Ethics part

Reflection note on ethics of data

Our main dataset is from Twitter, a social network which allows users to post and interact using short messages known as Tweets. These messages can contain more than text, they can contain photos or links, and they can have more information attached, such as a geographical location or the time at which it has been published. All this mentioned information may or may not include personal data, but at least it can give hints of personal details like basic information and routines.

If we look at the Twitter dataset we have been given, we can see different columns. One is “tweet_id” which is a unique code given to each Tweet. We also have the text, the Tweet date, the Tweet name (refers to the user name) and the coordinates expressed in latitude and longitude. Hence, we have access to a lot of personal information about the users. We could make some assumptions; for example, if we see that an individual has tweeted many times from the same place, we could imagine it is their home, their workplace, or at least, that you can find them there easily.

It is debatable whether using specific information on individuals is violating individual privacy or not, as when signing up on Twitter users read and accept Twitter's mandatory conditions where they essentially give it up. Nonetheless, it is common to overlook and just accept these without reading them. Therefore, it is not strange to feel later that your privacy has been violated although you consented these actions. However, in a broader perspective than violating the privacy of individuals, data analytics enables sensemaking of small to vast volumes of collected data from diverse fields facilitating their clustering into meaningful groups or classes (Mittelsdaldt, 2017). This is accomplished by identification and/or understanding of individuals through the recognition of small patterns or causal correlations. For instance, individuals can be linked by offline identifiers (i.e. ethnicity, age, location, etc.) and inferred behavioural characteristics, making it possible to make decisions and predictions at a group rather than individual level (Grindrod, 2014; Mittelsdaldt, 2017). The chosen parameters for categorization (e.g. behaviour, preferences or gender) vary depending on the pursued end and condition the final outcome and gained insight.

Our current project must make an in-depth exploration of the potential ethical boundaries that may be crossed to ensure that no violation is committed. Despite the lack of information on the identities of the social media users and the seemingly innocuous potential that the datasets present, the use we have in mind for them can still turn out to pose ethical threats. Social media platforms make explicit the uses that will be given to their users´ data: however, they fail to prevent third parties from gaining further information on users by combining datasets from different sources. These combinations can result in the drawing of potentially harmful inferences on individuals and the groups to which they belong (Kammourieh et al. 2016); even though users have not given consent and may not want this information to be of public dominion; hence, they are subject to privacy violations.

By combining meteorological and Twitter data, our project will allow us to cluster passers-by and inhabitants of Amsterdam´s centre according to various variables. These include: behaviour under different moods and meteorological conditions; moods under different meteorological conditions and days of the week; and activities over the week and under different moods and weather. Our datasets do not include the users´ identities, so any privacy or consent violation will only affect group ethics. Third parties, such as companies and politicians, may use the gained knowledge for personal interests. Regarding possible unethical use of our results: companies will gain a better knowledge of the market that can enable them to exploit customers financial resources more effectively; political parties may adjust their messages´ frames and way of delivery (as our insight allows them to forecast peoples´ activities/locations) depending on the weather and day of the week.

In both individual and group data usage, the definition and clarity of consent is crucial. Voluntary, specific, and informed consent is the foundation of proper personal data privacy practice (Cheung, 2017). Under conservative understanding, it is given on a one-on-one basis, for a single study and for a specific purpose. Personal data use and sharing within research projects are governed by it and the right of withdrawal which is increasingly neglected in big data projects. As researchers would have to ask thousands to millions of people for consent and it is difficult to find and remove a person's data from an anonymized/pooled data set, it becomes clear why this is unfeasible. Because our project concerns big data harvested from Twitter, we need to shift from the conventional consent paradigm to a framework of accountability by taking into consideration risk assessment and potential harm to Twitter users making up our data set. Although their information is already in the public domain, our analyses could still potentially amplify it in a way users have not agreed with. Therefore, it is our goal to execute this research while preserving the values behind the consent including autonomy, fairness, and propriety. To ensure this, we assessed Twitter's 'Privacy Policy' and determined if it aligns with our research goals.

