GEORGIA DOT RESEARCH PROJECT 20-13

Final Report

COMMUNITY AUGMENTED RAPID-RESPONSE TO EVENTS (CARE) INTEGRATED CRISIS COMMUNICATION SYSTEM

GDOT

Georgia Department of Transportation

Office of Performance-based Management and Research

600 West Peachtree Street NW | Atlanta, GA 30308

December 2023

TECHNICAL REPORT
DOCUMENTATION PAGE

1. Report No. FHWA-GA-23-2013	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A
4. Title and Subtitle Community Augmented Rapid-response to Events (CARE) Integrated Crisis Communication System		5. Report Date February 2024 6. Performing Organization Code N/A
7. Author(s) Christin Salley (Graduate Research Assistant) Jiajia Xie (Graduate Research Assistant) Mathias Bermeo (Graduate Research Assistant) Maithly Bishnoi (Graduate Research Assistant) Amanda Ehrenhalt (Undergraduate Research Assistant) Neda Mohammadi (Co-PI), PhD John E. Taylor (PI), PhD		8. Performing Organization Report No. 20-13
9. Performing Organization Name and Address Georgia Institute of Technology School of Civil and Environmental Engineering 756 W Peachtree St NW, Atlanta, GA 30308		10. Work Unit No. N/A 11. Contract or Grant No. PI#0017428
12. Sponsoring Agency Name and Address Georgia Department of Transportation (SPR) Office of Performance-based Management and Research 600 West Peachtree St. NW Atlanta, GA 30308		13. Type of Report and Period Covered Final Report (August 2020-February 2024) 14. Sponsoring Agency Code N/A
15. Supplementary Notes Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.		
16. Abstract This project explores the potential of integrating social media data (i.e., Twitter/X) with data from community awareness applications (i.e., Waze) to bolster crisis communication and enhance the accuracy and promptness of incident identification, assessment, and reporting during emergencies. To navigate the challenges posed by varied data formats and the online nature of this information, the research employs a multi-pronged approach. This includes; (1) interviews with multiple state DOTs for insights on event detection software implementation and experience with social media and community data, (2) the development of a weighted competency matrix built from factors identified in the literature and from the aforementioned state DOT interviews to assess existing event detection software solutions, and (3) the development and application of two unique machine learning frameworks aimed at optimizing data integration and amplifying social media to augment DOT crisis event detection. This fusion promises heightened situational awareness and improved response times, establishing a foundation for equitable, real-time crisis detection with an approach that significantly boosts the potential of topic models, as well as improves information management, anomaly detection, and resource allocation, which are critical for adaptive decision-making in evolving crisis event conditions.		
17. Key Words Emergency Communication Systems, Hazard Analysis, Social Media, Data Mining, Geospatial Data		18. Distribution Statement No restrictions.

19. Security Classification (of this re- port) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 119	22. Price Free

GDOT Research Project 20-13

Final Report

COMMUNITY AUGMENTED RAPID-RESPONSE TO EVENTS (CARE) INTEGRATED CRISIS COMMUNICATION SYSTEM

By

Christin Salley
Graduate Research Assistant

Jiajia Xie
Graduate Research Assistant

Mathias Bermeo
Graduate Research Assistant

Maithly Bishnoi
Graduate Research Assistant

Amanda Ehrenhalt
Undergraduate Research Assistant

Neda Mohammadi, PhD
Research Faculty, City Infrastructure Analytics Director

John E. Taylor, PhD
Frederick L. Olmsted Professor

Georgia Tech Research Corporation, Georgia Institute of Technology

Contract with
Georgia Department of Transportation

In cooperation with
U.S. Department of Transportation, Federal Highway Administration

December 2023

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

ATMS	Advanced Traffic Management System
CMS	Crisis Management Systems
CSV	Comma-Separated Values
GeoJSON	Geographic JavaScript Object Notation
ITS	Intelligent Transportation Systems
JSON	JavaScript Object Notation
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
TMC	Transportation Management Center
TSMO	Transportation Systems Management & Operations
XML	Extensible Markup Language

EXECUTIVE SUMMARY

Digital forms of citizen communication with response organizations through social media continue to be widespread during disasters and will continue to be used for the foreseeable future. Public agencies can use this information to examine community sentiments and discussions to assess, determine, and prioritize critical areas in need of assistance. However, there are privacy and data volume concerns, along with limitations on harnessing precise geolocation information from social media coupled with a need to mitigate bias of machine learning models used during such events. These limitations can restrict emergency management personnel's ability to locate and promptly delineate actionable insights and discourages stakeholders from fully harnessing the potential of social media to provide valuable information in crisis response.

This project explores the potential of integrating social media data (i.e., Twitter/X) with data from community awareness applications (i.e., Waze) to bolster crisis communication and enhance the accuracy and promptness of incident identification, assessment and reporting during emergencies. To navigate the challenges posed by varied data formats and the online nature of this information, the research employs a multi-pronged approach. This includes; (1) interviews with multiple state DOTs for insights on event detection software implementation and experience with social media and community data, (2) the development of a weighted competency matrix built from factors identified in the literature and from the aforementioned state DOT interviews to assess existing event detection software solutions, and (3) the development and application of unique machine learning frameworks aimed at optimizing data integration and amplifying social media to augment DOT crisis event detection.

The interviews reveal that among state DOTs interviewed RITIS is widely used for event detection. Although RITIS is highly versatile, it was learned that social media data is not being ingested into its current processes or systems for event detection. The competency matrix developed is used to evaluate promising event detection software, which was found to lack functionality to automate the fusion of social media data into community data streams for event detection. To address this gap, two applications were developed by the research team to test the feasibility and precision of integrating social media into DOT systems that ingest community data (i.e., Waze) with tailored models. The first application counters the aforementioned limitations by proposing a semi-supervised machine learning model that utilizes Transfer Learning, Topic Modeling (i.e., LDA), and Natural Language Processing. By merging historical social media data with community-driven alerts, this model augments the understanding of emergency event locations and contexts while mitigating biases using the Wells-DuBois Protocol. The efficacy of the model is illustrated through a case study application on hurricanes. This fusion promises heightened situational awareness and improved response times, establishing a foundation for equitable, real-time crisis detection. The second model and application delves into the role of probabilistic topic models analyzing online data in real time. Applying such a model to online user-generated content poses challenges due to sparse relevant data. To address this, a novel approach was executed that integrates variational lower bounds with a linear reward function, enhancing model interpretability and precision. An empirical application validates this enhancement, showcasing improved data labeling and similarity metrics. This second advanced modeling approach significantly boosts the potential of topic models, as well as improves information management, anomaly detection, and resource allocation, which are critical for adaptive decision-making in evolving crisis event conditions.

1 INTRODUCTION

Growing levels of social media engagement among urban communities provide organizations such as the Georgia Department of Transportation (GDOT) with growing opportunities for raising situational awareness in their operation and response decision making, particularly in the event of a crisis. Communities are increasingly utilizing social media for providing local information and sharing their personal experiences online. Those in need reciprocally expect rapid response from organizations to issues that arise and seek attention and/or help through social and community applications in the event of emergency in return. Crisis events, whether it be in the course of routine operations (e.g., road maintenance) or in the event of an emergency, (e.g. hurricane, tornado) can result in substantial social, environmental, and economic impact on the life of Georgia residents. In order to use social media for emergency operations, however, data collection is heavily dependent on citizen participation and location information. With social media platforms, such as Twitter/X, the amount of posts made by citizens can fluctuate depending on the crisis event.

1.1 Social Media Applicability to GDOT

Waze, is a crowd-sourced navigation community application that utilizes real-time data from its user community to provide dynamic and efficient route suggestions, incorporating features like gamification, voice navigation, and social interaction. Waze is currently being utilized in GDOT's existing Advanced Traffic Management System (ATMS). Waze uses crowdsourced data from drivers, including geolocation and timestamp information, to provide real-time and location-specific insights about road conditions with a high degree of accuracy (Amin-Naseri, 2018). Waze users submit data through the application, including hazard type and descriptors, and an ATMS is employed to reconcile and aggregate this information by clustering and similarity-matching user reports, making Waze a leading system in detecting traffic incidents, although its reliability diminishes during nighttime with lower road activity. For instance, Iowa's ATMS found that Waze initially recorded 13.4% of congestion and crashes in 2018. It was, however, found to be less reliable during the nighttime, when fewer people were on the road (Amin-Naseri, 2018).

While Twitter/X has been explored for identifying congestion and traffic, challenges persist in filtering Tweets for pertinent text and precise geolocation data, despite ongoing improvements in this aspect (D'Andrea, 2015, and Gu, 2016). Moreover, there is a growing scarcity of posts featuring precise, geotagged location data. Instead, many provide a general place or bounding box location of the post's origin. This underscores the necessity to enhance social media data containing community awareness information with a supplementary platform. This platform should complement and address the gaps in both data volume and location information currently present in social media. Only then can this augmented system be seamlessly integrated into existing Department of Transportation (DOT) crisis identification and response management processes. Community augmented processes and systems that can detect and track crises in near-real-time would be a critical component of rapid crises identification and response deployment decisions. Crucially, a more immediate identification and response management system can provide information to reduce potential casualties and damages and improve allocation of scarce resources.

1.2 Georgia-specific Crises

Global weather patterns are becoming more erratic and challenging to predict, with Georgia experiencing the impacts of increasingly intense and frequent extreme weather events (Noy, 2016). Despite Georgia's historical susceptibility to hurricanes and tropical storms, the city of Atlanta, located hundreds of miles inland, found itself under its first Tropical Storm warning in 2017 due to Hurricane Irma. Hurricane Irma had significant consequences, resulting in three fatalities and prompting inland evacuation orders for 540,000 coastal residents. This resulted in a range of emergent crises, including stranded residents in rising storm surge, shortages of essential resources such as gas, water, and food, power outages, infrastructure failures, fires, traffic jams, traffic incidents, evacuation barriers, looting, and other situations necessitating rapid response and emergency assistance.

In such time-critical circumstances, where traditional communication channels such as 911 may prove ineffective (e.g., people unable to make phone calls, emergency telephone hotlines jammed, or emergency responders unable to assess the relative gravity of different crises), residents increasingly turn to social media for assistance, sometimes even resorting to posting their full addresses in desperation.

Additionally, in terms of meteorological events, Atlanta is challenged by having a large population and infrequent snow and ice. Previous winter storm events have shut the city down, generated large amounts of panic, and trapped citizens on icy roads for hours at a time. As most Atlanta citizens have relatively little experience driving on icy roads and understanding the dangers associated with them, it can be critical for GDOT to address icy roads before a driver can encounter the hazard.

Ultimately, social media is currently a relatively untapped resource with respect to its ability to identify on-road emergencies, provide additional information such as severity and human impact for existing incidents, and its generalizability for many different kinds of events. In detecting and responding to emergencies, any information that can increase reaction time or level of preparedness is valuable, and social media and other forms of community application have been repeatedly shown to be able to provide it. For Georgia, and for Atlanta in particular, there is a need for improving the identification of specific risks during ice/snow events and for flooding events. These are both events with widespread impact with geographic pockets of extreme severity and risk, and thus two of the most suitable applications for the widespread network of "human sensors" that can be tapped through social media. Currently, non-traditional forms of information and communication that afford social and community awareness are lacking from the GDOT operation and response decision making structure.

As emerging digital forms of citizen communication with response organizations and between citizens is becoming more widespread, augmenting social media for community awareness capabilities with existing GDOT crisis identification and response management systems is critical. The focus of this research project, thus, is to address the shortcomings that exist in our ability to more rapidly, and effectively, communicate and respond to crisis events through investigating the need for, the current state of practice of, and the design and development of a community augmented approach to identify and respond to events occurring on the Georgia highway system. Through comprehensive communications and collaboration with the GDOT Technical Implementation Team over the course of the preliminary GDOT Research Project (RP 18-34) Social Media-Informed Urban Crisis Detection by the Principal Investigators (PIs), potential utility of social media data was identified as an untapped resource to identify, track, evaluate and visualize geographically constrained crises within two large emergency event types (Ice in Winter and Flooding in Spring) in the state of Georgia.

1.3 Social Media-Informed Urban Crisis Detection (RP 18-34)

To address the critical need for enhanced crisis identification and response management, a preliminary research project (RP 18-34) was conducted by the research team. The primary focus of this project was to explore the potential utility of social media data as an information source for GDOT during extreme events. Through a systematic approach, the research team developed and executed two case studies—one centered on a winter storm event and the other on an episode of severe flooding. Within the framework of each case study, the primary objectives were to identify pertinent information within social media data, assess its relevance using Natural Language Processing (NLP) techniques, and assess the potential for a tracking and visualization method for GDOT. The chosen case studies included the winter storm that impacted north Georgia on January 16^{th} and 17^{th}, 2018, and the flooding across the state of Georgia resulting from the impact of Tropical Storm Irma from September 10^{th} to the 17^{th}, 2017.

Drawing insights from these case studies, our research team developed a method to filter geolocated Tweets from the extensive volume generated daily in Georgia. Collaborative efforts between the GDOT Technical Implementation Team and our researchers led to the identification of concerns and the establishment of effective filtration methods. Recognizing the limitations of keyword filtration alone, we incorporated sets of stop words and semantic analyses, emphasizing the importance of keyword pairs specific to GDOT concerns. In addition, our evaluation of the Twitter/X data and feedback from GDOT personnel highlighted the potential value in geolocated images associated with Tweets, offering specific information on the magnitude and exact location of events. As a result, we concluded that incorporating Twitter/X images into alerts could potentially enhance the value of the system.

Following the assessment of topic detection, sentiment analysis emerged as a valuable method to rank the value and criticality of posted data. We observed the utility of sentiment analysis, particularly within major metropolitan areas, and its effectiveness in identifying relevant Tweets. In the design phase, recognizing the need for a supplementary system for GDOT, our research team concluded that Twitter/X data could best complement incoming community application data used by GDOT. The primary focus was on reducing the time for community data to generate alerts viewed by GDOT employees. We proposed the inclusion of the Twitter/X data stream with the existing Waze data stream into the ATMS, maximizing added value while minimizing additional training and software requirements. Conversations with the GDOT Social Media Coordinators further informed our conclusions, emphasizing the importance of GDOT personnel reviewing social media data alerts before inputting them as incidents. Building on these insights, we developed a framework for converting Twitter/X data into the Waze data format. In conclusion, based on the outcomes and components of our preliminary research project, and in consultation with both the GDOT social media team and the technical team, we generated recommendations for further research on a social media augmented framework and system design.

1.4 Summary

In this Community Augmented Rapid-response to Events (CARE) Integrated Crisis Communication System project, we performed a more comprehensive exploration by first conducting interviews to gain insights into the utilization of social media by various Department of Transportation (DOT) entities across the nation. Social media emerged as the primary avenue for information dissemination among these DOTs, yet their involvement was predominantly confined to established platforms like Waze. This limited engagement stemmed from concerns surrounding the reliability of information circulating on alternative social media channels. The insights gained from these interviews reinforced the project's imperative, revealing an unmet demand for a system that not only provides trustworthy information through social media but also reliably filters out the inherent noise associated with social media data.

To systematically address this need, we developed a competency matrix outlining the essential functions required for the envisioned system. As part of our exploration, we evaluated various commercially available software packages, even those not currently in use by DOTs. This comparative analysis aimed to discern the features and capabilities offered by different solutions, enriching our understanding and contributing to the formulation of our proposed system design. A significant aspect of our work involved exploring the design of proprietary algorithms and the design and testing of a community augmented rapid-response to events crisis communication system for the Georgia Department of Transportation (GDOT). Our focus centered on the seamless integration of Twitter/X and Waze data sources, aligning with the project's core objective of harnessing the wealth of information generated by citizens across digital platforms to augment rapid response to events.

In the concluding section of this report, we present our recommendations derived from the culmination of our extensive research efforts. Furthermore, we outline suggested next steps, building upon the foundations laid in this project. These collective efforts aim to establish a framework for the seamless incorporation and automated assessment of the growing data generated by citizens through various digital platforms, notably social media. Our vision is to enhance crisis communication capabilities and empower timely, informed responses to events, leveraging the vast potential embedded in the digital landscape.

2 SOCIAL MEDIA STRATEGIES AND INSIGHTS

2.1 Literature Review

Regular natural and man-made crisis events such as hurricanes, tornadoes, severe storms, floods, infrastructure failures (e.g., the 2017 Interstate 85 bridge collapse in Atlanta), and major social events (e.g., sporting events), can generate substantial social, environmental, and economic damage to communities. These events range from more routine crises such as a faulty traffic signal, to storm-related emergencies such as trees down, landslides, potholes, power line failures, and road erosion. Having timely crisis identification and response plans in place to address these requires situational awareness of an event, access to social data-rich information, effective communication, and engagement with local communities.

Increasing accessibility to mobile services has enabled many social communications to move to social media platforms. This formation of virtual communities has become a critical source of information and medium for communication for both citizens and organizations in the event of a crisis. Social media (e.g., Twitter/X) applications are currently being used for improving emergency situational awareness (Yin et al., 2015) for a range of crisis events such as flooding (de Albuquerque et al., 2015), winter storms (Wang et al., 2017), hurricanes (Wang and Taylor, 2014), earthquakes (Sakaki et al., 2010), and power outages (Jennex, 2012). Many organizations such as the American Red Cross and the United Nations Office for the Coordination of Humanitarian Affairs (Imran et al., 2014) have adopted these practices. However, there is often a tradeoff between information timeliness and information accuracy or relevance.

The scope, generalizability, and direct relevance of social media analysis to state and federal personnel beyond catastrophes is still underdeveloped. The research community has focused on determining automated methods of removing the extraneous, irrelevant information and condensing the relevant information into formats that decision-makers in response organizations can effectively use. This includes a range of event detection techniques founded on clustering-based approaches that use co-occurrences of keywords for semantic examinations (Schubert et al., 2014 and Zhang et al., 2016), probabilistic topic models/LDA-based (Chaney and Blei, 2012) approaches, and Natural Language Processing (NLP)-based methods (Liu, 2011) in detecting and characterizing subjective information such as emotions, opinions, and sentiment intensity in textual data.

In the event of crisis, however, it is critical to identify events with respect to associated spatial and temporal patterns, relevance and proximity to major infrastructure, geographic dimension, and the intensity of negative sentiments. Transportation Research International Documentation (TRID) studies related to this research, including the records from Transportation Research Board (TRB) and Transportation Research Information Services (TRIS), include several TRB projects. For example, "Social Media Guidebook for Emergency Management" (ACRP 04-23, RiP 01642763) (Barich 2019) developed a guidebook to help airports leverage social media for emergency management and crisis communication. "Utilize Crowd-Sourced Data and Machine Learning Technology to Enhance Planning for Transportation Resilience to Flooding" (RP 01674188) (Zhang and Pan 2019) extended from this to develop a decision support system (DSS) that combines non-traditional, crowdsourced big-data with traditional data to enhance transportation readiness for quick response decisions in urban flooding. And "Emergency Management Agencies: Pilot for a Crisis Communication Analysis Assessment Test" (RiP 01460030) (Guth 2019) examined issues such as attitudes toward the use of social media, public information, and Emergency Management Agency (EMA) websites in emergency management.

TRB projects have also been completed in the area of crisis response and social media. "Social

Media Practices in Traffic Safety" (DOT HS 812 673) (Sack et al., 2019) explored how State Highway Offices (SHOs) may use social media to promote safety through qualitative and quantitative scans of social media platforms as well as interviews with nine states." "Improving Emergency Preparedness and Crisis Management Capabilities in Transportation" (RP 01467319) (Howitt 2009) explored whether and how one significant functional area—surface transportation—developed the capabilities to effectively fulfill US commitments for developing a comprehensive, integrated emergency management system. Expanding on the use of social media, "Modeling Disaster Operations from an Interdisciplinary Perspective in the New York-New Jersey Area" (RP 01566476) (Ozbay et al., 2016) used social media in addition to survey data to understand overall demand, destination type choice, and route choice decisions in the aftermath of Hurricane Irene. Finally, "Big Data During Crisis: Lessons from Hurricane Irene" (RP 01556674) (Sheffi and Goentzel, 2015) characterized the potential of big data from social networks and NLP methods in creating actionable information in a crisis in the event of Hurricane Irene.

These projects have established the initial steps toward integrating social media information and communication data into various dimensions of DOTs' crisis and emergency management strategies; however, low integration of social and community data with current DOT crisis identification and response communication processes limits the efficiency and inclusiveness of responses within and across operations and emergency management teams. Crisis identification and response information that can increase reaction time or level of preparedness is valuable and social media and community-driven applications data streams have been repeatedly shown to be able to provide this capability. Nonetheless, the need for integrating such data into current systems for near-real-time crisis communication and response, and, by extension, the lack of a validated approach for social and community-driven data fusion in transportation information systems, thus, remains unfulfilled.

For the design of our own algorithmic model solution, we explored the integration of the social media platform, Twitter/X, with the community-driven platform, Waze (i.e., community-driven application, as we define it, is a platform that seeks input from users for a particular situation or circumstance.). Twitter/X posts (i.e., tweets) can include information such as images and text descriptions, replies, retweets, favorites, and geographic information about where the user posted in the form of place reference or, occasionally, exact location information. While this can be made useful in crisis situations, an observed challenge with this is being able to extract relevant information to assess and gain actionable insights. Other typical challenges when dealing with social media pertain to trust, privacy, volume of data, availability of geotagged posts, and “rumors” or fake news that spread when people misuse social media (Rossi et al., 2018). As mentioned previously, many studies have analyzed how social media can be used to better protect people, property, and the environment in various phases of emergency management. Community-driven applications have been used in studies involving the Waze navigation application to assess its validity and coverage (Amin-Naseri et al., 2018), examining real-time traffic flow data from Waze in comparison to Twitter/X data congestion (Sidauruk and Ikmah, 2018), and performing incident detection that models uncertainty of spatiotemporal aspects in crowdsourced reports (Senarath et al., 2020). Waze has been shown to be superior at times to Twitter/X for event detection (Amin-Naseri et al., 2018). Community-driven applications such as Waze can address select shortcomings that most social media platforms currently possess, regarding lacking interactive features where users can send reports and update other community members specifically on certain harmful events through pre-established categories. Waze also has more precise location data and interactive geographical visualizations. Social media, however, adds the community individual voice and sentiment of users that Waze lacks. It has also been revealed that despite platforms to aid disaster management using social media data, few are designed for citizen connectedness or use both social media and another platform (i.e., different type of input data such as community-driven applications) (Chair et al., 2019). In essence, Waze enhances Twitter/X with the higher volume of accurate coordinates related to events beyond the bounding box Twitter/X provides with its current API, while Twitter/X enhances Waze with adding more context to the categorized alert types (e.g., a Waze alert deemed "Accident" when paired with a tweet in the same area could potentially show how many cars are involved, if someone was seriously injured or needs help, and possibly images related to the event). Thus, the two data streams complement one another.

