14210231 – Felipe Guth

Assignment 1 – COMP30120

The table 1 resumes all the tests performed in the Weka framework, where the correctly classified instances are depicted in percentage.

The better results were obtained using the method of training set as a test set, in most classifiers. This is expected since the classifier is overfitted in the data. In other words, it is optimized to classify the data it was trained on, but will have a bad performance in classifying new data. It will not be able to generalize in a proper way for new samples.

Excluding the overfitted classifiers, the best result was obtained with the 60%/40% training/testing method with the Naïve bayes classifier. However, the 10-fold cross validation method generated more consistent classifiers in general, compared to the 60-40 method.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Training set as test set | 60% training set 40% test set | 10-fold cross validation |
| Naïve Bayes | 67.58 | **68.96** | 63.44 |
| KNN-1 | **100** | 55.17 | 60.68 |
| KNN-3 | 77.24 | 53.44 | 65.51 |
| SVM | 73.79 | 63.79 | **66.89** |

Table 1 - Percentage of correctness of each classifier per method

Table 2 shows the round mean square error that is the distance, on average, of a data point from the fitted line, measured along a vertical line. The round mean square error once more reinforces the results depicted in Table 1.

Root mean squared error

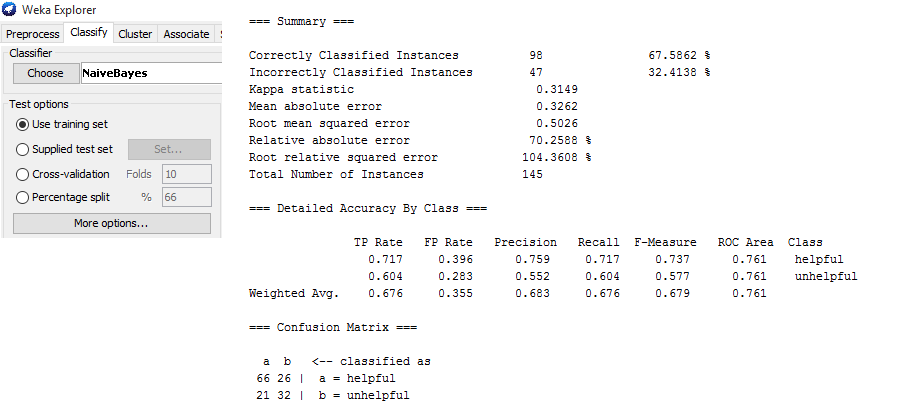
|  |  |  |  |
| --- | --- | --- | --- |
|  | Training set as test set | 60% training set 40% test set | 10-fold cross validation |
| Naïve Bayes | 0.50 | 0.53 | 0.55 |
| KNN-1 | **0** | 0.66 | 0.62 |
| KNN-3 | 0.38 | 0.54 | 0.49 |
| SVM | 0.51 | 0.60 | **0.57** |

