Climate Change Pulse: A Comprehensive Web Application Framework to Analyze Twitter Data Sentiment in Various Disaster Scenarios

Alan Zheng[1], Carlos Gonzalez[2]

[1]West-Windsor Plainsboro High School North, 90 Grovers Mill Rd, Plainsboro Township, NJ 08536
alanzheng240@gmail.com
[2]Computer Science Department, California State Polytechnic University, Pomona, CA 91768
softcom.lab.cpp@gmail.com

ABSTRACT

Climate change is an urgent global issue, with natural disasters becoming more severe and frequent due to human activities [1]. Understanding public sentiment around these events can inform climate awareness and policy. We developed ClimatePulse, a web-based tool that visualizes natural disasters alongside Twitter data to analyze how proximity and time influence climate-related sentiments [2]. Using the Climate Change Twitter Dataset, we examined over 15 million tweets, mapping them with disaster data through an interactive UI.

Challenges included missing geospatial data and sentiment classification limitations, addressed by refining data filters and leveraging embedded tweets. Our experiments tested how distance and time around disasters affect sentiment, revealing that proximity intensifies negative emotions, and climate change deniers exhibit surprisingly strong negative sentiments. Compared to prior methodologies focused on data collection or basic sentiment analysis, our approach emphasizes user interactivity and behavioral analysis. ClimatePulse offers a dynamic way to understand climate discourse, bridging data insights with public engagement.

KEYWORDS

Natural Language Processing, Machine Learning, Sentiment, Disasters

1. INTRODUCTION

Climate change is one of the most prominent, terrible issues we are facing right now. Extreme heat and climate change induced natural disasters directly and indirectly impact people worldwide.

Climate change from human activities causes disasters to be more intense and frequent. Hot seasons now keep breaking record temperature and worse heat waves, floods and droughts are normal in many countries. Ice sheets are melting, ocean levels are rising and warmer oceans create bigger hurricanes. Importantly, with the rising temperature's tipping points, changes in systems are passed which are irreversible. From The Climate Book by Greta Thunberg "IPCC estimates that global warming will reach 3.2°C by 2100". When Hurricane Sandy hit New York, 8 billion dollars of damage from the storm surge were because of climate change. In 2003, a heatwave in Europe caused more than 70000 premature deaths, and climate change has doubled the chance of occurance. If global warming is under 2°C over 50 years, it could prevent 4.5 million premature deaths in the U.S alone.

Technology can give us information about peoples ideas–tools like data analysis can assess people's perception of different issues. For climate change, how would being near a disaster impact a person's emotions, or their sentiment [3]? And how can we use data from social media platforms such as X, formerly known as Twitter? This initiative works on these tasks, aiming to inform the general public and drive change in policy-making in our society.

Present examples in tweets show potential areas of interest that we aim to further explore. We aim to observe whether natural disasters amplified by climate change impact people's sentiment. Compared to surveys that evaluate people's stance and awareness on current issues like climate change, people tend to express themselves more intensely and personally. This advantage is because social media is a less formal environment. Additionally, social media has more younger users and can be more diverse than sample populations used in surveys. This advantageous presentation of information is what allows us to extract more meaningful insights from the readily collected data.

This approach used a Bi-directional LSTM model for sentiment classification on climate-related social media data [4]. It effectively categorized sentiments but relied heavily on API data collection, which we lacked. Its focus was limited to model training, while our project advanced this by analyzing already labeled data for deeper insights.

This method visualized social media sentiment during disasters, using geographic data from the Ebola Twitter dataset [5]. While effective in mapping sentiment, it lacked user interactivity and broader analytical factors. Our project improved on this with a user-friendly UI and an emphasis on behavioral differences between climate change deniers and believers.

Mouronte-López and Subirán analyzed climate change sentiments on Twitter using sentiment models like VADER and topic modeling [6]. Although comprehensive, it didn’t explore user behavior or disaster proximity effects. ClimatePulse builds on this by integrating interactive visualizations and focusing on how disasters influence sentiment over time and space.