Immediately upon opening this document we were faced with a number of highlighted statements informing us that Tweets are public, "viewable and searchable by anyone around the world", and that user's personal information including IP address is collected by Twitter (Twitter, 2020a). Moreover, the information visible on user profiles such as username, Tweets and their date/time stamp, interactions with other users, and geo-location if they choose so, is public and visible to everyone. Users give their consent to this upon signing up by accepting Twitter's mandatory terms of service. Twitter also reserves the right to disclose publicly visible information broadly for various purposes through its APIs. Therefore, users are encouraged to be responsible in sharing personal and sensitive information in their Tweets. In theory, they should be aware that their Tweets and personal information could be used for research purposes although this is nowhere stated as an example. Furthermore, there are a couple of things to keep in mind regarding our research goals and the 'Developer Agreement' which a researcher accepts in order to harvest the Twitter data (Twitter, 2020b). If we would want to provide the government with the results of our analyses as the End User with the purpose of e.g. urban planning, we would have to acquire permission from Twitter first by submitting the case for reviewing. In this way Twitter protects its users by ensuring that the data will not be used by the government for gathering intelligence or surveillance of individuals. Another potential hurdle would be that a researcher has to delete the location data and/or a Tweet when the user removes these as required by the 'Removals' clause. This is not possible to track for our large data set, which is why we would mitigate this by removing such data upon request of Twitter or users themselves. Ultimately, as Twitter is ensuring user protection and keeping those in possession of their data accountable, we determined that we could proceed with our analyses.

Using "publicly available data" for a study can result in ethical challenges regarding privacy and consent, as is the case with our project. It is important to try to minimize and avoid these possible challenges in an early stage of the research, for it to not erupt at a late stage as witnessed at the case of the OkCupid Study (Zimmer, 2016). One way to cut the stem at an early stage can be done by de-identifying the used data. De-identification can be described as a "the process used to prevent someone's personal identity from being revealed" (WikipediaEntery, n.d.a). This method thus deals with the privacy aspect of the ethical challenges. In our Twitter dataset, de-identification could be done by censoring out all the usernames. But the combination of the text, location and day of the Tweet could in some cases still be a risk for re-identification. Re-identification is "the practice of matching anonymous data with publicly available information, in order to discover the individual to which the data belong to" (WikipediaEnter, n.d.a). So, to prevent such re-identification more of the data must be de-identified. For our sentiment analysis all words which include characters will be removed from the text column of the dataset. Afterwards, the remaining words will be brought back to their stem forms (e.g. eaten becomes eat and cafes becomes cafe). This will be done during the data cleaning parts; one of the earliest stages of the study. Removing and transforming words of the Twitter text, could help prevent this re-identification, as we believe it will be harder to trace back the person behind the text when the text is incomplete. This will hopefully be enough de-identification to avoid re-identification for our study. Therefore, with the method of de-identification in an early stage of our study, we try to ensure the privacy of the Twitter users as we are not able to ask all the participants for their consent on this specific study.

Reflection note on ethics of algorithms

A literature study by Mittelstadt et al. (2016) constructed a map of possible ethical problems caused by algorithmic decision-making. This map contains six different kinds of ethical concerns; three epistemic, two normative and one overarching concern. Below we will argue which of the concerns of each type are the most relevant for our project.

Epistemic concerns

The study article mentions three types of epistemic concerns: inconclusive, inscrutable, and misguided evidence. All three must be considered in our project as we want to find out whether we can create knowledge and actionable insights for end-users such as businesses and the government. However, as for now our research is of exploratory nature and we intend on being fully transparent in showcasing algorithms used, we find the possibility of misguided evidence the most concerning. Any algorithm we employ using flawed data can only produce flawed outputs or misguided evidence. Using unreliable low quality and incomplete data from questionable sources without proper prior assessment would further pose a risk of making incorrect decisions rather than enhance the decision making of end-users.

The use of Twitter data for research purposes comes with several limitations and biases in sampling. Firstly, the Twitter user representativeness of the general public is questionable due to its unequal usage distribution across different age groups as it is mainly attractive to the extroverted and younger technology-literate crowds (Bright, Margetts, Hale, & Yasseri, 2014). Nonetheless, Twitter data in general provides a great population coverage of the younger crowd which represents at least half of the permanent residents of Amsterdam (UrbiStat, 2020) and likely even more so the tourists. It should be noted that our data includes only geo-tagged Tweets and it is known that only a small portion of users activate the geo-tagging function (Sloan & Quan-Haase, 2017). This all should be made clear to the end-users as it would not be fair to generalize the results generated by algorithms to the whole population as this would inevitably lead to bias. Furthermore, we are aware that there are multiple languages present in our data as is expected in a multicultural and tourist-popular city as Amsterdam. Due to time limitation we will only focus on Tweets written in English, but we are aware that this will put constraints on generalizability and utility of our findings.

Another way our data could affect the workings of algorithms is during sentiment analysis, as the algorithms will fail in detecting jokes, sarcasm, and irony in the Tweets. There is also a real concern about the extent of the sentiment measured being of “real” humans since many organisations, governments, and businesses are actively trying to influence it. If these are abundant in our data set they would “pollute” it by adding noise and skew the credibility of the generated insights. Therefore, we should assess our data and try to remove their content as much as reasonably possible.

