Tweeting Through the Flood: Application of BERT Topic Modeling for a Comparative Flood Communication Analysis

Christin Salley
Michigan Institute for Data Science
cjsalley@umich.edu

Nathan Fox
Michigan Institute for Data Science
foxnat@umich.edu

Alyssa Schubert
Michigan Institute for Data Science
alschub@umich.edu

ABSTRACT

Floods are prevalent disasters in the United States, posing escalating risks due to climate change-induced factors like rising sea levels and erratic rainfall patterns. Despite governmental efforts, flood risk communication remains inadequate, hindering preparedness and response capacity. While governmental agencies predominantly employ traditional, one-sided information dissemination approaches, social media platforms pose as crucial early warning indicators and sources of real-time information. This study conducts the first part of our intended comparative analysis of social media messages between governmental agencies and communities. We perform a case study on a flooding event in Michigan from May 17-20, 2020. Utilizing advanced topic modeling, we examine Twitter/X message content and sentiment from community-based posts. Insights aim to inform more effective flood communication strategies, bridging the gap between official information and community needs during events. Future work will compare social media messages' content and sentiment from governmental agencies relative to those from the community.

Keywords

Crisis informatics, flooding, natural language processing (NLP), social media, Twitter/X

INTRODUCTION AND BACKGROUND

Floods are one of the most common weather-related natural disasters in the United States (NOAA National Severe Storms Laboratory, 2024). Indeed, seventy-five percent of all Presidential Disaster Declarations in the United States can be attributed to flooding (National Weather Service, 2018). As climate change continues to contribute to rising sea levels and unstable rainfalls, the risk of floods is increasing (Tabari, 2020). In fact, the U.S. National Flood Insurance Program experienced a 660% increase in dollars paid out for flood insurance claims from 2000 to 2020 compared to the claims from 1980 to 2000 (Federal Emergency Management Agency, 2021). Beyond flood damage to physical systems, floods have been documented to substantially impact communities' physical, mental, and financial health, causing long-term changes to quality of life that can persist for years after the flood event (Van

Ootegem & Verhofstadt, 2016). For example, studies have shown increased instances in anxiety, depression, and post-traumatic stress disorder in people who have experienced a flood compared to a control group (Van Ootegem & Verhofstadt, 2016). Further, those affected by floods have been reported to experience long- or short-term injuries, respiratory and gastrointestinal illness due to exposure to pathogens via contaminated drinking water, unsafe conditions for food preparation, or lack of access to hygienic practices, and, in some cases, carbon monoxide poisoning (Burger & Gochfeld, 2014; Mulder et al., 2019; Sampson et al., 2019; Van Ootegem & Verhofstadt, 2016; Waite et al., 2014). Moreover, these effects are not experienced independently of one another; a systematic review of factors related to one's vulnerability before, during, and after a flood event determined that the effects of floods can be compounded by other risk factors, such as the frequency and severity of floods experienced or health, socioeconomic, and educational status (Lowe et al., 2013; Sampson et al., 2019). In particular, urban flooding has been shown to occur disproportionately due to racial and spatial segregation (Hughes et al., 2021). Midwest states, including cities on the Great Lakes, are particularly prone to urban flooding, which tends to occur in cities with high percentages of impervious surfaces and outdated or aged infrastructure, such as combined stormwater/sewer systems (National Academies of Sciences, 2019).

The magnitude and widespread effects of floods, as well as their increasing risk of occurrence, has motivated a number of studies examining the information available to those experiencing a flood before, during, and after the flood event. The primary means through which communities receive information about a flood is via flood risk messages from a governmental agency, such as public health, weather service, or emergency management agencies on a federal, state, or local level. However, studies investigating the effectiveness of flood risk messages from these sources have found that they are inadequate for improving individual self-efficacy, adaptive capacity, and flood preparedness (Forsyth et al., 2023; Haer et al., 2016; Scott & Errett, 2018). For example, a study examining the information available on social media via government during the 2016 Louisiana floods found that, while most messages had information about recovery resources and updates about the status of the disaster and response, few contained 'actionable requests,' or requests that were substantial and relevant enough to those experiencing the flood that they increased self-efficacy (Cooper et al., 2022; Mostafiz et al., 2022; Scott & Errett, 2018). Further, while some agencies release 'Get Ready' checklists as a tool for preparedness before a flood, they do not adequately cover the long-term impacts on mental health and overall well-being that may be experienced (Forsyth et al., 2023). Current flood risk communication practices are still primarily based on a knowledge- or information-deficit model which perpetuates one-sided communication from a governmental agency to the community (Maidl & Buchecker, 2015). These models are based on the assumption that the public does not understand the risk and that simply providing more information will result in 'better' decision-making (Maidl & Buchecker, 2015; O'Sullivan et al., 2012).

