Hot or Not

AN ENQUIRY INTO CLIMATE CHANGE SENTIMENT IN SOCIAL MEDIA

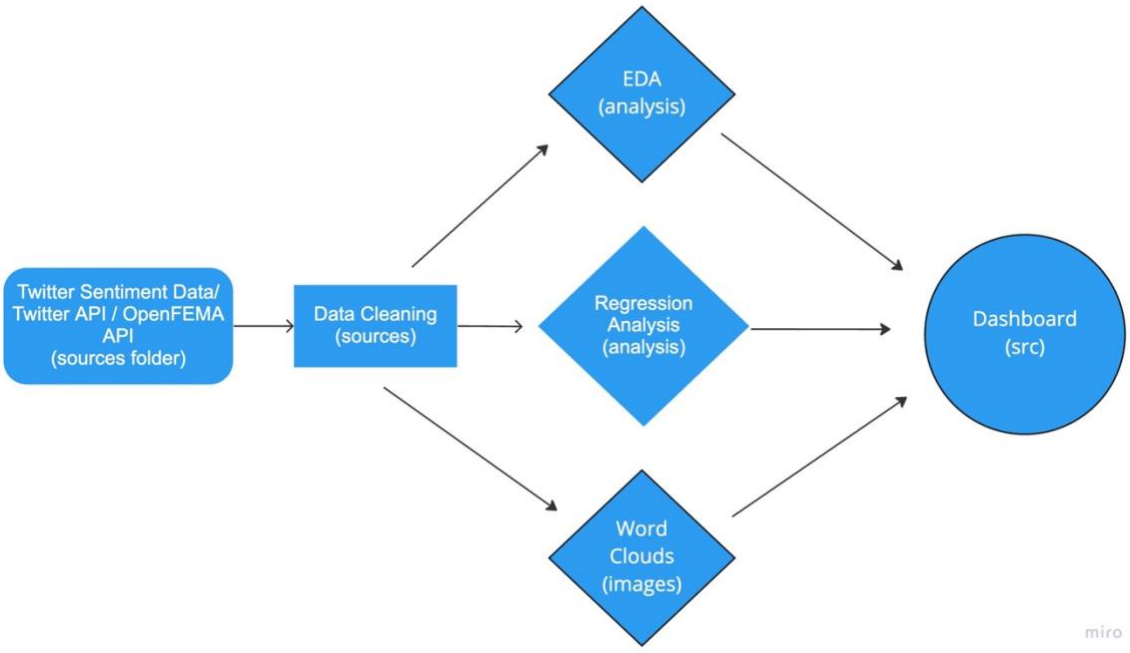
Grey Xu (greyxu), Jaskirat Kaur (Jaskiratk), Jonathan Juarez (jonathanjuarez), Ridhi Purohit (ridhi)| CAPP 122 | March 2023

Abstract

In recent years, climate change has been prominently discussed across social media. Even with an increase of extreme weather events occurring at a higher frequency globally, there is still a portion of the population who are unaware that their lives are being impacted by climate change or are ignorant of the magnitude and reality of the issue. Thus, the purpose of this project was to analyze climate change patterns to determine people's sentiments of climate change using Twitter tweets data. We attempt to correlate people's awareness with the intensity of climate change's impact (natural weather disasters) within their area of residence. The data set we used is a twitter collections dataset regarding climate change and human opinions from the study "Exploring climate change on Twitter using seven aspects: Stance, sentiment, aggressiveness, temperature, gender, topics, and disasters"[i]. We used a FEMA (Federal Emergency Management Agency) API for extreme weather events to merge with the twitter data. In our final application, we created a visualization that presents the sentiment data across the United States, the words often used on Twitter relating to climate change, and a graph that shows the sentiment levels before and after a severe weather event.

Program Structure

The diagram below shows our workflow in creating the application and related features and in parenthesis are the modules in our GitHub repository that relate to each given step.



We started the project by focusing our efforts in analyzing the dataset given by the paper named in the abstract. Our second step was acquiring related data using two APIs, OpenFEMA for weather disasters and using "twarc" a third-party Python Twitter API to collect tweet text data. The data cleaning phase required cleaning this data and merging the three data sets. We created several data frame samples depending on our needs. The exploratory data analysis portion focused on understanding our merged data to begin formulating a hypothesis of climate change sentiment attitudes in the United States. The regression analysis portion attempts to answer these questions. The word clouds aim to demonstrate the language used between climate change believers and deniers. The dashboard application combines some of these features to create an interactive experience for our project. These features are all accessible in our hot-or-not GitHub repository.

