



Monitoring and mitigating misinformation that disrupts adaptation to climate driven health threats

Students: Roberto Lopez-Trigo, Colleen McCurdy, Nathan Chiche, Brian Rodriguez, Kyle Ghani, Kyle Zimmer, Javier Oliva, Jeffrey Quispe, Khizar Nasir, Gabriel Prieto

Mentor: Sam Malloy, Mowafak Allaham, Ayse Lokmanoglu

Instructor/Faculty: Masoud Sadiqi, Florida International University

PROBLEM

Misinformation has a great impact on the public since people are obtaining their information mostly from social media and other online sources. Navigating this chaotic space is overwhelming. People unknowingly spread information from these questionable sources, magnifying the problem. One of the most concerning areas of misinformation is climate change.

In order to shed light on the issue, our team has compiled engagement data related to climate news articles from Newswhip. Using this data we are able to analyze trends and point out topics of concern.

CONTRIBUTION

Dataset:

- tokenized text, removed stop words, common nonsensical text, punctuation, and numerical values
- converted and manipulated data as dataframe using the pandas library
- created queries activated by interactions in Plotly
- tested hypotheses, deduced from exploratory data analysis, by using BERTopic modeling
- tested and compared multiple machine learning algorithms

Database:

- created database in PostgreSQL
- developed relationships between sets of data
- established secure and reliable connections from application to database

```
import re
import string

def clean_text_round1(text):
    text = text.lower()
    text = re.sub('[\.\?!\,\',\']', '', text)
    text = re.sub('[\w*\d\w*', '', text)
    text = re.sub('[^a-zm-]', '', text)
    text = re.sub('\n', '', text)
    return text

round1 = lambda x: clean_text_round1(x)
```

```
pca = PCA(n_components=3)
data['x'] = pca.fit_transform(data['vec'].values.tolist())[:,0]
data['y'] = pca.fit_transform(data['vec'].values.tolist())[:,1]
data['z'] = pca.fit_transform(data['vec'].values.tolist())[:,2]
```

```
topic_model = BERTopic()
topics, probs = topic_model.fit_transform(X.headline.to_list())
```

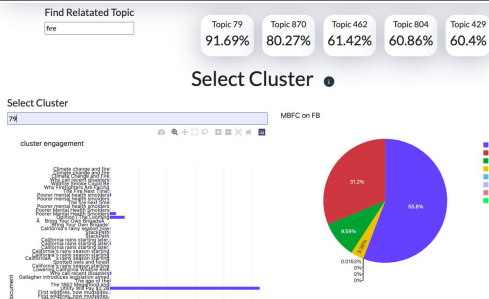
CURRENT SYSTEM

The monitoring and mitigating information project aims to spread awareness about climate misinformation and the public perception of climate related issues.

User's can focus on multiple sources and filter the data according to their needs. This data could be filtered by engagement type, publication date, and top articles. That data can then be analyzed using pie charts, histograms, bubble charts and line graphs.

We used BERTopic to cluster the data into meaningful groups allowing us to find topics related to a search term and thus the articles related to that topic.

Various climate disaster types can also be found on a United States map with the number of incidents separated by states. Estimates of the public's opinion on certain climate questions can also be seen on the survey data page. A breakdown of public interest into climate news can be found on the climate relevancy page.



We compiled the work from the previous capstone iteration and turned it into a DASH multi page application with analytical functions and visual aids in the form of dynamic graphs. We used BERTopic to cluster the articles in meaningful groups to find relevant documents to the user's searched term. All the Newswhip data static files that were being used as a reference for the scripts were put into a Postgres database. We tested the database by building a proof of concept using optimized queries.

We created a multithreaded processing scraper using beautiful soup. All UI features were added during this iteration. BERTopic used to cluster articles based on their text and headlines. Github pages were deployed and used for collaboration. Database was created, connected to and queried to from application. Further iterations will make an application that is more depended on the database.

REQUIREMENTS

Hardware:

- Mac
- Linux

Software:

- Python
- PostgreSQL
- BERTopic
- DASH

Version Control:

- GitHub (<http://www.github.com>)



Google Docs - Used to keep track of everyone's contributions to the project. Multiple meetings and their information has been recorded through Google Docs.



GitHub - A site we used to access our code. For the group to test and add changes to the project.



Plotly - Allows a point & click interface. Assists in creation of our dashboard.



PostgreSQL - Relational database management system.



Python - The main language that we used for this project.



BERTopic - Topic modeling technique we used to cluster articles based on their text and headlines.

REFERENCE



ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by Sam Malloy. We thank Sam Malloy, Mowafak Allaham, and Ayse Lokmanoglu for their mentorship and assistance that we received throughout the senior design project.