Group Social 2 – Assignment on Data Ethics

Lisa Pleijt, Luc van Dijk, Wouter de Ronde, Karl Scheifinger

Assignment 1 - Ethics of Data

In this assignment, we focus on consent, privacy and risk before describing our internal project guidelines.

Introduction

In this chapter we reflect on legal and ethical issues evolving around our project. Firstly, we assess whether we satisfy the legal requirements Twitter concerning data usage. Secondly, we investigate the ethical issues that are associated with our project. Legal and ethical domains are not completely overlapping, although they arguably would be in an ideal world. After investigating whether we meet the legal requirements, we will reflect more on the "grey area" of legal but unethical conduct. Finally, we then establish what we can and need to do to stay out of that area.

Twitters Policy & Informed consent

Twitter is very transparent about their data and open in sharing it, both to users of Twitter and the collectors of the data.

Besides the legally required privacy policy, Twitter also offers a user-friendly policy summary, in which they highlight key takeaways from their privacy policy in plain, easily understandable language. They clearly explain what data is shared with whom, and how the user can turn on or off certain setting related to data-sharing, such as geo-location. Based on their transparent and honest communication, we conclude that Twitter users are able to make an informed decision about whether to use Twitter.

Of course, there still might be ethical issues hidden in the legal-lingo privacy policy. If there are, these have not raised significant outrage among experts.

In order to be able to collect and use Twitter data, it is obligatory to comply with the Developer Agreement, therefore we also agreed to these terms of use. We now have Twitter's permission to:

- *Use the Twitter API to integrate Twitter Content into your Services or conduct analysis of such Twitter Content, as explicitly approved by Twitter;*

- *Copy a reasonable amount of and display the Twitter Content on and through your Services to End Users, as permitted by this Agreement;*

(Twitter)

We can conclude that there are no legal problems regarding the collection of our data.

Consent on usage for analysis

We do not have the explicit consent of people to use their data for analysing the general sentiments of tweets. However, we applied for and received an official permission from Twitter that allows us to harvest a certain amount of data. Therefore, it can be concluded, that people allow third parties, like us, to use their data when they agree with Twitter's terms of use (also see above).

The question, whether people could have foreseen that their personal data is used for research purposes, is indeed hard to answer. On the one hand, Twitter users might not be fully aware neither about the amount of data that they generate on Twitter, nor about the research possibilities that are offered by this data. On the other hand, our study basically consists of reading a lot of Twitter posts. The idea that others read and interpret the content of your tweets is not out this world, but rather at the heart of Twitter. Therefore, we conclude that both, users and the general public, would not be surprised to learn that tweet-content was used for a city-wide sentiment analysis.

Privacy & Risk

Social media information in general can be used to create a digital profile of someone including sensitive and personal information (Lampoltshammer, 2019). Determining what actions could be harmful to one's privacy depends on the cultural and societal perception of privacy within society (Lampoltshammer, 2019) and regulations that are in place. Actions performed in this project could potentially be harmful to one's privacy because the data that is gathered can be linked to separate individuals. For example, the username of every tweet could be used to reveal a person's true name. Most usernames are very similar to people's names (Solberg et al, 2017) which would make identification relatively easy. Real names of Twitter users could be used to link that person to other databases which include their name. Secondly, the location of where a tweet was sent from can be determined very accurately. Depending on whether people have allowed the Twitter app to collect their GPS location, information about where people have been, can be retrieved very accurately (Solberg et al, 2017). This information could potentially be used to identify places which are regularly visited by a person such as workplace, a shop or home (Lampoltshammer, 2019). Moreover, the specific date and time of every tweet posted can be retrieved. Combining this information with the personal location can enable to create an individual timeline of a person. Furthermore, the contents of the tweets could potentially contain sensitive information depending on what the Twitter user tweets. Many types of information such as political orientation, personal beliefs, habits, social behavior and personal feelings can be mapped with use of an in-depth sentiment study. Therefore, it has to be noted that the purpose of this project is to explore general sentiments within the Twitter community of four Dutch cities which means that the aim is not to identify or highlight any individuals. An attempt will be made to classify different types of sentiments per city and examine the differences.