Social media data will continue to be used in various fields for analysis and detection within communities pertaining to emergency events. This suggests the necessity to improve social media data sets through integration with applications that are more equipped for real-time event detection (e.g., Waze) because, as mentioned previously, social media lacks certain features community-driven applications possess. The related research described above presents a gap with social media analysis methods related to emergency events that investigate their integration with location-based applications that capture incidents relating to emergency preparation and response. Furthermore, there is a gap regarding methods that determine an effective augmentation of location-specific social media data with community-driven data to address the shortcomings that exist in the ability to more rapidly, and effectively, communicate and respond to crisis events. Intervening and alleviating disasters as they occur in real-time poses an issue for many emergency responders. Therefore, fusing data from social media platforms, such as Twitter/X, and community-driven applications, such as Waze, is critical for heightened emergency management capabilities.

2.2 US DOT Organizational Insights: In-depth Interviews and Perspectives

To thoroughly examine the integration of social media within existing DOT systems and gather insights from operators, we conducted interviews with six DOTs (District of Columbia DOT, Maryland DOT, Massachusetts DOT, North Carolina DOT, Pennsylvania DOT, and South Carolina DOT), along with consultations with two affiliated organizations (the University Maryland CATT Lab and the University of Florida McTrans Center). In our initial internal research team assessment, we identified potential applications, including Data Capable, RITIS, Swarco, and Yunex Traffic as candidate solutions for viably augmenting and integrating social media with current GDOT systems. We then discovered RITIS is the most widely adopted, and composed interview questions. Upon completing our interviews, we directly compared the six DOTs and the findings from the two related transportation organizations. Our interviews provided insights into how several state DOTs are using transportation software and analysis tools for data integration. We found the DOTs employ the platform RITIS to enhance their real-time data analysis and traffic management capabilities. While most DOTs express contentment with their current tool, some are in the process of evaluating alternatives. Key considerations in these evaluations encompass costs, user preferences, data integration capabilities, and privacy-related factors. The University of Maryland's CATT Lab was instrumental in the development and support of RITIS, and multiple DOTs collaborate to enhance traffic flow and quickly identify crisis events. We formulated the interview protocol (provided in Appendix A), making adjustments as necessary during the interview based on the real-time responses we received. The results of our interview discussions are summarized in Table 1.

Table 1. Summary of DOT interviews.

Primary Functions Used and Benefits	Year Adopted	Rationale for Implementation	Waze Community Data Usage	Twitter/X Social Media Usage*
-Communication with other DOTs -Verify Waze alerts with RITIS -Understand events without physically sending operators	2017/ 2018	-Collaboration with other state DOTs about traffic incidents -Verification of Waze alerts with RITIS maps -Improve sense of awareness	-To understand live incidents without physical investigation	-None
-Timeline feature: visualization of lane closures, speed ratings, impact of traffic, time stamps, etc. -ITS Planner: can load data onto RITIS for evacuation plans, emergency management tools, etc. - Accident reports - Data archives	2008	-Partnership with University of Maryland CATT Lab to develop: -Real-time detection and comparative speed analysis -Connection between multiple states DOTs	-To validate RITIS alerts	-None
-Collection of road data without field visits -Congestion and traffic data -Safteygroup: speed tracking -Post-incident reporting -Gives information across entire state -Provides clear information to share with senior leadership	2018	-Collection of road data without field visits	-Interested in the use of Waze to communicate DOT alerts	-None
-PDA- Suite: congestion scans, bottleneck ranking, causes of congestion graph, trend map: movie of traffic, massive data downloader, etc. -Automatic calculators feature with Mandatory Federal reporting	2008	-Included within a package purchased for probe-speed data	-To analyze congestion scans	-None
-Consolidates ATMS information -PDA-Suite analysis for planning -Origin and destination studies	2015	-Real-time data analysis -Visualization of routes	-For the basics found within the connected citizens program	-None
-Real time data analysis -Historic data archives -HERE incident data	2019	-Real-time data	-For speed maps and real time data	-None

**Twitter/X Social Media Use was narrowly discussed as ingesting social media postings (as is common practice with Waze alerts). DOTs did describe using social media to communicate uni-directionally to the traveling public. Overall, the DOTs described concerns regarding integrating Twitter/X social media into their current systems anticipating issues of overburdening operators with false alerts and the potential distraction of large volumes of social media data from which it would be difficult to identify relevant information.*

The DOTs emphasize the ongoing necessity of ATMS systems in conjunction with RITIS to efficiently catalog detailed reports. They recommend a dual approach, advocating for the continued use of both ATMS and RITIS while advising a cautious approach to directly integrating social media within DOT systems. Their concern lies in the need to determine the root problem that a DOT aims to address in adding social media data, and they caution against the risk of data overload and the creation of additional issues stemming from social media feedback.

Furthermore, they highlight the value of thoroughly exploring all the capabilities of RITIS when making software purchasing decisions. Their advice underscores the importance of evaluating existing DOT systems to identify gaps and effectively fill them with RITIS or the desired software to be employed. Additionally, one DOT makes a noteworthy observation regarding the user-friendliness of another system, noting its role in streamlining the flow of information up the management chain. Overall, these insights shed light on the nuances and considerations that shape the decisions made by DOTs when it comes to adopting and integrating transportation software and analysis tools.

We also met with two university research centers developing related software for DOTs, including GDOT. Here we summarize the discussion with CATT Lab, a self-funded nonprofit organization from the University Maryland. We include results from that discussion primarily as RITIS was identified as being used by most of the DOTs we interviewed and the CATT Lab originally designed RITIS 20 years ago to improve the communication of real-time situational awareness between Maryland, Virginia, and Washington D.C. RITIS ingests various data, such as crash and work zone data, from states to analyze and improve traffic flows. Since its creation, it has expanded and offers real time analysis for approximately 13 state DOTs, including the Georgia Department of Transportation. RITIS requires a single point integration to implement the platform in a new

DOT. Employees of DOTs can learn from RITIS’ training sessions and video archives to familiarize themselves with the platform. Once implemented, RITIS can fulfil various actions including creating archives—which ATMS systems are not designed to do—overlaying information to the state, as well as the capability to write reflections on incidents after their occurrence. The team within the CATT Lab frequently updates RITIS, including a daily technical refresh to fix bugs and make functional improvements that are heavily guided by users through feedback committees and RITIS-hosted Eastern Transportation Coalition meetings that take place four times a year. RITIS integrates crowdsourced data from Waze, but not Twitter; the CATT Lab described being not interested in integrating Twitter/X because of its lack of reliability and difficulty of identifying relevant information among a large volume of communications, similar to sentiments shared in interviews with state DOTs. Although RITIS ingests every input reported in Waze and can filter real-time content by geography and reliability, 511 and 911 reporting are the preferred choice of data. An interviewed representative of the CATT Lab noted that the ATMS system is the primary system used in urban settings because of its precision and shorter delay compared to crowdsourced data; whereas crowdsourced data is more beneficial to fill in the gaps in rural settings where it is less likely for a government official to observe and report an incident.

3 COMPETENCY MATRIX DESIGN

3.1 Background

Having completed the DOT interviews and learned a great deal about how DOTs evaluate and implement systems, we developed a competency matrix to assess the diverse capabilities among potential software solutions and to inform the design and development of a social media augmented crisis communications system with features that extend beyond the current state of the art. Developing the competency matrix for crisis informatics and emergency response requires an understanding of the methodology and theories involved and an assessment of the information technology applications used by various stakeholders of an organization. It is crucial to consider the most advanced technological solutions and their feasibility and potential for success. The Institute for Defense Analysis developed a study where seven principal categories were fundamental for a prompt software acquisition process and sustainment.

Although software development and deployment time are important factors in the success of IT (Information Technology) projects in the public sector and there are several factors (Table 2) to be considered to accelerate the process it is important to note the early warning signs of IT project failures (Garrison, Tate, & Bailey, 2019). Early warning signs would be defined as risk indications of a project's future problems and potential failure. Kappelman et al. (2006), conducted a study seeking to determine from the participation of 138 experienced IT managers the early warning signs of IT projects. The results indicate two principal causality groups: People related early warning signs and Process related early warning signs (Table 3). Considering and evaluating these dominant warning signs during the initial 20 percent of an IT project would permit the identification of indices that would contribute to long-term project failure. Nonetheless, it is vital to also understand the importance and role of public procurement for IT solutions.

Table 2. Speed limiting factors in software acquisition (Garrison, Tate, & Bailey, 2019)

Factors	Description	Considerations
Required Functionality	Defined scope of the software program.	• Assess the negative requirements of a system. • Achieve Minimum Viable Product (MVP) to gather feedback and refine issues. • Review the modularity of systems for parallelization of development efforts.
Architecture	Program organization and operating environment.	• Develop system to support agile improvements. • Consider the lifespan of the system and the need for upgrades in the future. • Contemplate the interoperability needs and capabilities. • Determine the degree of inclusion of current avant-garde technologies. • Evaluate the need for the rapid implementation of new capabilities or a rapidly upgradeable system.
Technology Maturity	The maturity level of innovative solutions implementation.	• Assess the level of maturity of software and hardware design processes. • The Department of Defense evaluates AI as immature technology given that there are issues such as the validation and maintenance of datasets or the achievement of testable requirements.
Resources	Ecosystem required for a successful implementation.	• Determine the experience of the implementation team. • Evaluate access to datasets and specialized IT infrastructure for system development and deployment. • Consider the level of funding stability for a program.
Testing Strategy	Defect detection and interactions throughout the development.	• Debugging, finding, and fixing defects within the system in a periodical and rigorous manner. • Implement the testing strategy as early as possible in the process. • Ensure correct testing methods are completed and results are adequately conveyed to the development process.
Contract Structure	Alignment of contractor and outsourcing organization.	• Consider the need for maintenance and future upgrades of software-intensive systems. • Existing law forbids making a condition of contract award from IP rights.
Change Management	Transformation of processes for adaptation.	• Align stakeholders controlling system requirements and developing/fielding the systems. • Establish the definition of system need, and place requirement thresholds.

Table 3. Warning Signs of IT Project Failure (Kappelman, McKeeman, & Zhang, 2006).

People Related Early Warning Signs	Process Related Early Warning Signs
• Absence of top management support. • Inefficient project manager. • Lack of stakeholder participation and involvement. • Fragile project team commitment. • Inexperience or knowledge gap from the team. • Experts are unavailable from extensive responsibilities and workload.	• Absence of scope requirements and success criteria. • Lack of change control process. • Weak management and scheduling. • Communication breakdown with stakeholders. • Limitation of resources and resources reassignment. • Shortfall of project business case.

Procurement of IT solutions. IT procurement is the acquisition of hardware systems, software programs, and upgrades along with other services through a series of steps that include a proposal, bidding, contract awarding, and contract management. The extensive process needs to be structured and organized to ensure program success. Therefore, several procurement maturity models have been developed to support the efficacy and capability of managing procurement challenges. Hua (Hua, 2022) proposed a procurement maturity model expanding from current models in the theory and subsequently applied four different maturity stages while taking into account political and managerial objectives. Hua determined that for organizations looking to improve procurement maturity, firstly organizations have to conduct a strategic assessment of how the procurement process integrates with the organization's leadership. Secondly, organizations should achieve increased awareness from the procurement team members on software development methods. Finally, the procurement team needs to comprehend the vendor and handle business negotiations during the process (Hua, 2022).

Academic and industry reports show that around 46 percent of software projects are challenged which signifies that although they are operational, they have problems with their budget, schedule, or even capabilities requirements (Johnson, 2018). Nguyen et al, (Nguyen, et al., 2022) conducted a study to determine the criteria for software acquisition, understand the timeline during procurement, and the variations in the requirements of Request for Proposal (RFP) from multiple software categories. Although the evaluation weights assigned for software RFP vary from software category and the type of organization, a common evaluation criterion was determined by Nguyen et al. (Table 4).

Table 4. Evaluation criteria weighs for different software categories (Nguyen et al., 2022).

Evaluation criteria	Average Weight (%)				
	Enterprise Resource Planning	Financial Systems	Asset Management Systems	Common Business Application	Specialized Business Application
Cost proposal	20	23	20	16	24
Response to RFP requirements	2	9	3	4	4
Implementation approach	22	23	28	29	26
Company qualifications	15	21	20	21	19
Project team qualifications	3	6	5	6	3
System capability	27	17	21	22	19
Software demonstration	9	0	2	0	2
Other criteria	2	1	1	3	2

For incident management systems the most relevant category would be the Asset column in Table 4. On average a larger weight percentage is given for the implementation approach (28%) which consists of assessing the system deployment methods proposed by the vendor and the customization capabilities regarding different scenarios. Secondly, with 21 percent, the system capability is evaluated, and the procurement team rates the capacity of the vendor to meet the functional, technical, and security requirements of the project while also understanding the potential integration of future modules within the system.

Thirdly, with 20 percent, is the cost proposal which refers to the overall financial burden including the installation, training, licensing, and maintenance of the software system. Equally important during RFP evaluation, with 20 percent, company qualifications reflect the experience and expertise of a vendor to handle the project. Lower importance is given to the project team qualifications, the response to the requirements of the Request for Proposal, and other criteria such as the business structure of the vendor or the past experiences between the two sides. Other significant findings from Nguyen et al., show that on average the implementation duration noted on the RFP is 265 days which corresponds to almost 70 percent of the time allocated in a project on the RFP (Nguyen et al., 2022).

Technology Choice in Procurement. Although technological choice and procurement is a formal process, the acquisition is not simply the result of a rational decision but the tension between sociocultural factors. A study by Pollock and Williams (Pollock & Williams, 2007) explores the sociology behind the procurement of software technologies by observing a joint venture between a city council and an IT company along with management and computer science experts. During the yearlong selection, the research team observed the procurement process and conducted interviews which allowed them to determine the following findings. The decision environment was described as having high uncertainty levels where the features of the vendors were negotiable within the procurement team, making the comparative measures flexible. The authors claim that the property of each system is not relevant in the comparison of similarities and differences but instead it is the validation of evaluating criteria that provided meaning to these properties. The scaffolding metaphor and the disentangling, framing, and overflowing framework converge towards a decision not only dominated by a formal process but as a sociological attempt to reach a common decision.

3.2 Methodology

With the consideration of the success and failure factors in IT solutions, the current practices in public software procurement for IT solutions and the implications of a formal process and a cultural sociological approach driving a decision, we designed a competency matrix to evaluate the range of capabilities and alignment of potential IT solutions for crisis informatics leveraging social media (Table 5). The vertical axis of the matrix addresses 9 dominant factors that were determined to be important from the technical and procurement perspectives (Garrison, Tate, & Bailey, 2019; Nguyen, et al., 2022), "Social Media-Informed Urban Crisis Detection" (No. FHWA-GA-20-1834) (Samuels, Mohammadi, & Taylor 2020), (Abel, Claudia, Houben, Tao, & Stronkman, 2012; Endsley, Bolte, & Jones, 2003; Jodoin & Austrich, 2020; Jin, Pang, & & Cameron, 2007; Kappelman, McKeeman, & Zhang, 2006; Thales, 2022). The horizontal axis of the matrix addresses the level of proficiency around each item where the scale includes 5, 4, 3, 2, and 1. Each cell in the matrix has been given a baseline that was determined from the current trends in technology, transportation authorities' requirements need, and considerations during the acquisition of a solution. Each solution is then evaluated with the designed competency matrix following the baseline of each factor.

Factors. The Architecture of a software system addresses the fundamental structure and behaviors of a system. Software Architecture is the basis for qualities such as modifiability and security. In the procurement process, it is essential to verify the feasibility and applicability of a potential solution. Garrison et al. (2019) explain that an efficient Architecture will determine the cost and maintenance effort of a system in the future. According to Nguyen et al. (2022), the weight given during RFP to the implementation approach is 28%, and system capability is 21% which indicates system architecture is given significant importance by the procurement team. In the context of crisis informatics for transportation authorities: "*5*" system Architecture would show high modularity to integrate with the existing system while being designed for long-term operations within the organization. *"4"* Architectures includes the integration of modular components within existing transportation authority's systems but lack the extended design mindset that factors system longevity from an absence of agile and lean methods in their improvement requirements. *"3"* Architectures include modular components towards a solution but are not designed towards integration with the existing system. These can be systems that parallelly support existing systems but have the potential to be integrated in the future from their modular nature. *"2"* Architectures are systems without modular components and interoperability concerns that could be integrated from their technical design despite not being modular. *"1"* Architectures do not consider modularity or address interoperability in their system design.

The Automation of a system covers the level of independent applications that reduce the need for human input to operate. Previous research showed the importance of automatic event detection for the reduction of response time and enhanced emergency response for road-related incidents "Social Media-Informed Urban Crisis Detection" (No. FHWA-GA-20-1834) (Samuels, Mohammadi, & Taylor 2020). In the context of crisis informatics for transportation authorities, it addresses the need for automatic event detection of road-related events and emergencies. *"5"* Automation is a reliable completely autonomous system to detect events without human intervention in the determination process. *"4"* Automation is achieved from autonomous systems that detect automatically detect events but require human intervention for validation. *"3"* Automation in event detection is achieved from advanced automation in event detection that corresponds to the integration of automated reporting methods not limited to social media such as phone calls. *"2"* Automation is achieved from the automatic integration of externally determined events with the transportation authority operating system. *"1"* Automation is attributed to solutions that do not consider automation in event detection.

Event Detection of a system encompasses the range of events a system can detect and the range of information received and transmitted to be considered during an emergency response. In the context of Transportation Authorities, the main requirement is to model the road network yet additional contextualization of events can provide additional information to support emergency response (Abel et al., 2012) from the use of metadata. "5*"* Event Detection identifies various types of emergencies not limited to the state road network but aggregated among a variety of public networks, for example, utilities. *"4"* score event detection is reached from the identification of a different type of event affecting the road network. *"3"* Event Detection is reached from the identification of various types of road-related events across the road network but not limited to vehicle incidence, for instance, flooding or ice on the road. *"2"* Event Detection is reached from the detection of vehicle-related events including predetermined road conditions such as planned road closures from construction. Finally, *"1"* Event Detection is given to single event detection across the road network such as vehicle crashes or traffic.

Situational awareness of a system is defined by Endsley, Bolte, and Jones (2003) as being sensitive and informed about the events happening around and understanding the meaning of the information now and in the future. Three levels are given for the obtention situational awareness: perception of elements in the environment, comprehension of the current situation, and projection of future events. In the context of Transportation Authorities situational awareness is achieved when more information is shared and complemented with the operator's training. *"5"* is given for systems that deploy FHWA views for NextGen TIM (Jodoin & Austrich, 2020). *"4"* is achieved by leveraging live video communication with individuals on-site to guide and support the emergency response. *"3"* is reached by the augmentation of situational awareness through live visual media from existing static or dynamic cameras. *"2"* is achieved from the inclusion of social media-filtered data streams to obtain metadata such as videos and cameras. The strength of social media relies on the derivation of emotions that can shape the strategic response of an organization (Jin, Pang, & Cameron, 2007). *"1"* is given from situational awareness only provided by current 511 phone reports.

Detection Speed covers the time between the time an event occurs to the time it is informed to the emergency operator. Getting the information as fast as possible is important to rapidly create situational awareness and support emergency decision-making response (Kwan & Lee, 2005). In the context of Transportation Authorities, the detection of events is critical on the road network across large networks. "5*"* is achieved by ideal systems that achieve detection in real-time or between 0 and 1 minute of the event happening. *"4"* is achieved by a system that has perfected near-real-time reporting of events or in a timeframe between 1 to 3 minutes. As the time for event detection increases the score achieved is lower such that a *"3"* is scored between 3 to 10 minutes, a "2*"* is scored between 10 to 20 minutes and a *"1"* is scored for more than 20 minutes. The detection speed is fundamental in crisis informatics and specifically social media has demonstrated to be a tool capable of leveraging near real-time detection of events.

System Maintenance is a requirement that considers the contract method by which a vendor or solution provider will design the system in the long term and its commitment to ensuring the appropriate level of operation of the system requirements. Given the speed at which technology is updating and how new challenges arise software usually undergoes corrective, preventive, perfective, and adaptive maintenance (Thales, 2022) but given constraints such as intellectual property (Garrison, Tate, & Bailey, 2019) affect the proposed contract regarding maintenance. Therefore, in the context of transportation authorities, it consists of the maintenance of ATIS systems. "5*"* score is received for a system that is operated by the user, in this case, the respective Transportation Authority. A self-dependent DOT is able to maintain and update its modules, saving costs and customizing the solution to their needs. *"4"* is achieved from regular maintenance of the system by a contractor, this includes current updates and system tests and checks that would detect any malfunction or failure. A score of *"3"* is achieved for maintenance as needed, the downsides of this aspect include the potential crashes and system integration problems which would compromise the effectiveness of the system, *"2"* is reached if maintenance is not provided as part of the solution provided. And a "1*"* is given for planned obsolescence where the system cannot be maintained and eventually becomes obsolete and needs replacement.

