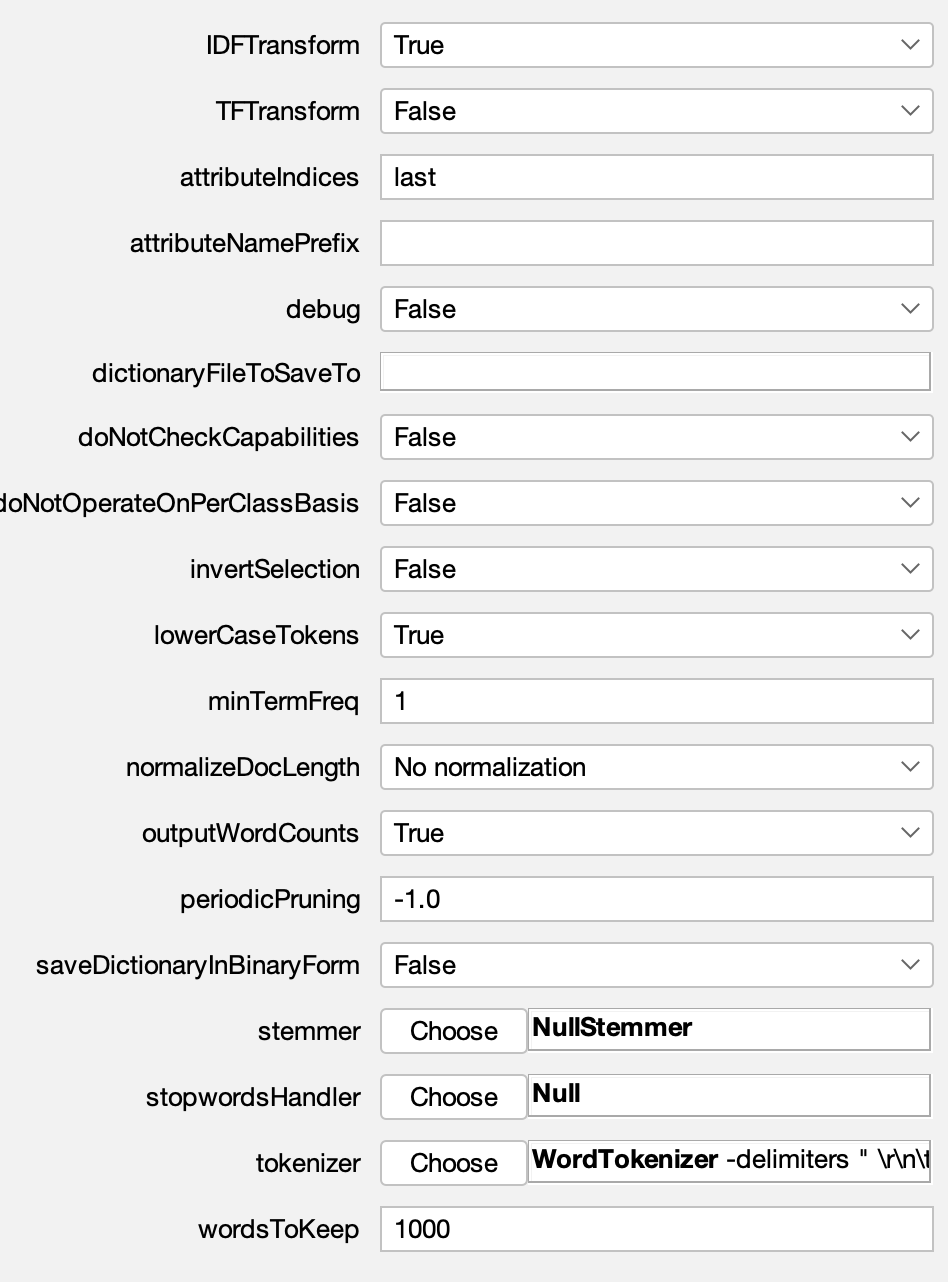
Homework 8

IST 707

Sana Khan

Pre processing:

I used Weka for this homework and choose the following parameters:



idfTransform= true to set the IDF weighting for vector values.