Normative concerns

Mittelstadt et al. (2016) describes two types of normative concerns; unfair outcomes and transformative effects. This first concern is about the fairness of the algorithms; specifically, the concern about 'unfair' outcomes causing discrimination. We do not consider this last part applicable for our project as we do not categorise the Twitter users ethnically. Although this still could be argued as we have selected the tweets on language; thereby excluding the users writing in other languages form our study. But we believe this is more of an ethical concern about misguided evidence that could possibly lead to bias, as described above, than a real (un)fairness issue. The unfair outcomes concern is not only about the fairness of the algorithm itself; the input, but also about the effects of the actions resulting from the algorithm; the output (Mittelstadt et al., 2016). In other words, there is a concern about the fairness of the effects of decision-making that is based on the used algorithms.

The results of our project are aimed to help (small) business owners to gain insight in the activities of their possible customers in relation to different weather types. So, the decision-making will in our case be, adapting a marketing strategy. We will not categorise the possible customers into different groups as we simply do not know the identity of twitter users, except of some of their usernames, but these were deleted from our dataset due to privacy reasons. Nor will we investigate the relation between groups and different activities. The fact that the Twitter users are not categorized, does not immediately imply that the outcomes of our study are applicable for the whole society. This is because we do not know 'our sample group', the Twitter users can be any mixed possible group of people, but they can still be dominated by a certain group. For example, tourists, adolescence or well-educated people are more likely to use Twitter and write their tweets in English, therefore they might dominate our sample group. In such a case other groups, such as children, elderly and locals will be excluded from the study. For a small business owner as an ice cream man, our study would seem quite interesting at first; as he is reliable on the weather, which we include in our study. But as his main target group is children; who might be underrepresented in our study, it might not be wise to base his marketing strategy on our study after all. This is just a small example in which the ice cream man is the one that would be disadvantaged by our study. But in a hypothetical situation where all small business owners of Amsterdam would use our study to base their marketing strategy on, all marketing campaigns will be addressed for 'our sample group', thereby excluding the children, elderly and locals. On the long run such a tourist focused marketing strategy can result in so-called Mcdonnaldization of the market; a homogenization of the market due to globalization (Ritzer, 1992). This because the small local businessowners will have their main focus on non-locals just as most of the bigger companies located in the city centre of Amsterdam. If this situation were indeed the case for our study, it could also be considered unfair as not all types of groups are included, which can be argued as a kind of favouritism of one group over another.

It is important to note that this unfairness in favouritism would be completely unintentional; as we are only provided with a certain dataset with limited tweets and a lot of noise (the different languages) and will not categorise our data. This makes it a good example of the real-world ethical algorithmic issue on fairness; raising the question can the researchers, ourselves in this case, be held accountable if their research leads to favouritism?

On the one hand we should be; as we are the ones presenting a research for decision-making that leads to favouritism. On the other hand, our research is not intended to be favourable in any way and we, as researchers, have not enough control over our used data to entirely be sure that one kind of group is not favoured over the other; making us as the researchers not aware of our deficiency. Hence, unfair outcomes can even in our project still be an ethical concern and we should therefore be critical about how to publish our results. As our results might just only be relevant for a certain part of society and not the whole society.

Traceability

Traceability is an important aspect to consider as it establishes the identity and provenance of a product, or in our case the code and output, thereby facilitating pinpointing of the agent to be held responsible when there is a failure (Mittelstadt et al., 2016). As the 'designers' or rather 'users' of algorithms designed by someone else we need to examine the algorithms' potential effects and consequences of their malfunctioning. There is a limit, however, to which we can do this as the Python libraries we use could sometimes be considered 'black boxes' due to the complexity and volume of their code.

In order to foster traceability in our work we will keep track of our metadata by documenting it in our personal Python notebooks including all data inputs, outputs, and processes taking place. We will also actively search for errors and possible bias-introducing code of algorithms. This will be done by investigating the different algorithms we use that come from 'pre-defined' codes; algorithms that are part of python libraries. To make our code and algorithms used open for scrutiny and criticism, we would upload it to the GitHub along with all the project data. We still have not decided whether we will use learning algorithms. In case we do, we must be cautious in our choices and discuss them with our lecturers, as these algorithms are black-boxes in which decision-making rules have been written by third parties; dispersing responsibility too much for anyone to be held accountable.

Final thoughts

Although it is clear that our data will be flawed to a certain degree and a perfect algorithm for sentiment analysis does not exist, Tweet content and geo-location data still provide a possibility of measuring instant reactions which might give us an idea of how users respond to certain weather conditions. This is advantageous when compared to conventional data collection techniques including one-on-one interviews and polls as Tweets come from individuals unaware of and unaffected by the researcher's observation. Additionally, we could compare the Twitter data to other data sources such as Flickr in order to check whether their findings reinforce each other and thereby instil some trust in the utility of Twitter data.

As for the normative concern; in order to reduce possible unfair outcomes, we as the researchers should investigate our data source; Twitter, more closely to identify its user group. If the user group is indeed not representable, the results must be published with a note on whom they apply on. This would then exclude a part of society, which can also be categorized as favouritism of a group, but at least it will make sure we as the researchers are as fair as possible.

For the traceability concern; we will continue tracking our data processing and if we decide to use the learning algorithms they will be handled extra carefully.

Reflection note on ethics of governance

Our project´s objective is testing the hypothesis that "people´s moods and activities are influenced by the weather". Close to the project´s end, we have not found any sound evidence supporting the claim, so it seems likely that proving the hypothesis to be wrong will be our work´s conclusion. However, it must be noted that, as literature suggests, we may well have arrived at the wrong conclusion. The lack of the experience and knowledge required to flawlessly carry out such an ambitious undertaking surely means that some steps may have been overlooked and that our statistical analysis can be improved. Hence, taking the former into account, we will reflect on the potential city insights that may be gained through our inquiries and what those profiteering from these gains must have in mind to ensure a responsible use of these project´s findings and prevent pitfalls.

The combination, coordination and integration of ICT with cities´ infrastructures is at the root of the "smart city" concept (Zook, 2017). Collection and analysis of the generated data makes it possible to gain insight into a vast amount of novel urban and environmental governance possibilities that were previously intangible and hold the potential to utterly transform lives within cities. Organizational changes mean that both private actors and citizens will have to adopt new roles (Zook, 2017). In our project´s case, finding out causal correlations between passers-by activities (inferred from the locations from where tweets are published) and moods, and weather (which moves over a wide yearly spectrum in Amsterdam) can render useful insights: which areas are more or less concurred during different periods/seasons of the year (depending on predominant climatology); which kind of establishments are more visited and thereby which commercial activities are more profitable; and to what extent does the weather´s influence over mood condition all the former. The latter knowledge can be taken advantage of in many ways; such as investors looking to start or expand a business and policy makers trying to improve the city´s organization and citizens´ lives.

Improved understanding of cities may increase the private sectors´ role in urban governance. People´s behaviours and moods under different circumstances can be very valuable for businesses and commercial operations´ strategies. A thorough understanding of what incentives people look for under different conditions is a useful asset for adapting one´s offer of services (e.g. if locals offering hot beverages in a comfortable space receive more affluence on rainy days, other shops can add that service to its current ones so that more customers feel encouraged to enter) or for maximizing profits by tailoring offers to peoples´ changing preferences. Concerning public governance, citizens´ activities under different weathers and moods may render a detailed portrait of passers-by behaviours. This knowledge gives insight into public areas´ potential and therefore contribute to better plan cities´ organization (e.g. public spaces where affluence is heavily conditioned by weather can be adapted/transformed to attend policy makers' needs, such as establishment of green areas or of public service infrastructure if the place does not have commercial potential because of its low affluence over a significant number of days per year. This knowledge can also be used for redesigning cities´ circulation systems (moving cars from places with high activity to those where it is less intense). It is also important to note that this use of information positions citizens as passive actors in urban governance, providing direct input to decision-makers.

Users of the project findings must keep in mind that the conclusions obtained from a project of these characteristics are not fully representative of society. The data used comes from Twitter active users (it cannot even be reduced to those who have the app, but to those who publish tweets). Taking into account that older generations are less acquainted with social media platforms and that Twitter does only account for a fraction of those who are active, the conclusions appear to be representative of a selective group of the population. Therefore, there is an overrepresentation of certain groups´ behaviours and activities. Besides, some activities may also be more visible than others; for example, if planners trying to redesign traffic routes depending on citizens´ daily movements should keep in mind that cyclists´ routes will be underrepresented in comparison to those who walk or take cabs). As data is performative, meaning that it defines what it describes, being aware of its limitations (i.e. its selective representativeness) is crucial to ensure that policy makers can attain better outcomes (by drawing on further sources when necessary).

Bibliography

Bright, J., Margetts, H., Hale, S., & Yasseri, T. (2014). *The use of social media for research and analysis: A feasibility study*. Crown.

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.

Ritzer, G. (1992). *The McDonaldization of society*. Pine Forge Press.

Sloan, L., & Quan-Haase, A. (2017). *The SAGE handbook of social media research methods*. SAGE.

UrbiStat. (2020). *Age classes by genderMunicipality of Amsterdam, old-age index and average age of residents*. Retrieved on 7th of April, 2020, from https://ugeo.urbistat.com/AdminStat/en/nl/demografia/eta/amsterdam/23055764/4

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