Deployment speed is one of the main concerns studied in the procurement of software solutions. Nguyen et al. (2022), determined the expected duration of the project durations in the RFP. On average the total time of implementation was 382 days while some projects were faster at 43 days others took up to 1384 days. In the context of Transportation Authorities, the faster they can deploy technology and put into effect crisis informatics systems the more impact they will have on the public. According to the SCCT crisis communication theory developed by Coombs (Coombs, 2007), the protection of assets is induced by the reputational threat to an organization posed by an event, motivating transportation authorities to be prepared for crisis scenarios. A "5*"* is achieved by an ideal system that has the capabilities of immediate implementation. A "4" is given for a faster-than-average deployment of 6 months while a *"3"* score is achieved for an average deployment rate. A *"2"* is reached for a slightly slower-than-average deployment while a "1*"* is a system that can be deployed in more than 2 years.

Contractual Needs are the basis of public procurement and different setups with the potential solutions providers will have repercussions in the short and long term. Procurement contracts scope the product selection the payment conditions among others. In the context of a transportation authority, a ranking is based on what would be more beneficial. A "5*"* score is achieved from an internally developed solution; this would avoid interaction with vendors and avoid risk from outside vendors. *"4"* is achieved by awarding a renewable contract, in the software development industry time-and-materials contracts are usually awarded from the uncertainty in the solutions. *"3"* is reached with a contract that includes the development and maintenance of the system on a fixed price contract that defines roles and responsibilities yet is more complicated to renew the contract on similar terms after the solution development. *"2"* is achieved for contracts that only cover the development scope of the project and do not include future interaction with the deployed system. On the Transportation Authority, a *"1"* is given for long-term concessions of a system and the complete dependence on a company to operate their crisis informatics solution.

Experience is determined by the impact on the field and the setup of an adequate team to design a solution. The relevance of experience has been linked by Kappelman et al. (2006), as a people-related warning sign to IT project failure. Along with this, experience factors the team commitment, knowledge, and quality of managers involved. In the context of Transportation Authorities, a team must be chosen to effectively implement crisis detection systems. For experience, scores are determined based on the National Institute of Health proficiency scale (US National Institute of Health, 2022) that examines teams based on involvement with past projects, impact in the sector, and market position as: expert, proficient, competent, average, and novice.

Competency Matrix Application. The competency matrix (Table 5) can be applied by transportation authorities to evaluate the factors determining the competency to implement crisis informatics solutions. The basics to determine a score are derived from knowledge on the corporate standing of vendors and technology maturity and refinement. Pollock and Williams (2007) determined that although procurement is a formal process it is also a decision process that engages the procurement team through a malleable process or negotiation. From this, the matrix has been designed such that the procurement team for crisis informatics solutions can be tasked with attributing a weight to the factors. The composition of the procurement team has to be diverse and group representatives across the divisions of a transportation authority. Pollock and Williams (2007) in their case study encountered the procurement team to be integrated with the primary end users of the system, IT personnel, project managers, a chairperson, and other parts of the organization. Although not part of the procurement team, academic professionals were present during the meetings and throughout the process. Conceptually, different members of the procurement team could weigh the different factors of the competency matrix based on their experiences, needs, and acceptable tradeoffs. The next section applies the competency matrix design to evaluate a set of different crisis informatics solutions.

Table 5. Competency Matrix for Evaluating Crisis Informatics Solutions by Transportation Authorities.

Factors	5	4	3	2	1
Architecture	High modularity and interoperability design for long term operations	Modular components and interoperability design	Modular components without interoperability design	Non-modular system with interoperability concerns	Non-modular systems and interoperability issues
Automation	Autonomous System without human intervention	Human-in-the loop ML powered system	Advanced Automation in event detection	Novice Automation in event detection	Lack of Automation
Event Detection	Supports various types of event emergencies across multiple networks	Supports various types of event emergencies across the road network	Supports various types of road related events across the road network	Supports only road related events across the road network	Supports unique road related events across the road network
Situational Awareness	High integration of data supporting first response	Supports live conference streams with responders	Supports various types of data streams including live cameras from response vehicles	Supports various types of data streams including social media	Supports phone call-based data streams
Interactional Capability	Allows two-way communication between users and emergency responders or traffic operators. There is regular interaction between users and operators.	Regular updates from responders and traffic operators. Also, users can respond to that, and their response gets occasional replies.	One way communication. The emergency responders and traffic operators can upload updates regularly.	Extremely limited interactive space, where occasionally operators and responders upload updates.	No communication option available.
Detection Speed	Real time	1 minute to 3 minutes	4 minutes to 10 minutes	11 minutes to 20 minutes	More than 20 minutes
System Maintenance	Maintained and operated by user	Scheduled maintenance and update by contractor	Maintenance and update by contractor as needed	No Maintenance	Planned obsolescence
Deployment speed	Immediate	Within 6 months	Within 1 year	Between 1 and 2 years	More than 2 years
Contractual Needs	Internal Authority Development	Renewable Contract	Single Development and Maintenance contract	Development Contract	Long Term Concession
Experience	Expert team from past projects and market position	Proficient team from past projects and market position	Competent team from past projects	Average teams without past projects	Novice team

3.3 Assessment

A total of 12 software applications/platforms with social/community data features, including RITIS, were compiled to assess potential vendors based on the information available online and on their website and available information about their previous work with DOT or related organizations. Each vendor was scored using the competency matrix designed to assess their abilities in crisis informatics and emergency response. The scores were determined based on online information about the vendors' technological capabilities, expertise, and experience. After each vendor was scored, a weighted average was used to compare overall competency. This approach allowed for an objective assessment of the vendors' strengths and weaknesses and is designed to help GDOT as a competency matrix for future software vendor evaluation. The competency matrix is designed to allow GDOT to adjust the weight of each factor to determine the overall weighted score. Table 6 shows the scoring results with equal weighting.

Table 6. Competency Matrix Scoring with Equal Factor Weighting.

	All the parameters of the matrix are scored out of 5.										
Software Vendor	**Architecture**	**Automation**	**Event Detection**	**Situational Awareness**	**Interactional Capabilities**	**Detection Speed**	**System Maintenance**	**Deployment Speed**	**Contractual Needs**	**Experience**	***Weighted Average Score***
Yunex Traffic	5	4	5	5	5	5	4	4	3	4	4.4
Swarco	4	5	5	3	5	5	3	4	5	5	4.4
RITIS	5	4	5	3	4	5	5	4	3	5	4.3
Rekor/Waycare	4	4	5	3	5	5	5	4	3	4	4.2
INRIX HELP	5	4	3	2	5	5	4	5	3	5	4.1
Data Capable	3	4	5	4	5	4	4	4	3	3	3.9
SMATS Traffic	5	4	4	3	4	5	3	4	3	3	3.8
TomTom	4	4	5	2	5	5	3	4	3	3	3.8
Hayden AI- SafeSense App	4	4	2	3	3	5	4	5	5	3	3.8
StreetLight Data	5	4	4	1	2	5	4	4	4	5	3.8
Castle Rock Associates	5	2	3	3	3	2	4	4	4	5	3.5
Transoft Solutions	4	4	4	1	1	5	4	4	3	5	3.5

A vendor checklist was supplied by GDOT which included instructions of how to further evaluate these vendors. The interviews with the DOTs about the vendors they are using provided valuable insights into the performance and capabilities of different vendors in the transportation industry. This information aided in the evaluation of potential vendors for GDOT, including RITIS, and providing recommendations. Gathering information from a trusted source can make the evaluation process more objective and comprehensive, leading to a more successful vendor selection. The competency matrix can be an effective tool for GDOT to evaluate purchases from future software vendors, and even the list of vendors noted in the matrix (although GDOT is already using RITIS). In neither our exploration of software in use at DOTs in interviews with 6 different state DOTs nor in the range of features provided by the software vendors analyzed in the competency matrix did we identify a solution that is being implemented by DOTs that ingests Twitter/X social media data. A gap remains in the inclusion of this functionality to aid in identification of crisis events on the highway system. The lack of such a functionality may be due to the expressed concern by DOTs over identifying the relevant information among a large volume of data, which might overburden transportation management center operators and result in false alerts. However, recent developments in artificial intelligence creates the possibility to pair Twitter/X social media postings with Waze community posts in an automated manner that would improve upon the reliability of crisis events identified. The following chapters explore this potential and develop and test algorithms to enable this functionality.

4 ASSESSING COMMUNITY NEEDS IN DISASTERS

Enhancements to emergency management systems are imperative to improve response execution and better serve society. The recording-breaking 2004 and 2005 hurricane seasons (e.g., Hurricanes Ivan, Katrina, Rita, etc.) exposed shortcomings in emergency management, especially in federal response capabilities (Schmidtlein et al. 2008). When a natural disaster event is deemed so severe that it exceeds the ability of both state and local governments to respond, the Federal Emergency Management Agency (FEMA) issues it as a major disaster declaration, however, there is no set definition of what "beyond the combined capabilities of state and local governments to respond" means in order to receive assistance (FEMA 2023c). Thus, subjective judgments have the potential to shape the outcome of declarations and resource allocation. In the majority of cases, before requesting a disaster declaration to receive aid, state and local officials must conduct a damage assessment. With this, emergency management responders can face challenges in providing immediate intervention and relief for ongoing disasters as they await the assessment and declaration of a crisis event. Many areas are underserved by this process, resulting in inequities with distribution of aid (Schmidtlein et al. 2008). Before necessary federal assistance is given, state and local emergency management personnel need to make decisions on potential resources needed to mitigate the effects of disasters, especially when there is little time to decide or wait for a drawn-out damage assessment. There needs to be alternative systems in place that can adequately and quickly assess community needs when hazardous events occur that can pose a significant threat to communities and hinder relief endeavors. There is also a need for emergency management personnel to have more effective communication with citizens during a disaster through a tool or interface such as social media (Lovari and Bowen 2019). This is where both social media and community-driven applications can further assist with identifying major disasters as they occur, and potentially speeding up the process of receiving assistance through enhanced context of community needs.

Utilizing social media for natural disaster assessment continues to trend in research studies since social media platforms, such as Twitter/X, first emerged in 2006 and gained increased popularity (Wu and Cui 2018). Twitter/X is one of the world's largest social media platforms, having over 368 million active users as of December 2022 (Tankovska 2022). Social media can be used for a multitude of activities and initiatives. As it pertains to disaster risk reduction, social media can be used in crisis response to serve as a listening function, to track events, for emergency planning and management, to foster connectedness and volunteering, to promote causes to raise donations or funds for those affected by disasters, and for academic research (Alexander 2014). Social media users can express their worry, relief, and other sentiments on such platforms during a disaster, or interact with various community members and stakeholders to share information. It is common during natural disaster events that affected citizens turn to social media for relevant updates, along with seeking help from other individuals or professional organizations during all phases of the disaster cycle (Roy et al. 2020), as social media is faster than traditional news outlets for the dissemination of information (Wu and Cui 2018). Social media platforms have the ability to extract pertinent information through crowdsourcing; benefiting emergency management agencies' protocols and practices when this knowledge is analyzed and modeled for detection, prediction, and other aspects of emergency management. Social media can be integrated into the emergency management process, particularly when it comes to decision making and assessing damages for major disaster declarations.

Platforms such as Twitter/X are social applications whereas a community-driven application, as we define it, is a platform that seeks input from users for a particular situation or circumstance.

Both types of platforms engage members of society, but social applications allow more creativity with content to build a unique network while community-driven applications result in more of a targeted network to share information. Waze, a popular community-driven navigation application, is used by over 151 million monthly active users across the globe as of December 2022 (Smith 2023) and employed in operations by U.S. State DOTs, collecting hundreds of thousands of high frequency data pertaining to traffic and incident events a day for a given state. Some U.S. State DOTs are using Waze data to enhance current communication systems already in place (The Eastern Transportation Coalition 2017). However, there is a need for augmented systems to be developed and deployed, as currently Waze data is typically input either in parallel with other systems or by itself into DOT feeds. The lack of integration of Waze and other communication systems, reveals a gap where Waze data can be merged with an outside source, such as social media data which is already popularly used in disaster research, to increase situational awareness and aid decision making. Community-driven applications such as Waze can address select shortcomings that most social media platforms currently possess, regarding lacking interactive features where users can send reports and update other community members specifically on certain harmful events through pre-established categories. Waze also has more precise location data and interactive geographical visualizations.

Social media, however, adds the community individual voice and sentiment of users that Waze lacks. It has also been revealed that despite platforms to aid disaster management using social media data, few are designed for citizen connectedness or use both social media and another platform (i.e., different type of input data such as community-driven applications) (Chair et al. 2019). In essence, Waze enhances Twitter/X with the higher volume of accurate coordinates related to events beyond the bounding box Twitter/X provides with its current API, while

Twitter/X enhances Waze with adding more context to the categorized alert types (e.g., a Waze alert deemed “Accident” when paired with a tweet in the same area could potentially show how many cars are involved, if someone was seriously injured or needs help, and possibly images related to the event). Thus, the two data streams complement one another. When information provided by active users on public platforms during crises are tagged with geolocation, it aids emergency responders in determining where people are located, evaluating community needs, and providing alerts and warnings to both citizens and first responders in regard to changing environments (Lindsay 2011). Georeferenced posts can strengthen situational awareness and aid in the four phases of emergency management (mitigation, preparedness, response, and recovery) by allowing agency officials to gauge and track community reactions and opinions in real-time related to a disaster.

As social media continues to play a significant role in disaster studies, it becomes crucial to develop approaches that effectively serve the population of people that can access and utilize social media during crisis events. It also becomes critical that when building computational models designed to be implemented into society, that machine learning bias is mitigated. While the effects of existing data analytics approaches and the fairness these techniques have on vulnerable and underserved populations during disasters remains relatively understudied (Yang et al. 2020), approaches such as qualitative measures to mindfully construct a machine learning model exist (Monroe-White and Lecy 2022). Constructing models with these demographics in mind is critical to integrate into any study. Following scholastic-based bias mitigating protocols can begin to bridge the knowledge gap of equitable measures in disaster informatics, advancing our understanding in this domain. To achieve augmented emergency management capacities with platforms such as Twitter/X and Waze, it is pertinent to understand past research that has been conducted on the use of historical data to enhance data sets, social media as a social sensor, the fusion of different data sets for natural hazards and disasters, and mitigating machine learning biases.

4.1 Use of Historical Data

Historical data can be used to provide additional content or background knowledge on a particular problem, or generate more robust models when trained on a previous event for tasks such as simulations of current or future events. The machine learning concept of Transfer Learning makes the use of past data easily capable of being integrated into prediction models, typically in situations where data is scarce or limited, as it is the ability of a system to provide the knowledge of the domain it is trained on (i.e., the source) to another domain (i.e., the target) (Neyshabur et al. 2020). The use of a pre-trained model on historical data for Transfer Learning is seen across various infrastructure research areas such as in the energy sector for models that have limited energy related data (e.g., wind power production) (Hooshmand and Sharma 2019) and to infer energy consumption and demand for buildings (Peirelinck et al. 2022; Ribeiro et al. 2018). Other infrastructure areas this approach is used in is smart city applications such as activity recognition and building dynamics (Pinto et al. 2022) and transportation for GPS record estimation on speeding (Yu 2019).

As it pertains to emergency events, Halse et al. (2019) generated a simulator system that emulates real time tweets from previous tweets based on their temporality with a crisis event. This was designed to replace collecting tweets directly from Twitter/X. The authors showed that use of historical tweets can be used for predicting current events and noted a recommendation that custom filtering should be used for training purposes (Halse et al. 2019). Other studies have made use of historical data for natural disasters through scenarios such as remote sensing for flooding (Pollard et al. 2018; Qi et al. 2009), predicting earthquakes (Yuen et al. 2005), and emergency management for validation of emergency vehicle travel times (Henchey et al 2014) and decision making (Romanowski et al. 2015).

4.2 Twitter/X as a Social Sensor

Infrastructure (e.g., bridges, power systems, etc.) can have physical sensors to monitor or detect damage but social sensors (e.g., Twitter/X) have been shown to detect events where physical sensors are lacking, such as providing detailed information about the failure (Tien et al. 2016). Twitter/X posts (i.e., tweets) can include information such as images and text descriptions, replies, retweets, favorites, and geographical metadata about where the user posted. Currently, about 1-2% of tweets are geo-tagged, and location information can either be a precise location or a Twitter/X "place" (e.g., bounding box) (Twitter/X Developer Platform 2023). While this can be made useful in crisis situations, it is challenging to extract relevant information to assess and gain actionable insights with precise coordinates. Other typical challenges when dealing with social media pertain to trust, privacy, volume of data, availability of geotagged posts, and "rumors" or fake news that spread when people misuse social media (Rossi et al. 2018).

Social media platforms, like Twitter/X, have been used in a wide range of ways in the field of civil engineering. Social media analysis has been used to improve traffic conditions (Athuraliya et al. 2015; Sujon and Dai 2021), detect emergency events via Natural Language Processing (NLP) (Verma et al. 2011; Wang and Taylor 2019), develop communication networks in the construction industry (Boddy et al. 2010; Du et al. 2020), and determine disaster-related impact assessments on the built environment (Hao and Wang 2021; Yuan and Liu 2020). Also, social media has been used to study human mobility by identifying city-scale patterns (Wang and Taylor 2016), user polarity of sentiments (Wang and Taylor 2018), and urban-level spatiotemporal energy demand prediction (Mohammadi and Taylor 2017). The use of social media integrated into other systems can improve situational awareness through augmenting communications and informing decision makers on resources and aid needed in affected areas (Yin et al. 2012).

Additionally, Twitter/X with community-driven applications have been used in research involving the Waze navigation application to examine real-time traffic flow data from Waze in comparison to Twitter/X data congestion (Sidauruk and Ikmah 2018). Twitter/X has been shown to be less reliable in comparison to other crowdsourced data feeds (Amin-Naseri et al. 2018) in terms of less tweets being made at night versus during the day, most being during peak traffic hours, and, while covering arterials well, most tweets come from the center of a city (i.e., providing less coverage from outside areas) (Gu et al. 2016). Twitter/X data will continue to be used in various fields for analysis and detection within communities, however, the number of tweets during a disaster can fluctuate depending on the disaster and how engaged community members are on the platform. There are also cases where tweets relevant to a disaster are smaller in volume than expected, necessitating more data points to be ingested into a model for further community perspective (Salley et al. 2022). This requires augmenting social media data sets with applications that are more equipped for real-time event detection (e.g., Waze), which social media largely lacks. Social media can be used to better protect people, property, and the environment in crisis events, however, relies on the interdependencies of different systems to enhance actions taken in the phases of emergency management.

4.3 Fusing Data for Natural Disasters

Data integration is critical for timely and effective crisis information collection and communication, data analysis, and emergency personnel decision making for disasters; however, data integration can be a challenging task (Peng et al. 2011). Within the field of disaster informatics, established research has highlighted significant challenges pertaining to data integration (Ogie and Verstaevel, 2020). Purohit et al. (2019) have identified three specific challenges associated with the integration of open-source data for disaster management. These challenges include the heterogeneity of data sources, where the diverse formats of multiple data sources can make merging difficult; the inconsistency of data sources, which results from different words or semantics used across data sources, making the establishment of an interpretable structure challenging; and the incompleteness of data sources, characterized by the scarcity of data or the lack of relevant information (Purohit et al. 2019).

Researchers have initiated efforts to tackle these challenges by devising data fusion methodologies. Some approaches aim to merge data from various sources to assess earthquake impacts, incorporating damage data from forecasts and remote sensing with field measurements (Loos et al. 2020; Loos et al. 2022). Additionally, they have been applied in situations such as the assessment of damage caused by Hurricane Matthew, where unmanned aerial vehicles (UAVs) and social media data, such as tweets, were integrated (Yuan and Liu 2018). Moreover, these fusion techniques have been employed in urban analytics by combining sensor data and social data (Psyllidis et al. 2015). With research emphasizing the intricate nature of data integration in disaster management, there is a continuous need for thoughtful solutions to address them effectively. Research also highlights the importance of approaching data integration responsibly by collecting "good data" (e.g., data that has quality content, truthful, etc.), that is unbiased (Nargesian et al. 2022). While integrating different datasets can help alleviate potential biases, it remains essential to mitigate bias through the implementation of some set of standards or well-defined parameters to ensure reliable computations (Albahri et al. 2023).

4.4 Mitigating Machine Learning Bias

Studies have shown how race, social class, and/or placement play a role in populations experiencing social and environmental injustices related to hazards and disasters (Adeola and Picou 2017; Bodenreider et al. 2019; Griego et al. 2020; Hamideh and Rongerude 2018; Nejat et al. 2022;Wright 2011). With the growing integration of machine learning into social decision-making and everyday routines, such as emergency management, there has been a call to control and assess fairness in computational efforts to avoid the risk of exacerbating bias. There is no consensus or widespread agreed upon definition of "fairness" as it pertains to bias and equity in machine learning; how fairness is determined depends on the research question and situation it is applied to. This paper defines fairness as the act of addressing bias with the objective of diminishing the potential adverse consequences upon societal integration. Research has established three ways to quantitatively perform bias mitigation before, during, and after model execution: in the training data, while training machine learning models, and on trained machine learning models (Hort et al. 2022). Previous research has investigated fairness through approaches such as fairness testing algorithms (i.e., inconsistencies between existing and mandated fairness requirements of a software), these are typically binary and divide the population into privileged and unprivileged based on a sensitive attribute that protects against unfairness such as age, race, gender, etc. (Chen et al. 2022). Issues with this type of quantitative testing is that it relies on sensitive attributes when in practice that information may not be available in a data set (Awasthi et al. 2021). For instance, Twitter/X does not provide such demographic information from its users to researchers. Additionally, studies report that current fairness algorithms and metrics cannot handle multi-class problems and non-binary problems (Hort et al. 2022). Therefore, if your data set does not have sensitive attribute data or has more than two labels, current models that assess fairness would not be adequate.

