

Twitter Sentiment Analysis

ECE-143 Group Project

Bochen Pan
Haoxing Du
Rishabh Patni
Sandeep J Nair
Xiaoran Zhou

December 2, 2022

PPT flow

- Problem Statement / Introduction
 - Data collection
 - Data cleaning
 - Sentiment analysis
 - Data visualization and insights

Goal

To examine the online discourse on climate change through measuring quantitative and qualitative (frequency of tweets and tweet sentiment) factors about tweets related to climate change.

Data collection

- Where is the data coming from? – Twitter API
- What type of data is used? – temporal and geographic properties of tweets
- keywords queried: ("climate change" OR "carbon dioxide" OR "fossil fuel" OR "carbon footprint" OR emissions OR "global warming" OR "greenhouse gases")

Number of tweets analysed per region

USA	100k
Asia	25k
Pacific	25k
Middle East and North Africa	15k
Africa	20k

Raw data

```
Unnamed: 0          0
text            @ABCPolitics @ABCNewsLive Praying that diploma...
                2021-12-31T23:52:41.000Z
created_at      0
like_count       0
retweet_count    0
reply_count      1
geo_place_id    60e2c37980197297
country_code     US
country         United States
name            St Paul
full_name        St Paul, MN
bbox             [-93.207783, 44.8907521, -93.003514, 44.992279]
Name: 0, dtype: object
```

Figure 1: Raw data from the API comes in JSON format, which is presented here as a dataframe for ease of visualization

Text data

```
['@ABCPolitics @ABCNewsLive Praying that diplomacy somehow works. The pandemic plus global warming for now is enough for both powerful nations to combat. A truce of somewhat even temporarily -wouldn't it be worth it.',  
 'And then...you have the cost of disposal of the fossil fuel stations... my guess...lots of billions!',  
 '@MaryHeglar Someone should do a remake of that Xmas Carol updated for climate change...']
```

Figure 2: Examples of text data before cleaning

```
['praying that diplomacy somehow works pandemic plus global warming now is enough both powerful nations combat truce somewhat even temporarily wouldn t it be worth it',  
 'then you have cost disposal fossil fuel stations my guess lots billions',  
 'someone should do remake that xmas carol updated climate change']
```

Figure 3: Examples of text data after cleaning

Textblob

TextBlob can automatically recognize and categorize opinions expressed in the text to determine overall sentiment. It will return text polarity within the range [-1.0, 1.0].

Text	Sentiment Score
Good point	0.7
Very good point	0.91
Very good point!	1.0
not a good idea	-0.35
dumb idea	-0.375
dumb idea :(-0.5625

VADER

It's a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments expressed in social media.

Text	TextBlob Result	Vader Result
This is great	0.8	0.6249
This is GREAT	0.8	0.7034
This is GREAT lol	0.8	0.7845
Love this idea	0.5	0.6369
Love this idea 😍	0.5	0.802
Today sucks!	-0.375	-0.4199
Today sux!	0	-0.4199