An increasingly recognized additional source of flood risk information are the messages, communicated via social media, by the community who is affected by the flood. Indeed, in an effort to improve flood mitigation and management practices, several studies have demonstrated the effectiveness of community-based social media messages as early warning indicators of floods (Carlos Villagrán De León et al., 2013; Cools et al., 2016). For example, a study analyzing the content of 60,000 tweets over a 5-day flood period found evidence of early and local detection of multiple-flood events (Shoyama et al., 2021). A method for identifying flood events globally using social media messages to build a real-time and historic database has been created (de Bruijn et al., 2019). A recent study developed a method for real-time sentiment analysis of social media messages to provide up-to-date information on community needs during a flood (Bryan-Smith et al., 2023); another designed a multilingual multimodal neural network to classify tweets containing information related to a flood with high precision ((de Bruijn et al., 2020). While these studies use a variety of approaches to analyze the content and sentiment of social media messages, machine learning methods such as natural language processing (NLP) can considerably speed up the process. Topic modeling is an NLP technique that generates structured text data from unstructured text data inputs, and in doing so, allows for the discovery of latent topics in the text. Analyzing social media messages for content and sentiment using NLP or topic modeling has been well-established in the literature. In the disaster space, for example, topic modeling has been used to identify the sentiments and thoughts of populations during flood events (Choirul Rahmadan et al., 2020), hurricanes (Mihunov et al., 2022; Zhou et al., 2023), and earthquakes (Vo & Collier, 2013).

Few studies have completed a combined analysis of the social media messages from governmental agencies and community networks during a flood event. However, given the observed gaps in flood risk communication from governmental agencies, there is an opportunity to learn about the community's needs and concerns from the content and sentiment of community-based social media messages in comparison to government-based social media messages during flood events. A better understanding of these needs can be used to build more effective communication practices and reconcile the gap between what is provided versus what is needed by communities before, during, and after a flood. To that end, this preliminary study conducts a comparative analysis of the content and sentiment of tweets from governmental agencies versus community residents for the state of Michigan during a flood event from May 17 - 20, 2020 using topic modeling. We aim to answer the following research questions: (1) What insights can be gained from a comparative study of these social media messages in the context and sentiments of flood response and communication strategies?, and (2) Do the frequency of specific topics, message content, and sentiments of social media messages during the flood event differ between governmental (i.e., state and federal) agencies and community residents?

RESEARCH DESIGN AND METHODS

Case Study

To investigate our research questions thoroughly, we performed a case study analyzing a flooding event that occurred in Michigan during the year 2020. Midland County, Michigan is located in close proximity to Saginaw Bay, which connects to the Great Lake Huron. On May 17, 2020, record rainfall in Midland County began due to a low pressure system and frontal boundary from the Great Lakes (National Weather Service, 2020). Consequently, the Saginaw and Tittabawasee Rivers flooded. Heavy rains, totaling 5 to 8 inches, in the Tri-Cities region (i.e., Bay County, Midland County, and Saginaw County) resulted in catastrophic failures of the Edenville and Sanford dams in Midland County. This prompted the issuance of Flash Flood Emergencies, signaling life-threatening flooding and the evacuation of 10,000 Midland residents. Significant flooding also occurred along several lake shores, affecting several other counties in southeast Michigan. The rainfall fell within thirty-six hours and the event was considered over on May 20, 2020. This flood event was the worst that had occurred in the region since 1986, prompting discussion by those throughout the state on social media.