Critiques have surfaced asserting that quantitative research undervalues equity, and when confronted with equity shortcomings, statistical measures are employed to defend the validity of such an analysis (Gillborn et al. 2018). However, with fairness testing there is no guarantee or empirical evidence demonstrating its applicability or effectiveness in real-life scenarios (Chen et al. 2022). Researchers further expose that the report of low bias scores using such quantitative approaches does not automatically equate to actual fair application of models (Hort et al. 2022). Social scientists strongly argue for the imperative of combining machine learning models with a qualitative approach to thoroughly assess bias mitigation efforts (Monroe-White and Lecy 2022). Protocols such as the Wells-Du Bois Protocol for machine learning biases could be deployed to overcome systemic inequities ingrained in data sets which historically sought to oppress marginalized communities (Monroe-White and Lecy 2022). Use of intentionally building machine learning models with qualitative protocols is a promising alternative for the limitations and discrepancies within current algorithms for bias control.

This research project addresses several research gaps: 1) for social media analysis methods: integration with community-based applications that may improve capture of incidents relating to emergency preparation and response (Chair et al. 2019), 2) incorporating equity-based practices to mitigate machine learning bias (Monroe-White and Lecy 2022; Yang et al. 2020), 3) creating a method to effectively augment location-specific social media data with community data to address the shortcomings that exist in the ability to more rapidly, and effectively, communicate and respond to crisis events (Lovari and Bowen 2019; Purohit et al. 2019), and 4) developing algorithms that can address imprecise location information in social media data when used to augment community data on crisis events. Intervening and alleviating disasters as they occur in real-time poses an issue for many emergency responders. Again, before necessary federal assistance is given, state and local emergency management personnel need to make decisions to prepare and respond to disasters to mitigate their effects with available resources. This can be facilitated through a more community focused, equitable approach to better understand local needs of citizens and engaging with community discussions that are occurring. To address these gaps, we investigated the following research question: *What is the impact of integrating social media with community-driven applications for the capture of incidents related to emergency management, mitigating machine learning bias, and validating its respective effectiveness (e.g., accuracy)?*

In the following chapter (Chapter 5), we develop and apply a model (i.e., Application I) that assesses community needs and provides context for emergency responders using machine learning techniques to train the model on previous events and fuse data from the social media platform Twitter/X and community-driven application Waze. We also mitigate machine learning bias of the framework using an equity-based protocol to show how our methodology integrated equity measures. Then, in the ensuing chapter (Chapter 6), we develop and apply a model (i.e., Application II) capable of addressing location inaccuracies that exist in many Twitter/X postings. Both models address the needs identified in our interviews with state DOTs (Chapter 2) and the key performance indicators established in our proposed competency matrix (Chapter 3). We anticipate these machine learning-enabled model frameworks can enhance event detection, provide further situational awareness about an emergency event, and thus improve crisis event response.

5 MODEL DEVELOPMENT AND APPLICATION I: TRANSFER LEARNING FOR FUSING LIMITED GEOREFERENCED DATA FROM CROWDSOURCED APPLICATIONS ON THE COMMUNITY LEVEL

The scope of this initial model framework is two-fold: 1) use historical data to develop a robust model and incorporate more community insights and 2) perform data integration across social media and community-driven platforms at the community scale. The reason this study is at the community-scale (i.e., neighborhood to city scale) is to correspond to the bounding box locations of Twitter/X, which will be explained in more detail later. To achieve these aims, we fuse Twitter/X and Waze data and propose machine learning approaches and spatiotemporal data fusion that utilizes labeling from Transfer Learning for Twitter/X and Waze data sets related to natural disaster events. Figure 1 illustrates the overall framework developed for the integrated approach with the goal of augmenting georeferenced social media data (i.e., Twitter/X) with corresponding data from a community-driven application (i.e., Waze). The framework overall utilizes the techniques of Transfer Learning, NLP, LDA, Semi-Supervised Learning (SSL), and spatial fusion to produce the output of an augmented data set that classifies each Twitter/X and Waze pairing to elucidate community conversations and issues. In Figure 1, the Source Domain Model is the component projected up and to the right from the Transfer Learning box, which produces the output of labels. The rest of the process occurs in the Target Domain Model which produces the output of a map of community needs. The following sections will explain in further detail the workflow of the framework outlined in Figure 1 below.

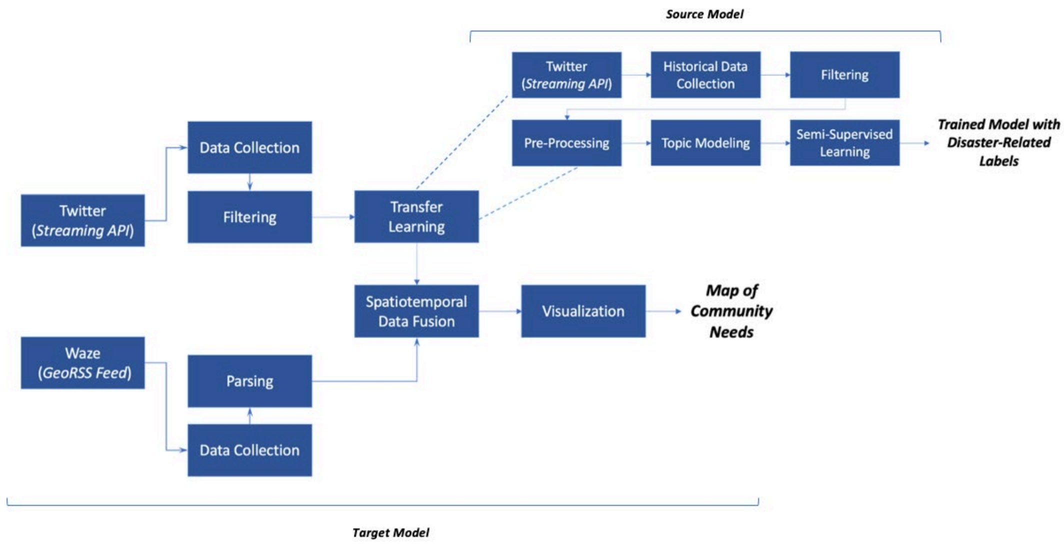


Figure 1. Framework. High-level framework for georeferenced data fusion (Twitter/X and Waze) workflow, including the process for transfer learning. The transfer learning process leverages pre-existing knowledge, which in this case is derived from historical tweets, to create the source model. Subsequently, the source model is trained and integrated into another domain, referred to as the target model. Here, the domain knowledge from the source model is effectively incorporated to amplify performance and understanding.

The evident biases of social media data should not discourage efforts to mitigate biases in models that utilize this data. Even if acceptable metrics in terms of precision, recall, and F1 score are achieved, it remains essential to assess the potential impacts of this work in practice through recognizing biases. The Wells-Du Bois Protocol is a tool designed to determine if research qualitatively achieves a baseline level of bias mitigation in social scientific research for neutral data collection and machine learning execution. It consists of three dimensions and seven items: Bad Data - 1) Inadequate Data and 2) Tendentious Data; Algorithmic Bias - 3) Harms of Identity Proxy, 4) Harms of Subpopulation Difference, 5) Harms of Misfit Models, and Human Intent – 6) Do No Harm and 7) Harms of Ignorance. In this study, these items were viewed through the domain of utilizing social media in crisis event/emergency management. Detailed information regarding each item and the corresponding steps undertaken in this study to assess the fulfillment of the specified standard is provided in a later section.

5.1 Source Domain Model

Historical Data Collection. To enhance the presence of the community's perspective, we incorporated historical data into our framework. We analyzed different major disasters of the same type (i.e., hurricanes), to track sentiments over time and to capture different communities who may have been engaged for one disaster but not another. The assumption posits that within the historical events under examination, varying geographical regions or demographic groups will be represented, as each catastrophic event attracts distinct audiences. Historical data were collected in the form of tweets from three hurricanes that occurred in Florida in 2020: Hurricane Eta (November 7th, 2020 - November 12th, 2020), Hurricane Isaias (July 31st, 2020 - August 4th, 2020), and Hurricane Sally (September 14th, 2020 - September 28th, 2020) (FEMA 2023a).

Filtering. In this process, tweets in the state of Florida were extracted and filtered based on location and keywords in the form of a disaster-based glossary we developed (see Table 7). Past studies have shown that the use of hashtags can limit the number of irrelevant tweets (Brunila et al. 2021). However, in this case the quality of data with hashtags was not sufficient, therefore restricted keywords were determined to be used after several tests were run and analyzed using one or the other (or both). Hashtags are also constantly changing and evolving. Therefore, for the model to be more generalizable the decision was made to use only keywords. Hence, we created a disaster-based glossary of common words related to natural disasters that could indicate a crisis event. The disaster-based word glossary with 103 words was developed based on the Emergency Events Database (EM-DAT), Federal Emergency Management Agency (FEMA), United Nations

Office for Disaster Risk Reduction (UNDRR), and Waze. Web scraping was performed to identify the keywords from the respective sites, and manual inspection was done to ensure there were no duplicate terms among the sources and words that were fully applicable or used commonly were represented from longer phrases (e.g., used the word "damage" instead of "estimated damage" in the EM-DAT database). The tactic was designed to maximize the number of relevant tweets that could be collected.

Table 7. List of keywords used to create disaster-related word corpus, and their source.

Keywords	Source
Affected, Airburst, Avalanche, Chemical, Climate, Coastal, Collapse, Damage, Death, Derecho, Disaster, Disease, Drought, Earthquake, Epicentre, Epidemic, Explosion, Famine, Fire, Flood, Flow, Fog, Food, Freeze, Frost, Hazard, Homeless, Hurricane, Ice, Impact, Injured, Injury, Lahar, Lake, Landslide, Lava, Lightening, Liquefaction, Loss, Missing, Niño, Poisoning, Power, Rain, Risk, Seiche, Shake, Sinkhole, Soil, Storm, Subsidence, Surge, Tornado, Transport, Tsunami, Typhoon, Volcanic, Vulnerability, Wave, Wind, Winter	EM-DAT (CRED 2009)
ARC, CDC, CERT, Community, Crisis, DHS, Drill, Emergency, EMS, EOC, EPA, Evacuate, Evacuation, FEMA, HAZMAT, IMT, Incident, JIC, JIS, NGO, NIMS, Procedure, Protection, Rescue, Responder, Response, Shelter, Structural, Threat, Tree, Warning, Watch, Water	FEMA (FEMA 2023b)
Building, Critical	UNDRR (UNDRR n.d.)
Accident, Alert, Construction, Jam, Road, Traffic, Weather	Waze (Waze 2017)

Pre-Processing. Textual data can be informal and not structured in a way to enable classification processes. Text from social media can be noisy containing special characters (i.e., emojis and symbols), slang and misspelled words, hashtags, URLs, and more (Salas et al. 2017). Text mining approaches ease the difficulties associated with the time consuming and inconsistent process of manually cleaning data, and have been proven to have higher accuracy than no pre-processing techniques being performed at all (Mhatre et al. 2017). In order to prepare all text for the classifiers, we removed these additional elements (e.g., extra URLs, white spaces, special characters, upper case words, and unnecessary words) using FastText (FastText 2023). This is done through standard techniques such as tokenization (i.e., breaking a sentence into words), stop words removal (i.e., to simplify text and remove words that add no meaning such as “a”, “the”, etc.), stemming (i.e., to find the root/stem of the word), and lemmatization (i.e., generating the base or dictionary form of a word) (Mhatre et al. 2017). After these pre-processing techniques, we had a clean corpus of words and the fused textual information was converted to vectors to be utilized in the LDA model to generate labels based on all text.

Topic Modeling. To obtain the labels for classifying the data set, LDA topic-based modeling was performed. We utilized different LDA-related Python packages to model our pre-processed tweets, running the model with different parameters (altering the number of topics and words within each topic), and using a standard deviation test to determine the number of topics. From the standard deviation test, 4-6 topics was identified as the preferable range, and running the model on these three different options, 5 topics was deemed as the most optimal. Also, to further refine the model, we went back and modified the parameters further to only include words that were nouns, adjectives, and verbs in the pre-processing portion. After running several tests, the number of topics was set to 5 (with 10 words in each topic). Interpreting each topic, the topics were: “0”: Broadcast/News (e.g., anything to do with the news, the government, alerts, etc.), “1”: Power (e.g., anything to do with power outages, power lines, power systems, lights, Wi-Fi, Internet, etc.), “2”:

Traffic Incidents (anything to do with car crashes, congestion on the roads, evacuation, etc.), “3”: Forecast/Weather (anything to do with the climate, flooding, etc.), and “4”: Miscellaneous (anything that does not fit into the categories and/or has nothing to do with a disaster). The last topic also acts as an additional filter to catch tweets that made it into the text corpus that may have a different interpretation of a word in the disaster-based glossary. Throughout the rest of this paper, the labels will primarily be referred to by their corresponding numerical identifiers as mentioned in the previous sentence. The top eight most frequent words identified in the LDA model were: watch (appearing 7,772 times), broadcast (appearing 4,910 times), storm (appearing 4,553 times), chance (appearing 3,293 times), tonight (appearing 2,960 times), live (appearing 2,391 times), forecast (appearing 2,323 times), and alert (appearing 1,960 times). These top words indicate discussion around a time-sensitive storm, and that needs pertain most frequently to the topics connected to weather and what is being outlined in news reports. This is beneficial to operators as it can help them with tasks immediately after a disaster such as crafting public safety messaging relevant to what people may or may not already know about the disaster, or emerging risks responders will face when dispatched.

Semi-Supervised Learning. The topics from the LDA model described in the previous section (i.e., “0”: Broadcast/News, “1”: Power, “2”: Traffic Incidents, “3”: Forecast/Weather, and “4”: Miscellaneous) were used as labels in this SSL approach. The model was generated using a Label Spreading package (Zhou et al. 2004). Roughly 1% of the historical storms data set was manually labeled, leaving 99% unlabeled. To determine the 1% of the fused data that would be manually labeled, the data set was randomized using a function in Python, then labeled with equal distribution of each topic classification. Labeling 1% of the data was determined to be the guideline for how much of the data should be labeled, as the aim is to limit the manual training of the data, and labeling 1% has been found to achieve high accuracy (Chen and Wang 2017). The annotators consisted of two members from our research team. Annotators divided the 1% of the data set that required labeling according to a well-defined and mutually agreed-upon set of label definitions. After each designated annotator completed their assigned portion, they collaboratively reviewed and discussed the labeling. In the rare event of any disagreement, a third team member was available to arbitrate. This internal validation protocol was implemented throughout the labeling process (Chowdhury and Zu 2023).

The data was split into 70% being the training set and 30% being the testing set. This was fed into the model, generating pseudo-labels for the entire data set based on the model's prediction. A validation set of 20% of the data was extracted from the training set prior to this analysis, to provide an unbiased evaluation of the model fit on the training set and to tune the hyperparameters. Once the model was completely trained, we ran the testing set to see if the model could predict labels on this data set with adequate accuracy through our evaluation metrics discussed in the next section, and labels were successfully generated from the model for our testing set. With this task completed, we then had a trained model that was ready to be used for the Transfer Learning process.

Evaluation Metrics for the Model. To assess the validity of the model, precision (i.e., true positives over all that was predicted as positive), recall (i.e., true positives over all that should have been predicted as positive), and F1-score (i.e., combination of precision and recall, the overall accuracy) were calculated (see Equations 1-3), along with creating confusion matrices. Table 8 shows the classification reports for the historical data set. "Support" outlined in last column of Table 8 and other classification report tables, is the count of occurrences experienced in each class. A confusion matrix was generated, Figure 2, based on these initial scores and a single-fold analysis. A 10-fold cross validation was then done again on the precision, recall, and F1-score metrics to generate a final accuracy, and a confusion matrix was also produced for this cross validation based on the label spread performance of the model (see Figure 3). These were done for the 5 topics (i.e., the 5 classes in the classification reports).

$$Precision = \frac{TP}{(TP+FP)} \qquad \text{Equation 5.1}$$

$$Recall = \frac{TP}{(TP+FN)} \qquad \text{Equation 5.2}$$

$$F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \qquad \text{Equation 5.2}$$

where TP and TN are the number of true positives and true negatives, and FP and FN are the number of false positives and false negatives.

Table 8. Classification report for label spread for historical data.

	Precision	Recall	F1-Score	Support
0	0.95	0.90	0.93	2270
1	0.82	0.59	0.68	2221
2	0.93	0.69	0.79	666
3	0.86	0.87	0.87	2476
4	0.81	0.95	0.88	5392
Accuracy			0.85	13025
Macro avg	0.87	0.80	0.83	13025
Weighted avg	0.85	0.85	0.85	13025

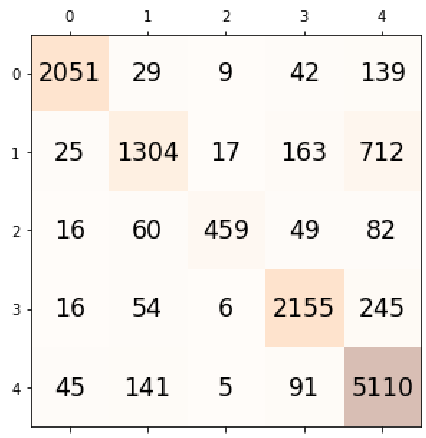


Figure 2. Matrix. Confusion matrix for label spread for historical data (no cross validation).

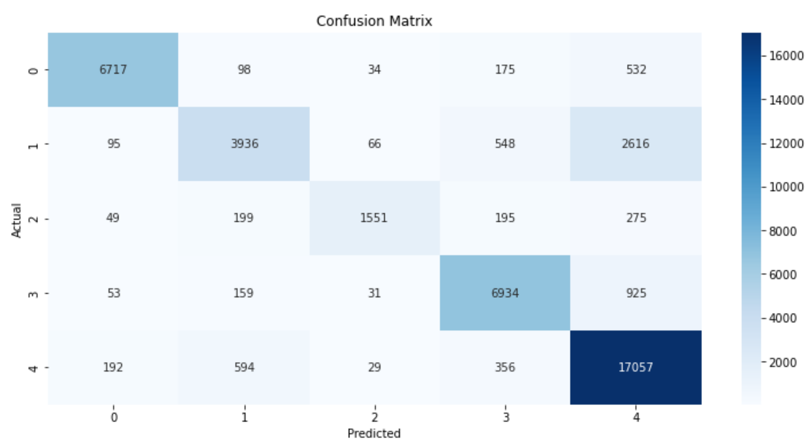


Figure 3. Matrix. Confusion matrix for label spread for historical data (cross validation). With average accuracy across folds: 0.833.

5.2 Target Domain Model

Case Study. According to FEMA, the state of Florida has experienced over a dozen major disaster declarations in the last decade alone, ranging from tornadoes to hurricanes, with one of the most recent major disasters being Hurricane Ian (FEMA 2023a). Despite Florida being a coastal state that experiences numerous natural disasters, historically it has just under a 70% success rate with being granted major disaster status for aid disbursement (Schmidtlein et al. 2008). Hurricane Ian is tied as the fifth strongest hurricane to hit the United States and began on September 23rd, 2022 in the central Caribbean as a tropical storm, and three days later on September 26th, 2022 became a hurricane (NOAA US Department of Commerce 2022). When Hurricane Ian approached southern Florida on September 28th, 2022 it was a Category 4 storm, and left Florida the next day, with intense winds and rainfall, as a tropical storm again heading to South Carolina (NOAA US Department of Commerce 2022). This case study, which we conducted to demonstrate our framework focuses on the "immediately after" part of the disaster cycle (i.e., meaning right after the disaster has left an area) to see community conversations based on the impacts of the hazard. This also aligns with when damage assessments would typically take place. Other studies have investigated two-week periods beginning at the landfall or origin of when the storm begins (Samuels and Taylor 2020) and showed that at two weeks the discussion gradually decreases (Zou et al. 2018). Since this study is focused on immediately after the storm exits an area, and investigates when people could be most engaged, a weeklong period was studied for Hurricane Ian, making the "Post-Disaster Period" September 29th, 2022 to October 6th, 2022. Hurricane Ian also was declared a major disaster on September 29th, 2022 (FEMA 2023a), emphasizing the importance of promptly understanding the ongoing situation with the expeditious declaration.

Data Collection. Data were collected from Twitter/X and Waze during this period for Hurricane Ian with 10,209 filtered tweets and 15,913 Waze alerts. Twitter/X data were collected through Twitter/X's public Application Programming Interface (API). Data were retrieved for Hurricane Ian from a live data collection stream developed in Python within our Lab. The data are collected by year, month, day, and hour and are stored in JavaScript Object Notation (JSON) format. The Waze data used for this study were Waze alerts, which were initially collected through the Waze

GeoRSS Feed that is shared with Connected Citizens Program (CCP) partners, such as the Georgia Department of Transportation (GDOT), for further configuration. The Waze data are collected in Extensible Markup Language (XML) format, showing pertinent information such as the date and time of an incident, precise coordinates, type and subtype of an alert, street name where the alert occurred, country, road type, report rating, confidence, and reliability of incident feeds within the bounding box of the state of Florida.

Parsing (Waze) and Filtering (Twitter/X). For Waze alerts, the provided GeoRSS feed collected data that needed to be transformed to a readable format for the model. The same filtering process in the Source Domain Model for tweets was executed here to maximize the number of relevant tweets on Hurricane Ian in Florida.

Table 9. Classification report for label spread for Hurricane Ian (with transfer learning from historical data).

	Precision	Recall	F1-Score	Support
0	0.93	0.73	0.82	205
1	0.85	0.49	0.62	89
2	0.92	0.87	0.89	146
3	0.77	0.92	0.84	968
4	0.91	0.85	0.88	1655
Accuracy			0.86	3063
Macro avg	0.87	0.77	0.81	3063
Weighted avg	0.86	0.86	0.85	3063

Transfer Learning. The Transfer Learning process is outlined in Figure 1 and the "Source Domain Model" section. The model built in the Source Domain Model is already trained and ready to be used at this point in the Target Domain Model. There is no more training or manual processes. It is fully automated since the Source Domain Model was saved and applied here. When the data for Hurricane Ian was run through the saved model, just as in previous evaluations, the predicted labels were assessed with precision, recall, and F1 scores (see Table 9) along with confusion matrices (see Figures 4 and 5). The outputs demonstrated that the model is a reliable model, even having a higher accuracy score than the Source Domain Model.