Wordclouds

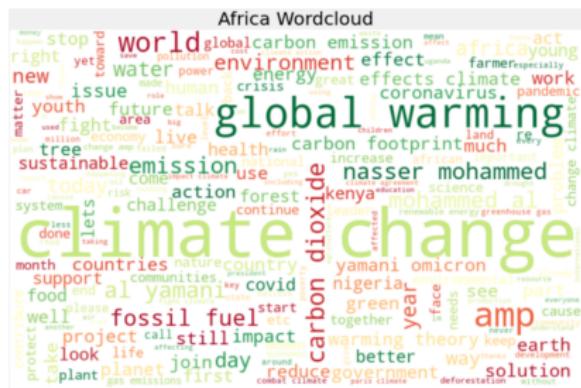


Figure 4: Africa Wordcloud

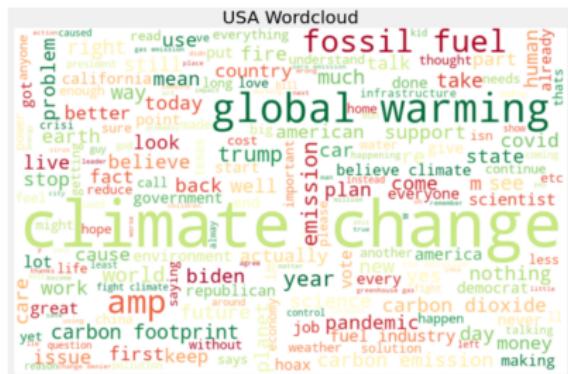


Figure 5: USA Wordcloud

Wordclouds

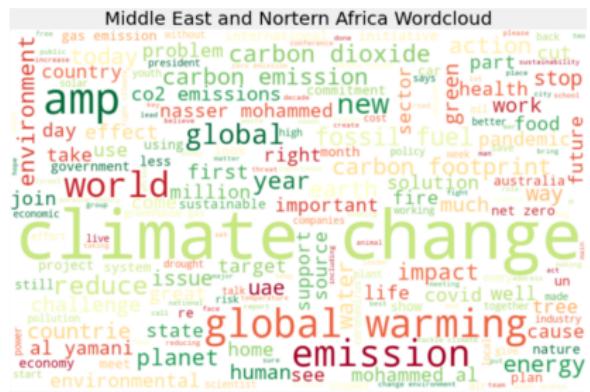


Figure 6: MENA Wordcloud

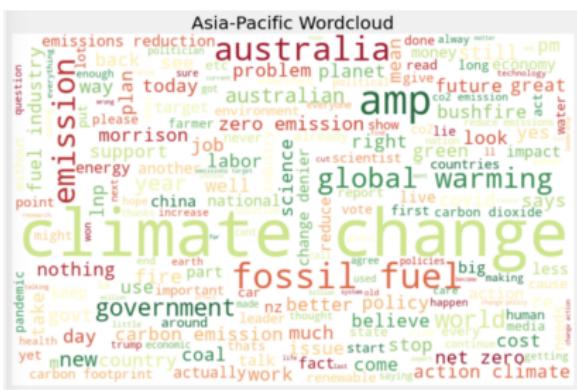


Figure 7: Asia Pacific Wordcloud

Wordclouds

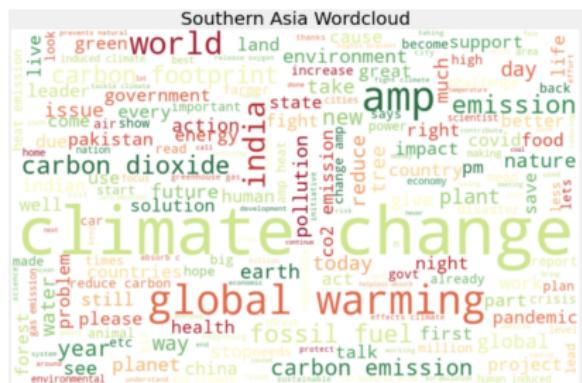


Figure 8: Asia Wordcloud

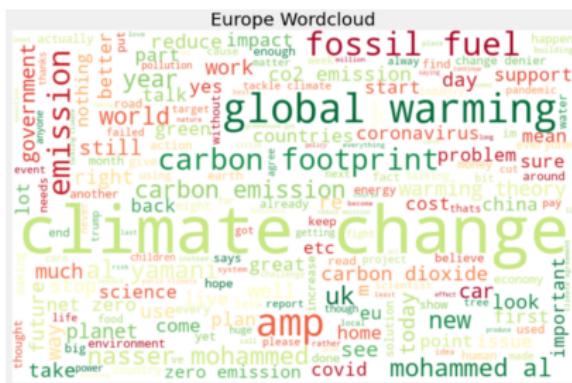


Figure 9: Europe Wordcloud

Pie Charts

