Analyzing Natural Disasters by Sentiment and Volume Generated on Reddit

Hayden Outlaw

Tulane CMPS 6160

Fall 2022

Introduction

- Natural disasters generate discussion about climate change
- Given some traits about a natural disaster, can the discussion be predicted?
- Are there any traits about affected countries that change the amount/sentiment of climate change discussion?

Combine Reddit API data, natural disaster database, CO2 emission database to answer

Tulane University

Data Sources

Three datasets to pull from:

- "Our World in Data" CO2 Emissions Dataset
- "Reddit Climate Change" dataset - compiled using SocialGREP scraping API
- "EM-DAT" Natural Disaster Databases

"Reddit Climate Change" dataset came with uncited sentiment analysis, ditched and rerun with *VADER* model

Tulane University

EDA

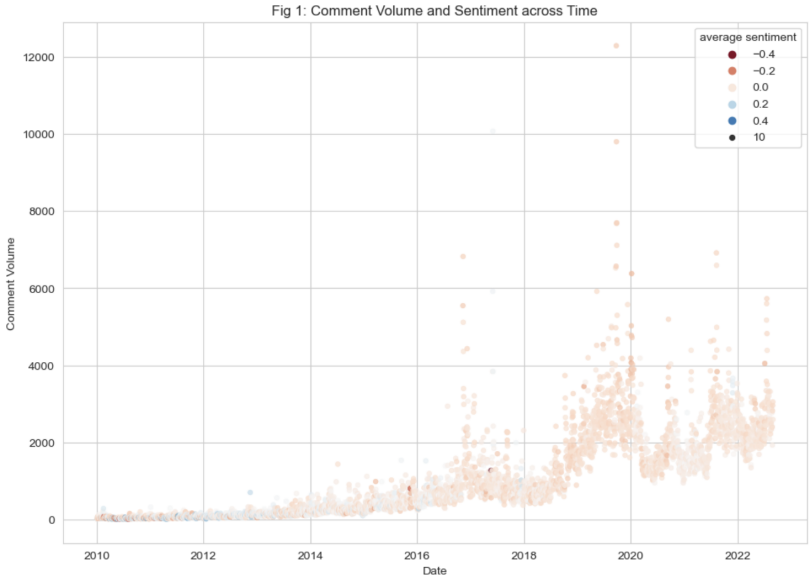


Fig 1: Comment Volume and Sentiment across Time
average sentiment
−0.4
−0.2
0.0
0.2
0.4
10
Comment Volume
Date

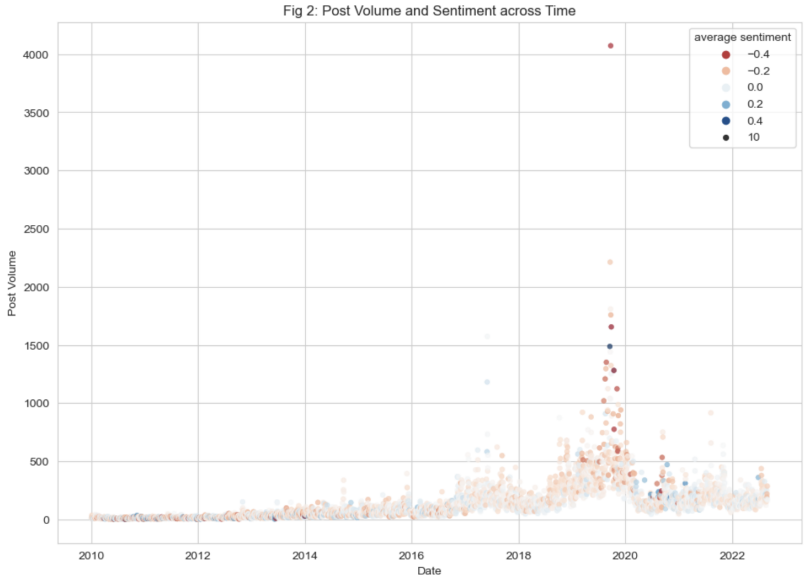


Fig 2: Post Volume and Sentiment across Time
average sentiment
−0.4
−0.2
0.0
0.2
0.4
10
Post Volume
Date

