

# Team abstract

Team 15

Risto Ilomäki 596116

Aron Schulman 589796

Liisa Koivusalo 997919

Our topic is sentiment analysis with BERT. We will classify tweets related to covid restrictions and vaccinations to hate speech and no-hate speech using different BERT models and compare, which model performs the best. There should be a vast amount of tweets including loads of hate speech available in the aforementioned theme, yet it's subject to change. The tweets we classify will be in English, as the models probably perform best with them. However, if the models seem to work well, we may extend the analysis to Finnish tweets. The analysis will require labeled training data. We will search, whether there is existing training data in the context of tweets available, and if not, use labeled data from other contexts. If that training data doesn't seem suitable, we will consider manually labeling tweets for training data. The actual data to be analyzed, the tweets, can be extracted from Twitter with Twitter's API for Developers and Tweepy Python package.