

# Exercises SW03

## Task

The task is to write a one paper report regarding the Lab3b and Lab3c for AnTeDe.

## Unsupervised Learning approach

With unsuperised learning the learning approach for analyzing the sentiment of a text is quite simple by comparing the number of positive and negative words given in a sentence to observe if the sentence has a positive or a negative sentiment. Therefore two sets of words are given containing the information if a given word is positive or negative. But the problem there is if some words are inverted or bag of concepts are used which used for thinks like **not bad**, there are one or two bad words which could in this context refer to good instead of bad sentiment. The benefit of using the unsuperised learning approach is that you don't have to label a lot of data in order to get a ok result. But the performance is significant worse than using the supervised learning approach.

## Supervised Learning approach

With supervised learning, you have the problem of the data which has to be labeled by a human. But there is the advantage of words which are not widly known as positive/negative which influence the model as well. In the exercise the non-sentiment words are even filtered out to increase the accuracy of the classifier. Afterwards a multinomial naïve bayesian algorithm is applied to the data to get a prediction of the sentiment. This could enhance the quality of the classifier with sort of **hidden knowledge** of words which don't have to be labeled explicit. In comparison with the unsuperised approach, the supervised approach gets much better results for guessing the sentiment of a given sentence.