

## Task 5: Sentiment analysis, again

### Data

A subset of *Yelp reviews* from this study:

Zhang, Xiang, Junbo Zhao, and Yann LeCun. "Character-level convolutional networks for text classification." *Advances in neural information processing systems* 28 (2015).

### Tasks

1. Install [SentenceTransformers](#) (you will also need [fasttext](#) and [stanza](#) libraries).
2. Use train/dev/test splits of the yelp data from the previous lab.
3. Use [fasttext](#) sentence embeddings (`get_sentence_vector`, see description [here](#)) to encode the data and train a [logistic regression](#) classifier. Select the best configuration based on the validation set, apply it to the test set. (20)
  - a. Apply *TfidfVectorizer* to the training data, use idf -weighted & normalized vectors to obtain a new representation. Train & test a logistic regression classifier. (20)
4. Substitute the fasttext model with a SentenceTransformers model from the [list](#), e.g. `all-MiniLM-L6-v2`. Use SentenceTransformers' `encode` method instead of fasttext's `get_sentence_vector`. Train, tune, and test a logistic regression classifier. (15)
  - a. Experiment with another SentenceTransformer model. Motivate your choice. (15)
5. Define a Stanza pipeline with a [sentiment analysis](#) model. The Stanza model performs three-way classification. Using the validation set, decide which class to assign to the texts classified as neutral. Apply the classifier to the test set in batch mode. (15)
6. Summarize the evaluation results (accuracy on the test set) of all tested configurations, analyze/compare results. (15)