This Climatepulse is a web tool that visualizes disasters onto a world map, overlaid by various different tweets. There is a scroll feature for the user to visualize which natural disasters occurred per year. Red dots on the map indicate a recorded disaster, and upon selecting a disaster, tweets within a 1000 mile radius to the disaster will be displayed. The world map combines tools such as javascript and d3 from the following github repository.

Our research aims to use the datasets to see if there are any connections between sentiment and disaster. Both the tweets and disaster data were pooled from the “Climate Change Twitter Dataset” research initiative by Dimitrios Effrosynidis et al. (2022) [7]. Over fifteen million data points spanning over thirteen years related to climate change were sourced from the social media platform, where data such as gender, stance, sentiment, and disaster type are included. The tweets are from different people and organizations but are all talking about climate change.

In our experiment, we aimed to test how distance and time windows around natural disasters influence public sentiment on climate change as expressed on Twitter. We systematically varied distance thresholds (500 km, 1000 km, 2000 km) and time windows (1, 3, and 7 days before and after a disaster) to identify patterns. Using the Climate Change Twitter Dataset, we filtered tweets based on proximity and time relative to disasters and calculated average sentiment scores.

Our most significant findings showed that tweets closer to disasters tend to have stronger negative sentiments, highlighting the emotional impact of proximity. Sentiment also became more negative as time progressed after a disaster, reflecting prolonged concern. Interestingly, climate change deniers exhibited the most negative sentiment, even more than believers, which was unexpected

[8]. This suggests emotional responses are influenced by both proximity to disasters and pre-existing beliefs. The experiment underscored the importance of parameter selection and revealed complex interactions between distance, time, and user stance.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. The lack of datasets

One of the primary challenges in implementing a machine learning model to predict lithium carbonate levels is the lack of available, high-quality datasets. Machine learning models require large, diverse datasets to train and validate the model's accuracy. However, data specifically correlating lithium concentrations with physiological factors such as temperature, blood pressure, and oxygen saturation is limited. Most available studies and reports focus on analyzed conclusions rather than providing raw data. This scarcity of usable data poses a significant obstacle to model training. To address this, collaboration with healthcare institutions or using synthetic data could be explored to build a more robust dataset.

2.2. Ensuring the accuracy

Another significant challenge is ensuring the accuracy of the machine learning model. Typically, machine learning models require a substantial amount of data to achieve reliable predictions, with a recommended minimum of around 70 datasets. However, in our case, we only had access to approximately 40 data sets. This limited amount of training data leads to a high divergence in the model's predictions, reducing the overall accuracy. To improve the reliability of the model, more data must be gathered, and techniques such as cross-validation or data augmentation could be employed to optimize the model's performance with limited data.

2.3. Personalizing predictions

A third challenge is the difficulty of personalizing predictions for individual patients. Since lithium concentration varies greatly from one person to another due to differences in metabolism, body composition, and sensitivity to the medication, creating a model that can generalize effectively across all patients is challenging [9]. A model trained on generalized data may not account for these individual differences, leading to inaccurate predictions for some users. To address this, personalizing the model through patient-specific data collection and utilizing techniques such as transfer learning could help tailor predictions to individual patients, improving both accuracy and effectiveness.

3. SOLUTION

The visualizations are a website with a user interface, users can change the year and check various countries. We fill in missing data points from our dataset using the Google Maps API [10]. Notably, we fill in the missing latitude and longitudinal coordinates by using the region / address a disaster has occurred–that way, we ensure that we can overlay every disaster onto the visualization. We used a Google Maps API key to find the locations based on the data and calculated the coordinates. We also worked more on the data, we removed tweets from the twitter data that didn't have coordinates. Due to the large sample size, and the lack of other geospatial information from the Twitter dataset, finding the location for these tweets was not possible. Though not used in the project, Pandas was used to graph the data for a better understanding. For example, one graph compared the mean sentiment of tweets and their aggressiveness with the worst disasters. The purpose of showing tweets when you click on a disaster is to see if there are changes in sentiment. Specifically, to see if being near a disaster around the time of it is related to sentiment on climate change.