Data Collection and Filtering

Twitter data were collected for the state of Michigan from between May 15, 2020 at 20:00:09 and May 23, 2020 at 19:59:57 using a Twitter/X streaming API built in the Network Dynamics Lab (Wang & Taylor, 2015). In the realm of disaster studies, flooding is a prime example of what is known as a short-notice disaster (Chiu & Zheng, 2006). This means agencies typically have only 24 to 72 hours to send out important risk communication messages, such as evacuation notices (Wolshon et al., 2001). Furthermore, upon our initial analysis, we observed a consistent decline in the frequency of tweets concerning our case study within three days following the conclusion of the event. Therefore, the date and time range of May 15, 2020 to May 23, 2020 were chosen to encompass the average timeframe for a short-notice disaster (i.e., 48 hours prior) prior to the beginning of the flood event and three days after the incident. To compare the content and sentiment of tweets between governmental agencies and the public, the Twitter data were filtered for twelve specific state and federal agency handles (Table 1). This list was determined by including the social media platforms for relevant Executive Branch departments in the state (State of Michigan, 2024), as well as common federal agencies employed for disastrous events such as floods. Tweets from these handles were treated as the messages from governmental agencies and all other remaining tweets from the data collection were considered to be from the public or community-based.

Table 1. List of state and federal governmental agencies for comparative analysis.

Agency Name	Twitter Handle
Environmental Protection Agency (EPA)	@EPA
Federal Emergency Management Agency (FEMA)	@fema
Michigan Department of Environment, Great Lakes, and Energy (EGLE)	@MichiganEGLE
Michigan Department of Health and Human Services	@MichiganHHS
Michigan Department of Natural Resources	@MichiganDNR
Michigan Department of Transportation (MDOT)	@MichiganDOT
Michigan State Police (MSP)	@MichStatePolice
National Weather Service (NWS)	@NWS
National Weather Service (NWS)	@NWSFlashFlood
National Weather Service (NWS)	@NWSSevereTstorm
State Emergency Management Agency (Michigan State Police, Emergency Management and Homeland Security Division)	@MichEMHS
State of Michigan government agencies	@migovernment

Topic Modeling

Next, to analyze the Twitter/X dataset collected, BERT (Bidirectional Encoder Representations from Transformers), an advanced NLP model introduced by Google in 2018 (Devlin et al., 2018), was used. BERT utilizes bidirectional encoding to generate contextualized word embeddings, revolutionizing various NLP tasks. For this project, we specifically employed BERTopics, an extension of BERT, that leverages its contextualized embeddings to perform topic modeling by aggregating document-level representations derived from BERT embeddings and subsequently clustering similar documents based on their semantic content. This method enables fine-grained topic differentiation and accurate thematic analysis, and has been used successfully to assess Twitter/X datasets for a range of environmental topics (Kastrati et al., 2023; Pampapura Madali et al., 2022).

We carried out hyperparameter tuning to optimize the performance of our topic modeling approach. During this process, we tuned the parameters of three key models: UMAP (Uniform Manifold Approximation and Projection) for dimensionality reduction, HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) for clustering, and BERTopic for topic modeling. Specifically, we adjusted parameters such as the number of neighbors and components for UMAP, the minimum cluster size and minimum samples for HDBSCAN, and minimum topic size for BERTopic. For a comprehensive list of the tuned parameters, please refer to the Supplementary Information at the end of the paper (SI 1).

To understand the meaning of each topic, we manually interpreted them and then labeled them based on their keywords to best capture their core meaning. The topic modeling tool assigns a c-TF-IDF (class-based term frequency-inverse document frequency) score to each word in the topic, indicating its relevance to the topic. These scores are ranked by importance, with the highest-scoring words most closely mirroring the central theme of the topic. While more topics can contain up to ten words, those with the highest c-TF-IDF score better reflect the overall meaning of the topic.

ANALYSIS

Results: Entire Dataset

We extracted 124,362 tweets from May 15 to May 23. The topic modeling clustered the tweets into 133 topics. The four most tweeted topics from the entire dataset were about “Food” (Topic 0, 4,583 tweets), “Thankfulness” (Topic 1, 3,411), “Jobs” (Topic 2, 2,537), and “Music” (Topic 3, 2,402) (Figure 1).

Figure 1. Top most tweeted topics from May 15 to May 23 in Michigan.

