

Machine Learning – Blatt 5
Julian Brummer, Alexander Petri, Hernando Amorocho

Task 01)

Pretty much naive Bayes, but we remove stemmed stopwords, numbers and other meaningless characters from the dataset and vocabulary.

Our vocabulary consists of the n most frequently used words in the training set, where n is a parameter, which we tuned using a third of the training set for testing.