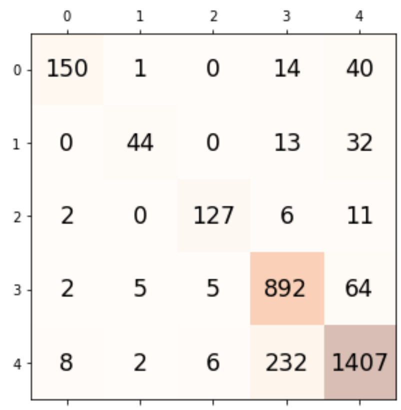


Figure 4. Matrix. Confusion matrix for label spread for Hurricane Ian (no cross validation).

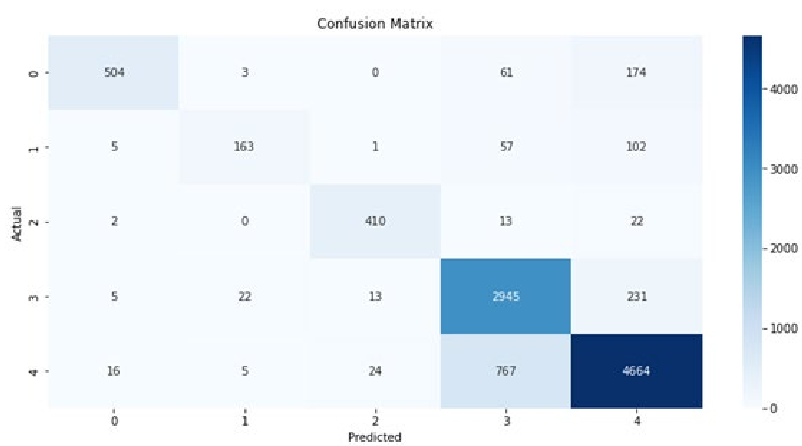


Figure 5. Matrix. Confusion matrix for label spread for Hurricane Ian (cross validation). With average accuracy across folds: 0.834.

Spatiotemporal Data Fusion. The benefit of data integration is being able to increase and strengthen data sets that complement one another. Both Twitter/X and Waze data sets have date, time, location, and textual information pertaining to event detection for a natural disaster. With

Twitter/X, the textual data is the tweet itself (i.e., what the user has posted) and the location is in the form of a precise location or bounding box with the current API (most are the latter). For Waze, the textual data is the alert type and subtype given to the report that the user assigns to the incident, and it provides a single coordinate pair. Waze alerts are classified with the following types: Accident, Jam, Weather hazard/Hazard, Miscellaneous, Construction, and Road closed. The subtypes provide more detail for each alert type such as Weather hazard/Hazard displaying subtypes pertaining to fog, hail, rain, snow, hurricanes, etc.

To fuse these data sets, we identified and paired the tweets and Waze alerts within minimum spatial proximity. This was achieved using the Haversine Distance between locations (see Equation 4), which can be used to calculate distance between latitude/longitude pairs for real-time classification (Zubiaga et al. 2017). With the current API's bounding boxes for tweets described as being able to be as large as 25 miles in width and height (Twitter/X Developer Platform 2023), in order to refine the spatial scale of the tweets collected they were further filtered to identify neighborhood or city information (i.e., shrinking the size of the bounding box). To obtain coordinates for each neighborhood or city bounding box the center of each bounding box was found, which has been done in previous work on a larger scale (Zubiaga et al. 2017). The crowdsourced data produced by Waze has been reported to be slightly inaccurate for location, with about a 30 second delay in reporting causing an incident to be recorded 0.8-km (i.e., half a mile) away (Amin-Naseri et al. 2018). The parameters for selecting the tweets closest to Waze reports were set within 1.61-km (i.e., 1-mile) of location to one another, to account for delays of up to 60 seconds in reporting an incident. The merge is based on location and date, displaying all attributes of both feature layers in one data set. The output is a fused data set, matching a tweet with the nearest Waze alert with each data point showing the paired data sets' information along with a classification label and distance from one another. Upon completion, there were 2,566 Twitter/X and Waze pairings. Note that the same tweet can be paired to multiple Waze alerts depending on proximity (see Figure 6).

$$d = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2-\varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \qquad \text{Equation 5.4}$$

where φ_1 and φ_2 are the latitude coordinates of two points, and λ_1 and λ_2 are the longitude coordinates of two points.

Figure 6. Illustration. Example of how one tweet, within a one-mile radius of a Waze alert, can make 3 different pairings depending on location.

Generalizability. To test generalizability of the proposed framework, the model was run again on data in a different state, for a different storm. Tropical Storm Zeta in Georgia was the storm used to test the generalizability of the model. The disaster began on October 21st, 2020 in the western Caribbean as troubled weather, being slow to develop, then on October 23rd, 2020 forecasts reported Zeta disturbance being brought into the southern Gulf of Mexico along with greater odds of this weather being an actual storm (NOAA US Department of Commerce 2020b). On October 25th, 2020 the tropical storm formed and strengthen over the next few days as it began approaching the U.S., with its highest strength being reported as a Category 3 hurricane (NOAA US Department of Commerce 2020b). Georgia news outlets, and other mediums (e.g., National Weather Service) reported on Tropical Storm Zeta with the storm striking and leaving the state of Georgia on October 29th, 2020 (FEMA 2021; NOAA US Department of Commerce 2020a). This test focused on dates pertaining to after this incident, the same “right after” timing with a weeklong period done for Hurricane Ian in the case study above. The “Post-Disaster Period” was October 30th, 2020 to November 6th, 2020. The results of the study (Table 10) show that the model is acceptable not only for the state of Florida but can be transferred and used effectively for Georgia. The results suggest the possibility of using this framework for other events in other locations as well, as it has an accuracy score that is within a 3% margin of the model outputs from Hurricane Ian. Confusion matrices with and without cross validation were also produced for the Georgia event (see Figures 7 and 8).

Table 10. Classification report for label spread for Tropical Storm Zeta in Georgia.

	Precision	Recall	F1-Score	Support
0	0.94	0.48	0.64	127
1	0.95	0.28	0.43	76
2	0.93	0.92	0.93	74
3	0.85	0.65	0.74	343
4	0.81	0.98	0.89	968
Accuracy			0.83	1588
Macro avg	0.90	0.66	0.72	1588
Weighted avg	0.84	0.83	0.81	1588

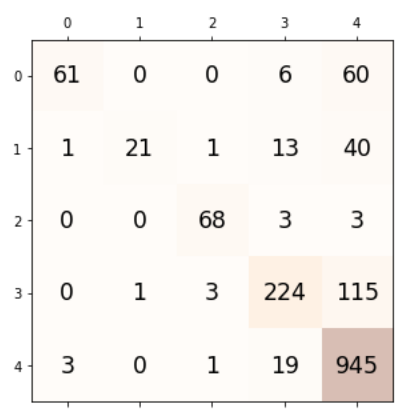


Figure 7. Matrix. Confusion matrix for label spread for Hurricane Ian (no cross validation).

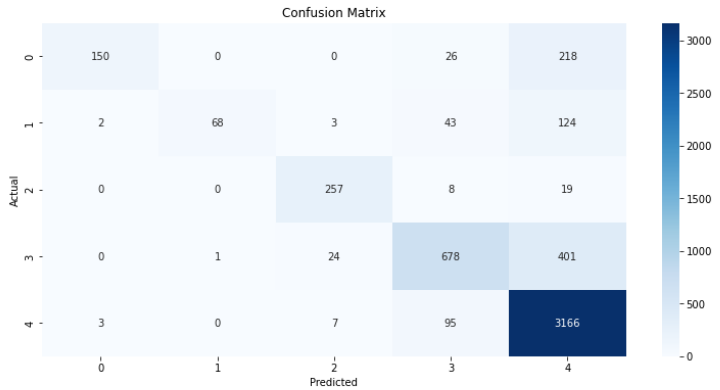


Figure 8. Matrix. Confusion matrix for label spread for Hurricane Ian (cross validation). With average accuracy across folds: 0.812.

Case Study Visualization. Figure 9 shows tweets fused with Waze alerts being spatially mapped (displayed with the labels for each pairing), with context embedded in each icon on the map for emergency personnel to have access to in a visual interface. An example pairing is also pictured in Figure 9 in the blue table, showing how a tweet can add further context to a Waze alert beyond its original classification. As the Miscellaneous label was noted as an additional filter, it is not displayed in the final visualization. The visualization shows a substantial amount of fused data points related to forecast/weather and traffic. The discussion of these topics in particular aids emergency operators and responders with actions such as feasibility of potential infrastructure repair, cleanup, or evacuation planning. Engaging in weather-related discussions enables responders to obtain crucial information, such as the extent of severe flooding in a building or instances of lightning striking trees. Traffic discussion allows them to know what major route or highways are jammed, through the augmented data from Waze. They can then select the proper evacuation routes that avoid congestion, with context of how long it might take for traffic to lighten up (e.g., accident, debris on highway, how many lanes are closed, etc.). Furthermore, with the location information, local governments will also be able to see exactly how a portion of a neighborhood might be affected by a disaster, which can help guide what preparedness plans or mitigation tactics can be explored. All this knowledge assists them in comprehending the specific impacts of the disaster and to tailor their response efforts accordingly.

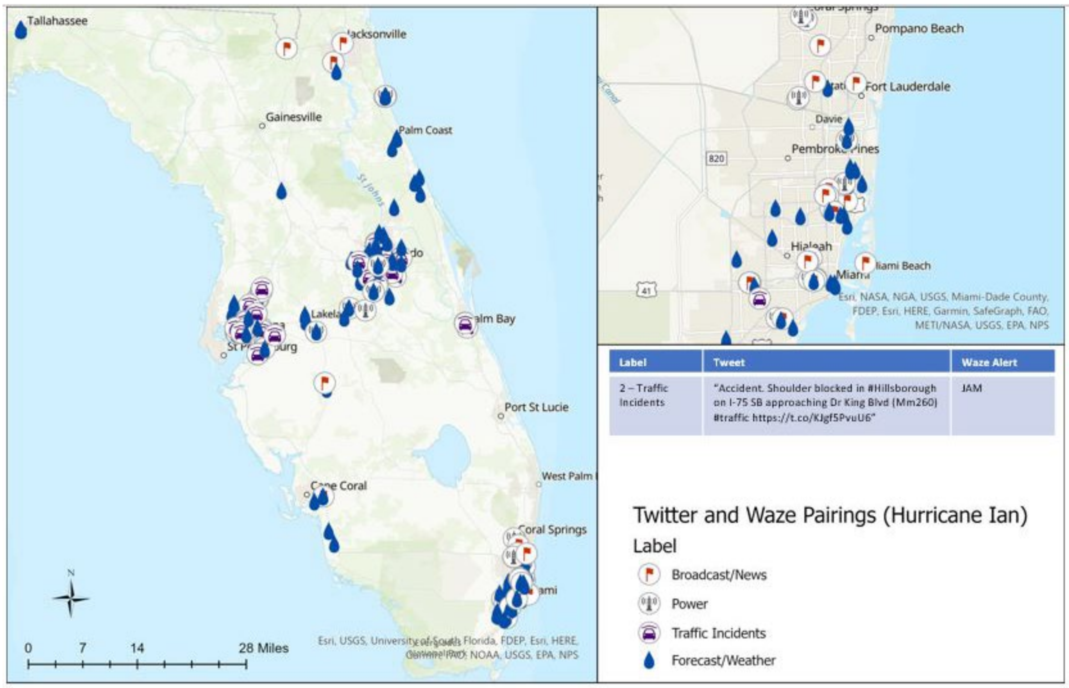


Figure 9 Maps. Example of a classified Twitter and Waze pairing output for Hurricane Ian in Florida, zooming in on an area displaying multiple topics from the study.

5.3 Mitigating Machine Learning Bias

Table 11 below outlines how the model was designed toward equitable practices. It is important to acknowledge that this process does not guarantee the model has no problems in terms of potential bias, but rather serves as a means to implement mitigation strategies and strive towards achieving a threshold for reducing biased research practices; this protocol emphasizes that mitigation efforts are not primarily aimed at solving the issue at hand, but rather at acknowledging and addressing the issue prior to the implementation of a model (Monroe-White and Lecy 2022).

By directing our attention towards the population we aim to serve, those with access to and who use social media and community-driven applications, we meticulously evaluated our data collection and model implementation and adhered to the seven items outlined in the Wells-Du Bois Protocol to actively work towards mitigating potential biases.

5.4 Discussion

The model developed and applied in this research contributes to the larger discussion of enhancing community perspective in disaster informatics. As discussed relative to FEMA's disaster declaration process, it is crucial for emergency management agencies to receive information, from models such as the one from this study that can represent and assess what communities need in near real-time from the people themselves. Communities, such as areas in Puerto Rico after Hurricane Maria, have been documented as being failed by federal agencies due to these organizations not being fully prepared to respond to disasters or being able to anticipate locals' needs (Sullivan and Schwartz 2018). Knowing what an area needs while a crisis occurs, can prevent missteps such as this.

It has also been shown that federal disaster relief falls short of equitable measures, leaving disenfranchised and historically marginalized communities at a disadvantage, with FEMA themselves stating "For disaster preparedness, mitigation, response and recovery to drastically improve in 2045, emergency management must understand equity and become equitable in every approach and in all outcomes" (National Advisory Council 2020). This is why some of their goals in their 2022-2026 FEMA Strategic Plan is to have more of a "people first approach" and "meet current and emergent threats" (FEMA 2023d). To address these needs, our framework is centered around community perspective and constructing a system that keeps equity at the forefront and acknowledges current disparities and potential impacts of machine learning efforts.

Table 11. Model Analyzed with the Wells-Du Bois Protocol (Monroe-White & Lecy 2022).

Dimension	Item	Actions Taken to Mitigate Bias
Bad Data	Inadequate Data - *Does the data exhibit systematic omissions or misclassifications of certain subpopulations?*	Reporting data sizes and metrics is needed to overcome this. Interpreting this to social media in disaster management, applicable descriptive statistics are given in the training data sets and are separated by group identities (i.e., the classification of labels).
	Tendentious Data – *Does the model reflect subjective decisions?*	Disclosure of human judgement is needed. We disclose in our model that 1% of our model is manually labeled, however, the labels themselves that were generated do not pose bias as they were constructed with an LDA model based on the textual information provided by the people and not influenced by the researchers.
Algorithmic Bias	Harms of Identity Proxy - *Is there a potential for the model to exhibit systematic biases towards specific races, genders, or social classes?*	The model did not consider race, gender, or social class as factors for the desired outcome. The way Twitter/X and Waze are both designed, it does not provide such demographic data per post, and only the text and location were needed in this study. This is because in the context of immediate emergency response, it is challenging to prioritize one life over another, as natural disasters can strike without regard for such distinctions. While indicators like community vulnerability would be crucial for assessing measures like these, the focus of this study was to identify and classify imminent needs of social media users.
	Harms of Subpopulation Difference - *Does the algorithm demonstrate varying performance outcomes among different subgroups?*	This study caters to the population that relies on these platforms for communication during a crisis, thereby considering them as a distinct demographic subset in itself. As mentioned previously, only the text and location of each data source was used, the final output does not note who the user was but solely what they said and where they are for enhanced context of the disaster located (i.e., maintaining consistency).
	Harms of Misfit Models – *How does the model assess error? What are the broader public and social implications of this research?*	The model undergoes cross validation to avoid over-fitting. The goal of this research is serving the population who uses such platforms to communicate during a crisis, to aid and enhance the decision making process for emergency management personnel. The impact of this work can improve allocation of resources for emergency events.
Human Intent	Do No Harm - *Are you ensuring transparency regarding the objectives and aims of your research?*	The goal is for this work to be implemented into agencies such as state DOTs, we make sure to document and share this work both with applicable stakeholders and the research community.
	Harms of Ignorance - *Have you carefully examined the potential unintended consequences of your research?*	We have examined the unintended consequences of our research. For instance, if a cyber-attack were to occur on this model as it was deployed into society, adversaries would have knowledge of communities that are currently at risk and what they proclaim to need. Adversaries could send phishing emails, tweets, etc. to try to take advantage of those impacted populations. Inclusion of protective measures should be done for such a system.

Studies such as this add to the growing body of knowledge of determining ways to more accurately, and effectively, recognize community needs during or after a disaster to better serve society with a community centered approach.

The framework also guides decision making towards equitable response to disasters. It is important that computational models work towards fairness as most are currently unfair due to training data that can disproportionally affect marginalized populations, and not considering the harmful effects a model can have when integrated in the real world (Monroe-White and Lecy 2022). Disparities such as wage gaps, mortality, and access to care can be seen in all areas of life and the built environment, and when exposed to natural disasters such disparities can be exacerbated when not accounted for properly. Research indicates that there are still few studies on infrastructure and social equity (Dhakal et al. 2021). Social equity systems research in emergency management and disaster research has employed analyses of social vulnerability to expose how disenfranchised populations recover at a slower rate back to their pre-disaster state (Kim and Sutley 2021). Often the occurrence of natural disasters is viewed as "equal opportunity" in the sense that storms, tornadoes, etc. do not intentionally target a certain population, they just occur haphazardly and can damage everyone just the same (Lieberknecht et al. 2021). While it is true the damage done by major disasters on the surface can be the same (e.g., power outages, extensive flooding, etc.), the post-disaster and recovery phase is not an "equal opportunity" when it comes to the dissemination of resources and the time it takes to re-build a community depending on its pre-existing conditions. This phenomenon may arise as a result of a limited conceptual framework that fails to account for the disparities inherent in contemporary machine learning techniques employed to assist communities, wherein the incorporation of equity benchmarks or the pursuit of fundamental bias reduction may be overlooked. In some cases, without adequate support a disenfranchised population that is met with an emergency event may never fully recover because they already began at a deficit. Addressing such disparities in the physical, economic, and social environments could improve infrastructure systems and approach equity to establish a culture that provides just assets, funds, policies, and education to communities that need it most.

5.5 Conclusion

This study was able to identify areas in Florida that were impacted by a disaster with augmented context of specific needs based on classification of a paired data set employing machine learning techniques. The final output for the historical data identified pertinent topics that could be transferred and applied for use on future hurricanes. The final output for the Post-Disaster Period of Hurricane Ian data showed extensive discussion related to the forecast and weather issues related to the storm, as well as the traffic occurring within communities due to the disaster. This research addresses the post-disaster period of a natural disaster, focusing on disasters deemed as hurricanes and tropical storms for emergency responders, to be able to aid civilians and distribute the necessary resources to specific areas more quickly and efficiently. The model addressed the research question: What is the impact of integrating social media with community-driven applications on improving the capture of incidents related to emergency management, mitigating machine learning bias, and validating their respective effectiveness? The comprehensive investigation demonstrated the integration of social media data with community-driven applications, thereby amplifying the efficacy in detecting and documenting incidents from communities relevant to emergency management. Additionally, our model was capable of accurately representing pertinent community needs while concurrently adhering to a baseline standard for equity through mitigation of machine learning bias. This was evidenced through an illustrative case study using a machine learning-based fusing Twitter/X and Waze through Transfer

Learning, NLP, and spatiotemporal analytics on the georeferenced data streams pertaining to emergency events to accurately detect the location and type (i.e., flooding, road closure, etc.) of an event. To the best of our knowledge, this research project represents one of the initial endeavors to integrate the Wells-Du Bois protocol in order to ascertain the attainment of a fundamental threshold of bias mitigation.

The practical contributions of our research include aiding emergency management decision making and situational awareness for disasters as well as improving allocation of resources to reduce the harmful effects of disasters. It adds to the growing body of knowledge on this topic addressing the shortcomings of Twitter/X and Waze applications for disaster detection and effective augmentation of platforms such as these. It establishes a foundation for 1) an integrative approach between social media and community-driven applications for crisis event detection towards further expansion of response capacity for real-time decision making and 2) including an equity appraisal through incorporating equity protocols into the data process. Understanding such potential disparities is crucial to discover equitable ways to alleviate the subsequent recovery process for those without the necessary resources and contribute to bolstering community resilience.

5.6 Limitations and Future Work

Although Twitter/X is the world's largest microblogging social media network and a popular platform used to extract information for research purposes, latitude and longitude pairs (i.e., precise coordinates) of tweets are no longer automatically attached to tweets, reducing the number of precisely located posts since about 2016 (Maurer 2020). It is optional for users to share their location, thus most tweets collected through Twitter/X's streaming API are not georeferenced with exact location but with bounding boxes from place information instead (Maurer 2020). For the model developed and applied in this chapter, both precise coordinates (when provided) and bounding box coordinates were utilized for tweet location information on the neighborhood and city level. However, more advanced algorithms are required to address location inaccuracies. This is explored and addressed in the next chapter.

As highlighted earlier as a challenge with social media, the volume of data is an ongoing and probable obstacle when dealing with Twitter/X data. Using social media or community-driven data in disaster research is heavily reliant on citizens participating and providing useful information on such platforms. While this information can be useful to measure other metrics or relationships, in real-time tracking when trying to assess the needs of a community, an extensive community voice is needed. Additional efforts can be made by relevant agencies and stakeholders to educate community members about leveraging these platforms as a means of meaningful interaction, fostering actionable outcomes. Alternatively, they can also prioritize the promotion of their existing systems to ensure greater engagement and effectiveness. However, we discovered through a disaster-based glossary for filtration and the use of Transfer Learning, more relevant tweets can be found than previous work (Salley et al. 2022). This framework accounts for scarcity of data and allows for a faster, more automated process when evaluating social media data.

