

Description of the data

This database provides the data used for a study into the debate on Black Pete ("Zwarte Pete") in the Netherlands from 4th of December 2017 until 7th of May 2019. Due to the sensitive character of this data (relating to political beliefs), we only share tweet-ids and the mention sentiment (positive, negative, ambiguous) expresses towards the mentioned user in each tweet. We derived this sentiment with a fasttext machine learning model, which we will not share in line with ethical guidelines.

The file shared is a csv file with the following columns:

- `id`: stores the tweet id
- `user_mentioned`: stores the username of the use that is mention in the tweet
- `via_mention`: a dummy that represents whether the mention to the user is a 'via mention' (e.g. via @youtube). These mentions are usually automatically generated when a user shares a tweet through another platform (e.g. youtube)
- `mention_position`: indicates whether the mention takes place in the begin, body or end of the tweet
- `mention_final_prediction`: indicates the polarity of the mention (positive, negative or neutral). The mentions that were ambiguous are also categorized as neutral

Notice that each row contains a unique combination of tweet id and user mentioned. Some tweets contain several mentions, thus for these tweets there will be several rows in the datafile.

Ethical guidelines

In line with Twitter's data sharing policy and ethical guidelines for Internet research provided by The Association of Internet Researchers (IRE) and by the British Sociological Association, we do not share any information about user's stance on the Black Pete issue. We do not share the tweet text nor the `user_id` of the tweeting or mentioned user. If other researchers would like to use this data for follow up research, they will need to rehydrate the tweets through the Twitter API.