Project internal guidelines

During our project we analyse publicly available Twitter data on a large scale. As our project concerns Twitter data we save, visualize and publish the data of individual Twitter users. Therefore, setting up a cautious procedure of data harvesting, saving and processing in advance of the project is essential. Furthermore, we need to consider whether the outcomes and visualisations of this study can impact individuals whose data was analysed. It is most important to constantly take *Prima Facie* Assessment, which means that we evaluate whether violation of contextual integrity could occur during processes as saving, visualising and publishing the data (Zimmer, 2018). In the process of harvesting and pre-processing the data we anonymise the data, as for our project the individual is not relevant. We are rather interested in the larger trend of people tweeting about corona in their city. So, we do not harvest the geo-location of people who have this option set available, as it is not purposeful. What we do collect is data that is made available by users themselves, like their username, the content of the tweet and the location put in the bio. However, the fact that we store this information does not mean that it will be visualized or published as such. When publishing or visualising data we will apply the concept of data minimisation, that is we fully de-identify the data, as the only aspect of relevance is the content of the tweet, the city which it was posted in and the date of posting (Metcalf and Crawford, 2016). In this way we try to prevent from potential attacks by adversaries, as we try to minimise the consequences of such harmful scenarios (Bhattacharjee et al., 2020). So, we assume to balance the utility and precision of the data for preferred outcomes and the privacy of the individual users (Bhattacharjee et al., 2020)

Assignment 2 - Social & Ethical Aspects of Algorithms

In this assignment, the focus is put on ethical issues evolving around our algorithms.

Introduction

Data mining algorithms, which can be understood as computational tools to make sense out great amounts of data, often play an important role in big data applications. Too often, such algorithms are depicted as perfectly objective tools, despite, in reality, they are inevitably 'value-laden' (Mittelstadt et al., 2016, p.1). Enhanced operative autonomy and complexity paired with potentially crucial impacts on societies make advanced algorithms critical sources of ethical concerns. In this chapter, we follow the map offered by Mittelstadt et al. (2016) to assess the ethical concerns of the algorithm that we develop in our project.

Epistemic Concern of Misguided evidence

In the context of our project, an algorithm is used to analyse and derive sentiments from tweets sent within four Dutch cities. Although the basic parts of this algorithm have been studied, our team is unsure about the effectiveness and outcomes of this algorithm and if the outcome shows reliable evidence of specific sentiments found in the tweets. Therefore, as one of the epistemic concerns, the use of the sentiment algorithm could lead to 'misguided evidence'.

Misguided evidence refers to the notion of 'garbage in, garbage out' (Mittelstadt et al, 2016, p.5). It means that complex algorithms as well as other data-processing tools potentially show 'better' results than what could be considered reliable based on the input data. In other words, the reliability of the results can only be as reliable as the input data (Mittelstadt et al, 2016).

The risk of misguided evidence explained here is a serious threat to the outcomes of our project. One could argue to what extent it is possible to derive sentiments from tweets. The sentiment of a person can be expressed in various ways which means that the algorithm should be able to recognize and assign a sentiment label to a tweet based on a wide variety of words. The question is whether the algorithm is able to do this correctly while the words used in tweets could trick the algorithm. Think of sarcasm for example. Whether the results of the sentiment analysis are reliable enough to accurately determine the way people feel and express themselves on Twitter also depends on the bias in the algorithm (Mittelstadt et al, 2016). Especially within a sentiment study, bias of the algorithm could heavily affect the labelling of tweets related to a certain sentiment type. Although bias is unavoidable in a sentiment analysis, it is important for us to take this into account when analysing the results of the sentiment algorithm.

Normative Concerns

Following Mittelstadt et al. (2016), normative ethical concerns focus on the potential impacts of actions that are explicitly or implicitly suggested by the outcome of algorithms. To anticipate normative ethical concerns of an algorithm, it is therefore helpful to forecast its potential fields of application.

In our case, we identify at least two possible applications for our automatised sentiment analysis.

The first potential application is policy related. City-wide anti-corona measures could be based partly on the general mood of its citizens. For example, our algorithm might enable the national government or city authorities to adjust local lockdown measures in case the frustration about the status-quo expressed on Twitter exceeds a certain limit. Democratically elected policy makers would be thereby enabled to balance the strictness of measures and the moral condition of citizens, limiting the damage done to their public image.