Team Responsibilities

Our team split programming responsibilities on several key components. Grey managed acquiring the original dataset from the study, cleaning and merging data frames from two APIs listed below and presenting the exploratory analysis of the acquired data. Jonathan used the OpenFEMA API to acquire the weather data, cleaned data, and wrote code to generate the word clouds for the application. Ridhi used a Twitter API to pull and clean Twitter text data to get the original tweet text using the tweet ids we acquired from the original study and created the statistical analysis for sentiment before and after weather events. Jaskirat implemented the Dash application for our visualization interface and led the team in planning what interactive features should be included in the dashboard. Most of this work was done through Python using Visual Studio Code, and we all contributed implementing our code through this software and GitHub.

Application Guideline

Running the hot-or-not GitHub repository will generate the Dash application. To do this follow the "Getting Started" steps in the README.md file.

First, the geographical map visualizes the average sentiment across the United States from 2009 to 2019. Hovering over each state shows the sentiment levels of each state for a given year. Year can be selected below the map. The stance dropdown allows a user to select sentiment values depending on the stance of a given group, which from our dataset includes those individuals that are either climate change deniers, believers, or those that are neutral. This visual allows us to understand how people's sentiment towards climate change has evolved over time in different parts of the country.

Second, below this portion are the word clouds for the most frequent words used in tweets related to climate change in 2009 compared to 2019 and can also be selected based of

stance. These word clouds help to differentiate how the public's vocabulary and attention towards climate change have evolved over the past decade. For example, we can see in 2009 the words most often used for both climate change deniers and believers were not as aggressive. By 2019, the tone in language was beginning to shift to a more aggressive nature.

Third, we visualize a graph that details the average sentiment of tweets before and after a disaster was declared. You can select the before and after sentiment values to the right of the graph. It appears people's sentiments regarding climate change did not drastically change after they experienced a disaster event. This analysis did not distinguish between climate change believers and deniers which indicates that whether people believe or deny climate change doesn't impact the connotations with which they speak about it.

The EDA and analysis code is accessible as a separate feature in the repository, which provides greater detailed explanation of our findings.

Accomplishments and Learnings

Our initial idea for the project was to study the negative impacts developing countries face when dealing with severe weather events due to climate change. We shifted our focus to working exclusive with data in the United States for several reasons. In the beginning phases of the project, we learned that using global data for several countries' social media sentiment and combining it to extreme weather events would be challenging. This is due to the potential language barriers we would have faced and acquiring the limited data developing countries have on weather events. A project of this scope would be better suited if given a greater time frame of completion. For these reasons, we shifted to U.S centric approach for the data accessibility.

A challenge that we faced was merging the data sets into a cohesive data frame- this phase of the project required days of trial and error to get useable csv files for data manipulation. This goes along with another major roadblock which was working through GitHub's constraint of file sizes. Since we were working with a massive dataset that included over 15 million data points, we had to find other solutions to store the data and make it accessible. Finally, understanding how to use a new Python module for dashboard visualization was a huge challenge and we believe given more time and resources we could have implemented more features in the final application.

Our findings demonstrate that certain political events have influenced twitter user's sentiment across years. For instance, the successful treaty of Paris agreement on COP21 encouraged believers to have positive attitude toward climate change as shown in our first visualization. Moreover, we found out that the common words applied by believers and deniers are becoming more aggressive in the year of 2019 comparing with 2009 as shown in the second visualization. Finally, we compare users' sentiment before and after the

extreme weather events listed in the FEMA dataset and discovered that twitter users are experiencing more negative attitude after the weather disaster regardless their stance.

In the end, we managed to acquire and manipulate several datasets from credible sources to create our own enquiries into this complex subject, learned how to generate a dashboard in Python to visualize the data, and learned how to work as a team within a programming project that taught us how to integrate everyone's coding efforts for a finished product.

---

[i] Effrosynidis, D., Sylaios, G., & Arampatzis, A. (n.d.). *Exploring climate change on Twitter using seven aspects: Stance, sentiment, aggressiveness, temperature, gender, topics, and disasters*. PLOS ONE. Retrieved March 4, 2023, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0274213