1 Description of Supplementary Material

This supplementary comprises the p-values for the Nemenyi post-hoc tests performed in the main paper. The NaiveBayes and LibLinear classifiers were evaluated separately for different combinations of feature groups. The train-test evaluation resulted in 90 samples for each model. The Nemenyi-test for the combinations of the baseline and other features for the NB classifier was not significant and is not reported. Consequently, the third stage in the evaluation process for NB was not performed. Here, the annotation * indicates low significance ($p < \alpha = 0.10$), while ** and *** represent medium ($p < \alpha = 0.05$) and high significance ($p < \alpha = 0.01$). A leading “+” in front of a feature in Table 3 and 4 indicates that the feature was combined with the baseline, i.e. the first feature in the table.
Table 1: P-values from the Nemenyi test for the Weighted F-Measure, the best char- and word-n-grams, and NaiveBayes

Table 2: P-values from the Nemenyi test for the Weighted F-Measure, combinations of word n-grams and LibLinear

Table 3: P-values from the Nemenyi test for the Weighted F-Measure, features extending the best baseline, and LibLinear

Table 4: P-values from the Nemenyi test for the Weighted F-Measure, the powerset of the best features extending the best baseline, and LibLinear