NN with N=3 and Manhattan distance*



*Figure 9. k-NN configuration panel.*



*Figure 10. CSV output of the Testing dataset.*



*Figure 11. Result obtained in Kaggle.*

**Conclusion**

A k-NN is a powerful algorithm classifier, it is a little bit basic and easy to understand. Maybe my results in Kaggle weren’t as high as I expected, but I’m sure that if I keep modifying my k-NN configurations, I can rank higher.