Lastly, as mentioned in Table 11, Twitter/X and Waze do not provide specific demographic information such as race, gender, or social class on a per-post basis. Consequently, the focus of this study was not on sociodemographic vulnerability but rather on the needs of populations affected by crises that rely on these platforms for communication. These populations can be considered a distinct demographic subset in themselves, highlighting Twitter/X's role as a social sensor. While it is important to note that these platforms do not represent the entire population, and recent, comprehensive demographic information has not been readily available since around

2013 (Wang and Taylor in 2016), it is worth mentioning that recent data suggests certain trends. For instance, among its multi-million users, approximately 37% of users are female, while 63% are male on Twitter/X; furthermore, users between the ages of 25 and 34 exhibit high activity, representing around 38% of users worldwide (Dixon 2023). The pursuit of representativeness of the data needs to be continuously asked and answered to further analyze any limitations of the research or further generalizability of its results (Kumar and Ukkusuri 2020) as no data set suits every single task and all can have some sort of limited scope (Nargesian et al. 2022). The most beneficial utilization of social media is achieved when it is used in conjunction with existing emergency management systems at local and government agencies, such as a Department of Transportation (DOT), as it does not holistically represent an entire community and other measures should be used to further contribute to decision making.

Future work related to this framework should adapt this framework to completely online machine learning labeling, negating any manual process. Future work should also further examine historical data in relation to the typical engagement that communities have with emergency management (e.g., good or bad relationships, levels of engagement on social media, etc.) on different spatial and temporal scales. This study explored neighborhood and city levels day by day, but exploration of county and census level data on an hourly or minute basis may reveal other insights. This could also reveal how a community already utilizes local agencies in these spaces and can provide a baseline for how useful social media networks may be for real-time tracking in a particular area. How citizens currently use social media should also be continually re-evaluated, as new platforms are merging and old ones are obsolescing and updates to current platforms occur often. Additionally, it is essential to undertake extensive quantitative and qualitative investigations when dealing with complex issues like these to effectively counter computational biases during model construction and deployment. Given the nature of these challenges, which rely on data and computational solutions, it becomes imperative to continue to investigate a range of bias mitigation methods. More approaches to determine the most appropriate strategy tailored to the unique demands of the research problem should be investigated.

This study can be taken further in the future through the development of a process that works towards fairness more and establishing measures for proper allocation of resources. Presently, there exists a paucity of scholarly investigation concerning the integration of equity metrics or protocols in the utilization of social media within the scope of emergency management. Further exploration is called for to thoroughly examine ongoing constraints as it relates to equity measures in this domain, such as with specific population subsets, thereby fostering a more comprehensive depiction of society.

6 MODEL DEVELOPMENT AND APPLICATION II: ONLINE CONFIRMATION-AUGMENTED PROBABILISTIC TOPIC MODELING

In this chapter we build upon the model developed and applied in the preceding chapter to improve the locational fusing of social and community data postings regarding crisis events. We consider a scenario in which online data streams are leveraged in emergency response, disaster management, or public health monitoring and disease control systems. Due to the real-time coordination in place in these systems, they rely on user-generated data (e.g., social media content, data generated through traffic-information applications, health-related data, and emergency hotlines) to monitor, detect, and respond to events and trends in a timely and effective manner. These systems, depending on the specific monitoring requirements, operate within a predefined spatiotemporal window, which is typically a short timeframe that encompasses real-time data collection and response. In this context, additional data becomes available online, helping to identify when and where specific relevant information may emerge after a certain delay. The objective is to develop a topic model based on the online data stream, with a focus on the topics of interest. The model continuously learns from historical and real-time data to enhance its detection algorithms and improve response strategies.

Streaming data analysis in this way is crucial, as it enables the discovery of relevant topics within the selected content, which in turn plays a pivotal role in tasks like information augmentation (Salley et al., 2022; Wang et al., 2012; Yi and Allan 2009), detecting traffic or crisis events (Fan et al., 2020; Tien et al., 2016; Wang et al., 2019; Wang and Taylor, 2018; Zhang et al., 2017), and more. The key component of the aforementioned studies is the topic representations inferred from short text-based data. Conventional topic models, such as Latent Dirichlet Allocations (LDA) (Blei et al., 2003) and Mixture of Unigrams (MUG) (Nigam et al., 2000), when operating in fully unsupervised settings, often exhibit suboptimal performance resulting from the sparse presence of relevant contents. The reasons for this can be attributed to two factors: (1) user-generated data being exceptionally brief (Lin et al., 2014; Yan et al., 2013), and (2) online data inherently containing noise (Morstatter et al., 2010). While various unsupervised topic models have been developed to address data sparsity (Qiang et al., 2020; Vayansky and Kumar, 2020), less emphasis has been placed on leveraging the data itself for model enhancement. Explicit methods for improving data quality through post-processing and annotations are often time-consuming and costly, especially given the online nature of the data (Aggarwal, 2011).

In this study, as we introduce this challenge, our approach aims to integrate online data with topic models. We do so by introducing a confidence score associated with the periodic arrival of content of interest, which is derived from the combination of various sources of information. To demonstrate the aforementioned scenario involving online data, we provide a real-world application example. We consider the data source as a stream of tweets within a specific geographic area, re-calling that the Twitter/X API V2 has enabled streaming via bounding boxes (Khalid, 2019). We are interested in posts related to emergency events within this area. Waze, one of the largest GPS navigation apps, offers interactive features that allow users to share real-time traffic information and report crisis events. For the purposes of this research, Waze can serve as a valuable information resource, as its data can be correlated with the presence of relevant tweets within a common spatiotemporal window. One valuable aspect is *nThumbsUp*, which reacts and directly reflects the event's significance within the online community. The primary concept here is to gather data that quantifies confidence within a smaller online community. If this score reaches a significant level, it may indicate that a larger online community, such as Twitter/X, is also discussing events of interest. See Figure 10 for the schematic demonstration.

The main problem addressed in this study is the design of a model capable of interactively confirming the presence of relevant information within the topic representations of interest, particularly in an online context. To the best of our knowledge, no prior studies have specifically addressed this setting leveraging a confidence score as a form of weak supervision to enhance an otherwise purely unsupervised model. We propose a novel online machine learning model that integrates a linear reward function linked to the confirmation confidence (e.g., *nThumbsup*) with the variational-Bayesian lower bounds of probabilistic topic models. The only modification applied to the topic model involves the incorporation of a variational distribution for document-topic assignments through a bilinear function that connects variational posterior parameters and confirmation parameters. Our experimental results, obtained using real-world data, highlight the following advantages of the entire framework:

1. The linear reward function for confirmation will eventually reveal topics linked to the events of focus.
2. Empirically, simple baseline models, LDA and MUG, when augmented with the confirmation model, yield improved se-mantic interpretations. The results imply that the framework can be extended to other topic models.
3. Our method can improve downstream tasks for event detection and data augmentations. Additional experiments demonstrate improvements in data labeling for classification and in measuring similarity/dissimilarity.
4. A real-world case study demonstrates a potential application of our model for augmenting Waze alerts using the *nThumbUps* feature.

6.1 Literature Review

To the best of our knowledge, this presents the first attempt at the interactive learning of a confirmation model and a topic model within an online setting, aiming to address the problem we have presented. Our work is related to probabilistic topic models with variational inferences and their applications in event detection and information augmentation. One of the earliest probabilistic topic models is the well-known Latent Dirichlet Allocation introduced by Blei, Ng, and Jordan (2003). The parameter estimation of LDA is challenging as the posterior distribution is computationally intractable (Sontag & Roy, 2011). Variational inference, where the posteriors are assumed to be multinomial and Dirichlet (Blei et al., 2003), has been one approach to address this issue. Hoffman et al. present the online variational inference of LDA (Hoffman et al., 2010). However, LDA and many of its extended versions struggle with learning topics from documents in the format of short texts, such as social media data from Facebook or Twitter/X, which typically contain only one or two topics, rather than a mixture of all topics (Vayansky and Kumar, 2020). One preliminary model for addressing the sparsity of topics in short text is the mixture of unigrams model (Qiang et al., 2020). This idea has been further extended by Lin et al. to a dual-sparse topic model (Lin et al., 2014). Other works address the sparsity issue by expanding the dimensionality, such as a bi-term model (Yan et al., 2013). We refer to two surveys on probabilistic topics addressing the issue of sparsity in short texts (Qiang et al., 2020; Vayansky and Kumar, 2020). While various modeling methods exist, limited attention has been paid to incorporating data and weak supervision into variational inference in topic models, which is the focus of what we propose.

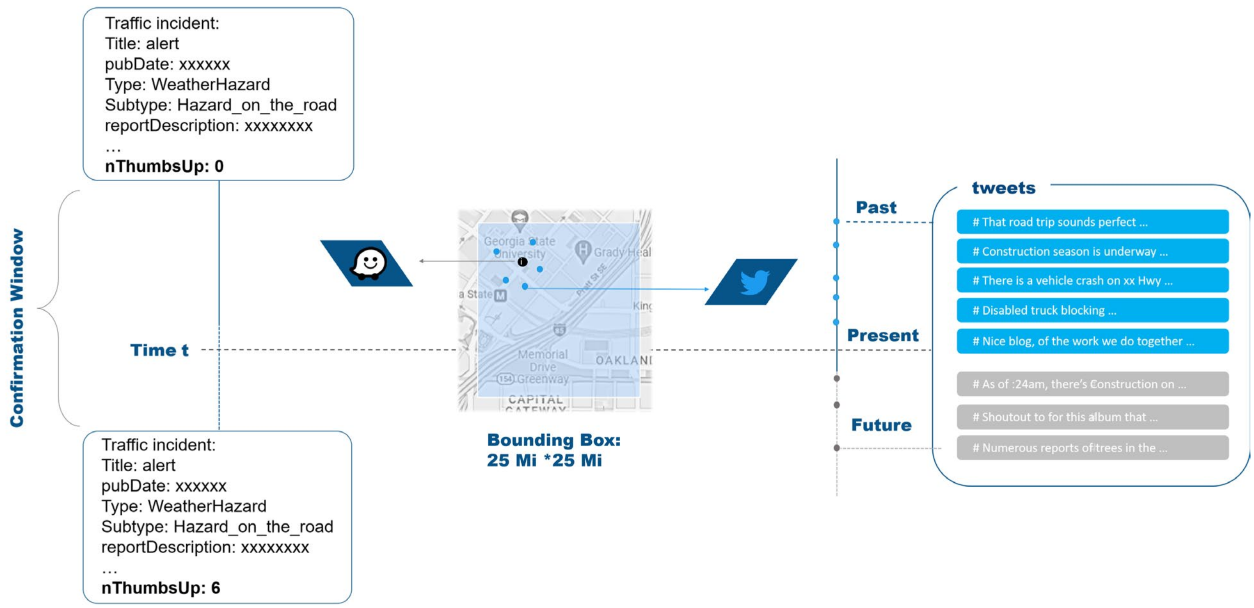


Figure 10. Illustration. A real-world scenario. The left-hand side is a Waze alert, which has *nThumbsUp* being constantly collected from online users to support the reliability of events. The right-hand side is the pool of tweets being posted within the same spatial-temporal window. In the real world, if an event has influenced a relatively small online community, the same event may have already influenced larger online communities, such as Twitter/X.

Effective emergency response relies on robust information management practices including the application of topic models to collect, process, and analyze real-time data streams from diverse sources. These analyses are vital for detecting events, anomalies, emerging patterns, and ensuring the seamless operation of these systems. Topic models are particularly effective in event detection and information augmentation. In event detection, the objective is to measure the uniqueness of identified patterns in the data. In information augmentation, the goal is to match similar content from different information sources to pro-vide comprehensive context and background for users.

Technically, both tasks often employ topic models for either (1) data labeling (Fan et al., 2020; Salley et al., 2022; Tien et al., 2016; Wang et al., 2012) or (2) similarity measurement between topic distributions of two documents, (Yin and Allan, 2009; Zhang et al., 2017). The former serve as sources of data annotations for supervised/semi-supervised (Salley et al., 2022; Tien et al., 2016; Wang et al., 2012) classifiers, while the latter aims to retrieve lower dimensional representations for clustering analysis. For example, by computing the cosine similarity between the target and a potential candidate, we can assess the relevancy (Wang and Taylor, 2018; Yin and Allan, 2009) or uniqueness (Zhang et al., 2017) of the candidate compared to the central topic.

6.2 Definitions and Preliminaries

Problem Statements. Let $t \in [T]$ represent discrete timestamps where $[T] = \{1, 2, . . .\}$. We assume that at each $t \in [T]$, we are provided with a set of tweets $Dt \subset$ D, where each $d \in D^t$ is a vector in the bag-of-words format. In addition, we are given a binary label $y^t \in \{0, 1\}$, but there is no guarantee that y^t will arrive at timestamp t. It is common for yt to have a certain delay. A value of $y^t = 1$ indicates the presence of some d of interest, while $y^t = 0$ indicates the absence of such d of interest. A real-world example of y^t is the $nThumbsUp$ illustrated in 1. We should manually tune a threshold τ such that if $nThumbsUp > \tau$ we set $y^t = 1$ and 0 otherwise.

We present the formulation of topic modeling in terms of dimensional reduction. That is:

Problem 1. Given D^t, learn/update the parameters of a topic model H: D $\rightarrow$ R^K, where K is the number of topics and Σ_k H (D^t) =1.

Therefore, the output is a K dimensional multinomial distribution over K topics of D^t. Meanwhile, the confirmation model can be defined as follows:

Problem 2. Given D^t, learn/update the parameters of a reward function f: D × Y → R so that it will gain more reward if $y^t = 1$ prior to the reveal of the ground truth y.

It is important to note that Sub-problem 1 is an unsupervised learning problem and Sub-problem 2 is supervised. Although, these two problems may initially appear independent of each other, the primary contribution of this study is to propose a model capable of effectively addressing both problems interactively in an online machine-learning setting. We will demonstrate that by simultaneously solving Sub-problem 2, which involves confirmation, along-side Sub-problem 1, the model can yield a more focused (or skewed) distribution of topics of interest.

Table 12 provides a summary of all the notations used in this study and their corresponding descriptions.

Table 12. Notations and Descriptions.

Notations	Descriptions
D^t	The set of documents (e.g. tweets) collected at time t
y^t	The binary label indicating the existence of targeted tweets at t
T	The maximum amount of timestamps
K	The Number of the topics.
W	The set of words.
β_k	The random latent variables of word-topic distribution.
γ_d	The random latent variables of document-topics distribution.
π	The parameters of the linear reward function for confirmation.
ϕ_{dwk}	The parameter of the variational posterior $q(Z_{dw} = k)$.
γ_d	The parameters of the variational posterior $q(\xi_d)$
λ_k	The parameters of the variational posterior $q(\beta_k)$
$\hat{\theta}^t$	All parameters of the topic model depends on t
$\bar{\theta}$	All parameters of the topic model independent of t

Variational Lower Bounds. Let Z^t represent all the latent variables of the topic models parameterized by $\theta^t = (\theta^{-}, \theta^{\wedge t}) \in \Theta$, the variational lower bound of a topic model with given D^t is:

$$\log p\,(D^t\,|\theta^t) \geq \mathrm{E}_q\,(Z^t)\,\{\log p\,(D^t, Z^t|\theta^t)\} - \mathrm{E}_q\,(Z^t)\,\{\log p\,(Z^t)\} := \ell\,(\theta^t\,|D^t) \qquad \text{Equation 6.1}$$

where $p\,(Z^t)$ is the prior distribution and $q(Zt)$ is a variational posterior distribution we select. For generality, θ^t has parts θ^{-} independent of t.

6.3 Proposed Method

Online Machine Learning. Our method further requires that the topic models must have a K-multinomial variational posterior. For instance, in LDA, there is a variational posterior of the per-word $w \in W$ topic assignment $q\,(Z_{d,} = k) = \phi_{dwk}$ [5]. In MUG, there is a variational posterior of per-document topic assignment $(Zd = k) = \phi dk$ [10]. The key idea of the interactive model is to introduce a linear reward function parameterized by $\pi \in \mathrm{R}K$ such that $\square k\ \pi k = 1$. We assume LDA is the topic model for the rest of the section. The linear reward function is defined as:

$$f(\pi|\phi^t, D^t, y^t) = \sum_{d\in D^t}\sum_{w}\sum_{k}\pi_k\phi^t_{dwk}y^t \qquad \text{Equation 6.2}$$

The linear rewards' parameters π interact with the LDA model via the posterior distribution such that:

$$q(Z^t_{dw} = k) = \phi^t_{dwk}\pi_k/(\sum_{k'} q(Z^t_{dw} = k')) \qquad \text{Equation 6.3}$$

Taking a closer look at the linear function, when $y^t = 1$, which happens when we have enough confidence to confirm the presence of relevant $d \in D^t$, the multinomial distribution $\phi^t{}_{dwk}$

temporarily assumed given to us must be distinct from $y^t = 0$. For maximizing $\Sigma f(\pi^t|\phi^t, , y^t)$ from t=1 to T' for every $T' \in [T\ ]$, the problem is equivalent to the well-known online learning problem: learn from K experts' advice (Abernethy et al., 2014). π is encouraged to assign more weights to the topics that frequently appear when $y^t = 1$.

The main idea is to maximize both the linear reward function and the variational lower bound of the topic models simultaneously. In the context of online convex optimization, we set $(\theta^t, \pi) = \ell(\theta^t \mid D^t) + f(\pi \mid \phi^t, D^t, y^t)$ and the benchmark of success is:

$$\mathcal{R}(\pi^t, \theta^t) := \max_{\pi, \theta^t} \sum_t g(\theta^t, \pi) - g(\theta^t, \pi^t) \qquad \text{Equation 6.4}$$

subject to constraints: $\Sigma_k \phi^t_{dwk} = 1$ and $\Sigma_k \pi^t_k = 1$. Recall that $\phi^t \in \Theta$ and R is the regret function.

An Online Algorithm. To solve the online problem, we derive an online algorithm with a theoretically guaranteed bound on the regret function R. Due to page limit, we only present the iterative updates on πt and ϕ^t and refer to (Hoffman et al., 2020) for the rest of the other parameters' updates.

For LDA, there are two additional prior distributions:

$\beta_k \sim$ Dirichlet (η) Equation 6.5

where $\beta k \in \mathrm{R}^{|W|}$ is a distribution over words for each topic. Besides, for each document d

$\xi_d \sim$ Dirichlet (α) Equation 6.6

where $\xi_d \in \mathrm{R}^K$ is a distribution over topics. $\eta, \alpha \in \mathrm{R}$ are two scalar hyperparameters of the model, which define the Dirichlet priors to be symmetric.

The variational inference of LDA also requires the variational posteriors of (β_k) and (ξ_d). They are:

$q(\beta_k)$ = Dirichlet (λ_k) Equation 6.7

and

$q(\xi_d)$ = Dirichlet (γ_d) Equation 6.8

where $\lambda_k \in R^{|W|}$ and $\gamma_d \in R^K$ are vector parameters. In terms of θ^t, $\lambda = \theta$ since it does not depend on time t, and $(\gamma^t, \phi^t) = \hat{\theta}^t$ since the document $d \in D^t$ depends on time t.

Algorithm 1

procedure OnlineModel($K, \alpha, \eta, \rho^0, \bar{\theta}^0, \pi^0$)
  for $t \in T$ **do**
    y^t has been revealed at time t.
    while Change of $|\frac{1}{K}\sum_k \gamma^t_{dk}| \leq 0.00001$ **do**
      $\phi^t_{dwk} \propto \exp(\mathbb{E}_q\{\beta_{kw}\} + \mathbb{E}_q\{\xi_{dk}\} - \frac{y^t}{\pi^t_k})$
      Update $\hat{\theta}^t$ based on [5].
    end while
    Solve $\bar{\theta}^t_\Delta, \pi^t_\Delta = \arg\max \quad g(\theta^t, \pi) - ||\pi||^2$
    s.t. $\sum_k \pi_k = 1$
    $\bar{\theta}^t = (1-\rho^t)\bar{\theta}^{t-1} + \rho^t \bar{\theta}^t_\Delta$
    $\bar{\pi}^t = (1-\rho^t)\bar{\pi}^{t-1} + \rho^t \bar{\pi}^t_\Delta$
    Update ρ^{t+1}
  end for
end procedure

Assuming a given learning rate ρ^t, our online algorithm relies on incremental updates from each sub problem's optimum at t. For each $t \in [T]$, the sub problem is:

$$\bar{\theta}^t_\Delta, \pi^t_\Delta = \arg\max \quad g(\theta^t, \pi) - ||\pi||^2 \quad \text{s.t.} \sum_k \pi_k = 1$$ Equation 6.9

where an ℓ-2 regularizer, $||\pi||^2$, is added to the objective. The incremental updates for these two time-independent variables are:

$$\bar{\theta}^t = (1-\rho^t)\bar{\theta}^{t-1} + \rho^t \bar{\theta}^t_\Delta \qquad \text{Equation 6.10}$$

$$\bar{\pi}^t = (1-\rho^t)\bar{\pi}^{t-1} + \rho^t \bar{\pi}^t_\Delta \qquad \text{Equation 6.11}$$

In addition, solving $\phi^t{}_d$ is different from the above online problem as the solution in nature depends on time t. The sub problem for ϕ^t given π^t and θ^t attained from the updates is:

$$\begin{aligned} \max \quad & g(\theta^t, \pi^t) \\ \text{s.t.} \quad & \sum_k \phi^t_{dwk} = 1, \ \forall d \in D^i, \ \forall w \in W \end{aligned} \qquad \text{Equation 6.12}$$

The above sub problem has a closed-form solution as well:

$$\phi^t_{dwk} \propto \exp(\mathbb{E}_q\{\beta_{kw}\} + \mathbb{E}_q\{\xi_{dk}\} - \frac{y^t}{\pi^t_k}) \qquad \text{Equation 6.13}$$

This equation implies that if $y^t = 1$ and the topic weight for confirmation $\pi^t{}_k$ is small, $\phi^t{}_{dwk}$ tends to be zero. If $y^t = 0$, we recover the same update as in (Blei et al., 2003; Hoffman et al., 2010). Algorithm 1 describes all computations for each $t \in [T]$. Overall, we repeat the computations of the two time-dependent parameters $\phi^t{}_d$ and γ^t until the convergence of γ^t is satisfied.