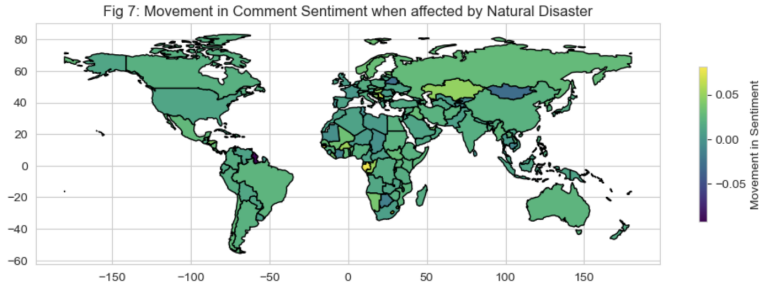


Fig 7: Movement in Comment Sentiment when affected by Natural Disaster
Movement in Sentiment

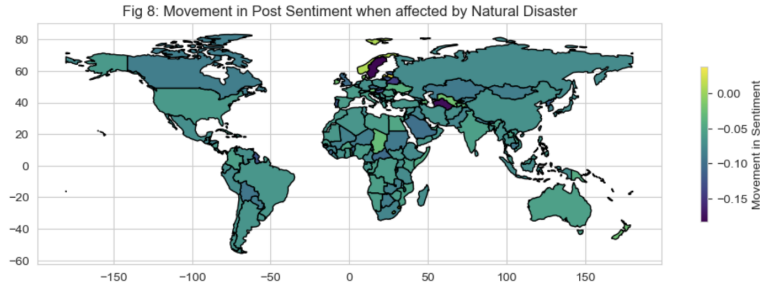


Fig 8: Movement in Post Sentiment when affected by Natural Disaster
Movement in Sentiment

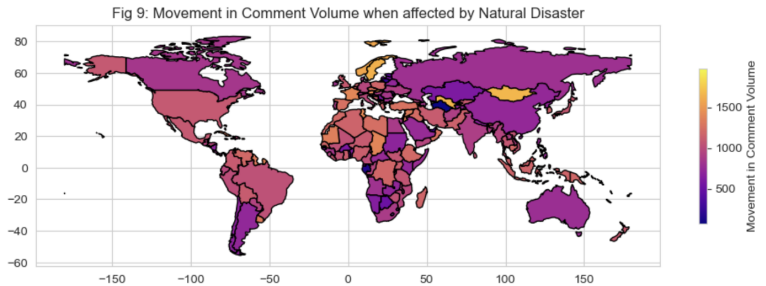


Fig 9: Movement in Comment Volume when affected by Natural Disaster
Movement in Comment Volume

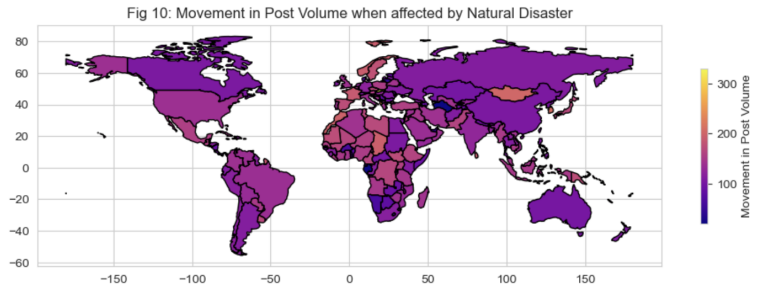


Fig 10: Movement in Post Volume when affected by Natural Disaster
Movement in Post Volume

Hypothesis

- Question: What features affect Reddit climate change discourse generation the most? Attributes of natural disaster? Attributes of countries affected?
- Hypothesis: The start date of each disaster for any given event is the strongest factor in determining discourse generation.

