

Machine Learning for Natural Language Understanding

12401625 Practical course 2021/2022

Exercise 7 - Challenge

sentiment classification

The task is to beat Simon's binary sentiment classifier.

Submission: **17.01.2022**

1. Implement a binary Sentiment Classifier (*neutral, not neutral*) on the given dataset. You freely decide how to prepare the data and which ML model to use for classification.
2. The challenge is considered passed if your model performs better than our baseline (a simple classifier; $F1 = 0.67$). Report at least the F1 score of your classifier.
3. If your model is better than our baseline, there is a bonus. For the best model another bonus on top.
4. The data consists of two files, a training set and a test set. Another part of the data unknown to you is held by us and is used to evaluate your trained model.
5. The .csv files contain 6 columns each: Tweet ID (the Twitter internal ID of the tweet), time (time and date of the tweet), tweet (the actual tweet; images are given as URL), user ID, language (english only), sentiment (neutral/non-neutral)
6. If you have any questions, please contact Simon Werner (werners@uni-trier.de)