When the user clicks on a disaster there is code that goes through twitter data to find tweets in a certain radius and time range close to it. To display the tweets we used embedded tweets, since Elon Musk acquired twitter and the data can't be accessed directly. The code is written in javascript, HTML & CSS.

The proposed lithium carbonate prediction system consists of a user-friendly web interface, a machine learning backend, and a database of collected data. The system's primary objective is to allow users to input physiological data such as temperature, blood pressure, oxygen saturation, and time since the last lithium dosage. This data is then processed through a series of machine learning models to predict the lithium concentration levels in the bloodstream.

Flow of the System:

1. User Input: Users access the web interface and input their physiological data.

2. Data Processing: The data is sent to the server, where it is preprocessed (e.g., normalized) to ensure compatibility with the trained models.

3. Model Prediction: The machine learning model, trained on physiological data, predicts the concentration of lithium carbonate in the blood.

4. Output: The predicted value is sent back to the user interface, where users can view whether their lithium concentration is within a safe range.

The system was developed using Python's Flask framework for the backend, with machine learning models implemented via libraries such as scikit-learn. Models include linear regression, random forest, support vector machines (SVMs), and neural networks [14].

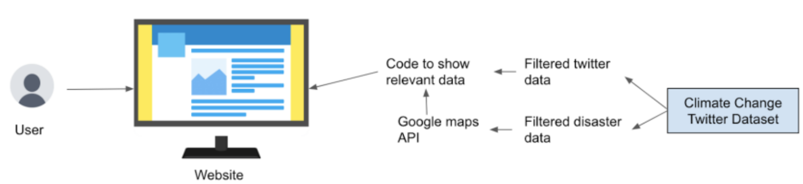


Figure 1. Overview of the solution

On the website cooler countries are in shades blue while hotter countries are in shades of red. The symbols used for the disaster are red dots, when the map is running they "flicker" on and off.

Hovering over a disaster also gives information of the number of deaths from the disaster. In order to show the tweets for disasters windows are used, the windows have a scroll feature to see all the embedded tweets. To get out of a tweet window click out of the window or on another disaster.

Figure 2. Screenshot of the map

```
function embedTweets(tweetIDs) {
  tweetIDs.forEach(function(id, index) {
      twttr.widgets.createTweet(
          id,
          tweetsElements[index],
          {
              conversation : 'none',
              cards        : 'hidden',
              align        : 'center',
              theme        : 'light'
          }
      );
  });

}
```

Figure 3. Screenshot of code 1

The embedTweets function expects a numpy array of tweet id's as its parameter. This function is immediately called upon execution of the program, when the website is first launched. It uses the tweetsElement array which is in code dealing with the twitter data. After that it has some attributes that affect the appearance of the embedded tweets. Setting conversation to none means it doesn't show replies to the tweet. Cards hide images and polls that the tweet would have. Align centers the tweet and theme makes the tweet in light mode.

```
[{'address_components': [{'long_name': 'Šiauliai',
    'short_name': 'Šiauliai',
    'types': ['locality', 'political']},
   {'long_name': 'Šiauliai City Municipality',
    'short_name': 'Šiauliai City Municipality',
    'types': ['administrative_area_level_2', 'political']},
   {'long_name': 'Šiauliai County',
    'short_name': 'Šiauliai County',
    'types': ['administrative_area_level_1', 'political']},
   {'long_name': 'Lithuania',
    'short_name': 'LT',
    'types': ['country', 'political']}],
  'formatted_address': 'Šiauliai, Šiauliai City Municipality, Lithuania',
  'geometry': {'bounds': {'northeast': {'lat': 55.9715387, 'lng': 23.4290861},
    'southwest': {'lat': 55.8410099, 'lng': 23.2276569}},
   'location': {'lat': 55.9349085, 'lng': 23.3136823},
   'location_type': 'APPROXIMATE',
   'viewport': {'northeast': {'lat': 55.9715387, 'lng': 23.4290861},
    'southwest': {'lat': 55.8410099, 'lng': 23.2276569}}},
  'place_id': 'ChIJiXzOSVzi5UYRANvpcIzRAAQ',
  'types': ['locality', 'political']}]
```