Table 2 – Root mean square error

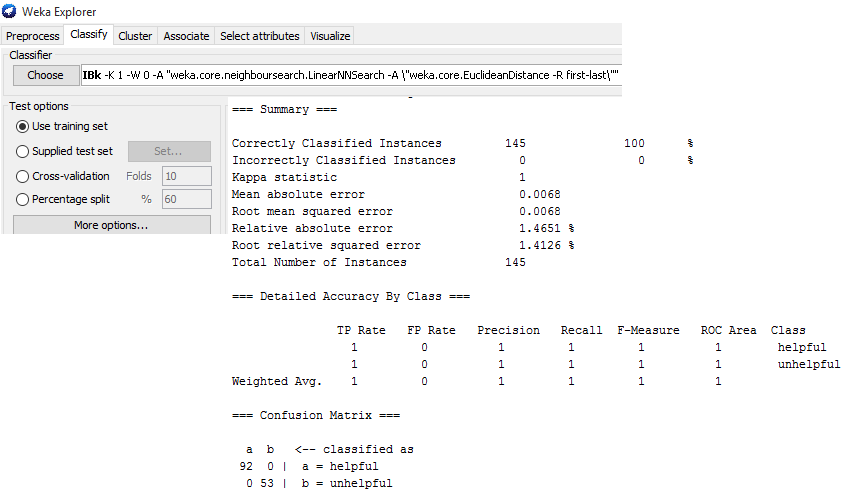
**Detailed Results**

1 – Training set as test set

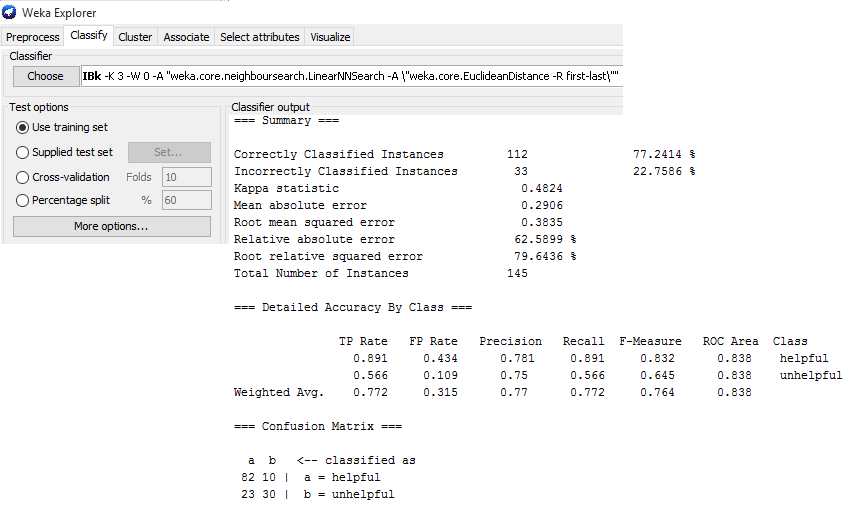
a – Naïve Bayes



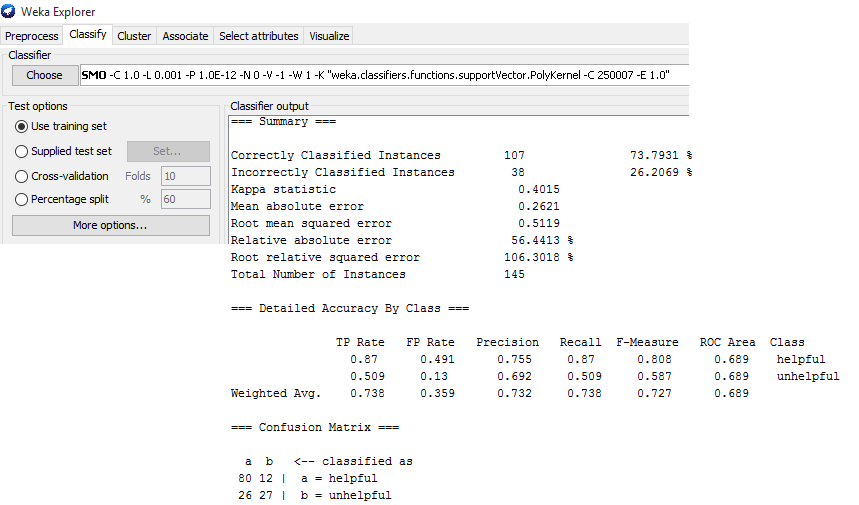
b – KNN-1



