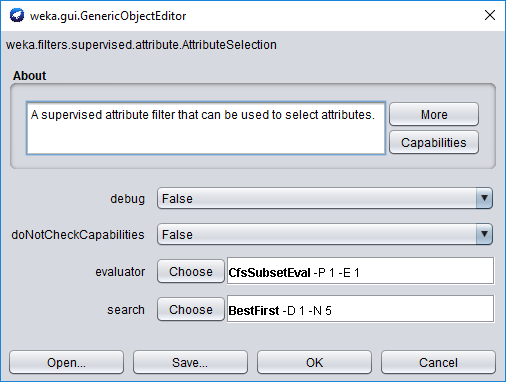
Patrick Austin

CS 491: Social Networks: Lab 5

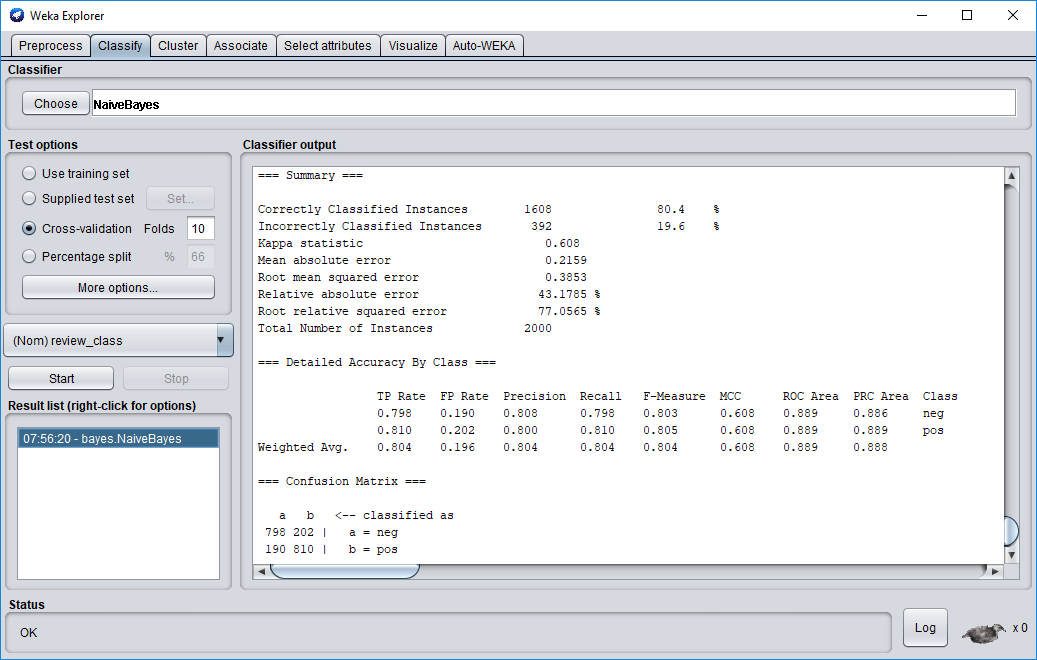
10 April, 2018

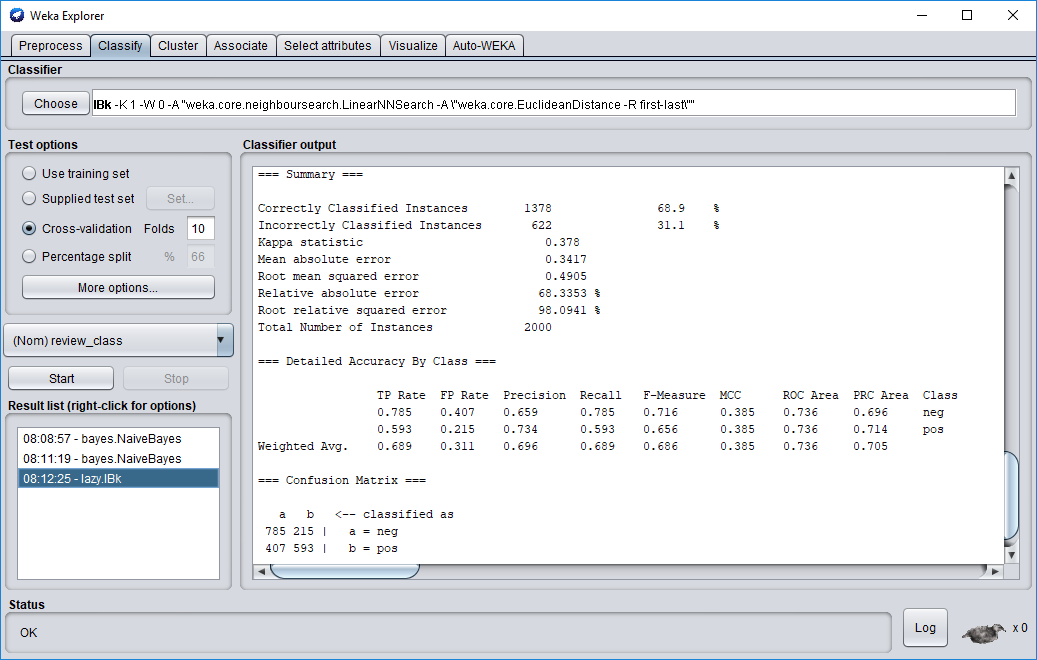
I will show WEKA output for Naive Bayes, 1NN, 3NN, and 5NN for each of the 8 combinations requested. At the end I have compiled a table of results.

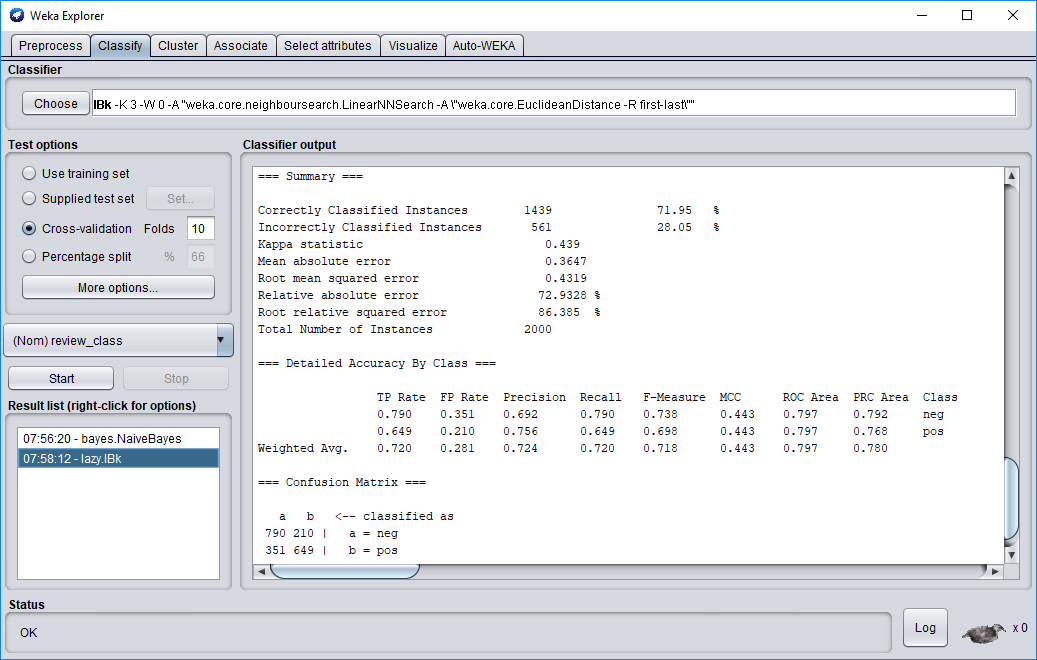
First I show the configuration options used for data transformation. The images below show selected options for all three required transformations; these were selected and un-selected as necessary for each combination. I used the snowball stemmer, the stopwords.txt file provided by the author, and the attribute selection options shown below.