Figure 4. Screenshot of the components

```python
def get_coords(location):
    output = gmaps.geocode(location)
    lat_lng = {'lat': [], 'lng': []}
    for loc in output:
        lat_lng['lat'].append(loc['geometry']['location']['lat'])
        lat_lng['lng'].append(loc['geometry']['location']['lng'])

    if lat_lng['lat'] and lat_lng['lng']:
        avg_lat = mean(lat_lng['lat'])
        avg_lng = mean(lat_lng['lng'])
        return avg_lat, avg_lng
    else:
        return None, None

def decode_special_characters(location):
    '''
    Decode our special characters in the string 'location'
    >>>byte_string = b'Cox\x92s Bazar'
    >>>decoded_string1 = byte_string.decode('cp1250')
    >>>print(decoded_string1)
    Cox's Bazar
    '''
    location = location.encode('latin1')
    location = location.decode('cp1250')
    return location

for index, row in filtered_df.iloc[:].iterrows():
    location = row['Location'] #location -> '\x8aiauliu'

    location = decode_special_characters(location) #location -> 'Šaiauliu'

    print(f"Processing row {index} out of {len(df)}")
    lat, lng = get_coords(location)
    df.at[index, 'Latitude'] = lat
    df.at[index, 'Longitude'] = lng
    print(f"Latitude: {df['Latitude'][index]}; Longitude: {df['Longitude'][index]}")

print("Coordinates added and saved to filtered_df")
```

Figure 5. Screenshot of code 2

We were particularly interested in answering the following question: When a disaster occurs, does this increase the discussion of climate change in the region? Combining the tools of programming with the large corpus of data, we aim to further explore the data itself by performing various exploratory data analysis techniques, such as aggregating, filtering, augmenting the data to paint a clearer picture of the granularity of the datasets.

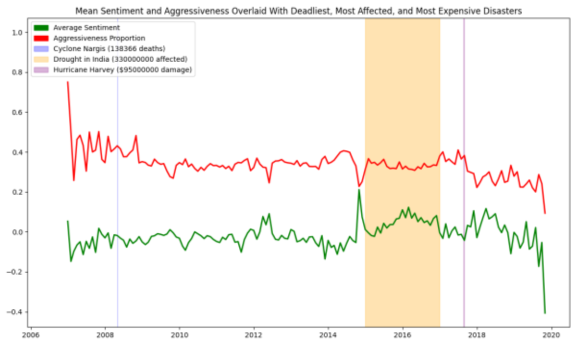
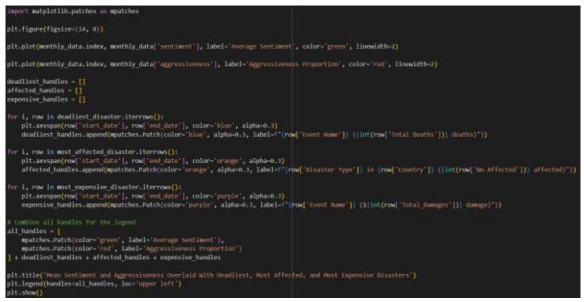


Figure 6. Screenshot of the figure



```
import matplotlib.patches as mpatches

plt.figure(figsize=(14, 8))

plt.plot(monthly_data.index, monthly_data['sentiment'], label='Average Sentiment', color='green', linewidth=2)

plt.plot(monthly_data.index, monthly_data['aggressiveness'], label='Aggressiveness Proportion', color='red', linewidth=2)

deadliest_handles = []
affected_handles = []
expensive_handles = []

for i, row in deadliest_disaster.iterrows():
    plt.axvspan(row['start_date'], row['end_date'], color='blue', alpha=0.3)
    deadliest_handles.append(mpatches.Patch(color='blue', alpha=0.3, label=f"{row['Event Name']} ({int(row['Total Deaths'])} deaths)"))

for i, row in most_affected_disaster.iterrows():
    plt.axvspan(row['start_date'], row['end_date'], color='orange', alpha=0.3)
    affected_handles.append(mpatches.Patch(color='orange', alpha=0.3, label=f"{row['Disaster Type']} in {row['Country']} ({int(row['No Affected'])} affected)"))

for i, row in most_expensive_disaster.iterrows():
    plt.axvspan(row['start_date'], row['end_date'], color='purple', alpha=0.3)
    expensive_handles.append(mpatches.Patch(color='purple', alpha=0.3, label=f"{row['Event Name']} (${int(row['Total_Damages'])} damage)"))

# Combine all handles for the legend
all_handles = [
    mpatches.Patch(color='green', label='Average Sentiment'),
    mpatches.Patch(color='red', label='Aggressiveness Proportion')
] + deadliest_handles + affected_handles + expensive_handles

plt.title('Mean Sentiment and Aggressiveness Overlaid With Deadliest, Most Affected, and Most Expensive Disasters')
plt.legend(handles=all_handles, loc='upper left')
plt.show()
```