6.4 Experiments

To verify the effectiveness of our augmented model, we experimented with real-world data. We considered two standard probabilistic models, LDA and MUG, due to their efficient variational inference (Blei et al., 2003; Hoffman et al., 2010). Importantly, our method is compatible with any probabilistic topic model featuring a multinomial per document-topic variational posterior distribution, a structure found in many existing models (Lin et al., 2014; Yan et al., 2013; Yin and Wang, 2014). Future research may explore integrating our approach with other topic models. We employed our model for two key downstream tasks in information science: data labeling and similarity measurement. The first task involves obtaining interpretable topic representations and assessing clustering correlation with human-generated labels. The second task leverages the representational space to reveal semantic content similarities and dissimilarities between documents.

Data and Ground Truth Labels

Hurricane-related Tweets. We will conduct experiments on a real-world Twitter/X dataset during Southern US hurricanes to evaluate disaster information augmentation in real-world applications (Salley et al., 2022; Tien et al., 2016). The dataset includes manually generated labels for five different classes. The data set consists of geotagged tweets collected from three hurricanes that occurred in Florida in 2020: Hurricane Eta (31/10/2020-14/11/2020), Hurricane Isaias (31/7/2020-4/8/2020), and Hurricane Sally (14/9/2020-28/9/2020). The entire data set consists of 10,210 tweets, which are evenly distributed over the first four categories of events as below:

- **(0)** Broadcast/News - Includes tweets related to news, government updates, alerts, and official sources information.
- **(1)** Power - Includes tweets related to power outages, power lines/systems, lights, Wi-Fi, Internet connectivity, etc.
- **(2)** Traffic Incident - Includes tweets related to car crashes, road congestion, evacuations, traffic updates and incidents.
- **(3)** Forecast/Weather - Includes tweets related to weather conditions, forecasts, rainfall, flooding, etc.
- **(4)** Miscellaneous - Includes tweets that do not fit into other categories or are unrelated to the disaster.

Waze. A Waze alert is in a standardized schematic of *Type* plus *Subtype* and *Description*. An example of a traffic alert is provided:

$alert: Traffic\ Accident, Minor\ Accident\ on I-75, Rear-end$

The content of these alerts tends to be similar due to the limited format and the specific categories used to describe the incidents. This categorization enables rapid identification of the alert's nature, including incident type (e.g., accident) and severity (e.g., minor).

Metrics

Perplexity. We use perplexity on out-of-sample data as a model fit measure (Hoffman et al., 2010). Perplexity is defined as the geometric mean of the inverse marginal likelihood of each word in the tweet set.

Topic Coherence. The two downstream tasks necessitate that topic representations are interpretable for readers. Topic Coherence (TC) measures the degree of semantic similarity among high-scoring words (top 15 in our case) (Röder et al., 2015). We employ the "Umass" version of TC (Mimno et al., 2011), which calculates the word-wise score function based on the document co-occurrence of the two words. The overall score is obtained by summing the score of every word-word pair and taking the average among all topics.

Adjusted Mutual Information. Normalized Mutual Information measures the agreement between two clusters and quantifies the similarity between two cluster assignments (Vinh et al., 2009). In our experiment, we have labeled tweets in 5 categories, denoted as $C \in [K]$. To match the number of classes, we set $K = 5$ resulting in an assignment score into 5 clusters. For each tweet $d \in D$, we assign it to the cluster with the highest score denoted as $C' \in [K]$. The mutual information is measured as:

$$MI(C, C') = \sum_{i=1}^{K} \sum_{i=1}^{K} \frac{|C_i \cap C'_j|}{|D|} \log_2 \frac{|D||C_i \cap C'_j|}{|C_i||C'_j|}$$

Normalized Mutual Information (NMI) is a normalization of mutual information, which scales the score between 0 and 1. A higher NMI score indicates a higher level of agreement between the two clusters. The Adjusted Mutual Information (AMI) is an extension of NMI that takes into account the size of clusters, making the score independent of size K.

Recall. The recall score of a binary classification model is computed as follows:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

Experiment Designs

We conducted our experiments in an online machine-learning set-ting using real-world data but with simulated streams for event types of Traffic Incident and Forecast Weather. In the simulation, at each time step t, a batch of tweets and a binary label y^t are sampled. The batch size is uniformly distributed between 10 and 15. If $y^t = 1$, relevant tweets related to the targeted event type were included in the batch, ranging from 5 to 1, while the remaining tweets were randomly selected from the Miscellaneous class. The label y^t was only revealed to the model after t. Additionally, to test the robustness of the model, the accuracy of y^t could be compromised to some extent. The overall expectation of the model augmented with the confirmation we proposed is that it will gradually outperform the baseline model in all evaluation metrics regardless of the perturbation in y^t and $D^{t.}$

Hyper-parameters. In our online machine learning framework we consider several hyper-parameters that impact the performance of the model: (1) K the number of topics, (2) ρ^t , the step size of updating parameters at each step, (3) N , the number of runs of each experiment, (4) eta, the hyperparameter of the prior β, and (5) α, hyperparameter of the prior ξ. Throughout the experiments, we take $\rho^t = (t + 2)^{-0.7}$ for $t \in [T]$ and $\alpha, \eta = 1/K$. We enumerate K from the set of

$\{5, 12, 15, 21\}$. Let N = 30. All experimental results are aggregated from the N samples. For y^t, we test a label accuracy of $\{70\%, 80\%, 90\%\}$. The experiments are performed on a computer with AMD Ryzen 7 5700G 3.80GHz CPU, 16GB memory, and NVIDIA Ge-Force RTX 3060 graphics.

Labeling Data. A core task in employing topic modeling in information studies is to label data based on the clusters inferred by the models (Salley et al., 2022; Tien et al., 2016). We adhered to the standard practice of randomly reserving 10% of the data as a test set for each run. We then evaluated the AMI between the ground truth labels and the cluster assignments obtained from the topic model, considering a total of K = 5.

Matching Similar Tweets. For each Waze alert, one of the experiments in this study is to match its content with tweets found in a nearby spatial-temporal window. We, therefore, treat a Waze alert as another tweet and compute the cosine similarity scores of every other tweet within the specified window.

For each y^t = 1 revealed afterward, we retrieve the top 5 tweets with the highest similarity scores, considering them as the predicted relevant tweets. We assess the success of this matching process by reporting the recall. Note that we do not compare the models based on precision. This is because the baseline methods employed in this study are inherently unsupervised, meaning they do not utilize y^t. Consequently, the unsupervised topic model consistently output matched tweets, regardless of y^t, which results in low precision.

Baseline. To demonstrate the effectiveness of the augmented models, we also include a baseline model with:

$\pi_k = 1 / K$

The parameters remain constant throughout the online experiment. We use Online Mixture of Unigrams (OMUG) and Online Latent Dirichlet Allocation (OLDA) to denote the baseline models.

6.5 Results

Convergence and Fit. The perplexity of all models is presented in Figure 11. Each sub-figure represents instances of LDA and MUG, along with their respective 95% confidence intervals. The results indicate that the convergence criteria are satisfied. The inclusion of the augmented linear reward function and the new posterior does not impede the training of the topic models. All models exhibit a good fit when measured on the held-out evaluation dataset. In general, when the value of K is large, we can anticipate that the model will require more computations to converge.

For $K = 21$, the LDA model augmented with the component we proposed takes 60 timestamps, denoted as t, to reach a similar perplexity level as a baseline LDA model using variational inference with 40 timestamps. However, for MUG model, when $K = 21$, the augmented version never reaches the same level of perplexity as the baseline within 100 iterations. In addition, LDA exhibits slower convergence when the accuracy of y^t goes down. In contrast, MUG demonstrates robustness under the uncertainty of y^t.

It is important to note that the convergence on the held-out dataset does not necessarily prevent overfitting of the models. This consideration holds significance when analyzing the other experimental results presented below.

Interpretability of Topics. Our second remark is that our method significantly improves the semantic interpretability of topics in terms of topic coherence. As shown in Figure 12, the augmented LDA significantly outperforms the baseline LDA after 50-60 iterations. Moreover, the advantage of our augmentation method becomes more prominent as K increases. When $K = 15$ and $K = 21$, the baseline LDA tends to be overfitted after 50 iterations as evidenced by a continuous decrease in TC. The augmented LDA does not have this issue under all settings.

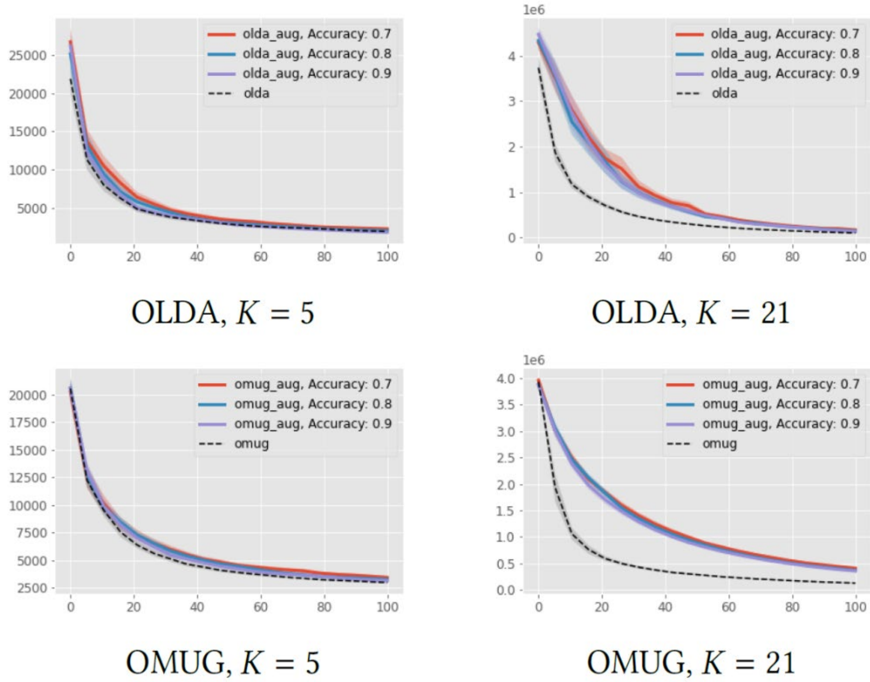


Figure 11. Graphs. The Perplexity of Models in t.

With regard to MUG, it is worth noting that a baseline MUG already outperforms LDA, with the optimal value ranging from -9 to −9.5. This observation aligns with previous studies on topic models for short texts (Qiang et al., 2020). The one-text-one-topic assumption of short text is more reasonable than a mixture of topics, especially for tweets (Yan et al., 2013). Nevertheless, our method results in faster convergence in terms of TC, usually within 10 iterations when K is large. For small K, the increase is marginal. The overfitting problem arises again as there is a drop of TC in augmented MUG $K = 21$ after 80 iterations.

In summary, our method transforms LDA, which was not initially considered suitable for short texts, into a competitive model compared to MUG.

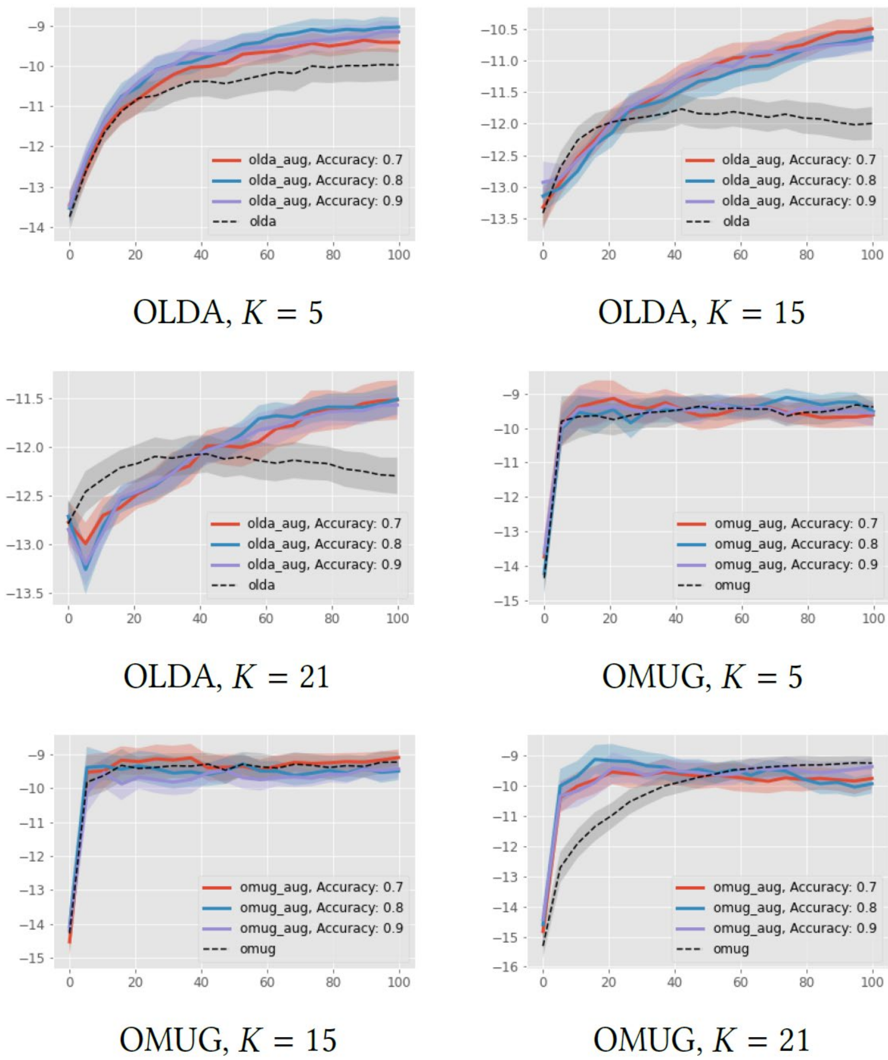
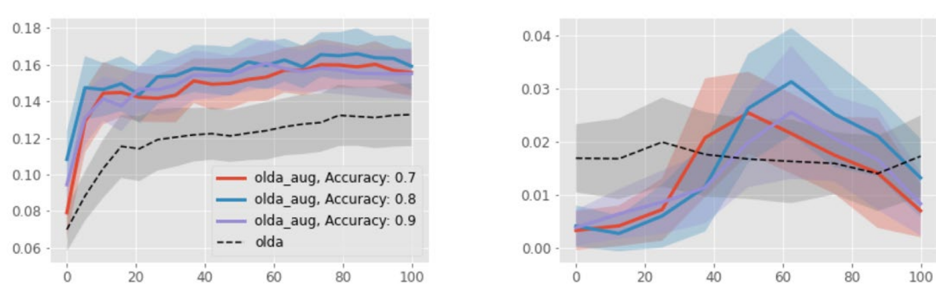


Figure 12. Graphs. The Topic Coherence of Models in t. The Y-axis is the Perplexity. X-axis is the number of timestamps t.

Labeling Data for Classifications. To evaluate the effectiveness of the proposed method on labeling data, we computed the AMI between ground truth labels and cluster assignments of the held-out data set for K = 5. Figure 13 presents the results of the AMI for all models. It is evident that augmented LDA consistently outperforms the baselines. Augmented MUG, however, peaks at 60-80 iterations, then dramatically decreases and becomes worse than the baseline model. We identify the drop as another evidence of the overfitting of MUG. The results suggest to use of LDA for data labeling, supported by empirical evidence of superiority, and is consistent with (Salley et al., 2022; Tien et al., 2016).

Interestingly, our method under the worst y, a 70% of accuracy, has the optimal AMI in both LDA and MUG. We investigated this observation in-depth and provided explanations for it. First, note that the augmented LDA with 80% accuracy is the worst case, which confirms that a higher accuracy does improve the model in terms of AMI. Both 80% and 90% suffer from overfitting as indicated by the nearly identical drop in MUG. The presence of 30% of noisy tweets accidentally expands the training data set, which mitigates the over-fitting issue and contributes to an LDA and a delay of drop in MUG. The analysis suggests that our results can be sensitive to data and hyper-parameters, such as the incremental update rate ρ^t. We will discuss it in the limitations section.

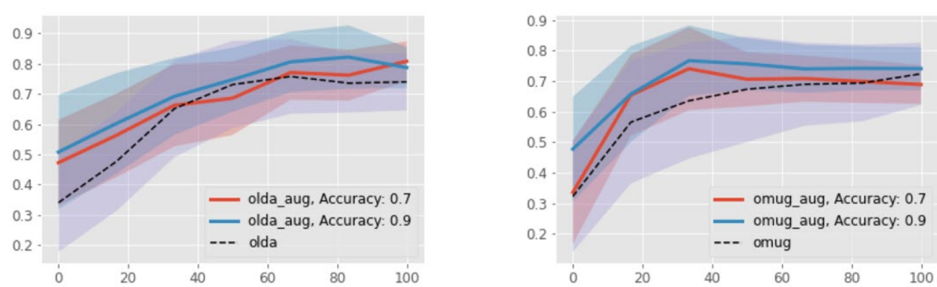
Similarity Measure. Figure 14 shows the recall of matching similar tweets with a single Waze alert. For better visualization, we only present two cases of accuracy, 90% and 70%. The case of 80% is omitted since it closely resembles the other cases in LDA and it overlaps with the baseline in MUG. The experiment of matching similar tweets is greatly dependent on chosen hyperparameters, particularly on K and label accuracy. As seen, the augmented LDA outperforms the baseline LDA across all uncertainties when K = 12.



OLDA, K = 5 OMUG, K = 5

Figure 13. Graphs. The Adjusted Mutual Information of Models in t.

The recall is compromised by the lower accuracy of y^t and it is below the baseline for a 70% of accuracy. Other K either have slightly inferior performance or marginal merit compared to the baseline. The augmented MUG dominates the baseline under 90% accuracy of y^t and it is marginally enhanced compared to the baseline. Overall, the experimental findings highlight the influence of hyperparameters on the performance of matching similar tweets and reinforce the advantages of our augmented models over baselines.



OLDA, K = 12 OMUG, K = 21

Figure 14. Graphs. Recall of Matching Similar Tweets in Models at t.

6.6 Case Study

To demonstrate the effectiveness and potential application of the proposed method in the real world, we also performed a case study using real-world data streams. We utilized an unlabeled data set consisting of Waze data and tweets from 10/18/2021 to 10/31/2021 within 115 bounding boxes covering Georgia, USA. The entire dataset consists of about 12,000 events and over 600,000 geo-tagged tweets with lengths greater than 5. For each Waze incident, we recorded the debut time, the current time of API calls, and the number of thumbs-ups from other users. We consider tweets as potential matches if their debut time is within a 30-minute time window and in the same bounding box. To label the matches, we set y^t = 1 if the number of thumbs-ups is above 3, and st = 0 otherwise. For each Waze alert with y^t = 1, we apply an LDA model, K = 21, augmented with the confirmation part to output the top four similar tweets. We randomly sampled five instances from the results, and they are presented in Table 13.

Table 13 shows successful cases of implementing the augmented LDA with real-world data. Nevertheless, it is important to note that the bounding boxes used in the filtering process do not completely eliminate other incidents that may be spatially/temporally close to the target incident.

6.7 Conclusion

Our research highlights the critical role of online probabilistic topic models in enabling the real-time analysis of complex data streams. These models empower infrastructure operators and decision-makers to extract actionable insights, detect anomalies, make precise predictions, optimize resource allocation, engage users, and leverage social feedback. However, traditional probabilistic topic models face challenges when applied to user-generated content, which is often sparse and dynamic. We propose a novel framework that integrates a linear reward function, guided by the confidence levels associated with relevant content, into the variational lower bound of the likelihood of Bayesian topic models.

This innovative approach enhances topic retrieval, improving interpretability and generalizability across various topic models. Our empirical experiments and case study, conducted on real-world datasets, showcase the effectiveness of our learning algorithm in enhancing topic models through two important downstream tasks: information augmentation and event detection. It significantly improves topic interpretability, data labeling precision, and similarity metric refinement, making it a valuable tool for processing and analyzing real-time data streams.

Table 13. Relevant Tweets of Bounding Box 1, Accident and Jam. Blue texts indicate a truth positive, Red texts describe incidents that occurred in nearby bounding boxes.

Waze Events	Relevant Tweets
alert ACCIDENT ACCIDENT MINOR I-575 S	Accident. left two lanes blocked in Cherokee on I 575 SB after Sixes Rd/Exit 11 ATLTraffic @MARTAservice there are multiple lights out on the walkway from Buckhead station towards Tower Place. He shoots under par and places at the national and regional tournaments ...
alert ACCIDENT ACCIDENT MAJOR I-85 S	Accident. left two lanes blocked in Hapeville on I-85 SB near Sylvan Rd/Central Ave/Exit 75, stop and go traffic b… Disabled vehicle, shoulder blocked in CollegePark on I-285 WB near I-85 (SW ATL)/Exit 61 (WB), stop and go traffic… Join the Lane Construction team! See our latest job opening here: https://xxxxxx Construction Craft-Workers ...
alert ACCIDENT ACCIDENT MAJOR I-285 E	Accident. right three lanes blocked in SandySprings on I-285 EB at Roswell Rd (GA-9)/Exit 25 (EB), stopped traffic… Accident. right shoulder blocked in Dekalb on I-285 SB at Lawrenceville Hwy (US-29)/Exit 38, stop and go traffic b… Cozy Cabin Overlooks the Suwannee River: A North Florida couple builds their family-friendly forever home along the… ...
alert traffic JAM JAM HEAVY TRAFFIC I-85 N	McDonough traffic might be the reason I actually atl cntrl delt Accident, right lane blocked in Norcross on I-85 SB at Jimmy Carter Blvd (GA-140)/Exit 99, stop and go traffic bac… Accident in Brevard on US 1 Both NB/SB between CO Hwy 502/Coquina Rd/Barnes Blvd and Eyster Blvd traffic ...
alert traffic JAM JAM HEAVY TRAFFIC I-75 S	Accident, left lane blocked in Snellville on Stone Mtn Fwy (Hwy 78) WB at Scenic Hwy (GA-124) ATLTraffic https://t.co/bABElTW6T2 Accident. two left lanes blocked. in Polk on I-4 EB before US 27 (MM 55) traffic Closed due to accident in Osceola on Poinciana Blvd SB south of US 17-92/Orange Blossom Trail and before Reaves Rd… Accident. left three lanes blocked in Jonesboro on I-75 SB after Tara Blvd/Old Dixie Hwy/Exit 235 ATLTraffic ...