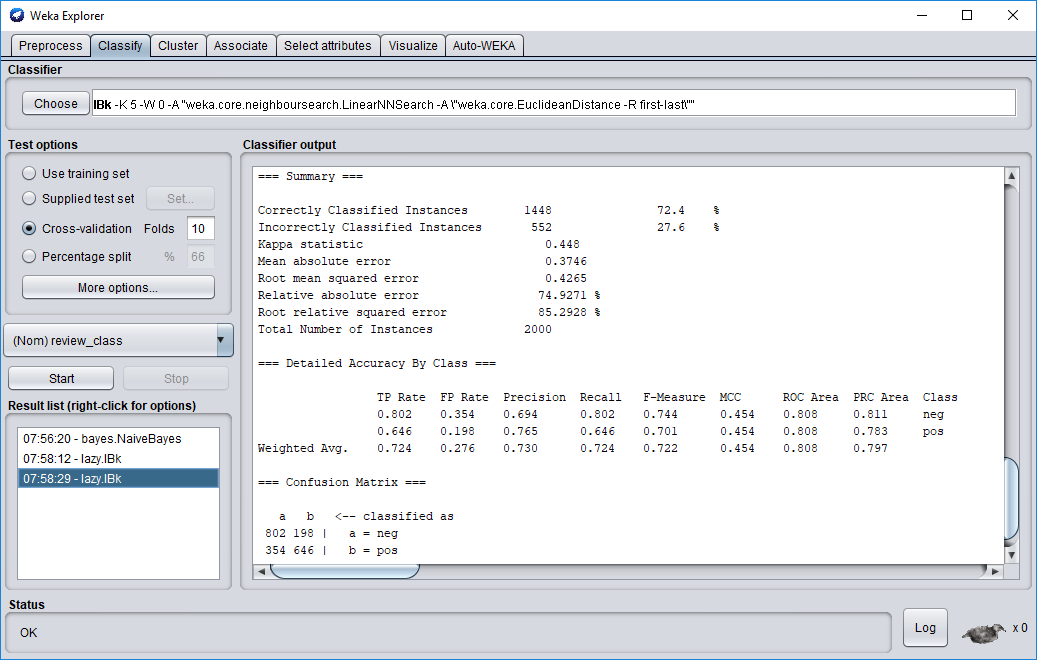


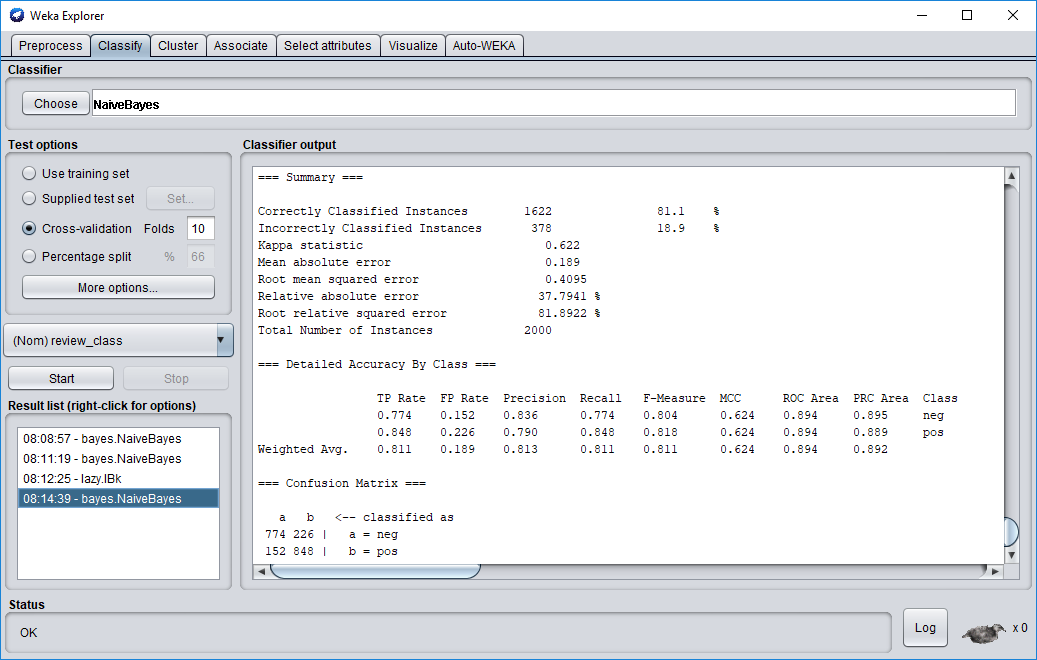
1. With stop-words removal, stemming, and attribute selection (ie all transformations applied):