The second potential application is to regulate police preparedness regarding anti-lockdown demonstrations or riots. Decision makers would be enabled to adjust police presence according to the general sentiment of citizens.

In general, normative ethical concerns encompass two aspects: unfair outcomes and transformative effects (Mittelstadt et al. 2016). In this assignment we put the focus on the first aspect.

Unfair outcomes describe actions which are taken based on algorithms that are widely understood as unfair and therefore lead to systemic discrimination against a certain social group. Both forecasted scenarios might lead to unfair outcomes.

Related to the first scenario, let's assume that majors want to avoid losing their face in the pandemic and therefore introduce less stringent covid measures if the Tweet-sentiment analysis reveals a high proportion of negative tweets. It is easily understandable that differences in lockdowns only based on Tweet content would be perceived as unjust by the public. Furthermore, in case the algorithm is open to wider scrutiny, citizens are likely to manipulate the sentiment analysis by purposefully posting extremely negative content, thereby trying to enforce more agreeable measures.

Related to the second, more realistic scenario, negative Twitter sentiment might lead to a higher probability of bans of demonstrations and might even encourage police force in certain cities, both of which are possible sources of discrimination. In any case, this use of our algorithm is very likely to feed the continuous public debate about the right balance between public safety and surveillance.

Traceability

Traceability in the context of using algorithms refers to the ability of a person to assess who or what is responsible for any damage in case the workings or outcomes of an algorithm did harm to something or someone (Mittelstadt et al, 2016). Potential harm caused by an algorithm is hard to track because technologies and networks of people behind an algorithm could be complex (Mittelstadt et al, 2016). In an ideal case, the algorithm is transparent enough allowing a person to identify the cause of harm and who or what is responsible for that (Mittelstadt et al, 2016). It must be noted that our algorithm has a great benefit in this respect. The accuracy of the sentiment analysis can be checked manually, because the original tweet in Dutch, as well as the English translation and the label allocated in the course of the sentiment are saved and are displayed in one row.

However, the data used in our project to carry out the sentiment analysis could be traced back to individuals. Although it is not the goal of the project to identify individuals it is important to be aware that this is possible and could possibly do harm to an individual. The algorithms used in this project have been used by many people. The details of how the algorithms works are freely available on the internet. One could argue that this can be seen as a first layer of transparency which enables a person to understand what these algorithms do with his or her data. Although obvious, it has to be stated that the outcomes of the sentiment analysis are only used for research purposes. There is no commercial interest which means that the network of people behind this project is relatively small. This also enhances traceability because the people using the algorithm are known. Not only can traceability be improved, the potential also to do harm can be reduced to. In this project, tweets used for the sentiment analysis are anonymized before the algorithm is used. This reduces the risk of harming people because information is much harder to be traced back to an individual.

Although the measures mentioned above enhance the transparency of our project and therefore improve traceability, it is to be noted that most people of whom Twitter data is used are unaware of the fact that their data is being used for such purposes. This is the case with many other algorithms as well. In case any harm occurs, it would be the responsibility of the algorithm-user or the data scientist.

Assignment 3 - Preventing Pitfalls and Responsible Data Use

In this final assignment, we discuss the potential governmental benefits and pitfalls of our algorithm.

Governmental contributions of our project

The outcomes of our project could potentially contribute to an improved city where the data could be a valuable addition to urban governance in the four cities we analysed in this project. The ultimate goal of our Twitter analysis is to identify patterns and therefore to identify new insights.

However, one must be well aware that Twitter users by itself are not at all representative for the whole population – that is every citizen living in the four cities. The Twitter dataset we use for our analysis certainly is a "selective representation of social reality" (Zook, 2017) while not everyone living in the four cities post tweets regularly. For example, some people may be too young, have no interest or do not know how to tweet. Therefore, deriving any sentiment patterns from this data comes with large uncertainty while only a small part of the population is included. This probably is the biggest remark to the outcomes of our project and important to be understood by users of our project data.

Moreover, the sentiment analysis performed in this study is designed in a way that it shows different types of sentiment in a very generic way. Sentiments of the tweets have been analysed using three types of possible sentiments being: "positive", "negative" and "neutral". This thus means that one cannot draw very detailed conclusions from this sentiment study and should therefore be careful to use this data as a justification for a change in urban governance for example.

