Assignment week 5 Text mining - Tanja Crijns s4204999

Inter-annotator agreement

Agreement table

		Joris	
		Yes	No
Tanja	Yes	19	9
	No	10	12

Kappa calculation

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where:

- Pr(a) = act ual (measured) agreement: percentage agreed
- Pr(e) = expected (chance) agreement

Pr(a) = (19+12)/(19+9+10+12) = 31/50 = **0,62**
Pr(e) =
Joris says 'yes' 29/50 times(0,58%) and 'no' 21/50 times(42%).
Tanja says yes 'yes' 28/50 times(0,56%) and 'no' 22/50 times(0,44%).
Pr(e,yes) = 0,58 *0,56 = 0,3248*
Pr(e,no) = 0,42 0,44 = 0,1848
Pr(e) = Pr(e,yes) + Pr(e,no) = **0,5096**

K = (0,62 - 0,5096) / (1 - 0,5096) = 0,1104 / 0,4904 = 0,2251

Difficulties

During annotation, it was hard to guess the sentiment of some of the comments as they seemed neutral. There were also some comments that were marked [deleted] and I guessed that this was because of a negative remark. However, perhaps this could also have been because of excessive swearing in a positive comment. The calculation of Cohen's Kappa did not induce any difficulties.

Classifier evaluation

Given table:

	Truth: positive	Truth: negative
Call: positive	10	4
Call: negative	14	22

- a

  Precision for the positive class:

  10 / (10 + 4) = 0,71

  Recall for the positive class:

  10 / (10 + 14) = 0,42

- b

  Precision for the negative class

  22 / (14 + 22) = 0,61

  Recall for the negative class

  22 / (4 + 22) = 0,85

- c

  Macro Precision

  (0,71 + 0,61) / 2 = 0,66

  Macro Recall

  (0,42 + 0,85) / 2 = 0,635

- d

  Micro Precision

  (10 + 22) / (14 + 36) = 0.64

  Micro Recall

  (10 + 22) / (14 + 36) = 0.64