Figure 7. Screenshot of code 3

We utilize the matplotlib module in python to create this comprehensive visualization, which is an example of several graphs we generated to better understand the underlying contents of the data. Several tasks needed to be completed, such as aggregating the sentiment and aggressiveness scores into months per year; labeling the extreme points in the disasters dataset, such as the most costly disaster (the amount, in USD, incurred in damages as a result to the disaster); most deaths occurred; and most people affected. We then overlaid a line plot of the mean sentiment and aggressiveness scores onto a graph, where the shaded in regions indicate our disaster points of interest. Once all these data points are collected, we simply combine all the data into this one graph.

4. Experiment

A potential blind spot in the program is the selection of distance and time windows around disasters. If the chosen thresholds (e.g., 500 km, 1000 km, or 2000 km) are too narrow or too broad, they might miss key sentiment patterns or dilute meaningful insights. It's critical to test these parameters to ensure the program accurately captures the relationship between disasters and public sentiment.

To test the impact of distance and time windows, the experiment systematically varies these parameters. Distance thresholds are set at 500 km, 1000 km, and 2000 km, while time windows are defined as 1, 3, and 7 days before and after a disaster. These ranges are chosen to reflect realistic geographical and temporal scopes of disaster impact.

The experiment uses the "Climate Change Twitter Dataset" by Dimitrios Effrosynidis et al. (2022), which includes over 15 million tweets related to climate change. Control data is sourced from the same dataset, ensuring consistency in tweet sentiment, stance, and disaster context. By filtering tweets based on distance and time, the program calculates average sentiment scores for each combination of parameters.

This setup allows us to observe how sentiment varies with proximity and time relative to disasters. For example, do tweets closer to disasters show stronger negative sentiment? Does sentiment become more negative as time progresses after a disaster? The experiment is designed to isolate these variables and identify optimal thresholds for capturing meaningful sentiment patterns.