Additionally, Twitter is just one medium where people can express their feelings and thoughts. There are countless possibilities to do this on other platforms as well, such as other social media or other types of participation which are not taken into account in this project.

In the end, one can argue that the outcomes of this project merely act as a valuable indication of sentiment in the population of Twitter users in the four cities. However, these outcomes cannot act as a standalone source of information to adapt urban governance to. Logical options for future research could be including other types of social media as well to increase the size of the dataset and amount of contributing people. Also, the sentiment analysis could be improved by identifying more types of sentiment. Both options would improve the value of this information.

Effects of our project on citizens

Just as many other urban projects that use big data sources, our outcomes could help the 'better run the city'. However, for users this is a well-known argument by various private and public actors in city management. So, what should users be aware of and what can be the pitfalls of using the Twitter data?

Though, citizens are often stipulated on the positive influence and improvement of the quality of life by more information-mediated cities, they should be aware about the consequences. Albino et al. (2015) point on some of these consequences, such as new roles for private actors, communities and citizens. And since smart city planning is not new and the use of more data driven technologies is rather a continuation of smart planning, it is important that policies and motivations to use data are transparent and specific. Goals are often to improve certain aspects, and to ‘run cities better’ in general. However, these motives are vague and do not show underlying motives of city data harvesting, which are often hidden for users. For users it is important to consider what ‘running better cities’ actually means? Better for what or whom? For example, they should consider if more visualized data metrics and dashboards are not just a methodology for urban consultants and planners to increase their authority (Zook, 2017). For our specific project, this entails that we should make the goals and partners of our project specific and to make people aware about the content and information we trace back from their ordinary everyday tweets.

Moreover, the purpose of using big data achieved from cities depends on the existing ideologies of specific place, such as free market, neoliberalism or welfare state ideas. And what project they are supporting by providing their individual data. Is the data used by private companies, by municipal institutions or by small actors who have a specific goal for the neighbourhood? These fundamental ways of thinking shape the way in which data can be used. Citizens are obviously not always aware about these underlying goals and incentives. Therefore, we should confront users about these aspects in using platforms (Zook, 2017).

Maybe it is not only the responsibility of individual project managers, but it should be noticed by the government as so many people are using data collecting services. Individual data collection has intensified largely by previous decades (CCTV controlling, Google maps/now, smart watches, etc). Citizens are intensively sensed and measured within smart cities by the increase of sensors and devices. Citizens should be made more aware about this.

In our specific case this mean that we must be open and transparent about the way in which metrics and individual data is used, and what conclusion can be drawn from it. We must stipulate the ideology and institution we use. Citizens can then decide individually whether they want to contribute to this movement.

Eventually, it is important to make the user mindful about the consequences of using platforms, what their personal data provides us and what movements or ideologies they support by using the services. Providing cities with private data is not wrong in essence, citizens should just be more mindful and informed about it.

References

Albino V, Berardi U and Dangelico RM (2015) Smart cities: Definitions, dimensions, performance, and initiatives. Journal of Urban Technology 22(1): 3–21.

Bhattacharjee, K., Chen, M., & Dasgupta, A. (2020). Privacy-Preserving Data Visualization: Reflections on the State of the Art and Research Opportunities. Computer Graphics Forum, 39(3), 675–692. https://doi.org/10.1111/cgf.14032

Lampoltshammer, T. J., & Eibl, G. (2019). Impact of Anonymization on Sentiment Analysis of Twitter Postings. In Data Science–Analytics and Applications (pp. 41-48). Springer Vieweg, Wiesbaden.

Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. Big Data & Society, 3(1), 205395171665021. https://doi.org/10.1177/2053951716650211

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Solberg Søilen, K., Tontini, G., & Aagerup, U. (2017). The perception of useful information derived from Twitter: A survey of professionals. Journal of Intelligence Studies in Business, 7(3), 50-61.

Twitter; https://developer.twitter.com/en/developer-terms/agreement [Access 01.04.2021]

Zimmer, M. (2018). Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity. Social Media + Society, 4(2), 205630511876830. https://doi.org/10.1177/2056305118768300

Zook, M. (2017). Crowd-sourcing the smart city: Using big geosocial media metrics in urban governance. Big Data & Society, 4(1), 205395171769438. https://doi.org/10.1177/2053951717694384