The topic most related to flooding events in Midland was Topic 15, which had 892 tweets in the cluster. The five most important words in Topic 15 were “Midland”, “Dam”, “Flooding”, “Water”, and “River” (Figure 2). Figure 2 also demonstrates the decline in score c-TF-IDF for each additional word included in the topic. The frequency of tweets categorized as belonging to Topic 15 varied over the search period. To start, there were virtually no tweets about the flood from between May 16 and May 19, followed by a rapid increase and peak posting on May 20, and then a decline between May 20 and May 23; however, during this period, the Tweet frequency remained higher than before the event (Figure 3).

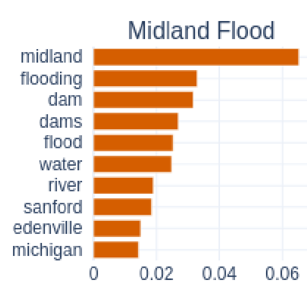

Figure 2. The top ten words associated with Topic 15, “Midland Flood”.

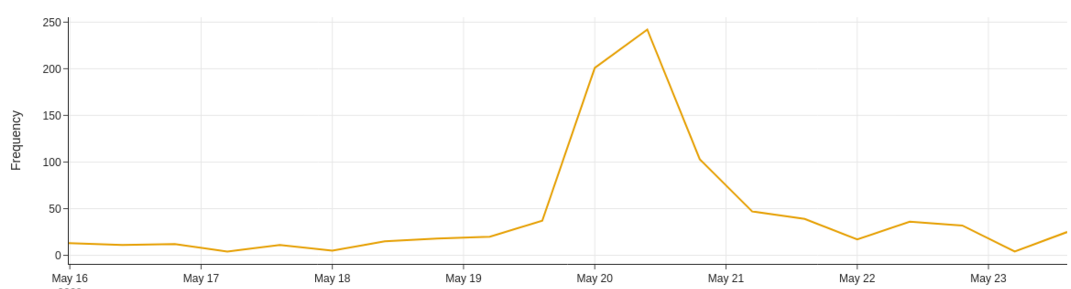

Figure 3. Frequency of Tweets associated with Topic 15 from May 16 to May 23, 2020.

The sentiment analysis results showed that for Topic 15 the most conveyed emotions were fear (347 tweets) and sadness (176) (Figure 4).

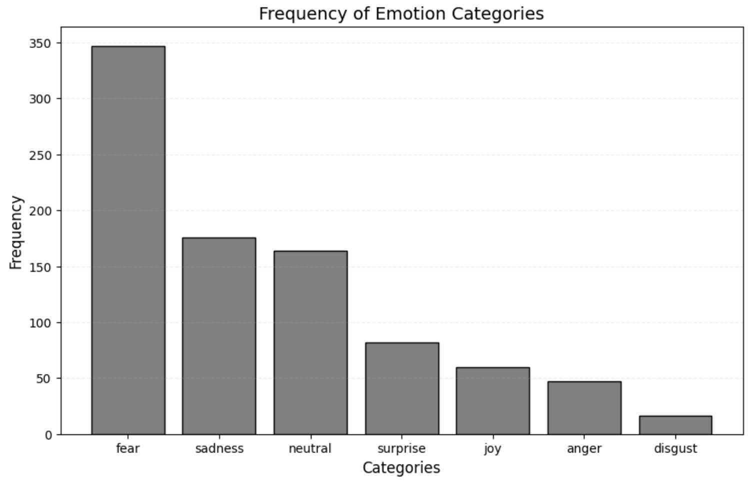

Figure 4. Sentiment Analysis of Topic 15 Tweets.

A hotspot analysis indicates that there is geographical variation in the location of tweets regarding the flooding events. Unsurprisingly, the location with the highest density of tweets was in and around Midland (Figure 5). There were also a large number of tweets occurring in other major Michigan cities such as Detroit and Lansing.