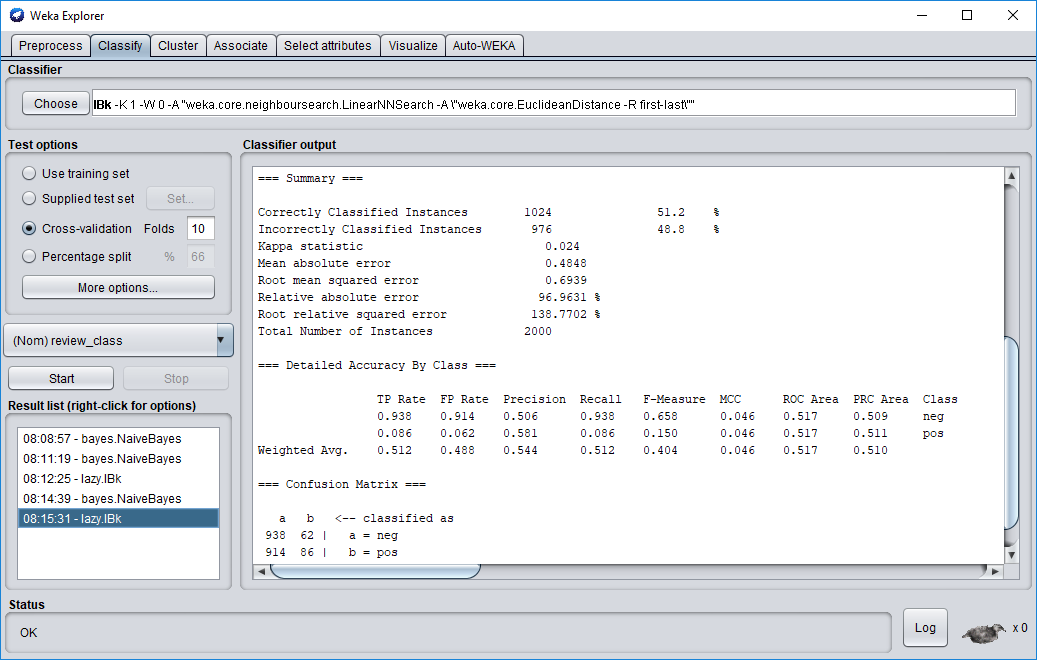


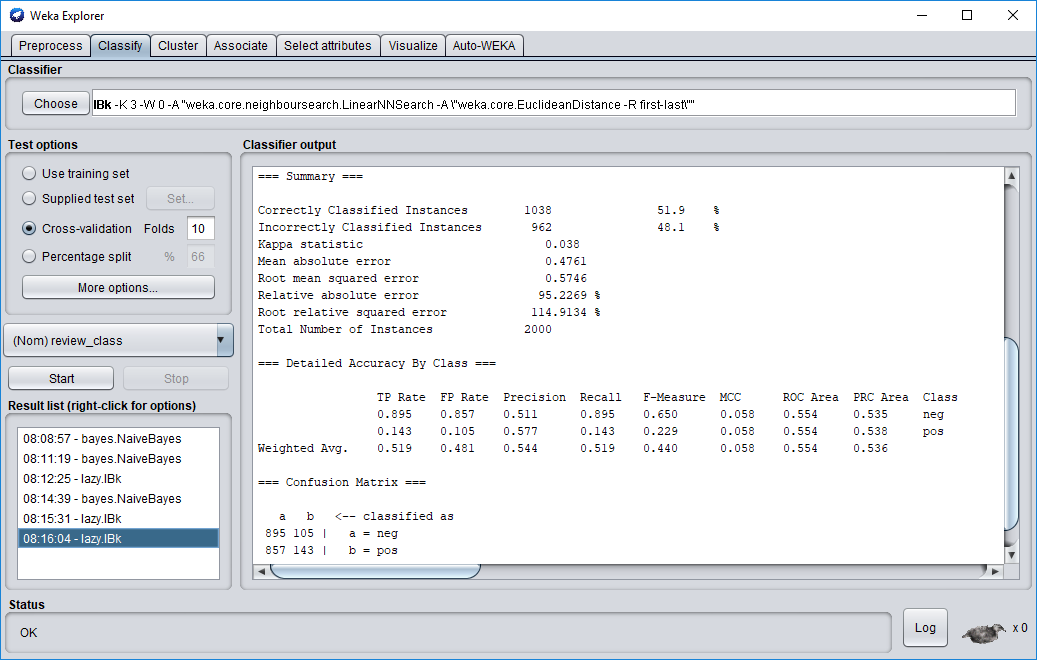


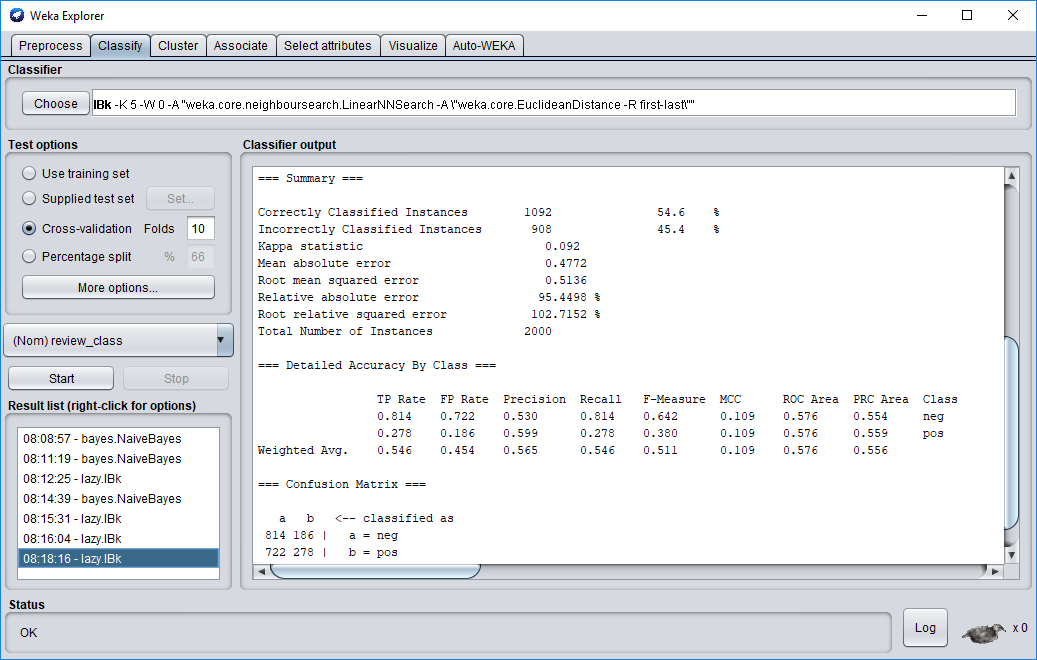




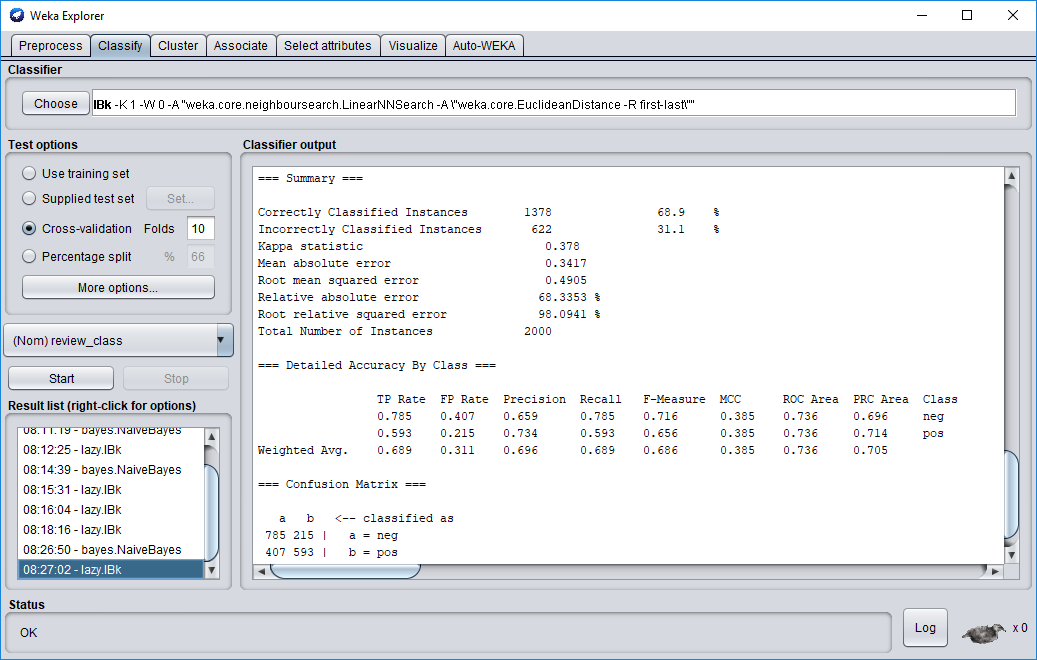
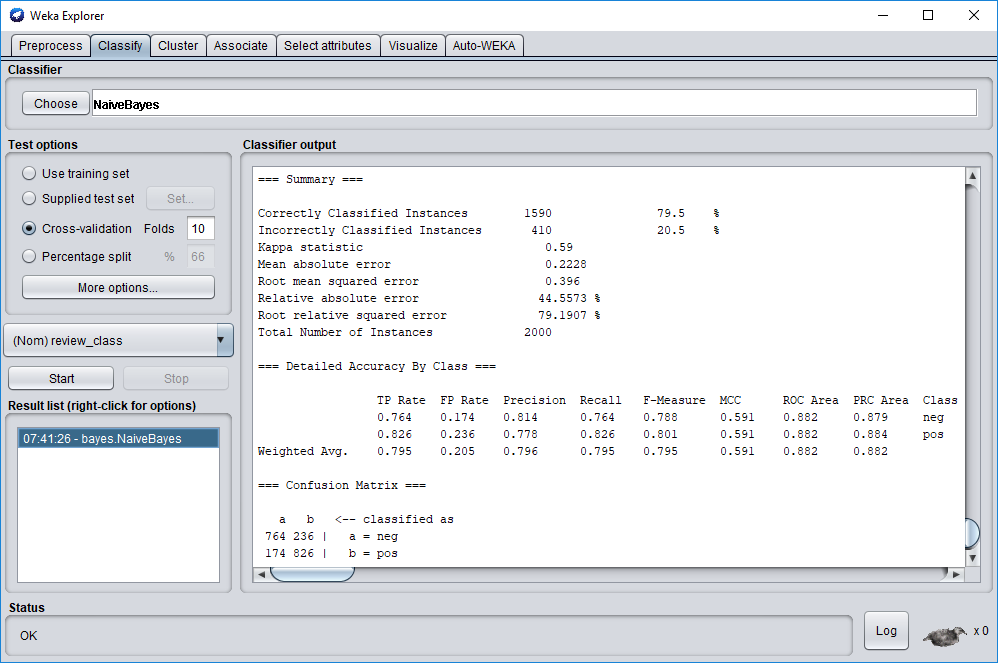
2. With stop-words removal and stemming only:

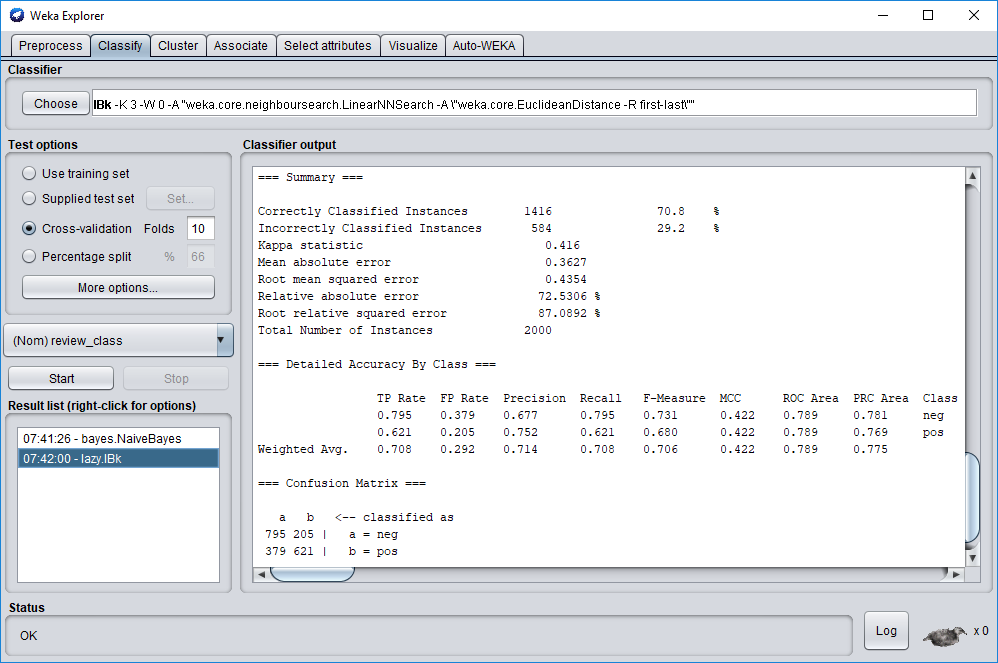






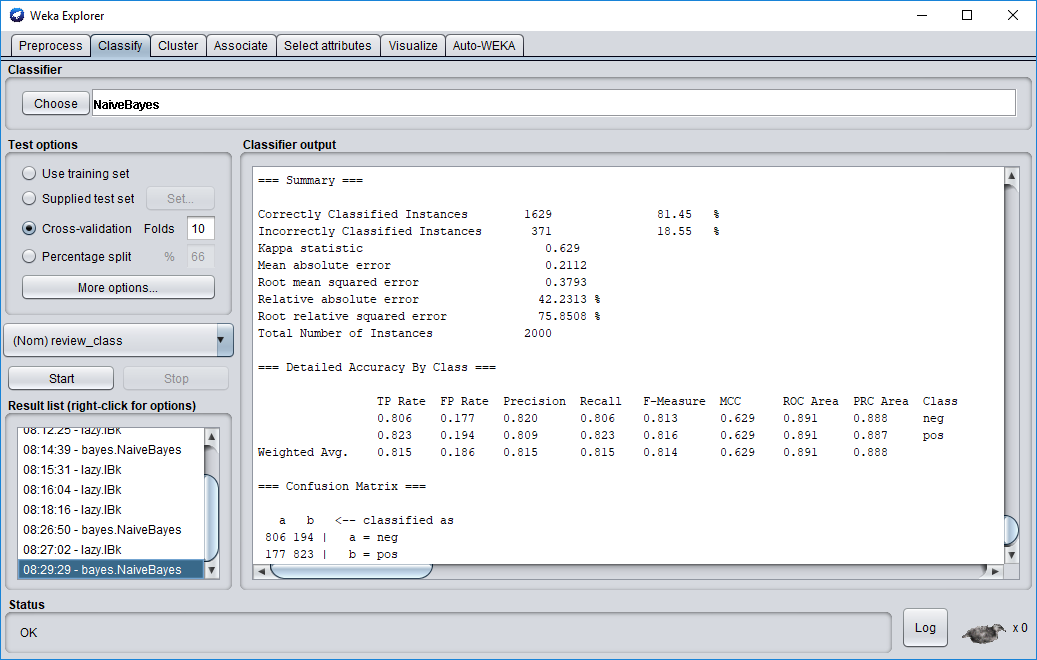
3. With stop-words removal and attribute selection only:

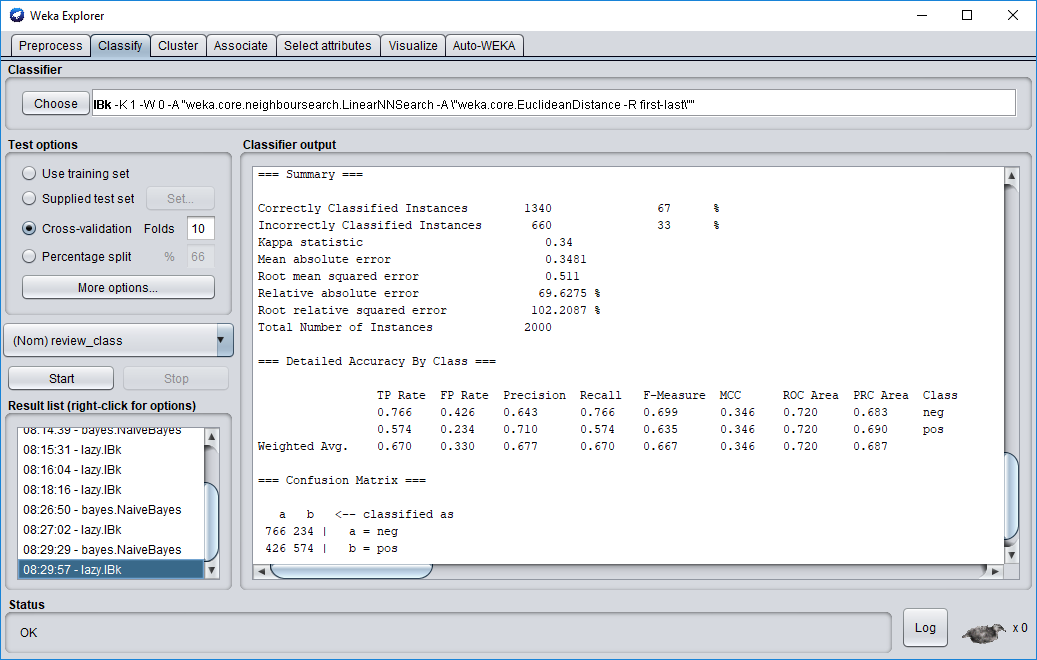


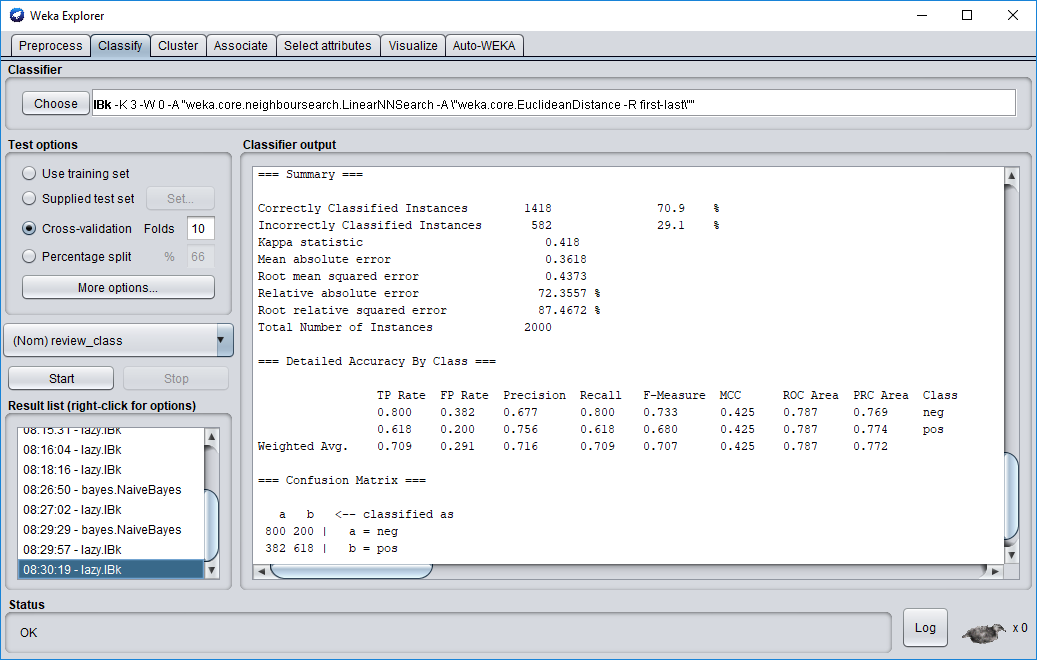


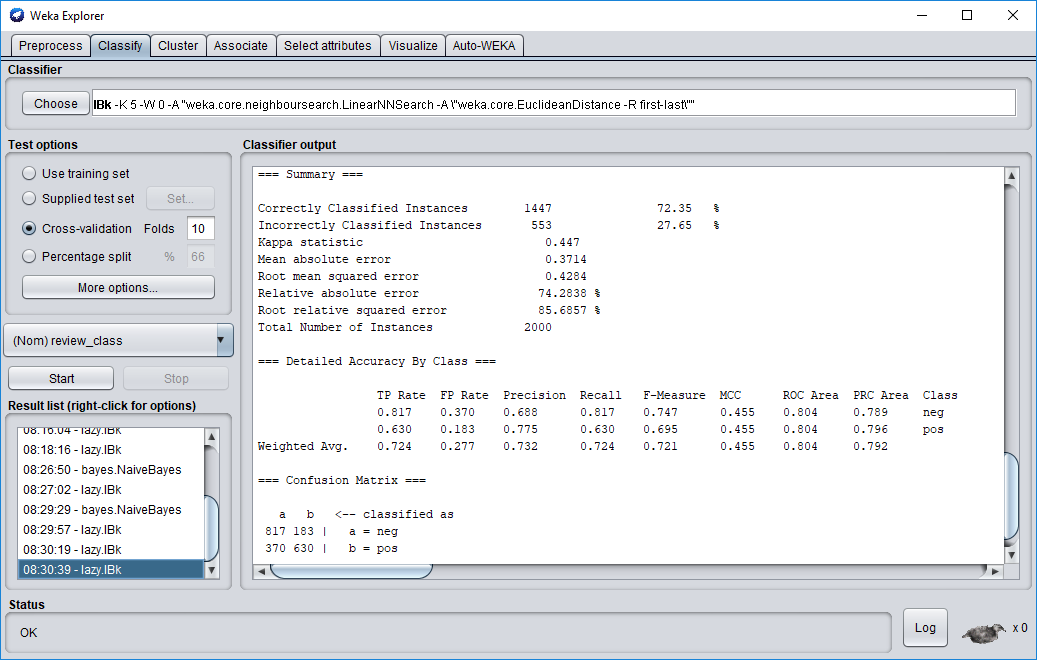


4. With stemming and attribute selection only:

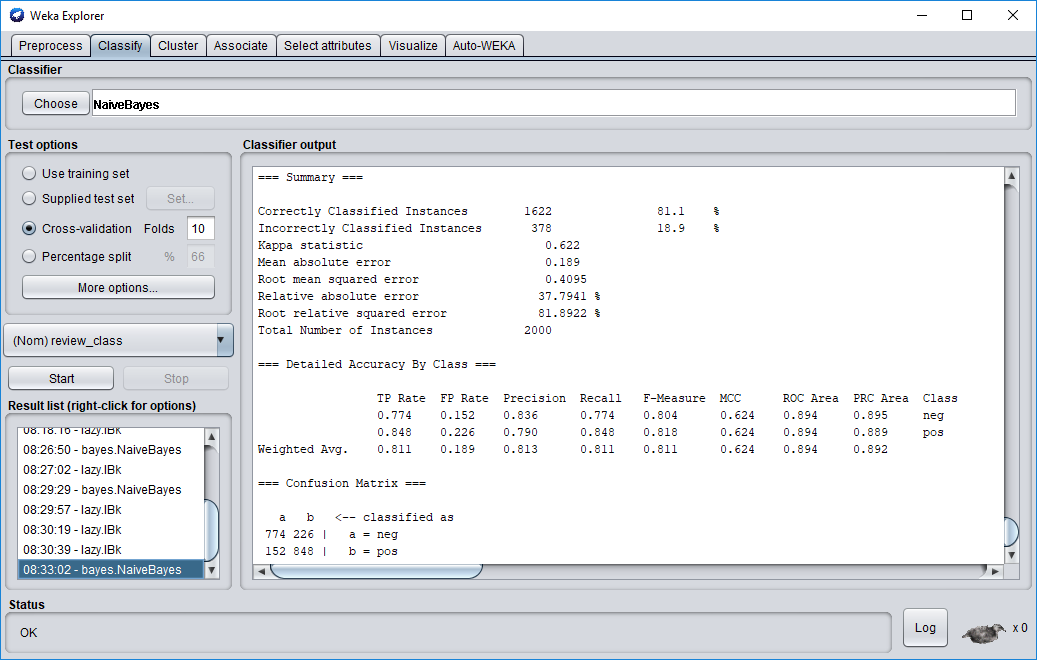


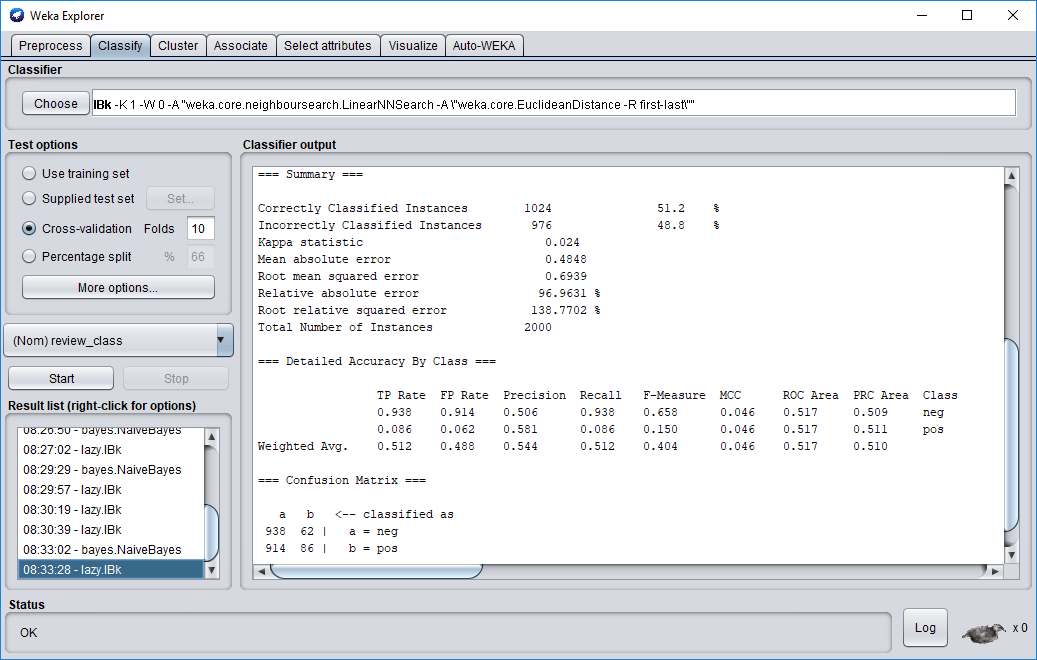
****

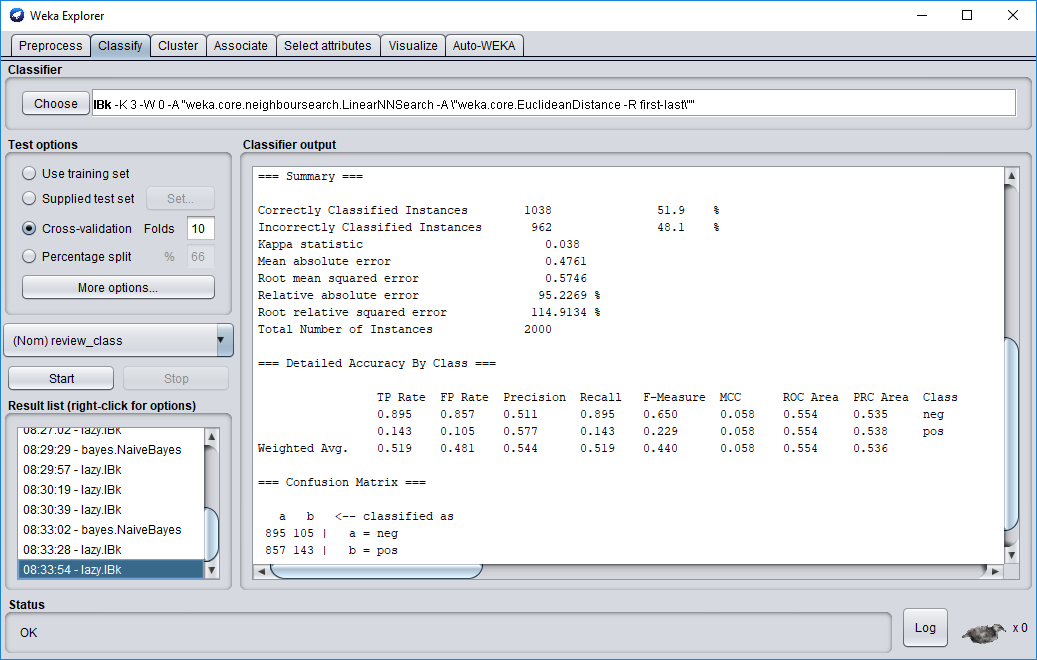
****

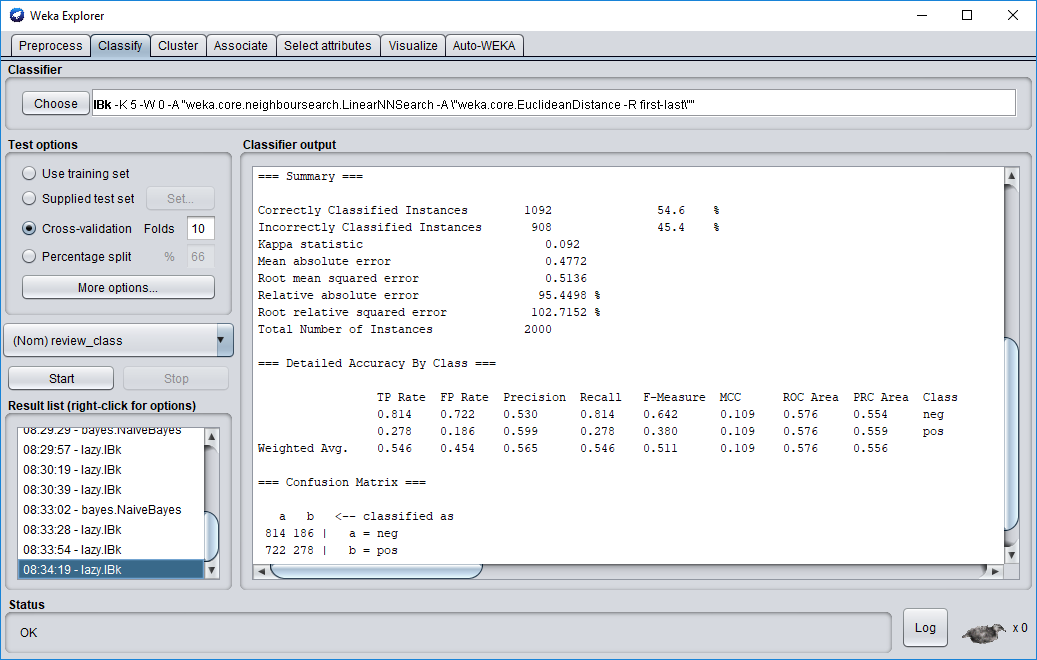
****

5. With stop-words removal only:

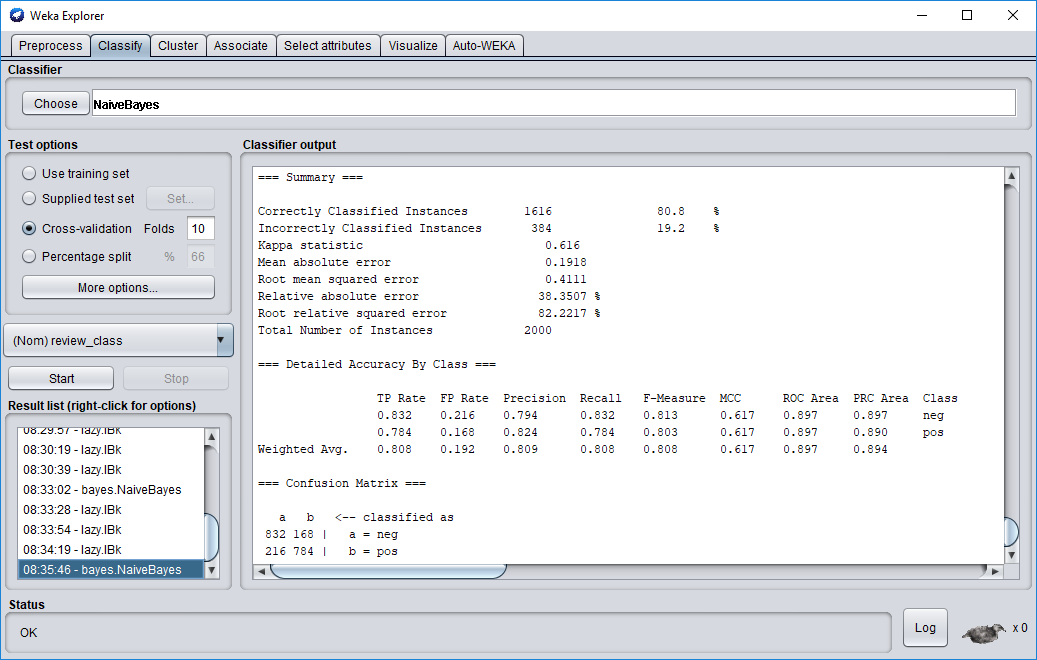