attributeIndices = last index to pick the correct value

the TFTransform = false so the frequency will not be a log value

minTermFreq = 1 to remove any values that occur less than once as those are often typos or irrelevant values

Lowercase = true so it can merge the lowercase and uppercase values and not count them as unique

SVM

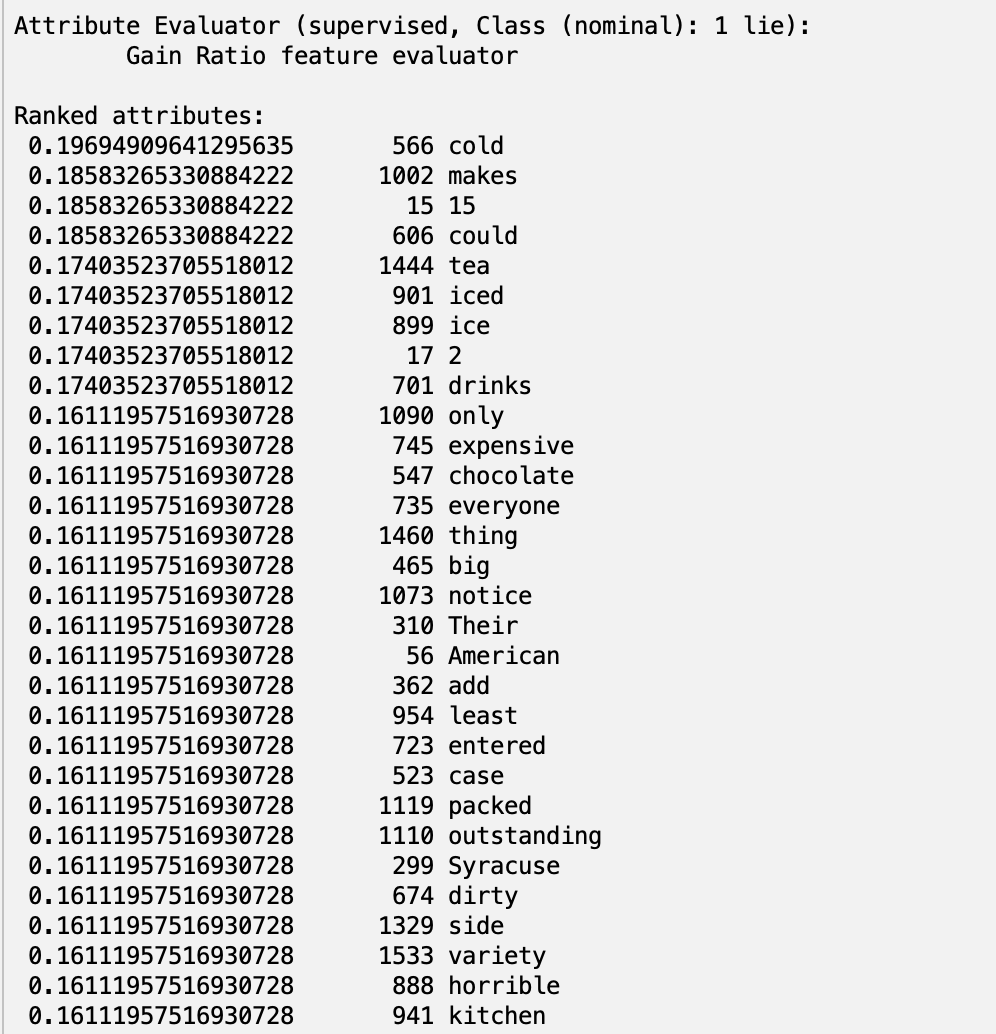
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Overall Accuracy** | **Precision in category I** | **Recall in category I** | **Precision in category II** | **Recall in category II** |
| sentiment classification | 79% | 0.846 | 0.717 | 0.755 | 0.870 |
| Lie detection | 41% | 0.409 | 0.391 | 0.417 | 0.413 |