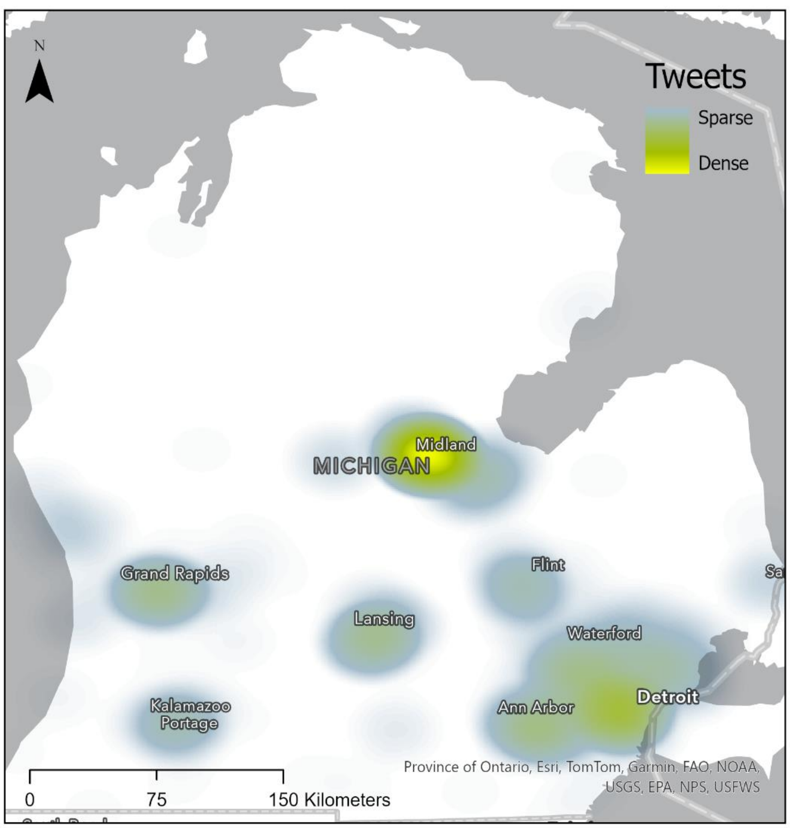

Figure. 5 Heatmap of tweets in the state of Michigan concerning the May 2020 flooding incident.

Results: Initial Comparison of Community vs. Government Tweets

Upon analyzing the dataset and dividing it into community and government tweets concerning the flooding event, we encountered a disproportionately small sample size of government tweets (13 out of 892 tweets; 1.4%) compared to the extensive volume generated by community members. All agency tweets were by either NWSDetoit (6 tweets) or NWSGaylord (5 tweets). Of these tweets, only three were directly related to Midland County, and a further three related to the associated flooding occuring in the village of Sanford, while all other tweets focused on a Gladwin County flooding event that occurred further north. This discovery perpetuates the ongoing notion of inadequate communication from key stakeholders at the local, state, and federal levels and the communities they serve, particularly on platforms as influential and interactive as social media, and identifies an opportunity to reach a broader audience through utilizing social media as a communication medium.

As our research remains ongoing, we intend to conduct another case study focusing on a different flooding event. Our aim is to either achieve a more robust sample size for both groups or pivot the study's focus to further comprehensively analyze the entire dataset, thereby uncovering novel insights in the forthcoming weeks. We anticipate presenting our findings at this conference upon acceptance and final submission of this work in progress paper.

DISCUSSION AND CONCLUSIONS

The particular insights gained from our study in the context and sentiments of flood response and communication strategies was the discovery that tweets surrounding the flood predominantly conveyed fear. Additionally, the frequency of tweets for Topic 15 (i.e., the topic most relevant to the incident itself) began rising mid-day on May 19, rapidly increasing on the 20th and peaking mid-day on the 21st. Overall, the highest volume of tweets for Topic 15 occurred between May 19 and May 22 with discussion continuing, to a lesser extent, on May 23. Rainfall began on May 17 and ended approximately thirty-six hours later. The temporal trends of tweet volume starting on May 19 shows, in near real-time, the community discussion and reactions to the impacts of the rainfall, including catastrophic flooding and dam breakages. Such insights are invaluable for stakeholders involved in risk communication, aiding in the development of initiatives and strategies to effectively disseminate information that promotes community safety during such disasters. This includes understanding not only the content to convey but also the timing when people are most likely to engage with and absorb such messages amidst the peak of discussion about the event.