c- KNN-3

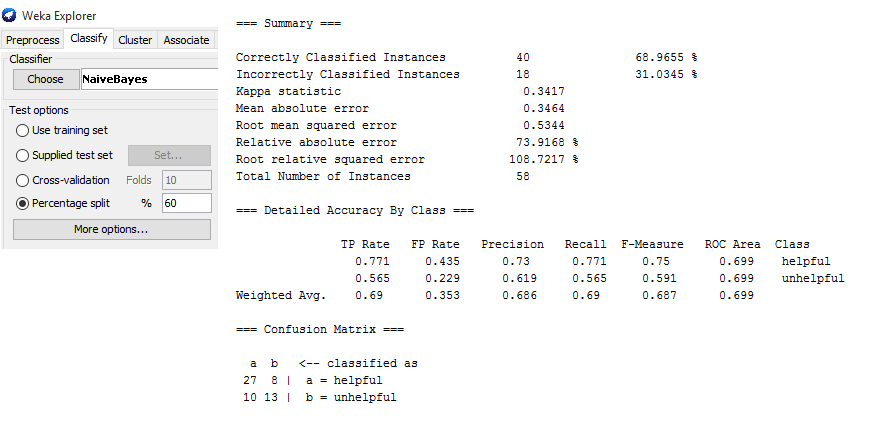


d) SVM

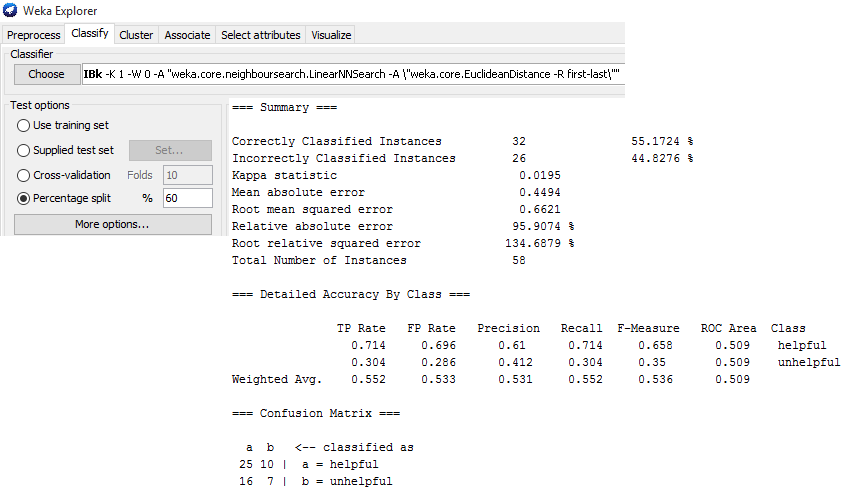


2 – 60% as training set 40% as test set

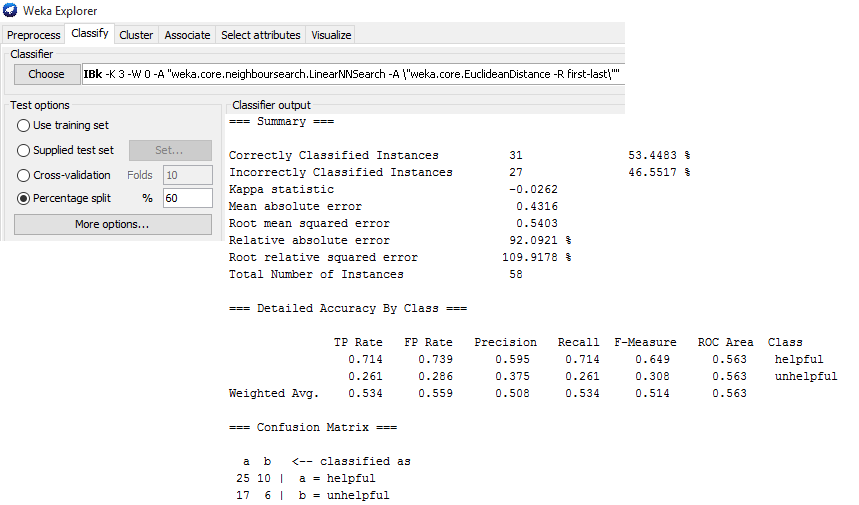
a- Naïve Bayes



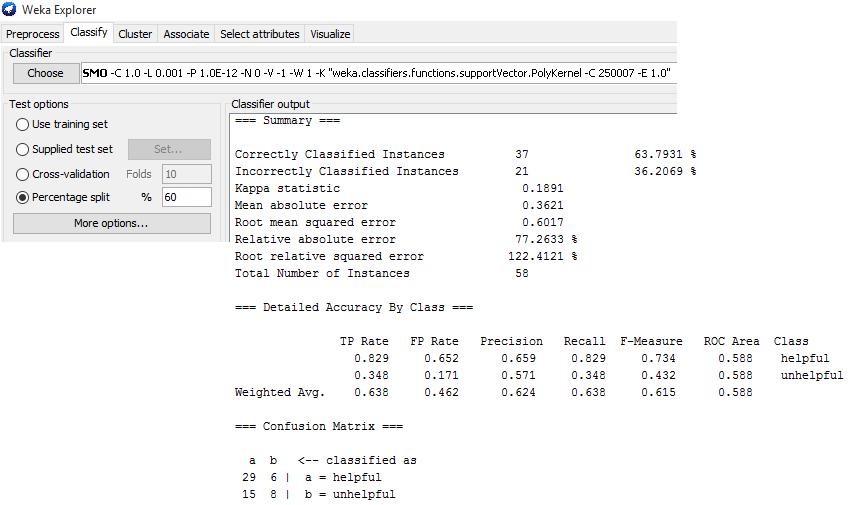