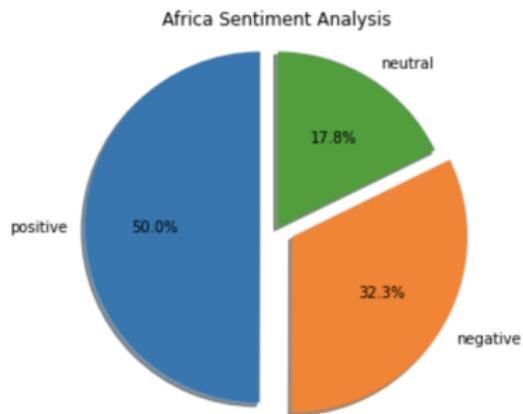


Figure 10: Africa Pie Chart

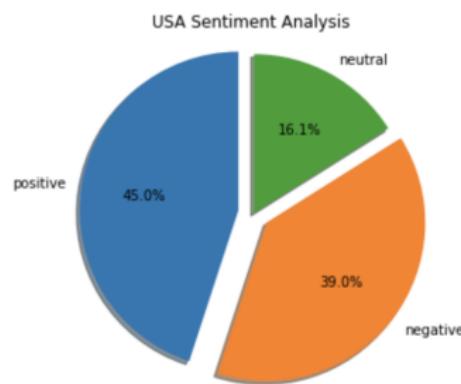


Figure 11: USA Pie Chart

Pie Charts

Middle East and Northern Africa Sentiment Analysis

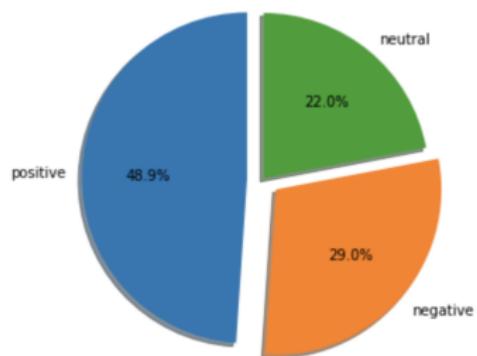


Figure 12: MENA Pie Chart

Asia and Pacific Sentiment Analysis

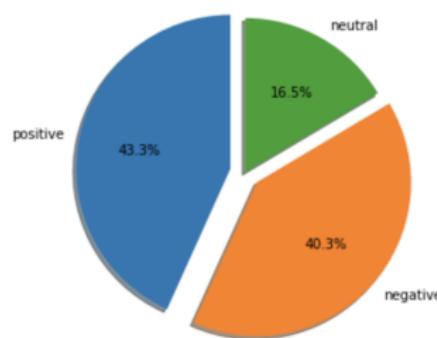


Figure 13: Asia-Pacific Pie Chart

Pie charts

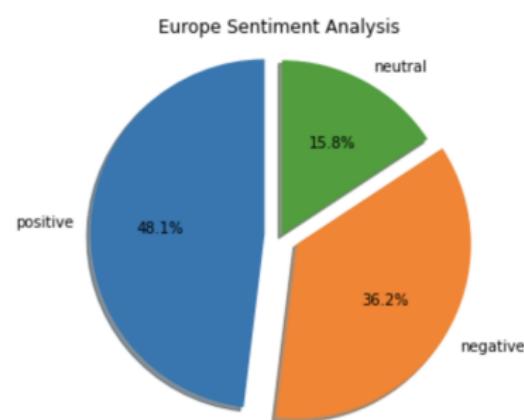
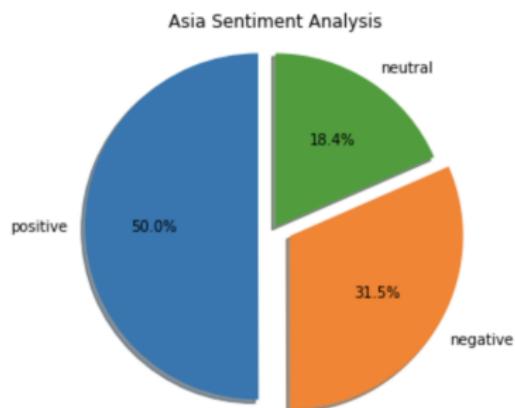