Although this study effectively utilized the BERT model to identify communication inefficiencies, certain limitations must be acknowledged. Firstly, as is common in social media research, there exists variability in the collected tweets, including differences in the frequency of tweets and how engaged users are. Notably, the volume of community-scale tweets significantly outweighed those from governmental sources. It is possible that governmental agencies were communicating about the flood on another social media platform other than Twitter. Given the volume of community Tweets during the flood event, the lack of tweets from governmental sources indicate that there may be a missed opportunity to communicate important information about the flood via this platform. Consequently, the lack of a balanced sample size may restrict the generalizability of the conclusions to other flooding events. Additionally, the filtering process (i.e., the selected state and government entities) implemented during data collection may have influenced the results obtained, potentially impacting the comprehensiveness of the findings. For example, we have not captured any tweets that may have originated from national agencies because we filtered for tweets from the state of Michigan only (i.e., EPA and FEMA).

In future research, it is imperative to explore additional social media networks for government-based tweets. For instance, investigative reporters or news journalists that work on the local, state, and federal level and their networks can be included in future works to improve the dataset for flood risk communication from government stakeholders. Additionally, expanding the analysis to encompass tweets from local agencies can be advantageous. We initially concentrated on state and federal entities, and finding more local agencies would provide a more comprehensive understanding of flood-related communication patterns and potentially increase our dataset. To this end, future investigations should delve deeper into other forms of classification in the content of flood-related tweets, following the approach outlined by (Scott & Errett, 2018). For instance, examining tweets within specific topics, such as Topic 15, could shed light on the discussion of resource requirements, incidents of downed power lines, and other relevant themes. Lastly, previous studies have found different predominant emotion from tweets surrounding other natural hazard events such anger associated with hurricanes (Berbère et al., 2023). This disparity may show that different natural hazard event types may evoke different emotion responses, or that public emotion may change throughout the timing of the event. As these studies only looked at the dominance of emotion and not temporal patterns, future analysis of such data should map the temporal distribution of emotion to best understand public perceptions. For example, fearful tweets may occur primarily before the event, whilst angry tweets may occur more frequently in the aftermath.

In conclusion, our work highlights how pressing challenges in flood risk communication persist, emphasizing the need for innovative approaches. While a more comparative analysis is forthcoming, the lack of government-based tweets generated again still show significant inefficiencies from these stakeholders to communicate with citizens. By analyzing social media messages, illustrated during a flooding event in Michigan, the study emphasizes the potential of social media platforms in enhancing real-time information dissemination and community engagement. These findings call for a paradigm shift towards more collaborative and responsive flood communication strategies to better address community needs and bridge the gap between official agencies and the public during flood events.

ACKNOWLEDGEMENTS

This project is supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors. We would also like to extend a special thank you to the Network Dynamics Lab at the Georgia Institute of Technology, for providing their archived Twitter/X postings for us to utilize in this research.

REFERENCES

Berbère, R., Elkefi, S., Layeb, S. B., & Tounsi, A. (2023). Exploring Cognitive Sustainability Concerns in Public Responses to Extreme Weather Events: An NLP Analysis of Twitter Data. *Cognitive Sustainability, 2(4)*.

Bryan-Smith, L., Godsall, J., George, F., Egode, K., Dethlefs, N., & Parsons, D. (2023). Real-time social media sentiment analysis for rapid impact assessment of floods. *Computers and Geosciences*, *178*. https://doi.org/10.1016/j.cageo.2023.105405

Burger, J., & Gochfeld, M. (2014). Health concerns and perceptions of central and coastal New Jersey residents in the 100days following Superstorm Sandy. *Science of the Total Environment*, *481*(1), 611–618. https://doi.org/10.1016/j.scitotenv.2014.02.048

Carlos Villagrán De León, J., Pruessner, I., & Breedlove, H. (2013). *Alert and Warning Frameworks in the Context of Early Warning Systems: A Comparative Review*.

Chiu, Y. C., & Zheng, H. (2007). Real-time mobilization decisions for multi-priority emergency response resources and evacuation groups: model formulation and solution. *Transportation Research Part E: Logistics and Transportation Review*, *43*(6), 710-736.

