Trends on Twitter: An analysis of the frequency in the use of 'Climate Change' on Twitter, outside of American election time and during American election time.

Sabin Safarova S5206162

Abstract

This study examines the dynamic language evolution on Twitter, with a focus on the sentiment and frequency of the words "climate change" outside and during US election time. The study examines language evolution using quantitative methodologies, utilizing a substantial dataset that spans over 13 years, and further examining Tweets via restricting the results. The predictions of the outcome of this research show an increased use of the words "climate change" during the American elections. Analyzing the frequency of words, like "climate change," on Twitter over time offers a valuable perspective on how the public perceives a topic of discussion, the media's coverage of it, and in this case, how the elections have affected public opinion and interest in climate change.

1 Introduction

Language on social media platforms changes quickly in our digital society, reflecting changes in our culture. Studying this change in language is important as it looks at how people incorporate popular terms from the media into their writing and speech. This, in turn, provides an insight into the dynamic relationship that exists between online communication and larger cultural trends in our society. Gaining an understanding of this change not only improves our understanding of the use of modern language and its effects on society, but it's also beneficial for understanding the impact that cultural trends can have. The evolution of language on social media can be interpreted as a lot of different things. This study will focus on the frequency of words during certain periods or societal happenings. More specifically, the topic of this research is the use of the words "climate change" on Twitter outside of American election time and during American election time. Examining the frequency of words, in this case, "climate change," on Twitter over time provides a useful insight into how to view a topic of public debate, its media coverage, and, in this specific case, the impact of the elections on the sentiment and interest people have with regards to climate change. This leads us to the research question: "How does the use of the words 'climate change' evolve on Twitter in terms of frequency when looking outside of American election time and during American election time?"

2 Related Work

Agwuocha (2019) studies language use on social media and it discusses the difficulties of assembling and interpreting data on social media platforms. This can be of importance to this research, in terms of evaluating the retrieved data. Their study is mainly focused on quantitative data and the evaluation of it, where several good programs are suggested for processing this data. "Chapter 9 focuses on the analysis of quantitative data. Computer software programs including SPSS, NVivo, and Microsoft Excel are suggested."(Agwuocha, 2019). This makes for good suggestions for the processing of this research's data, which can, in turn, be evaluated. The paper further states that ethical matters ought to be looked over. It suggests that researchers cannot just take people's posts off of social media and use it, without taking into account certain guidelines. For example, a partici-

pant's identity should not be disclosed and consent is required in case it will be. (Agwuocha, 2019). The study by Agwuocha (2019) shows it's relevance to this paper's research question when seeing how it emphasizes the importance of assessing the quality of data and following ethical standards, especially when doing a social media based research, considering many privacy concerns tend to come with this sort of research. When obtaining data for this research, it's important to ensure that it's relevant, accurate to this research and reliable to use, in order to reach insightful conclusions. Furthermore, the research should either not disclose the identity of the people who post the Tweets that will be used, or obtain consent from these people when necessary.

Research 2 examines patterns in language on Facebook, highlighting the alarming rise of nationalism in online debatesAkayoglu (2016). This study focuses on a different topic than this paper and different social media platforms, however, its nature is similar to this research. I believe that the way the research is conducted can provide helpful insight to continue this study. Namely, research 2 discovered that topics like crime, violence, war, elections, and insurgency are addressed online during this social media revolution, which is a correlation that is relevant to this research. Although the study by (Akayoglu, 2016) focuses on a different topic and platform, it lays emphasis on the value of examining linguistic trends on social media platforms and it, arguably, implies that the nature of their research, is important when it comes to researching language dynamics on Twitter, the focus of this paper. Overall, research 2 provides the reader insight about social discussions online, which can be relevant when examining the frequency of the use words "climate change" on Twitter, during and outside of American election time, considering a lot of sentiment comes with social discussions about political and social topics like climate change.Source 3 is the largest dataset project on people's opinions on Twitter regarding climate change, spanning more than 13 yearsEffrosynidis et al. (2022). The data set is based on more than 15 million tweets worldwide. This data set will be discussed further, later in this paper, as it will be the basis data set of this research. Overall, it's significance to this study's research question becomes clear when seeing what a robust foundation for data it can serve as for this study. By utilizing this vast data set as one of the ways to collect data for this study, it is possible to examine trends and patterns in the frequency of the words "climate change", during and outside of American election time. This, in turn, can contribute to gaining more insight into how the public debate on climate change changes periodically on Twitter.

