

Project Proposal: Climate Change Deniers in Social Media: Analyzing Linguistic Patterns of Climate Change Related Tweets

Ashkan Aleshams, Jiajin Wu, Michael Galloro, Ehsan Emadi

Friday, November 6, 2020

Problem Description and Research Question

Many experts believe that global warming and climate change is happening. According to the information that we have collected from the United Nations website, the global average temperature raised by 0.85 centigrades from 1880 to 2012, and the sea's average level increased by 19 cm from 1901 to 2010. Every day we observe new climate events such as ocean acidification, droughts, floods, Greenland and West Antarctic ice melting, and many other undeniable climate phenomena that prove that climate change is real. Although all industries, companies, governments, and scientists know the impact of climate change in our lives, many of them still ignore this phenomenon and its harmful effects. Unfortunately, governments and industries are not the only ones who refuse to solve the climate change problem; many people also do not believe in climate change. Therefore, we decided to explore these deniers' opinions, and social media provides us with a convenient way to analyze a large sample size of these opinions.

Nowadays, most people are spending a vast amount of their time on social media. Based on the data that Clement has gathered in Statista, in 2019, on average, people spent 144 minutes per day on social media, and this number has been continuously increasing over the past few years (Clement, 2020). Since it is incredibly challenging to investigate all social media types and all kinds of comments about climate change, we decided only to check tweets and find trends among climate deniers in social media. In this project we are only considering tweets containing the following hashtags, #climatechange, #climatechangeisreal, #actonclimate, #globalwarming, #climatechangehoax, #climatedeniers, #climatechangeisfalse, #globalwarminghoax, #climatechangenotreal.

Solving humanity's most significant challenge, climate change, is crucial because our survival depends on it, and every small action counts. Solving this global issue requires consensus, and we cannot do it if part of society disagrees that action ought to be taken. Therefore, this idea motivated us to investigate the attitudes of climate change refusers.

Therefore, through this twitter data we then aim to answer to following: **What linguistic patterns can we observe in social media amongst climate deniers?**

Dataset Description

The dataset we found consists of ID's for tweets about climate change. This dataset alone is enough for our purpose.

Source of dataset: Harvard Dataverse Organization Website.

Link: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/5QCCUU>

The format of this data set is text format; which is just ID's of tweets, each piece of data is on one line. This dataset alone doesn't give us any informations but these ID's are all for tweets about climate change which we can use.

Here is a sample of the data:

981117605141532673
981117608325013504
981117614347948033

981117614876516353
981117615929049089
981117616122204160
981117617409810432
981117619754315776
981117622191296512

In order to work with the data and get access to the hashtags about , we will convert it into JSON format through the Hydrator Platform which uses these ID's and Twitter API to get access to these tweets and put them into a dictionary. After converting, the each data is a dictionary with strings as keys and strings, int and dictionaries as values. These keys represent tweet data; such as date of tweet, name of user, title of tweet, text body, font, hashtags, etc.

Computational Plan

- **Computational Plan:**

The data set provides us with a list of twitter IDs, and instructions on how to take these IDs and collect the original tweet text and other related information of each tweet into a JSON file. We plan to first **filter** this data by collecting only the information that is relevant to us: the text of each tweet, and their identifiers, or hashtags. We will then need to **transform** the tweet text data in order to get rid of re-tweets, so that each tweet in our set is original, and remove capitalization, and URLs. From there, we plan to **aggregate** all of the tweets with the following identifiers into one set: "#climatechangehoax", "#climatechangeisfalse", and "#climatechangenotreal". We can then make comparisons of this new set of tweets with climate-denying hashtags, and the rest of the data set. We will **aggregate** the most popular words of each set, and then after **filtering** out common words such as "the" or "climate", we can compare the most popular words of each set. The open source py-readability-metrics or a similar tool provides us with an **algorithm** which aims to assign a rating that describes how readable a string of textual data is. We can **aggregate** the rating for our two sets of tweets to attain two ratings total, one that describes the readability rating of climate denying tweets, and one that describes the readability rating of the rest of the tweets. We will also use the VADER sentiment analysis tool to a similar effect; this provides us with an **algorithm** which aims to assign a rating that describes how positive or negative text is. We will again **aggregate** the rating of our two sets of tweets to attain to ratings total, one for each set.

- **Presentation Plan:**

To give an idea of how we split up our tweet data, we will provide random examples of the set of tweets with climate denying identifiers, and the set of tweets with out. We will report our list of most popular words for each set by creating a two word clouds on plotly using the WordCloud function, one word cloud for each set of tweets, where the more popular words are larger. We think this will be a great way to visualize the differences in word usages between the two sets, should there be any. Finally we will provide the two readability ratings, and the two positive ratings.

References

Cdimascio. (n.d.). Cdimascio/py-readability-metrics. Retrieved November 04, 2020, from <https://github.com/cdimascio/py-readability-metrics>

Clement, J. (2020, February 26). Daily social media usage worldwide. Retrieved November 03, 2020, from <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>

Climate Change – United Nations Sustainable Development. (n.d.). Retrieved November 05, 2020, from <https://www.un.org/sustainabledevelopment/climate-change/>

Cjhutto. (n.d.). Cjhutto/vaderSentiment. Retrieved November 04, 2020, from <https://github.com/cjhutto/vaderSentiment>

Littman, J., & Wrubel, L. (2019, May 20). Climate Change Tweets Ids. Retrieved November 04, 2020, from <https://doi.org/10.7910/DVN/5QCCUU>