Ethical review of Social Media Sentiment of the Dutch Nitrogen Crisis

**Polarisation and Sensitivity of the Issue**

The nitrogen debate has turned out to be very polarizing in Dutch society. The political left is mainly in favour of measures limiting nitrogen deposition, including buying out farmers. The right is mainly against the measures, saying that farmers are necessary for sustenance of life. The reliance on social media posts for sentiment analysis may skew the results into the direction of the loudest social media appearances and may not reflect the opinion that is most common throughout the entire Dutch society.

*Challenge:* Social media posts have to be selected for fairness across groups. We should also map out the biases in our research in a clear manner so that the limitations can be understood by the interpreter.

**Privacy**

When using this data from social media the concern is whether this data is public or private. The question is whether data from social media can be considered data in the public or private domain. Within the ethics of research, a critical component is getting consent from the participants, however with big data sets retrieved from social media this can give more difficulties compared to research based on interviews/questionnaires. It is a given fact that all users on any platform have agreed with the terms and conditions, however participants might not be aware of their participation “using their online data” in research (Townsend & Wallace, 2016). This is an important part to take into account.

*Challenge:* Protecting the identity of participants and anonymise the data.

**Collection bias**

Within our project research is carried out on the sentiment of the Dutch nitrogen crisis, however there are several biases that can occur. Researchers using data from social media mostly use small portions of data as sample to represent larger populations. However, it is likely that a data bias can occur. Especially within these kind of data sets it can sometimes be hard collecting a representative dataset since obtaining uniform random sampling is difficult when acquiring social data. Furthermore, certain social media platforms use different algorithms and some social media platforms focus only on text based, and the other one could be focused on visualization. People have different preferences on type of media information or may behave different across platforms or contexts which is referred to as population and behavioural bias (Olteanu et al., 2019). During our sentiment analysis it is possible that the sentiment is different for each social media platform, whether it based on preference media or political ideology which could create the corresponding bias in our results.

*Challenge:* Be aware of the different types of biases that can occur during the process of social media data scraping/acquisition and how these biases can affect our results.

**Reproducibility and re-publication**

Ensuring that the results and the models of our project of data science are reproducible by others is of importance. Therefore, within scientific research sharing datasets can be considered an (ethical) obligation to ensure transparent methodologies, replicable research process and open data which contribute to scientific integrity. However, within this process there are also ethical considerations that researchers have to consider. Although there is a moral obligation to share data, there are not many social media datasets publicly available for reuse, this is due to the lack of clarity around the legality of exchanging social media data and the knowledge on repercussions (Weller & Kinder-Kurlanda, 2021). Within our research project the ethical and legal issues related to reproducibility and re use is to a lesser extent relevant since our datasets are needed for our own project assignment but afterwards are deleted.

*Challenge:* We should focus on the replicability of project results to ensure a transparent methodology but have to take into account ethical aspects of possible re-use of our dataset.

**Informed consent**

Web scraping is the automatic retrieval of data from the Web. With scraping information can be obtained from for example social media sites, this information can end up in databases and can containing users' public and private information. The subjects of this study have not given informed consent. We do not aim to publish the results. However, the results will be submitted for grading and will therefore enter the university database.

Besides, it is a serious ethical concern that web scraping goes against the terms of Facebook and NU.nl. The users could therefore expect their data not to be used for the purpose of research by third parties.

The /robots.txt of the websites mentions the following:

Facebook

# Notice: Collection of data on Facebook through automated means is prohibited unless you have express written permission from Facebook and may only be conducted for the limited purpose contained in said permission.

Nu.nl

# All copyright, related and database rights vested in the content and layout of the DPG Media websites and DPG Media apps are expressly reserved by DPG Media BV. The content of the DPG Media websites and apps is for personal, non-commercial use only and it is not permitted to gather data from the website or apps by screen scraping or any other automated method).

Furthermore, individuals can always decide to delete their posts from their account, but cannot do so from our database if wanted. This does not comply with the right to be forgotten. Privacy and anonymity are related to this ethical challenge.

*Challenge*: These restrictions don't allow us to publish our results.

**References**

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. Frontiers in big data, 2, 13.

Townsend, L., & Wallace, C. (2016). Social media research: A guide to ethics. University of Aberdeen, 1(16).

Weller, K., & Kinder-Kurlanda, K. (2021). Uncovering the Challenges in Collection, Sharing and Documentation: The Hidden Data of Social Media Research?. Proceedings of the International AAAI Conference on Web and Social Media, 9(4), 28-37. <https://doi.org/10.1609/icwsm.v9i4.14687>