Figure 14: South Asia Pie Chart

Figure 15: Europe Pie Chart

Bar graph

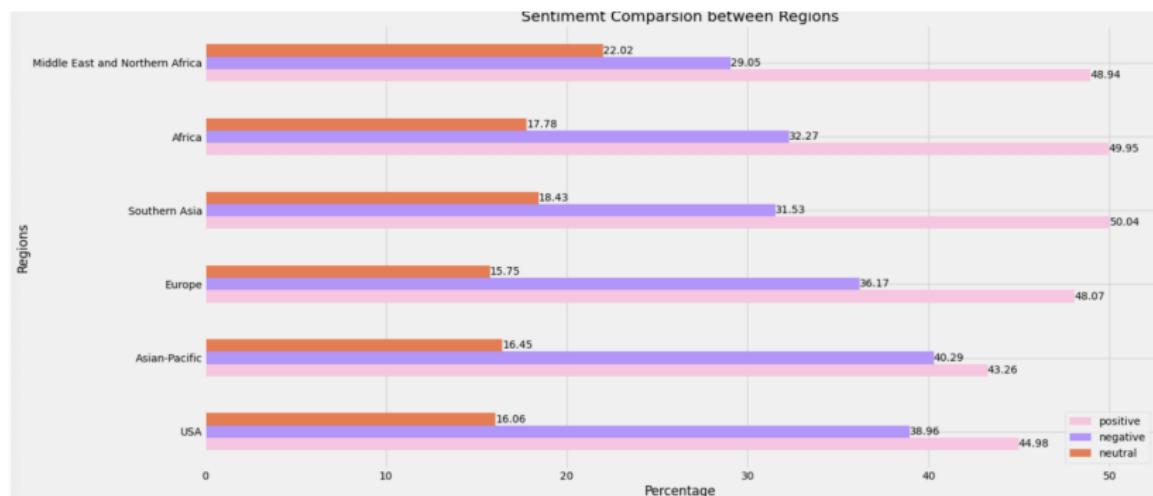


Figure 16: sentiment distribution across regions

Heat Maps

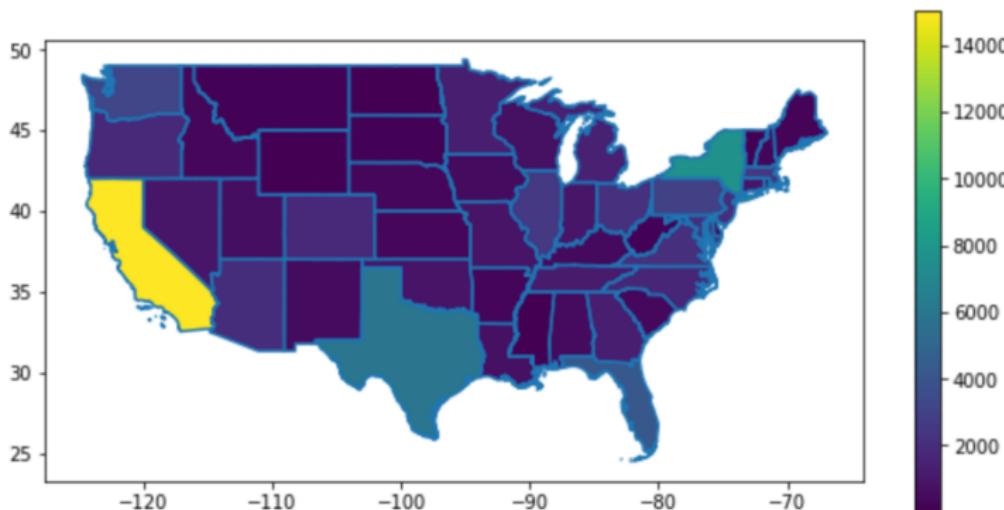


Figure 17: Heatmap of tweet count across different US states

Heat Maps

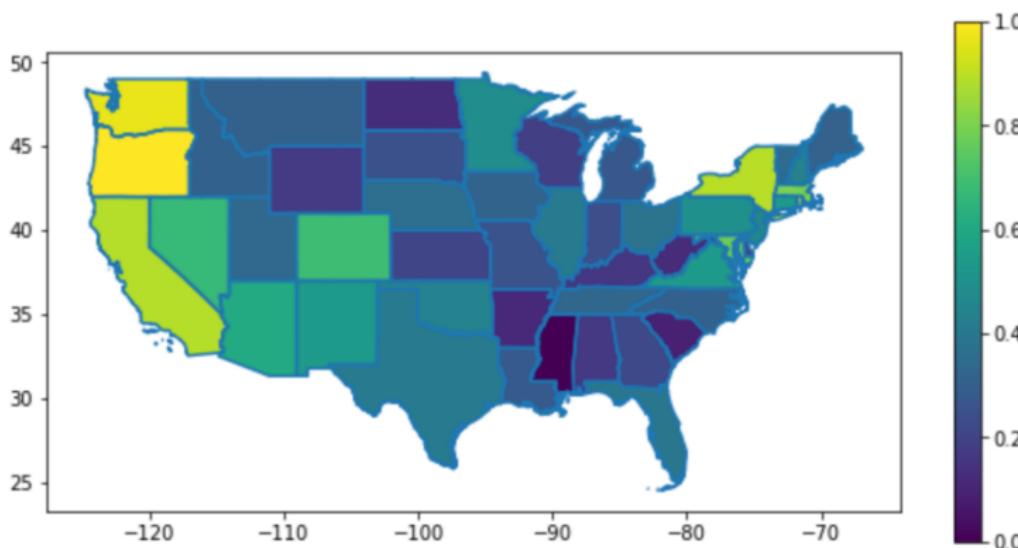