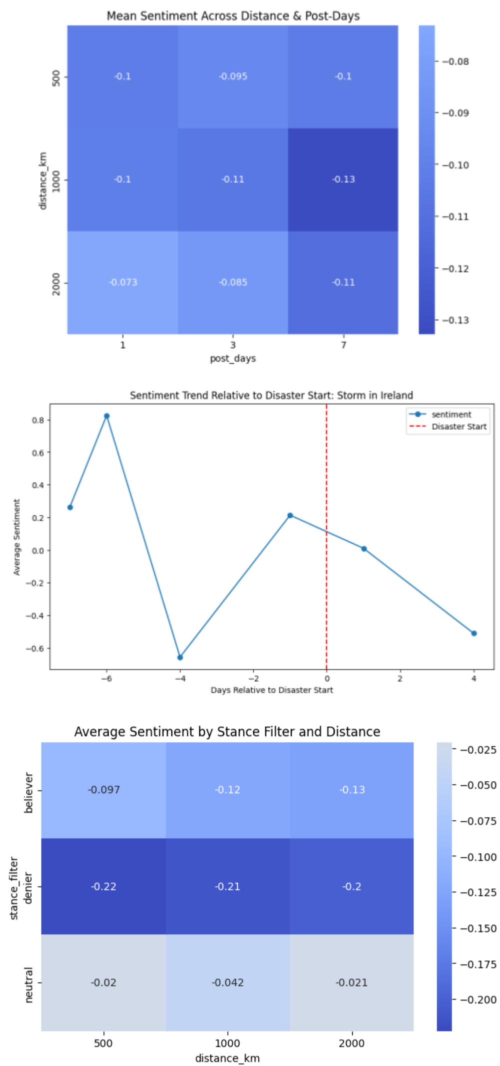


Figure 8. Figures of experiment

The analysis reveals several key insights about the relationship between disasters and public sentiment on Twitter. First, sentiment becomes more negative as the distance threshold increases, suggesting that tweets closer to disaster locations reflect stronger emotional responses. This aligns with the expectation that proximity to a disaster intensifies public concern and emotional expression.

Second, the time window significantly influences sentiment. Pre-event sentiment becomes more negative as the number of days before a disaster increases, potentially reflecting growing anxiety or anticipation. Post-event sentiment also trends more negatively over time, possibly due to prolonged discussions or the accumulation of negative news.

A surprising finding is the stark contrast in sentiment between different stances. Deniers exhibit the most negative sentiment, which may reflect frustration or skepticism toward climate change discourse. Believers, while also negative, show less extreme sentiment, possibly indicating a more measured or concerned tone. Neutral users, as expected, exhibit the least negative sentiment, suggesting a lack of strong emotional investment.

The biggest effect on results appears to be the combination of distance and stance. Tweets closer to disasters from deniers show the most negative sentiment, while neutral users remain relatively unaffected. This suggests that emotional responses to disasters are not only influenced by proximity but also by individuals' pre-existing beliefs about climate change.

Overall, the experiment highlights the importance of carefully selecting distance and time parameters to capture meaningful sentiment patterns. It also underscores the role of stance in shaping public discourse around disasters, offering valuable insights for targeted communication strategies.

5. RELATED WORK

A similar project was “A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change.” by Barachi, May El, et. al. The authors had access to data similar to the climate change twitter dataset and some api for data collection(which we didn’t readily have) [11]. It employs a Bi-directional LSTM that assigns and categorizes sentiments into different categories. Today state-of-the-art models would improve the quality of the performance. Their task was training a model specifically to assign labels, which we build off of.

We used already labeled data from a similar model in order to drive analysis and understanding of the data, and draw conclusions from it. We have different scopes, as they were trying to collect data while we focused on data analysis. They reported the accuracy for different emotions and positions. Our application has the potential for internet-users and experts alike to offer critique on the issue.

Visualizing Social Media Sentiment in Disaster Scenarios

A project (Lu, Yafeng, et. al.) Addresses a similar problem of using social media to see how disasters impact sentiment [12]. Uses sentiment modeling to model social media data and utilizes geographic visualization for it. Focused on Ebola Twitter dataset instead of climate change induced disasters.

Our project is building off this, because we have a UI and more factors in our analysis. Since we have a UI it is more accessible. Also, we were concerned about investigating the behaviors of deniers and believers.

- Uncertainty of disagreement, emphasizing sentiment classification through an entropy-based metric

- Use pretrained classifiers to label raw tweet data with a sentiment score, majority vote assigns label to this tweet, likely assigning continuous value (0-1) to tweet

- Take the distribution of data (with assigned labels) and pass into entropy function -> quantifies the level of disagreement/chaos/spread

Lu, Yafeng, et. al. addresses a similar problem of using social media data to see how disasters impact sentiment. Using the Ebola Twitter dataset, they are primarily concerned about investigating sentiment trends and behaviors in geo-located twitter data. Their approach follows by using pre-trained classifiers to label the raw data with a sentiment score, assigning labels based on a majority vote, and then taking the twitter data classes and passing into the entropy function. They aim to primarily investigate the quantifiable levels of disagreement among the classes.

Our project is building off this, because we have a UI and more factors in our analysis. Since we have a UI it is more accessible. Also, we were concerned about investigating the behaviors of deniers and believers.