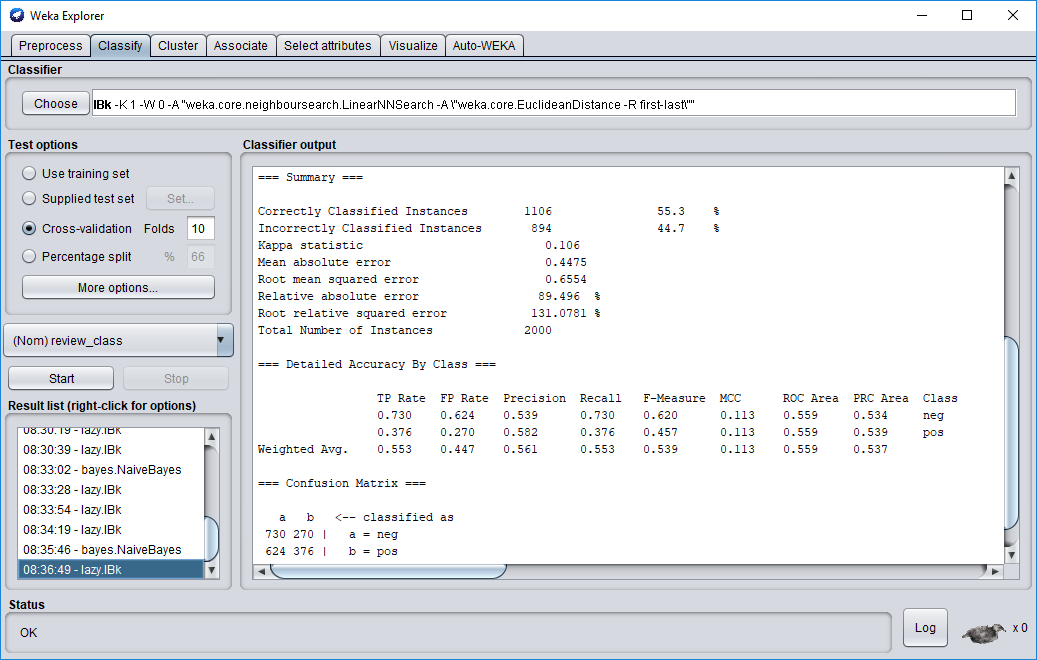


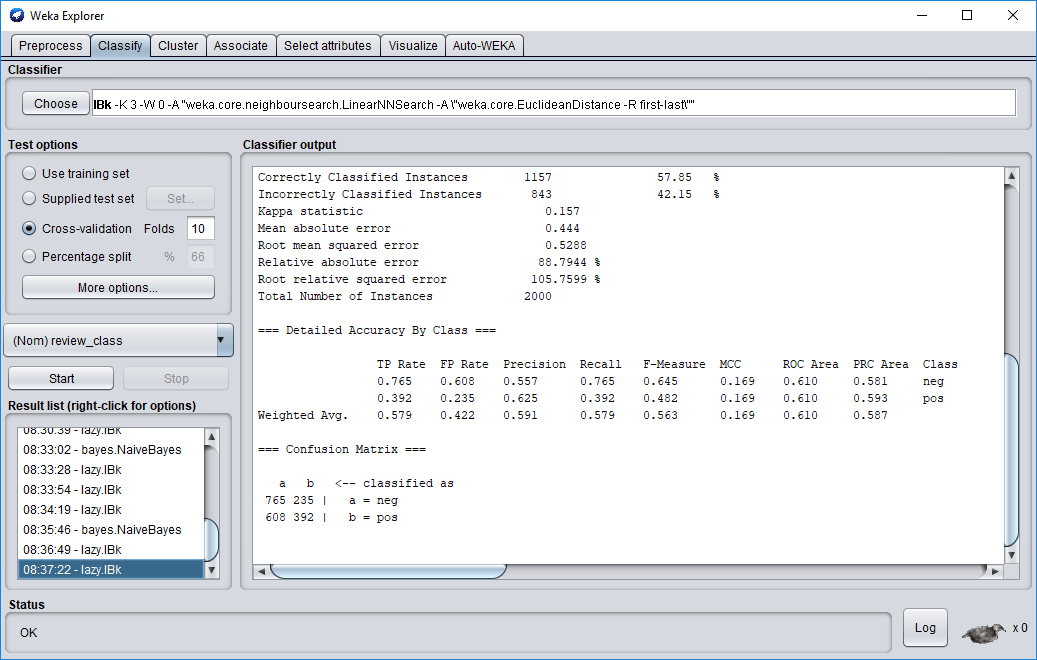


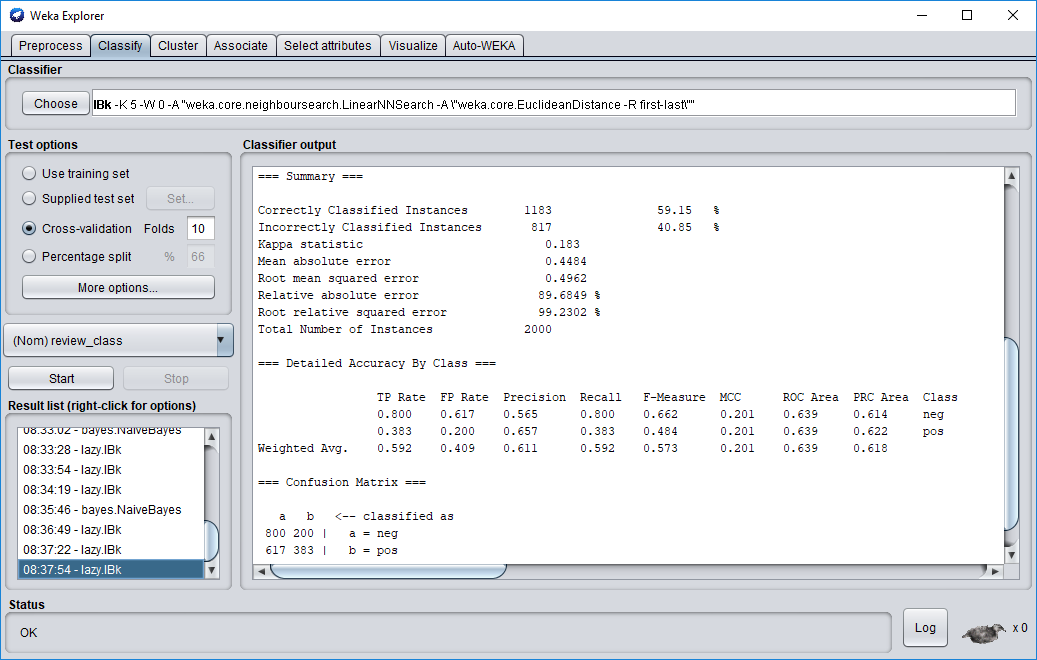


6. With stemming only:

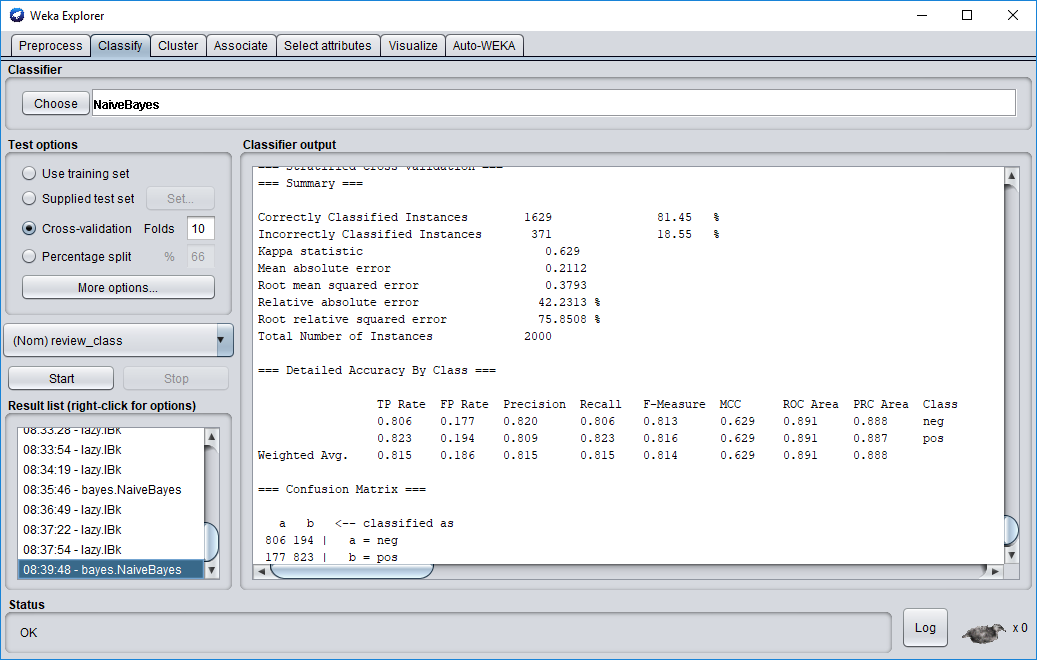


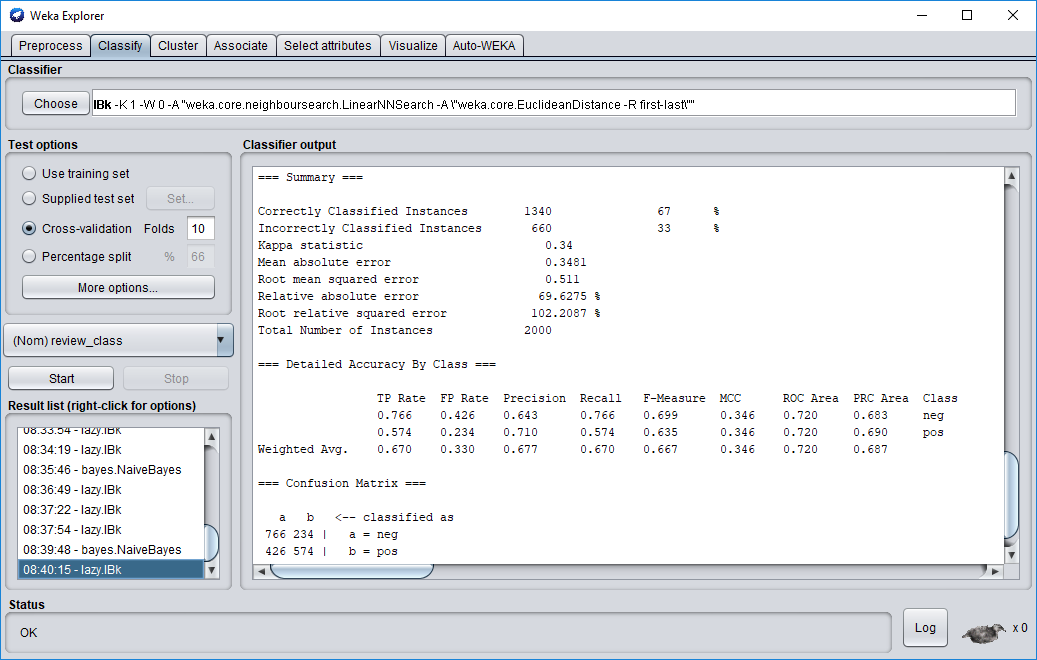


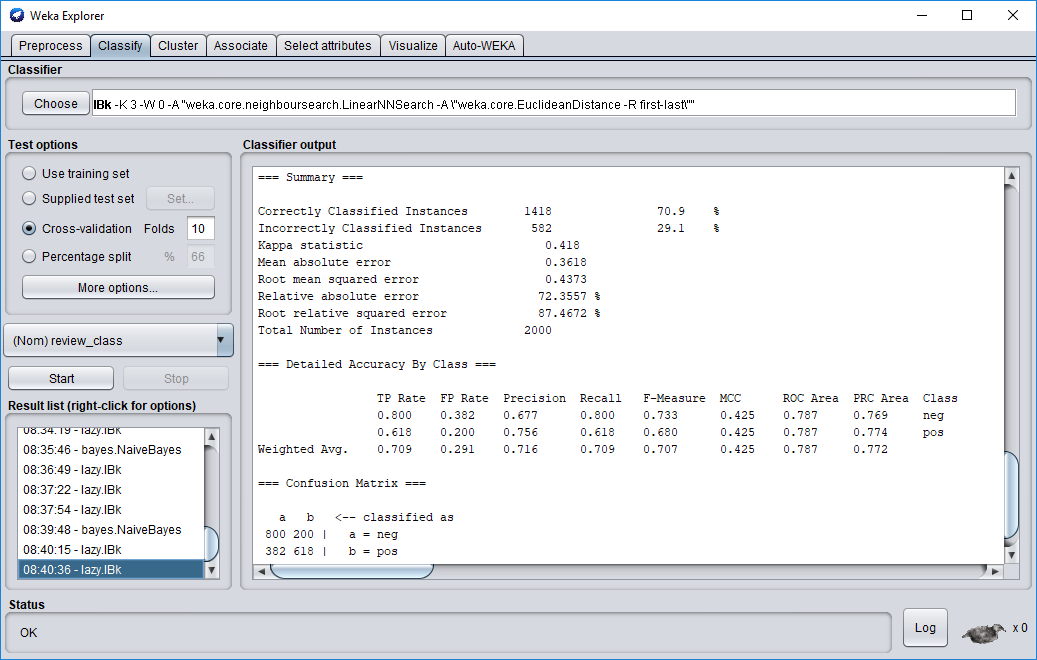


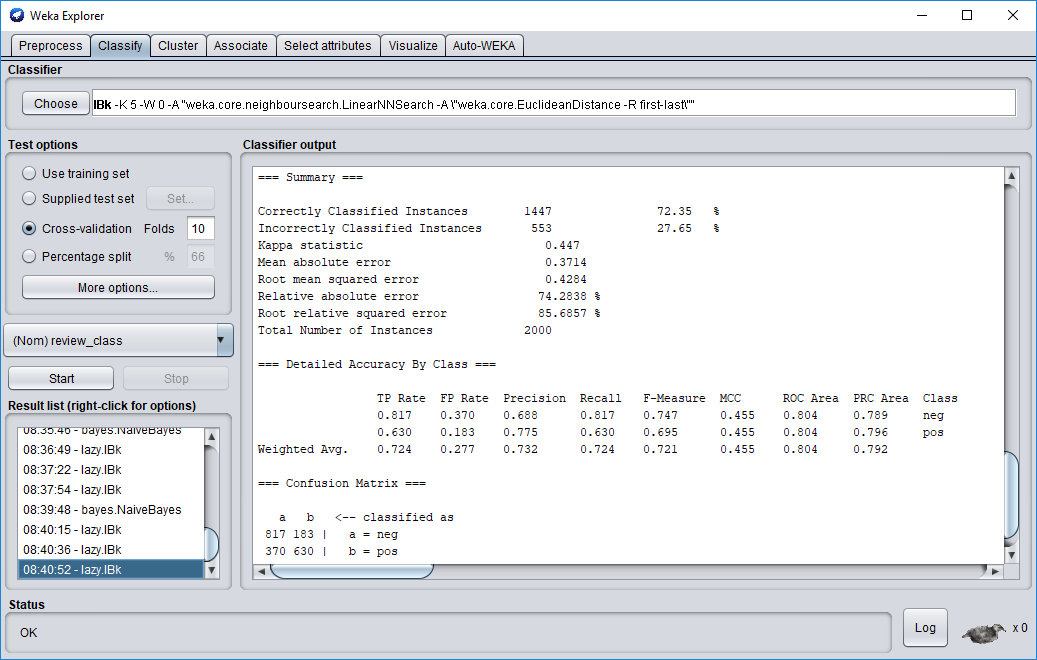


7. With attribute selection only:

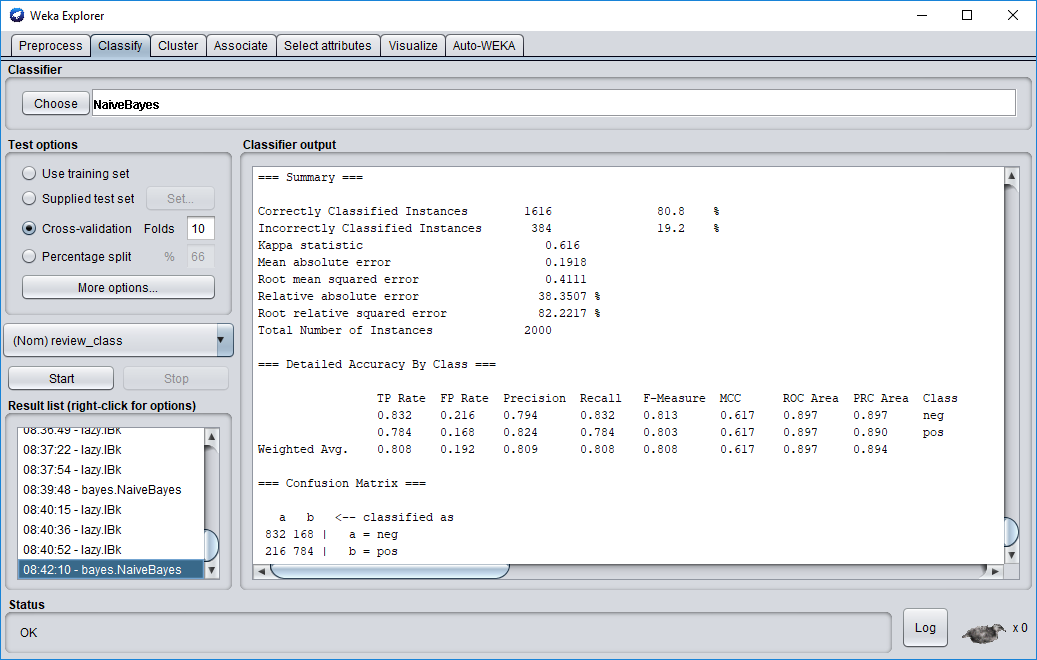


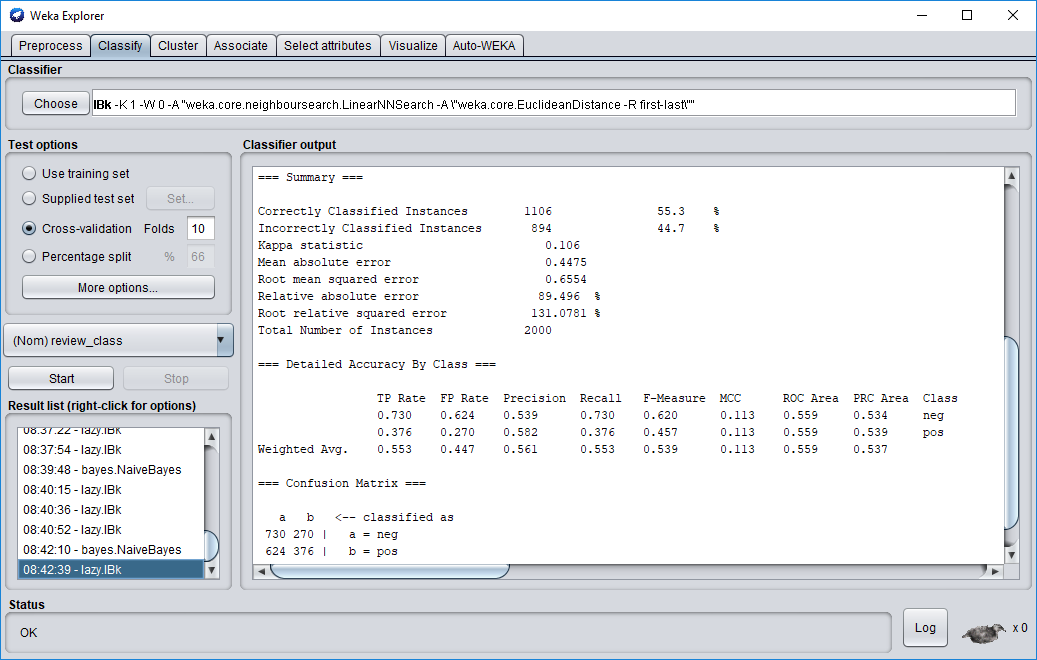