Choirul Rahmadan, M., Nizar Hidayanto, A., Swadani Ekasari, D., Purwandari, B., & Theresiawati. (2020). Sentiment Analysis and Topic Modelling Using the LDA Method related to the Flood Disaster in Jakarta on Twitter. *Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, 126–130. https://doi.org/10.1109/ICIMCIS51567.2020.9354320

Cools, J., Innocenti, D., & O'Brien, S. (2016). Lessons from flood early warning systems. *Environmental Science & Policy*, *58*, 117–122. https://doi.org/10.1016/j.envsci.2016.01.006

Cooper, C. M., Sharma, S., Nicholas, R. E., & Keller, K. (2022). Toward More Actionable Flood-Risk Information. In *Earth's Future* (Vol. 10, Issue 11). John Wiley and Sons Inc. https://doi.org/10.1029/2022EF003093

de Bruijn, J. A., de Moel, H., Jongman, B., de Ruiter, M. C., Wagemaker, J., & Aerts, J. C. J. H. (2019). A global database of historic and real-time flood events based on social media. *Scientific Data*, *6*(1). https://doi.org/10.1038/s41597-019-0326-9

de Bruijn, J. A., de Moel, H., Weerts, A. H., de Ruiter, M. C., Basar, E., Eilander, D., & Aerts, J. C. J. H. (2020). Improving the classification of flood tweets with contextual hydrological information in a multimodal neural network. *Computers and Geosciences*, *140*. https://doi.org/10.1016/j.cageo.2020.104485

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. http://arxiv.org/abs/1810.04805

Federal Emergency Management Agency. (2021). *Climate Change Increases Risk of Flooding*. https://www.floodsmart.gov/climate-change#:~:text=The%20risk%20of%20flooding%20is,and%20rising%20risk%20of%20flooding.

Forsyth, W., Roberts, T., & Brewer, G. (2023). Conceptualising risk communication barriers to household flood preparedness. *Urban Governance*, *3*(2), 116–129. https://doi.org/10.1016/j.ugj.2023.02.001

Haer, T., Botzen, W. J. W., & Aerts, J. C. J. H. (2016). The effectiveness of flood risk communication strategies and the influence of social networks-Insights from an agent-based model. *Environmental Science and Policy*, *60*, 44–52. https://doi.org/10.1016/j.envsci.2016.03.006

Hughes, S., Schwarz, K., Lemee, G. L., Dobie, S., Lane, M., Gonzalez, A., Turner, K., Norton, R., Baja, K., Howrani-Heeres, J., & Stults, M. (2021). *Centering Racial Justice in Urban Flood Adaptation: Planning and Evaluation Tools for Great Lakes Cities*.

Kastrati, Z., Imran, A. S., Daudpota, S. M., Memon, M. A., & Kastrati, M. (2023). Soaring Energy Prices: Understanding Public Engagement on Twitter Using Sentiment Analysis and Topic Modeling with Transformers. *IEEE Access*, *11*, 26541–26553. https://doi.org/10.1109/ACCESS.2023.3257283

Lowe, D., Ebi, K. L., & Forsberg, B. (2013). Factors increasing vulnerability to health effects before, during and after floods. *International Journal of Environmental Research and Public Health*, *10*(12), 7015–7067. https://doi.org/10.3390/ijerph10127015

Maidl, E., & Buchecker, M. (2015). Raising risk preparedness by flood risk communication. *Natural Hazards and Earth System Sciences*, *15*(7), 1577–1595. https://doi.org/10.5194/nhess-15-1577-2015

Mihunov, V. V., Jafari, N. H., Wang, K., Lam, N. S. N., & Govender, D. (2022). Disaster Impacts Surveillance from Social Media with Topic Modeling and Feature Extraction: Case of Hurricane Harvey. *International Journal of Disaster Risk Science*, *13*(5), 729–742. https://doi.org/10.1007/s13753-022-00442-1

Mostafiz, R. Bin, Rohli, R. V., Friedland, C. J., & Lee, Y. C. (2022). Actionable Information in Flood Risk Communications and the Potential for New Web-Based Tools for Long-Term Planning for Individuals and Community. *Frontiers in Earth Science*, *10*(February). https://doi.org/10.3389/feart.2022.840250

Mulder, A. C., Pijnacker, R., De Man, H., Van De Kassteele, J., Van Pelt, W., Mughini-Gras, L., & Franz, E. (2019). 'Sickenin'' in the rain" - Increased risk of gastrointestinal and respiratory infections after urban pluvial flooding in a population-based cross-sectional study in the Netherlands'. *BMC Infectious Diseases*, *19*(1). https://doi.org/10.1186/s12879-019-3984-5