b- KNN-1



c- KNN-3

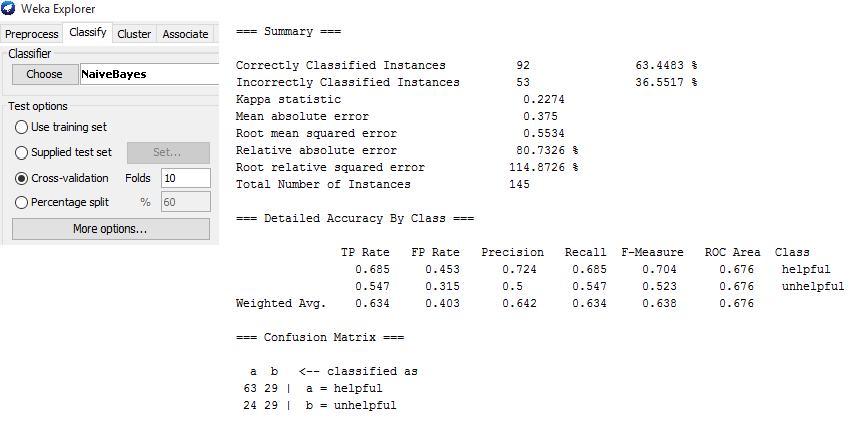


d- SVM

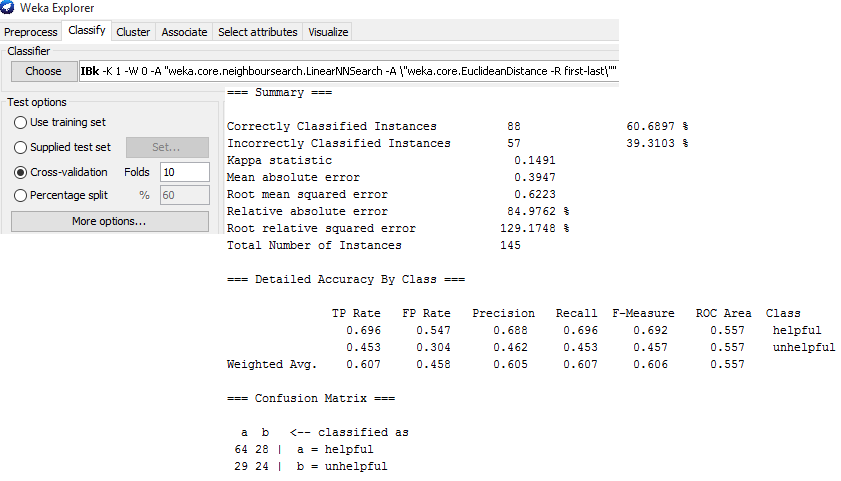


3 – 10-Fold Cross validation

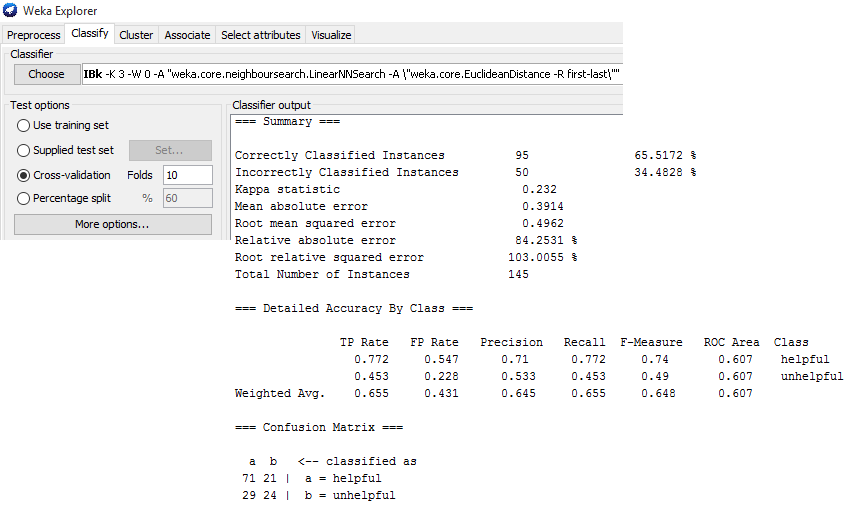
a- Naïve Bayes



b- KNN-1



c- KNN-3



d- SVM