Figure 18: Heatmap of tweet count across different US states normalised by state population

Heat Maps

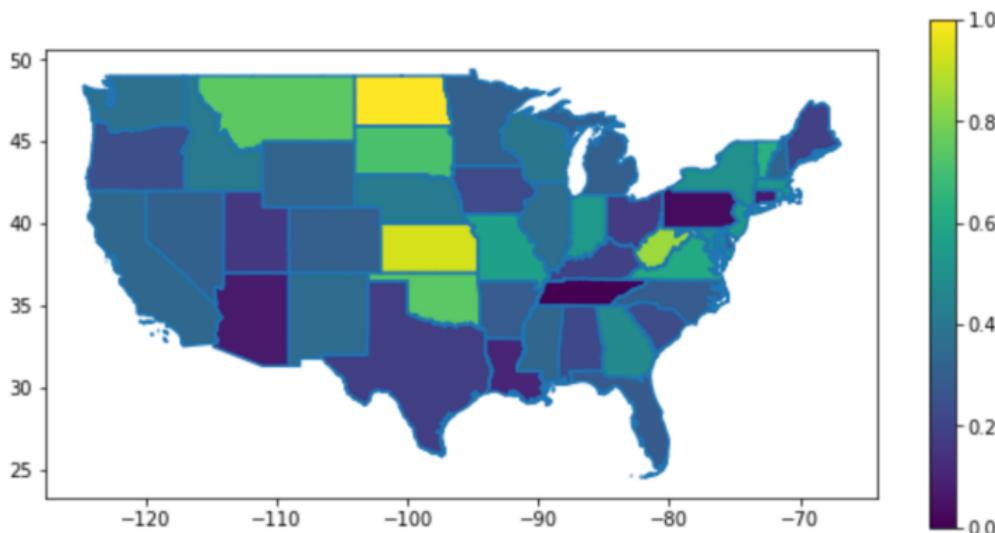


Figure 19: Heatmap of net sentiment across different US states using the textblob sentiment analysis model

Heat Maps

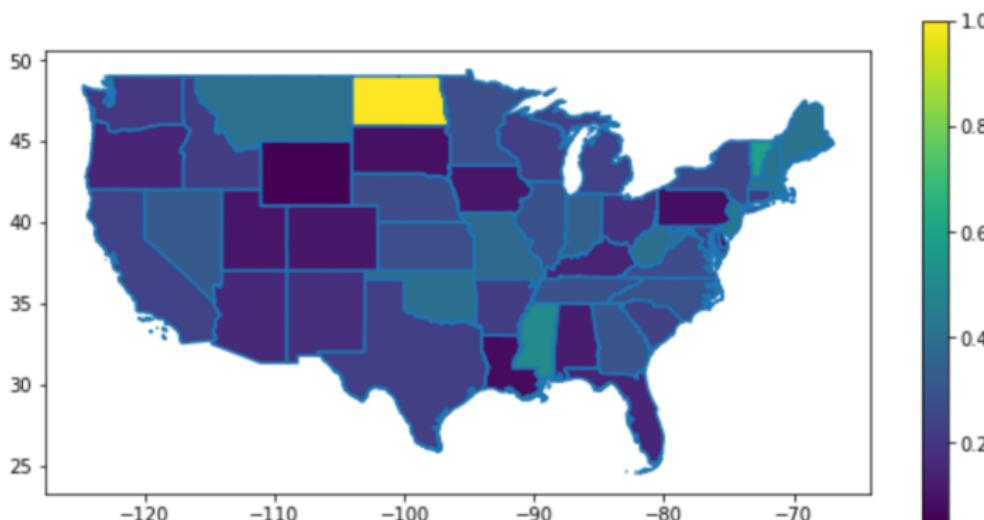


Figure 20: Heatmap of net sentiment across different US states using the VADER sentiment analysis model