A relevant study which aims to address the analysis of climate change discourse on social media examines sentiments expressed in Twitter interactions related to climate change, analyzing 92,474 tweets to assess sentiment polarity and underlying topics (Mouronte-Lopez and Subiran, 2022) [13]. The study employs algorithms such as VADER and TextBlob, alongside unsupervised machine learning techniques, to determine sentiment polarity and utilizes Latent Dirichlet Allocation (LDA) for topic modeling. The findings reveal that discussions on climate change are predominantly negative across various topics, including activism, biodiversity, and sustainability. Not only this, but they also explore differences in sentiment by geography, gender, and account type. Our project builds upon this foundation by providing an interactive web tool that visualizes natural disasters on a world map, overlaid with relevant tweets. Users can explore disasters by year and location, with red dots indicating recorded events.

6. CONCLUSIONS

One limitation of our project is the reliance on pre-labeled data from the Climate Change Twitter Dataset, which may contain inherent biases from the original labeling process. Additionally, many tweets lack geospatial information, limiting the scope of our disaster-to-sentiment correlation analysis. The embedded tweet feature, while useful, restricts access to older or deleted tweets, potentially omitting valuable data. Furthermore, our current sentiment analysis does not fully capture nuanced emotions or the context behind tweets, which could affect the depth of our insights.

If given more time, we would improve the project by incorporating more robust natural language processing techniques, such as fine-tuned transformer models, to enhance sentiment classification. We'd also explore geolocation inference methods to fill in missing location data. Expanding the dataset to include real-time data collection via APIs would provide more dynamic insights. Finally, adding advanced data visualization tools would improve user interaction and data interpretation.

ClimatePulse bridges the gap between climate change discourse and real-world disaster impacts through data visualization and sentiment analysis [15]. By mapping tweets alongside natural disasters, we provide unique insights into public perception. Our tool fosters awareness and can inform policy discussions, highlighting how climate events influence global sentiment.

REFERENCES

[1] National Research Council, et al. Advancing the science of climate change. National Academies Press, 2011.

[2] Li, Ning, et al. "Review on visualization technology in simulation training system for major natural disasters." Natural hazards 112.3 (2022): 1851-1882.

[3] Okuyama, Yasuhide. "Economic modeling for disaster impact analysis: past, present, and future." Economic Systems Research 19.2 (2007): 115-124.

[4] Sunny, Md Arif Istiake, Mirza Mohd Shahriar Maswood, and Abdullah G. Alharbi. "Deep learning-based stock price prediction using LSTM and bi-directional LSTM model." 2020 2nd novel intelligent and leading emerging sciences conference (NILES). IEEE, 2020.

[5] Kim, Erin Hea-Jin, et al. "Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news." Journal of Information Science 42.6 (2016): 763-781.

[6] Borg, Anton, and Martin Boldt. "Using VADER sentiment and SVM for predicting customer response sentiment." Expert Systems with Applications 162 (2020): 113746.

[7] Effrosynidis, Dimitrios, et al. "The climate change Twitter dataset." Expert Systems with Applications 204 (2022): 117541.

[8] Rothenhäusler, Luis. d3. js and its potential in data visualization. Diss. Technische Hochschule Brandenburg, 2022.

[9] El Barachi, May, et al. "A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change." Journal of Cleaner Production 312 (2021): 127820.

[10] Hu, Shunfu, and Ting Dai. "Online map application development using Google Maps API, SQL database, and ASP .NET." International Journal of Information and Communication Technology Research 3.3 (2013).

[11] Lu, Yafeng, et al. "Visualizing social media sentiment in disaster scenarios." Proceedings of the 24th international conference on world wide web. 2015.

[12] Mouronte-López, Mary Luz, and Marta Subirán. "What do twitter users think about climate change? Characterization of twitter interactions considering geographical, gender, and account typologies perspectives." Weather, Climate, and Society 14.4 (2022): 1039-1064.

[13] Lu, Yafeng, et al. "Visualizing social media sentiment in disaster scenarios." Proceedings of the 24th international conference on world wide web. 2015.

[14] Hearst, Marti A., et al. "Support vector machines." IEEE Intelligent Systems and their applications 13.4 (1998): 18-28.

[15] Hao, Ming, et al. "Visual sentiment analysis on twitter data streams." 2011 IEEE conference on visual analytics science and technology (VAST). IEEE, 2011.