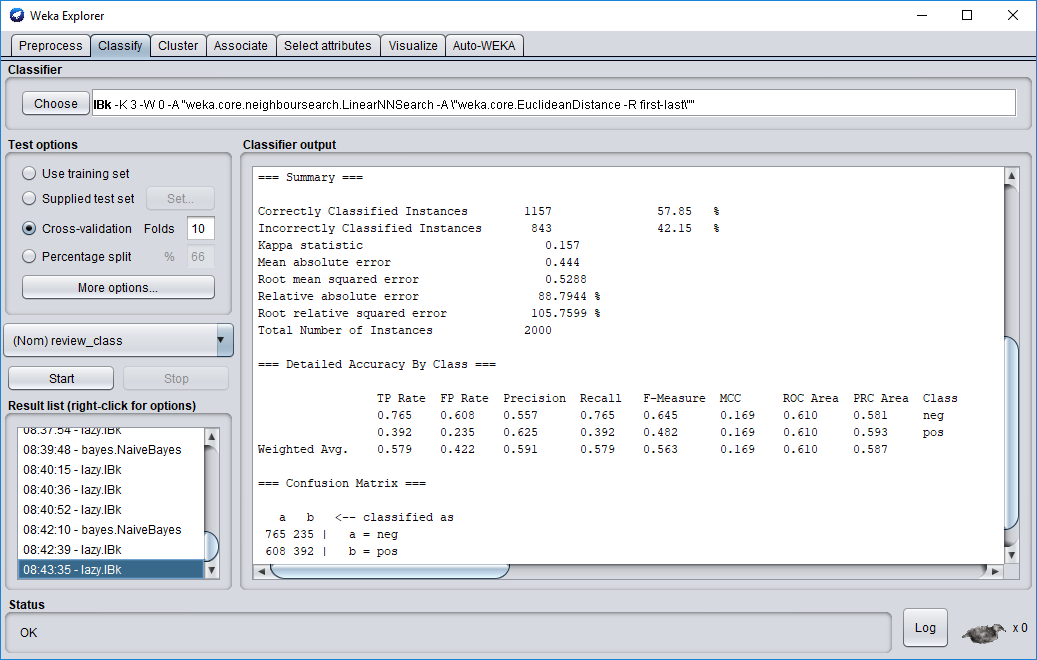


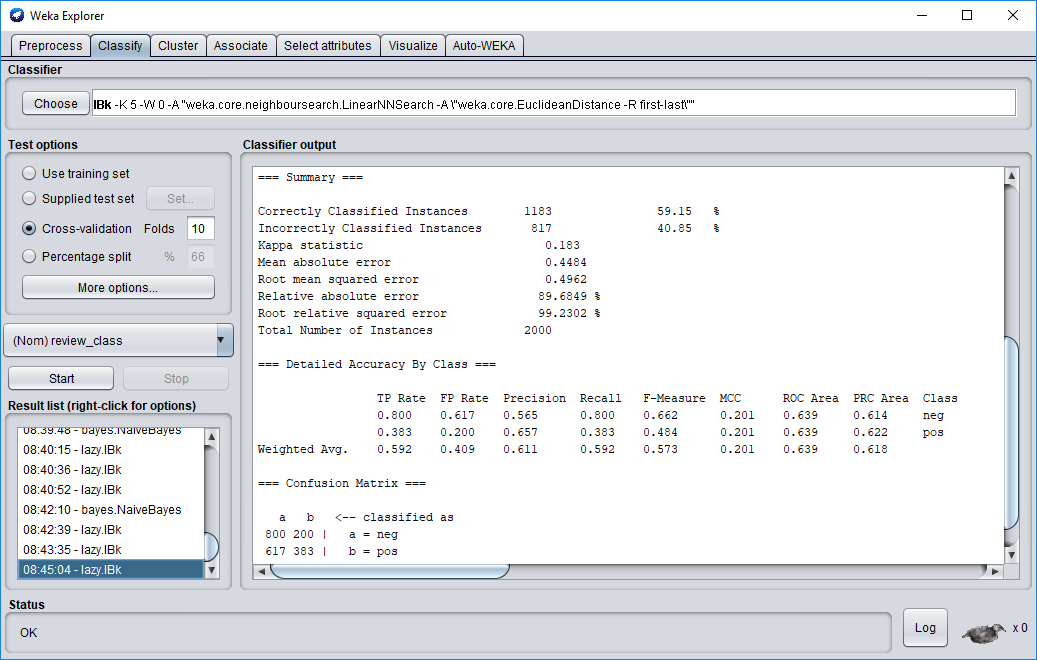


8. With no stop-words removal, stemming, or attribute selection (ie the raw tokenized data):









**Results:**

Here are the results collected in the foregoing screenshots compiled into a single table:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| # | Stop-words removal | Stemming | Attribute Selection | Naive Bayes | 1-NN | 3-NN | 5-NN |
| 1 | Y | Y | Y | 80.4 | 68.9 | 71.95 | 72.4 |
| 2 | Y | Y | N | 81.1 | 51.2 | 51.9 | 54.6 |
| 3 | Y | N | Y | 79.5 | 68.9 | 70.8 | 70.55 |
| 4 | N | Y | Y | 81.45 | 67 | 70.9 | 72.35 |
| 5 | Y | N | N | 81.1 | 51.2 | 51.9 | 54.6 |
| 6 | N | Y | N | 80.8 | 55.3 | 57.85 | 59.15 |
| 7 | N | N | Y | 81.45 | 67 | 70.9 | 72.35 |
| 8 | N | N | N | 80.8 | 55.3 | 57.85 | 59.15 |

Interestingly, I was able to recreate the author’s results for Naive Bayes, but my KNN results differ slightly for reasons I was not able to determine- perhaps version-related, as this guide was likely written for WEKA 3.7 and I am using WEKA 3.8. In general, Naive Bayes performs well for this dataset even with minimal preprocessing, whereas KNN really requires attribute selection to get acceptable results. This makes sense, as greatly reducing the dimensionality of the data should assist KNN in achieving better results.