National Academies of Sciences, E. and M. (2019). *Framing the Challenge of Urban Flooding in the United States*. The National Academies Press. https://doi.org/https://doi.org/10.17226/25381

National Weather Service. (2018). *Flood Related Hazards*. https://www.weather.gov/safety/flood-hazards

National Weather Service. (2020). *Historic Flooding May 17-20, 2020*. https://www.weather.gov/dtx/HistoricFlooding-May-17-20-2020

NOAA National Severe Storms Laboratory. (2024). *Severe Weather 101*. https://www.nssl.noaa.gov/education/svrwx101/floods/

O'Sullivan, J. J., Bradford, R. A., Bonaiuto, M., De Dominicis, S., Rotko, P., Aaltonen, J., Waylen, K., & Langan, S. J. (2012). Enhancing flood resilience through improved risk communications. *Natural Hazards and Earth System Science*, *12*(7), 2271–2282. https://doi.org/10.5194/nhess-12-2271-2012

Pampapura Madali, N., Alsaid, M., & Hawamdeh, S. (2022). *Social media and people's perception of global warming during critical environmental events: the impact of misinformation through the lens of social noise*.

Sampson, N. R., Price, C. E., Kassem, J., Doan, J., & Hussein, J. (2019). "We're Just Sitting Ducks": Recurrent Household Flooding as An Underreported Environmental Health Threat in Detroit's Changing Climate. *International Journal of Environmental Research and Public Health*, *16*(1). https://doi.org/10.3390/ijerph16010006

Scott, K. K., & Errett, N. A. (2018). Content, Accessibility, and Dissemination of Disaster Information via Social Media during the 2016 Louisiana Floods. *Journal of Public Health Management and Practice*, *24*(4), 370–379. https://doi.org/10.1097/PHH.0000000000000708

Shoyama, K., Cui, Q., Hanashima, M., Sano, H., & Usuda, Y. (2021). Emergency flood detection using multiple information sources: Integrated analysis of natural hazard monitoring and social media data. *Science of the Total Environment*, *767*. https://doi.org/10.1016/j.scitotenv.2020.144371

State of Michigan. (2024). *State of Michigan Executive Branch*. https://www.michigan.gov/som/government/branches-of-government/executive-branch

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-70816-2

Van Ootegem, L., & Verhofstadt, E. (2016). Well-being, life satisfaction and capabilities of flood disaster victims. *Environmental Impact Assessment Review*, *57*, 134–138. https://doi.org/10.1016/j.eiar.2015.12.001

Vo, B.-K. H., & Collier, N. (2013). Twitter Emotion Analysis in Earthquake Situations. *International Journal of Computational Linguistics and Applications*, *4*(1), 159–173.

Waite, T., Murray, V., & Baker, D. (2014). Carbon Monoxide Poisoning and Flooding: Changes in Risk Before, During and After Flooding Require Appropriate Public Health Interventions. *PLoS Currents*. https://doi.org/10.1371/currents.dis.2b2eb9e15f9b982784938803584487f1

Wang, Q., & Taylor, J. E.. (2015). *Process Map for Urban-Human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms*. https://doi.org/10.1061/(ASCE)CP.1943-5487

Wolshon, P. B., Urbina, E. A., & Levitan, M. L. (2001). *National review of hurricane evacuation plans and policies*. Baton Rouge, LA: Louisiana State University Hurricane Center.

Zhou, S., Kan, P., Huang, Q., & Silbernagel, J. (2023). A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. *Journal of Information Science*, *49*(2), 465–479. https://doi.org/10.1177/01655515211007724

APPENDIX A. UMAP, HDBSCAN, and BERTopic Model Parameters

```
umap_model = umap.UMAP(n_neighbors= 10,
                       n_components= 10,
                       min_dist=0.0,
                       metric='cosine',
                       random_state= 632
                       )


hdbscan_model = HDBSCAN(min_cluster_size= 622,
                        min_samples=5,
                        metric='euclidean',
                        prediction_data=True)


topic_model = BERTopic(language="english",
                       min_topic_size = 622,
                       top_n_words = 10,
                       calculate_probabilities=False,
                       verbose=True,
                       umap_model=umap_model,
                       hdbscan_model=hdbscan_model)
```