3 Data

This research aims to investigate the use of the words "climate change" and the way it evolves in terms of frequency on Twitter, during and outside of American election time. To illustrate the relationship between the independent variable, the American election time period, and the dependent variables, the frequency of "climate change" mentions in Tweets, a contingency table is provided below: In order to be able to conduct the investigation, two data sources will be utilized. Source 3, the data set by (Effrosynidis et al., 2022) that is mentioned earlier in this paper, lays a great basis for conducting this research. It contains a data set on the exact words that need to be examined in this paper. Using this data base, we will extract Tweets that contain the words "climate change", in order to observe the frequency of the use of these words. Additionally, data will directly be gathered from Twitter by searching the Twitter API for tweets that contain certain keywords and fall within a specified time frame. The way keywords will be used to collect the data directly from Twitter will be the following. Firstly, the words "Climate Change" will be typed in the search query, after which relevant hashtags will be used. These will include all variations of the word "election" and the year "2020". Examples are Elections2020, AmericanElections2020, Elections + 2020. This will allow for viewing the number of results for each search query. The results of the shown Tweets will need to be evaluated, in order to limit them to a certain time frame; the American election period. For the analysis, solely the previous election period will be used in order

-	During elections	outside of them
High frequency	xx	xx
Low frequency	xx	xx

Table 1: Contingency table: Frequency of "climate change" in Tweets during and outside of American election time

to limit complications. The mentioned limitations in the search query will ensure that the results are relevant to the needed political context. Similarly, data outside of the American election time will be collected. The results that we will obtain in this way will also include retweets, which is positive, as it provides more insight on the frequency on the use of the words "Climate Change". We will preprocess the data after collecting it, in order to be able to analyse it. This involves handling retweets by de-duplicating the dataset to prevent double-counting the same Tweet. In order to only concentrate on Tweets significant to our study. we will also exclude any irrelevant Tweets from the data set. This is done by, as mentioned earlier, restrictions in the search query, but the Tweets will also need to be looked over by people for evaluation. Databases, files and cloud storage will all be used to store the data for convenient access and backup. We'll utilize Git and other tools to track changes and organize it in a neat manner. Additionally, there will be security precautions in place to protect the data. Furthermore, the data will protect the privacy of the people posting the Tweets by excluding their identities before storage and analysis.

4 Predicted Results

The expected outcome of this research is that the words "climate change" are used more often by Twitter users during the American elections and the sentiments of these Tweets with regards to climate change vary. Hence, the contingency table will look similar to the following example: Of course, we have to be aware of possible shortcomings in the way we collected the data. A potential limitation can be that, if data is used exclusively from Twitter, the the data could add biases that are present in the social media platform itself. Examples of this are the representation of the general population and a demographic bias. Another important possible shortcoming of this research is that the keyword-based method might miss Tweets that address climate change in another language, different than English.

-	During elections	outside of them
High frequency	400	200
Low frequency	100	300

Table 2: Contingency table: Frequency of "climate change" in Tweets during and outside of American election time

Discussion The expectation of the more frequent use of the words "climate change" during the American elections, points to an increased public interest in global issues during political events. It is safe to assume that there will be a range of opinions reflected in these Tweet results and these varying results emphasize the significance of studies like this one, as this research makes it possible to comprehend climate change concerns by showing the impact of how a political context can influence debate.

5 Conclusion

In conclusion, this research aimed to explore the evolution of language on Twitter, with a particular focus on the use of the words "Climate Change" and their frequency during American Election time and outside of it. The primary source of data for this analysis is Source 3, a large dataset that spans more than 13 years and includes over 15 million tweets globally. Using this dataset, the study intends to achieve the examination mentioned above. Anticipated results suggest an increased usage of the term "climate change" during American elections, with a diverse range of sentiments expressed in the examined tweets, considering the American Elections are a time of discussing climate change. This study mainly contributes to a better understanding of how language can evolve in online debates and how it can connect to social happenings. Further studies might examine how Twitter discussions regarding climate change, or other global issues, evolve over time in different contexts, not limited to the American election time. This could offer some new perspectives on the way social media discussions and communications affect sentiment on political matters and global issues.

https://tinyurl.com/6yw5nj27

References

Agwuocha, U. A. (2019). Language use in the social media and national integration: Current trends and challenges. *International Journal of Development and Management Review 14*(1), 265–285.

Akayoglu, S. (2016). Book review for research-

ing language and social media: A student guide. ResearchGate.

Effrosynidis, D., A. I. Karasakalidis, G. Sylaios, and A. Arampatzis (2022). The climate change twitter dataset. *Expert Systems with Applications 204*, 117541.