Tulane University

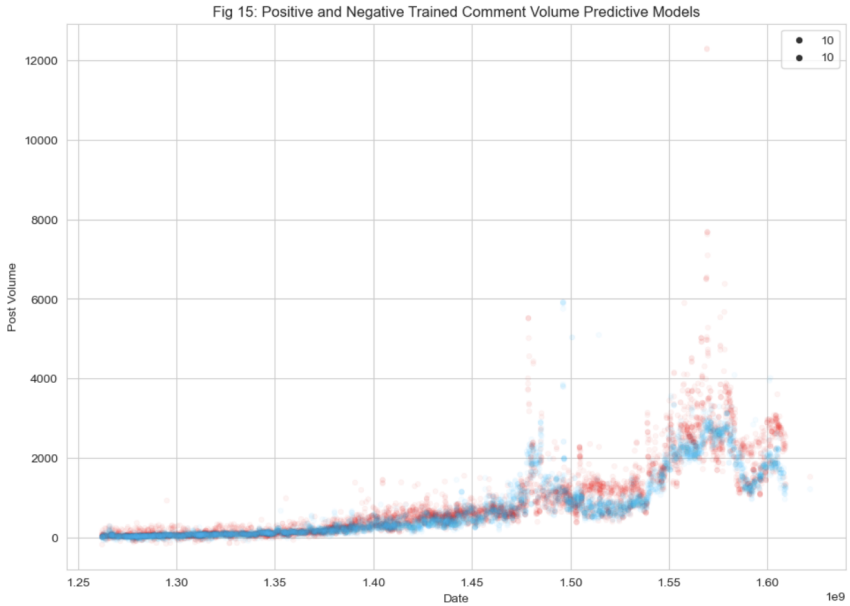
Model Selection

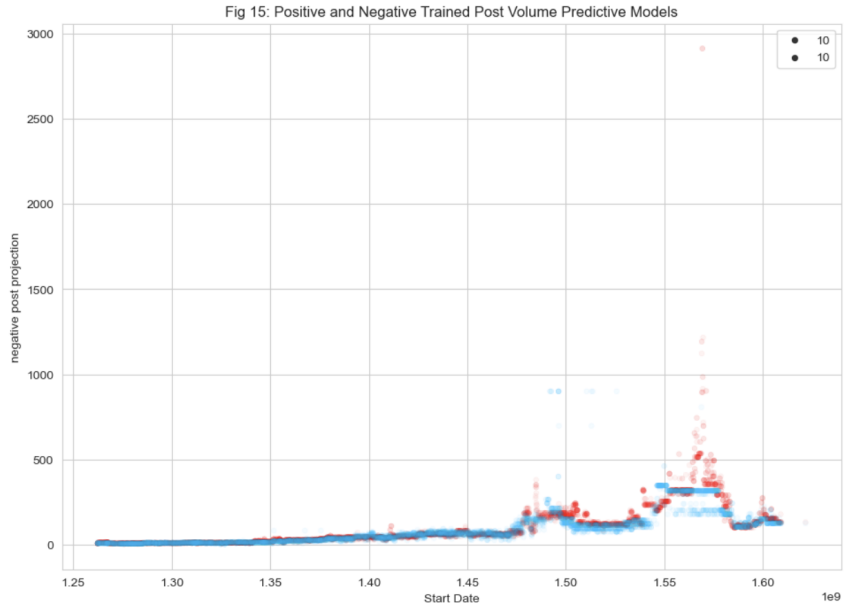
- XGB Decision Tree Regression Model - extreme gradient boosted decision tree, predict comment/post volume/sentiment as a function of presence/absence of features
- Selected due to ability to handle missing data, computational efficiency, able to handle mixed data types well

Tulane University

Model Results/Analysis - What's the Story?

- Can the social media movement be predicted? Sometimes - volume model is uncertain, but predictions made by XGB sentiment regression models are statistically significant
- What was important? Start date, specific nation vs any nation?
- Models trained on only positive or negative sentiment data





Tulane University