

## Python exercise sheet B

This exercise sheet uses the corpora of Bob Dylan and Taylor Swift lyrics provided on the Moodle page of the course.

### Exercise B.1: Preprocessing

1. Merge the corpora and preprocess the data.
2. Train a naive Bayes classifier to differentiate between Taylor Swift and Bob Dylan lyrics.
3. Analyze the performance of the model on test data using the confusion matrix.

Hints:

- The Python package 'sklearn.model\_selection' provides the 'train\_test\_split' function.
- The Python package 'sklearn.naive\_bayes' provides the 'MultinomialNB' function.
- The Python package 'sklearn.metrics' provides the 'confusion\_matrix' function.