NB

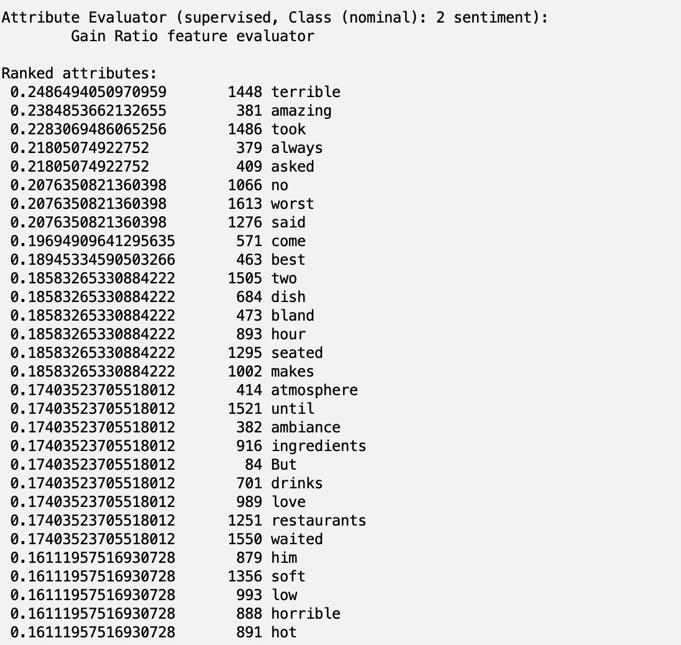
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Overall Accuracy** | **Precision in category I** | **Recall in category I** | **Precision in category II** | **Recall in category II** |
| sentiment classification | 78% | 0.906 | 0.630 | 0.717 | 0.935 |
| Lie detection | 46 % | 0. 464 | 0.565 | 0.444 | 0.348 |

Both models had higher accuracy when it came to sentiment classification. Both models had very similar results for accuracy compared to each other for both sentiment classification and lie detection. I believe sentiment classification is easier because it’s weighing each individual word as positive vs negative which can be done using a dictionary and assigning the value to the word. Lie detection is more difficult because it needs to use the context of the text to classify if it's a lie or not.

Gain Ratio for Lie detection:



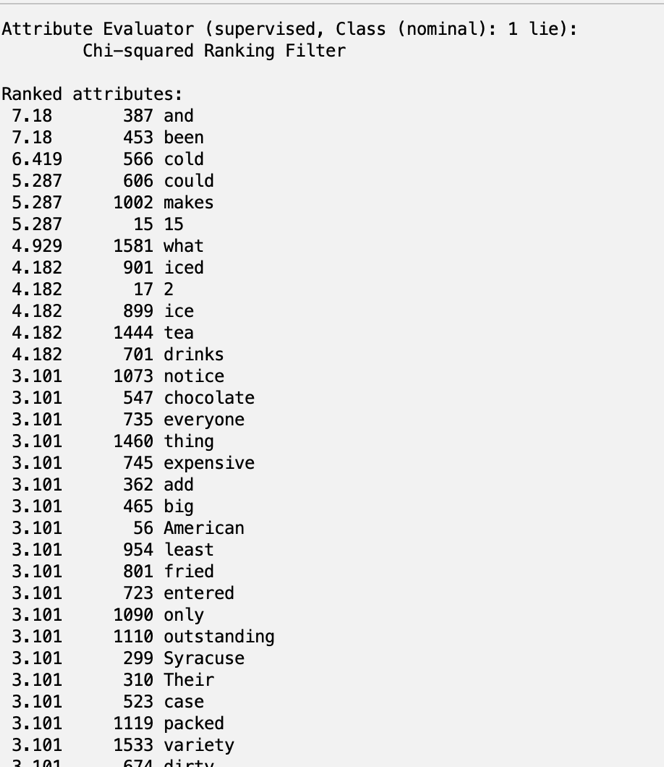
Gain Ratio for sentiment:



Based on the gain ratio, the sentiment analysis has learned the words and their sentiment better than the lie detection was able to learn. This also helps explain why the accuracy for sentiment analysis is higher.

Chi2

Lie:



Sentiment:

