

Leveraging Digital Trace Data for Electoral Predictions: A Critical Analysis

In a time characterized by the influence of developing technologies, the potential of digital trace data in predicting electoral outcomes has gotten some attention. Digital footprints left behind by individuals navigating the World Wide Web offer a promising avenue for understanding voter behavior and preferences. The allure of real-time insights into voter sentiments and trends is undeniable. Supporters believe that leveraging digital trace data could revolutionize electoral forecasting, providing nuanced understandings of voter behavior while democratizing access to information. Conversely, skeptics caution against the use of digital trace data, citing concerns regarding privacy, and algorithmic biases.

With this context, this paper will be reviewing a study that attempts to do exactly what has been proposed: predict voters in an upcoming election by using digital trace data. This paper is structured into four distinct sections. The first section, Analysis of Methods, will delve into the methods used throughout the paper while offering novel insights via simulated statistics through similar methods. By understanding the strengths and limitations of these methods, I argue against the usage of digital trace data for these purposes. The second section, Analysis of Normative Considerations, will evaluate the ethical dimensions of utilizing digital trace data in predicting electoral outcomes. Drawing upon deontological principles discussed in class, I will assess the implications of such practices in terms of the rights of persons, and the universality of these practices.

By engaging with both methodological nuances and normative considerations, I hope to contribute to informed discussions surrounding the intersection of technology, politics, and philosophy. This analysis underscores the imperative of adopting a responsible and ethically grounded approach to the utilization of digital trace data in electoral predictions and the

importance of adopting ethical frameworks for the implementation of new statistical methods in any context.

Analysis of Methods

The authors attempt to predict how people will vote in an upcoming German election using digital trace data. Additionally, they utilized survey data provided by the respondents to compare the results of their algorithm to the results of typical predictive measures about voters. The researchers started by having respondents install add-ons to their computers and phones that would report back the complete URL, the domain of the website, the device they used to access the website, and the amount of time spent on the website. Respondents were able to turn these off at any time, however, the researchers suspect based on the responses that they did not. Data was obtained from July to October. They created 4 different Boolean variables based on survey results:

Undecided - Did they respond that they were undecided in surveys?

Voted - Did they respond that they voted in surveys?

AFD - Did they respond that they voted for this party?

Greens - Did they respond that they voted for this party?

They chose the most polarized parties because they believe having more polarization between the two parties will make it easier to classify an individual based on their digital trace data. They have three different block predictors with varying predictive abilities. The first includes information about the device and when it is used. The second includes the duration and frequency with which respondents access the 50 most used news domains. The third block of

predictors features websites/apps that were used more than 80 times for over a minute. After combining everything there is a total of 11999 predictors used by their algorithm from the digital trace data.

When creating their model, the researchers decided on using XGBoost because it can filter through predictors to select only the best variables in the model-building process. These create various decision trees, then use the issues from the previous tree to create a new tree that should be an improved model. This model is extremely good at classification and ranking. Better put, the XGBoost is capable of trimming down a model so that it is small enough to be interpretable and retains strong predictive powers. Allowing for users to reach the simplest model that will still produce accurate predictions. The researchers created, “a total of 35 XGBoost models, 6 for each of the political outcome variables (undecided, voted, AfD, and Greens) and 1 for each of the sociodemographic outcome variables (age, gender, net income, marital status, federal state, childless, number of children in household, and employment status).” (Bach, L. R., et al.2019). They add that these groups allow them to gain insights at different levels. Group 1 just has demographic data, group 2 just has digital trace data, groups 3-5 have different combinations of each, and group 6 contains everything.

The researchers find that their models are unable to perform at the level of typical socio-demographic models made currently. They add that digital trace data struggles greatly at predicting how undecided voters will vote, but it does do a better job with more populist parties. They find that they are unable to predict how someone will vote by using digital trace data. They add that they can consistently predict the age and gender of respondents using their digital trace data. Additionally, they believe that if they had more resources to devote, like a major company like Google, they may have better luck at predicting these measures.

Analysis of Normative Concern

As mentioned in the introduction, I will analyze this paper from a deontological perspective. Before I can apply this perspective to the discussions from this paper, however, I will outline some of the core principles of deontology. Deontology is a moral framework emphasizing the importance of rules, duties, and principles in guiding moral decision-making. We learned in class that the word "deontology" comes from the Greek word "deon," meaning duty, and "logos," meaning science or study. This is to say that a deontologist finds there is a duty to only care about the positive or negative reasoning behind any action, rather than just the outcomes of those actions. Deontology has a strong emphasis on the ability to universalize an action as being morally permissible. This means that doing some action can be allowed in all contexts. The example that always comes up is lying. You may run into an instance where lying to somebody will offer them better results, like telling a kid their dog is not dead, but went to a farm somewhere. This outcome is superior to upsetting the young kid, but a deontologist would argue against lying in any context. This is because lying cannot be universalized. Sure, in some situations you can keep a kid from crying, but if you tell yourself that lying is universally permissible, then you are sure to run into trouble. Another important piece of deontology is having respect for other people. Meaning, understanding that people are important as ends in themselves, and should not be used purely as means to get to an end.

In the context of the paper I am looking at, a deontologist would likely be against the usage of digital trace data for predicting electoral outcomes because using these data utilizes persons as a means to an end, and because you cannot apply, nor allow everyone access to these types of data. As discussed previously, in this study, people are asked to install an add-on to their

computers and phones that would report back the complete URL, the domain of the website, the device they used to access the website, and the amount of time spent on the websites.

Using digital trace data violates individual autonomy. In the case of this study, their data was acquired from consenting individuals, however, in many cases, the digital trace data you leave are being accessed without informed consent. Using these data is undoubtedly an invasion of an individual's privacy. Using these data for this purpose treats people as objects meant for political analysis, and that is it. As previously stated, using individuals' data as a means to get your prediction, is not treating those people as though they are ends themselves. Even if you can properly predict how someone will vote, using their data for prediction does not provide them with any benefit. A deontologist would never argue that you could universalize this process.

Finally, there are transparency concerns with the usage of this data, specifically as it applies to machine learning. As far as the user is concerned, there is no way of knowing how exactly their data is being acquired, nor what it is being used for. If there is no informed consent regarding the collection of individual online data, then there should not be collection of their online data. Additionally, there are transparency concerns with how the data is used after it has been collected. Thinking about it, an individual does not typically understand how machine learning algorithms work, let alone what they are attempting to do. Even if the only thing the algorithm is doing is predicting how you will vote based on your online activity, if the individual does not know this is what their data is being used for, then there is a lack of informed consent. A deontologist would agree that there is a moral duty to be honest, and in this context, honesty exists in the form of transparency with those whose data is being implemented.

The universality concern can manifest itself in a couple of ways in the context of digital trace data. As I have mentioned you could not universalize a process that involves acquiring

these data without informed consent. Without that consent, you are, in effect, using those people and their data as a means to an end. Another way to phrase the universality concern is in the context of everyone having access. This means that if everyone were to have access to digital trace data and these predictive methods. In this instance, a deontologist would find that if you do not believe there should be universal access to this information, then you could not approve of its usage. The other example, the lack of transparency concern, also could not be universalized because you are effectively withholding important information from those whose data is being used. A deontologist would only advocate for the usage of digital trace data if there were substantial explanations of how the data was acquired, what it will be used for, and how that will benefit the individual whose data is being acted upon.

Conclusion

Digital trace data is already being used to fundamentally shape every user's online experience. Algorithms use this data to personalize content, recommend products, and tailor advertisements. However, the ethical implications of this practice are increasingly being questioned, especially regarding issues of privacy, and consent. While digital trace data has been utilized for predictive tasks, including electoral outcomes, The paper's assessment highlights significant flaws in using such data for this purpose. The inability to build a model that outperforms current strategies raises doubts about the effectiveness and reliability of using digital trace data for electoral predictions. Additionally, using deontological viewpoints, I found that using digital trace data without explicit informed consent is not permissible. Considering the limitations in predictive accuracy, and ethical concerns regarding consent, I conclude that digital trace data is not worth using for electoral predictions. This position reflects a cautious approach

grounded in ethical considerations and a critical evaluation of the practical efficacy of relying on digital trace data for sensitive tasks such as electoral forecasting.