Temporal tweet count visualization

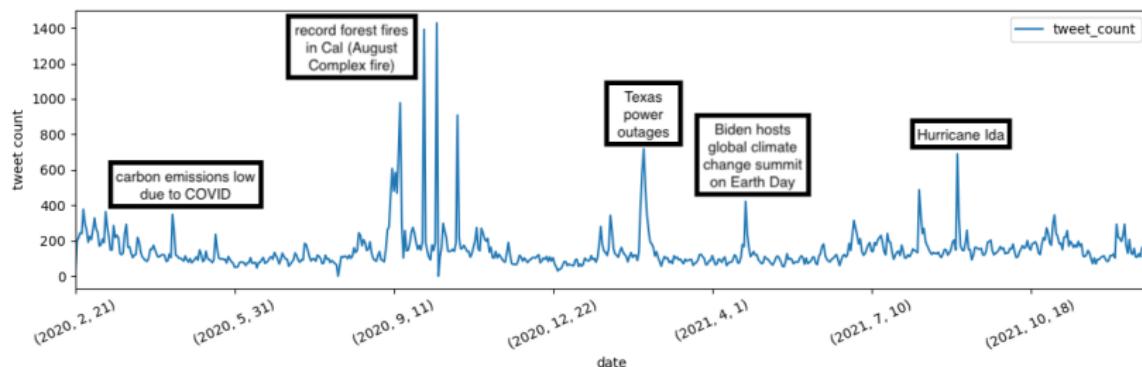


Figure 21: number of tweets per day across time, the trend of which can be correlated with different climate change related events

Temporal tweet sentiment visualization

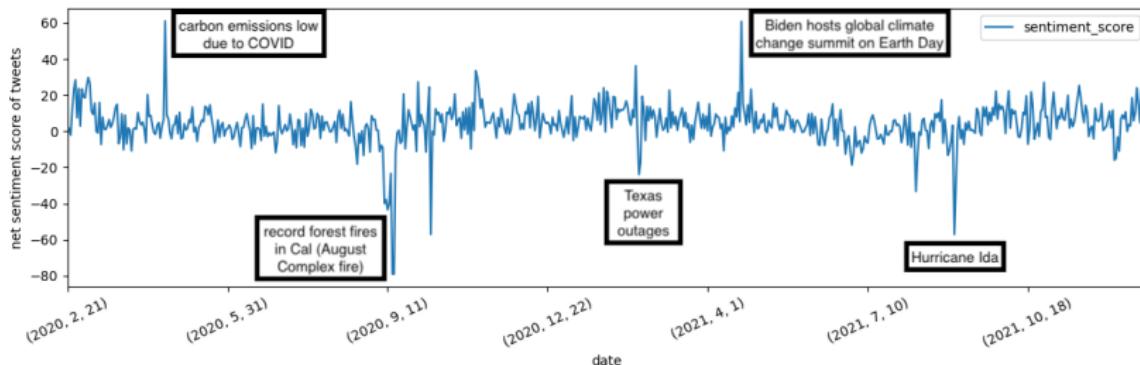


Figure 22: Net sentiment score of tweets per day across time, the trend of which can be correlated with different climate change related events

Insights

- The wordclouds of different regions reflect topics of local conversation around climate change in the region, such as politician names in the US
- The USA heatmaps show a general idea of the distribution of feelings towards climate change across different states. The state which tweets most turns out not to be the most positive one; it actually displays mixed feelings.
- From the sentiment pie charts and bar graphs, we see that those regions affected by climate change greater have a more positive discourse around it
- From the temporal analysis, we can correlate an increase in discourse and the net sentiment of discourse to real-life climate change events.

Thank You

Questions?