The effectiveness of our online confirmation-augmented probabilistic topic modeling approach for processing real-time data streams contributes to informed decision-making, efficient infrastructure management, and proactive engagement with evolving conditions. Our approach shows potential for unlocking new insights and addressing integration challenges.

6.8 Limitations and Future Work

One of the limitations of this model is the relatively limited exploration of hyperparameters. We acknowledge that the choice of hyperparameters, including the update weight $\rho^t = (t + 2)^{-0.7}$, is somewhat arbitrary and not optimized for all models. This update rate is not suitable for MUG as it resulted in overfitting within just 50 iterations. The update rate should decay at a much faster rate compared to the one used for LDA. Additionally, other initialization hyper-parameters, such as ρ_0, η, and α, can all impact the performance of the two downstream tasks. In general, these hyperparameters should be fine-tuned for each specific model augmented with our proposed component. However, we adopted the default settings from (Blei et al., 2003; Hoffmanet al., 2010). In addition, our exploration was limited to a small parameter set, with K values chosen from 5, 12, 15, 21, due to the considerable time required to complete each experiment (around 1.5 hours in average). We recognize the need for a more thorough investigation of hyperparameters as a future endeavor. This would entail refining, parallelizing, and addressing numerical issues in the current code implementation.

Another perturbation that our method is sensitive to is the accuracy of y. Initially, we anticipated that our method could still perform accurately with a 55% accuracy. However, the results indicate that an accuracy less than 70% will significantly compromise the performance of our method, see Figure 14 for the case of OMUG. We suspect that the bi-linear function may not be robust enough to handle noise in y^t. Exploring alternative stable functions as potential augmentation components is an avenue for future research.

7 CONCLUSION

The accuracy and timeliness of crisis communications depends on the scale of participation and coverage of the user population. Community awareness applications (e.g., Waze) allow individuals to raise concerns, gather insight, and report targeted information. Social media platforms (e.g., Twitter/X) cover large numbers of active users and can be an additional source to identify and describe ongoing incidents. Matching multiple streams of crisis communication (e.g., Waze alerts and tweets) can be challenging considering the heterogeneous formats and the online nature of the data, however, harnessing the powers of both can be advantageous and complement one another. To address this gap in knowledge and functionality, we: 1) conducted interviews with state DOTs to gain insight on current systems and potential barriers to implementing social media, 2) designed and tested a competency matrix on potential current applications and solutions, and 3) designed our own Machine Learning based applications for data integration and augmentation of community data with social media for DOT event detection.

Our project highlighted challenges faced by state DOTs in integrating social media into their systems. Despite the potential advantages, they face challenges due to the complexity of processing large volumes of data and extracting pertinent information. We then investigated existing event detection software that could integrate social media data streams. We evaluated 12 event detection software applications using a weighted competency matrix developed for this project. This assessment revealed a lack of adopted software within DOTs for processing Twitter/X and other social media data. Additionally, our examination of disaster response strategies in published reports on completed and in-progress transportation research projects showed promise for but a general lack of application of AI/ML models for rapid crisis event detection and providing essential contextual information for emergency responders. To address this gap, we developed and applied solutions involved training models using combined data from social media platforms (e.g., Twitter/X) and community-driven applications (e.g., Waze). The first model showcased the effectiveness of crisis event detection when social media data is fused with community-driven applications. This first model notably implemented the Wells-Dubois protocol, contributing to bias minimization in event detection outcomes. The second model introduced an enhanced approach for real-time online probabilistic topic modeling. By integrating a linear reward function into the variational lower bound of Bayesian topic models, when applied to a real world disaster scenario, this model exhibited improved effectiveness in topic retrieval, interpretability, and generalizability, thereby enhancing information augmentation and event detection. This project lays the foundation for an integrative approach to augmenting community-driven applications with social media data, providing transportation agencies with enhanced crisis event detection capabilities in terms of the speed of and confidence in alerts generated by community-driven applications such as Waze. The findings presented here offer practical insights and innovative solutions, contributing to a more resilient and responsive future for transportation crisis event management.

8 APPENDICES

8.1 Appendix A. Interview Protocol

Thank you for meeting with us. We are conducting a research project for the Georgia Department of Transportation to explore technological solutions that enable GDOT to take advantage of social media interactions (e.g., Twitter and Waze) to improve the speed and accuracy of identifying potential issues needing to be addressed on the highway system, while also providing the ability to interact with the public through social media about identified incidents or issues. We understand you adopted *(insert system adopted)* and would like to discuss that experience with you.

First, we have a few questions about when you were making the purchasing decision to implement *(insert system adopted)*:

- What aspects of your operations were you hoping to improve that drove the need to look for a solution?
- In a sentence, what was your goal or objective?
- What specific functionalities were you interested in and why?
- What social media features of *(insert system adopted)* interested you, if any?
- Did your DOT have security/privacy concerns with such a solution and, if so, how were they addressed?

Second, we have a few questions on the implementation process for *(insert system adopted)*:

- How long did it take from the decision to purchase until the vendor began installation? Was it more or less than expected?

- How long did it take to install *(insert system adopted)* into your systems and processes? Was it more or less than expected?
- Can you describe any technology challenges associated with the implementation? Were there more or less challenges than expected?
- What internal DOT resources were required to support the implementation of *(insert system adopted)*? Was it more or less than expected?
- What internal DOT resources are required to run *(insert system adopted)* now that it is up and running? Is it more or less than expected?
- How were and are DOT employees trained on the utilization of *(insert system adopted)*? Is it more or less training requirements than expected?
- What metrics to you use to evaluate the success of *(insert system adopted)* in your DOT?
- How would you describe the vendor's role in supporting implementation and on-going operations? Does it meet your expectations?

Finally, we have a few high level questions about how it is working out:

How does *(insert system adopted)* help you to identify incidents or issues on the state highway system?

How does identifying issues with *(insert system adopted)* differ from how you did it before?

All-in-all, after implementing *(insert system adopted)* do you think the process of identifying issues on the state highway system:

- Is faster with *(insert system adopted)* (why or why not)?
- Is more accurate with *(insert system adopted)* (why or why not)?
- Better engages the public (why or why not)?

- The functionalities you were originally interested in work as expected (why or why not)?
- The overall goal or objective of implementing *(insert system adopted)* was met (why or why not)?

Did you consider any of these other solutions? If you considered any of them, are you able to share why you did or did not purchase it?

Do you know of other DOTs that have implemented *(insert system adopted)* and would you be willing to put us in touch with them to discuss their experience adopting and implementing *(insert system adopted)*?

Is there anything we should have asked in the interview or that you think GDOT should know?

9 REFERENCES

Abel, F., Hauff, C., Houben, G. J., Stronkman, R., & Tao, K. (2012). Semantics+ filtering+ search= twitcident: Exploring information in social web streams. In *Proceedings of the 23rd ACM conference on Hypertext and social media* (pp. 285-294).

Abernethy, J., Lee, C., Sinha, A., & Tewari, A. (2014). Online linear optimization via smoothing. In *Learning Theory* (pp. 807–823). PMLR.

Adeola, F. O., & Picou, J. S. (2017). Hurricane Katrina-linked environmental injustice: Race, class, and place differentials in attitudes. *Disasters, 41*(2), 228–257. https://doi.org/10.1111/disa.12204

Aggarwal, C. (2011). *An introduction to social network data analytics*. Springer.

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., Albahri, O. S., Alamoodi, A. H., Bai, J., Salhi, A., Santamaría, J., Ouyang, C., Gupta, A., Gu, Y., & Deveci, M. (2023). A Systematic Review of Trustworthy and Explainable Artificial Intelligence in Healthcare: Assessment of Quality, Bias Risk, and Data Fusion. *Information Fusion, 96*, 156–191. https://doi.org/10.1016/j.inffus.2023.03.008

Alexander, D. E. (2014). Social Media in Disaster Risk Reduction and Crisis Management. *Science and Engineering Ethics, 20*(3), 717–733. https://doi.org/10.1007/s11948-013-9502-z

Amin-Naseri, M., Chakraborty, P., Sharma, A., Gilbert, S. B., & Hong, M. (2018). Evaluating the Reliability, Coverage, and Added Value of Crowdsourced Traffic Incident Reports from Waze. *Transportation Research Record, 2672*(43), 34–43. https://doi.org/10.1177/0361198118790619

Athuraliya, C. D., Gunasekara, M. K. H., Perera, S., & Suhothayan, S. (2015). Real-time natural Language Processing for Crowdsourced Road Traffic Alerts. In *Proc., 2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)* (pp. 58–62). Colombo, Sri Lanka: IEEE. http://doi.org/10.1109/ICTER.2015.7377667

Awasthi, P., Beutel, A., Kleindessner, M., Morgenstern, J., & Wang, X. (2021). Evaluating Fairness of Machine Learning Models Under Uncertain and Incomplete Information. In *Proc., FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 206-214). New York, NY: Association for Computing Machinery. http://doi.org/10.1145/3442188.3445884

Barich, F. (2019). *Social Media Guidebook for Emergency Management*. Transportation Review Board. In progress research at the time.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bodenreider, C., Wright, L., Barr, O., Xu, K., & Wilson, S. (2019). Assessment of Social, Economic, and Geographic Vulnerability Pre- and Post-Hurricane Harvey in Houston, Texas. *Environmental Justice, 12*(4), 182–193. https://doi.org/10.1089/env.2019.0001

Boddy, S., Rezgui, Y., Cooper, G., & Wetherill, M. (2010). Activity Awareness as an Enabler for Communication and Network Building in Construction Design Teams. *J. Comput. Civ. Eng., 24*(5), 430–440. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000046

Brunila, M., Zhao, R., Mircea, A., Lumley, S., & Sieber, R. (2021). Bridging the Gap Between Supervised Classification and Unsupervised Topic Modelling for Social-Media Assisted Crisis Management. *ArXiv*, 1–17. https://doi.org/10.48550/arXiv.2103.11835

Chair, S., Charrad, M., & Saoud, N. B. B. (2019). Towards A Social Media-Based Framework for Disaster Communication. *Procedia Computer Science, 164*, 271–278. https://doi.org/10.1016/j.procs.2019.12.183

Chaney, A. J. B., & Blei, D. M. (2012). Visualizing topic models. In *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*.

Chen, X., & Wang, T. (2017). Combining Active Learning and Semi-Supervised Learning by Using Selective Label Spreading. In *Proc., 2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 850–857). New Orleans, LA: IEEE. http://doi.org/10.1109/ICDMW.2017.154

Chen, Z., Zhang, J. M., Hort, M., Sarro, F., & Harman, M. (2022). Fairness Testing: A Comprehensive Survey and Analysis of Trends. *ArXiv*, 1-27. https://doi.org/10.48550/arXiv.2207.10223

Chowdhury, S., & Zhu, J. (2023). Investigation of critical factors for future-proofed transportation infrastructure planning using topic modeling and association rule mining. *Journal of Computing in Civil Engineering, 37*(1), 04022044. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001059

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corporate reputation review, 10(3)*, 163-176.

CRED. (2009). EM-DAT Glossary. *EM-DAT The International Disaster Database Centre for Research on the Epidemiology of Disasters – CRED*. Accessed March 9, 2023. https://www.emdat.be/Glossary

D'Andrea, E., Ducange, P., Lazzerini, B., & Marcelloni, F. (2015). Real-time detection of traffic from twitter stream analysis. *IEEE transactions on intelligent transportation systems*, *16*(4), 2269-2283.

de Albuquerque, J. P., Herfort, B., Brenning, A., & Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science, 29*(4), 667-689.

Dhakal, S., Zhang, L., & Lv, X. (2021). Understanding Infrastructure Resilience, Social Equity, and Their Interrelationships: Exploratory Study Using Social Media Data in Hurricane Michael. *Nat. Hazards Rev., 22*(4): 04021045. https://doi.org/10.1061/(ASCE)NH.1527-6996.000051

Dixon, S. J. (2023). Distribution of X (formerly Twitter) users worldwide as of January 2023, by gender. *Statista*. Accessed October 6, 2023. https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/

Du, J., Zhao, D., Issa, R. R. A., & Singh, N. (2020). BIM for Improved Project Communication Networks: Empirical Evidence from Email Logs. *J. Comput. Civ. Eng., 34*(5): 04020027. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000912

Endsley, M. R. (2016). *Designing for situation awareness: An approach to user-centered design*. CRC press.

Fan, C., Wu, F., & Mostafavi, A. (2020). A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. *IEEE Access, 8*, 10478–10490.

FastText. (2023). "Fasttext." *PyPI*. Accessed April 12, 2023. https://pypi.org/project/fasttext/

FEMA. (2021). FEMA Issues Public Notice for 21 Georgia Counties. *FEMA*. Accessed March 9, 2023. https://www.fema.gov/press-release/20210211/fema-issues-public-notice-21-georgia-counties

FEMA. (2023a). Declared Disasters. *FEMA*. Accessed March 9, 2023. https://www.fema.gov/disaster/declarations

FEMA. (2023b). Glossary. *FEMA*. Accessed March 9, 2023. https://training.fema.gov/programs/emischool/el361toolkit/glossary.htm

FEMA. (2023c). How a Disaster Gets Declared. *FEMA*. Accessed March 9, 2023. https://www.fema.gov/disasters/how-declared

FEMA. (2023d). 2022–2026 FEMA Strategic Plan: Building the FEMA our Nation Needs and Deserves. *FEMA*. Accessed March 9, 2023. https://www.fema.gov/about/strategic-plan

Garrison, K., Tate, D. M., & Bailey, J. W. (2019). Factors limiting the speed of software acquisition. *Institute for Defense Analyses*.

Gillborn, D., Warmington, P., & Demack, S. (2018). Quantcrit: Education, Policy, 'Big Data' and Principles for a Critical Race Theory of Statistics. *Race Ethnicity and Education, 21*(2), 158–179. https://doi.org/10.1080/13613324.2017.1377417

Griego, A. L., Flores, A. B., Collins, T. W., & Grineski, S. E. (2020). Social Vulnerability, Disaster Assistance, and Recovery: A Population-Based Study of Hurricane Harvey in Greater Hoffman, M., Bach, F., & Blei, D. (2010). Online learning for latent dirichlet allocation. *Advances in Neural Information Processing Systems, 23*.

Houston, Texas. *Int. J. Disaster Risk Reduct, 51*, 1–11. https://doi.org/10.1016/j.ijdrr.2020.101766.

Gu, Y., Qian, Z., & Chen, F. (2016). From Twitter to Detector: Real-time Traffic Incident Detection using Social Media Data. *Transportation Research Part C: Emerging Technologies, 67*, 321–342. https://doi.org/10.1016/j.trc.2016.02.011.

Guth, D. (2019). Emergency Management Agencies: Pilot for a Crisis Communication Analysis Assessment Test. Transportation Review Board. In progress research at the time.

Halse, S. E., Grace, R., Kropczynski, J., & Tapia, A. (2019). Simulating Real-Time Twitter Data from Historical Datasets. In *Proc., 16th ISCRAM Conference*, edited by Z. Franco, J. J. González, and J. H. Canós (pp. 780–787). València, Spain: ISCRAM.

Hamideh, S., & Rongerude, J. (2018). Social Vulnerability and Participation in Disaster Recovery Decisions: Public Housing in Galveston after Hurricane Ike. *Natural Hazards, 93*, 1629–1648. https://doi.org/10.1007/s11069-018-3371-3.

Hao, H., & Wang, Y. (2021). Assessing Disaster Impact in Real Time: Data-Driven System Integrating Humans, Hazards, and the Built Environment. *J. Comput. Civ. Eng., 35*(5): 04021010. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000970.

Henchey, M. J., Batta, R., Blatt, A., Flanigan, M., & Majka, K. (2014). A Simulation Approach to Study Emergency Response. *Journal of Simulation, 8*(2), 115–128. https://doi.org/10.1057/jos.2013.20.

Hooshmand, A., & Sharma, R. (2019). Energy Predictive Models with Limited Data Using Transfer Learning. In *Proc., e-Energy '19: Proceedings of the Tenth ACM International Conference on Future Energy Systems* (pp. 1–5). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3307772.3328284.

Hort, M., Chen, Z., Zhang, J. M., Sarro, F., & Harman, M. (2022). Bias Mitigation for Machine Learning Classifiers: A Comprehensive Survey. *ArXiv*, 1-27. https://doi.org/10.48550/arXiv.2207.07068.

Howitt, A. M. (2009). *Improving Emergency Preparedness and Crisis Management Capabilities in Transportation* (Report No. HVDR 19-8).

Hua, S. Y. (2022). Procurement maturity and IT failures in the public sector. *Transforming Government: People, Process and Policy*.

Imran, M., Castillo, C., Lucas, J., Meier, P., & Rogstadius, J. (2014). Coordinating human and machine intelligence to classify microblog communications in crises. In *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management* (pp. 712–721).

Jennex, M. E. (2012). Social media–viable for crisis response?: Experience from the great San Diego/Southwest blackout. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM), 4*(2), 53-67.

Jin, Y., Pang, A., & Cameron, G. T. (2007). Integrated crisis mapping: Toward a publics-based, emotion-driven conceptualization in crisis communication. *Sphera Publica*, 81-95.

Jodoin, P., & Austrich, J. (2020). Next-Generation TIM: Integrating Technology, Data, and Training (No. FHWA-20-CAI-011). *Federal Highway Administration (US)*.

Johnson, J. (2018). CHAOS report: decision latency theory: it is all about the interval. *Lulu. com*.

Kappelman, L. A., McKeeman, R., & Zhang, L. (2006). Early warning signs of IT project failure: The dominant dozen. *Information Systems Management, 23*(4), 31-36.

Khalid, A. (2019). Twitter removes precise geo-tagging option from tweets. https://www.engadget.com/2019-06-19-twitter-removes-precise-geo-tagging.html.

Kim, J. H., & Sutley, E. J. (2021). Implementation of Social Equity Metrics in an Engineering-Based Framework for Distributing Disaster Resources. *Int. J. Disaster Risk Reduct, 64*, 1–11. https://doi.org/10.1016/j.ijdrr.2021.102485.

Kumar, D., & Ukkusuri, S. V. (2020). Enhancing Demographic Coverage of Hurricane Evacuation Behavior Modeling using Social Media. *J. Comput. Sci., 45*, 1–15. https://doi.org/10.1016/j.jocs.2020.101184.

Kwan, M. P., & Lee, J. (2005). Emergency response after 9/11: the potential of real-time 3D GIS for quick emergency response in micro-spatial environments. *Computers, environment and urban systems, 29(2)*, 93-113.

Lieberknecht, K., Zoll, D., Jiao, J., & Castles, K. (2021). Hurricane Harvey: Equal Opportunity Storm or Disparate Disaster? *Local Environment, 26*(2), 216–238. https://doi.org/10.1080/13549839.2021.1886063.

Lin, T., Tian, W., Mei, Q., & Cheng, H. (2014). The dual-sparse topic model: mining focused topics and focused terms in short text. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 539–550).

Lindsay, B. R. (2011). Social Media and Disasters: Current Uses, Future Options, and Policy Considerations. *Congressional Research Service*.

Liu, B. (2011). Opinion mining and sentiment analysis. In *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data* (pp. 459-526).

Loos, S., Lallemant, D., Baker, J., McCaughey, J., Yun, S. H., Budhathoki, N., Khan, F., & Singh, R. (2020). G-DIF: A geospatial data integration framework to rapidly estimate post-earthquake damage. *Earthquake Spectra, 36*(4), 1695–1718. https://doi.org/10.1177/8755293020926190.

Loos, S., Levitt, J., Tomozawa, K., Baker, J., & Lallemant, D. (2022). Efficacy of Damage Data Integration: A Comparative Analysis of four major earthquakes. *Nat. Hazards Rev., 23*(4): 04022026. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000581.

Lovari, A., & Bowen, S. A. (2019). Social Media in Disaster Communication: A Case Study of Strategies, Barriers, and Ethical Implications. *Journal of Public Affairs, 20*(1), 1–9. https://doi.org/10.1002/pa.1967.

Maurer, S. M. 2020. "Evolving Approaches to Place Tagging in Social Media." *J. Urban Plan. Dev.*, 146(3): 04020023. https://doi.org/10.1061/(ASCE)UP.1943-5444.00005.

Mhatre, M., D. Phondekar, P. Kadam, A. Chawathe, and K. Ghag. 2017. "Dimensionality Reduction for Sentiment Analysis using Pre-processing Techniques." In *Proc., IEEE Int. Conf. on Computing Methodologies and Communication (ICCMC)*, 16–21. Erode, India: IEEE. http://doi.org/10.1109/ICCMC.2017.8282676.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Empirical Methods in Natural Language Processing* (pp. 262–272).

Mohammadi, N., and J.E. Taylor. 2017. "Urban Energy Flux: Spatiotemporal Fluctuations of Building Energy Consumption and Human Mobility-Driven Prediction." *Applied Energy*, 195, 810–818. https://doi.org/10.1016/j.apenergy.2017.03.044.

Monroe-White, T., and J. Lecy, 2022. "The Wells-Du Bois Protocol for Machine Learning Bias: Building Critical Quantitative Foundations for Third Sector Scholarship." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 34(1), 170–184. https://doi.org/10.1007/s11266-022-00479-2.

Morstatter, F., Kumar, S., Liu, H., & Maciejewski, R. (2013). Understanding Twitter data with TweetXplorer. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1482–1485).

Nargesian, F., A. Asudeh, and H.V. Jagadish. 2022. "Responsible DataIntegration: Next-Generation Challenges." In *Proc., SIGMOD '22: Proceedings of the 2022 International Conference on Management of Data,* 1–7. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3514221.3522567.

National Advisory Council. 2020. *National Advisory Council Report to the FEMA Administrator*. FEMA.

Nejat, A., L. Solitare, E. Pettitt, and H. Mohsenian-Rad. 2022. "Equitable Community Resilience: The Case of Winter Storm Uri in Texas." *Int. J. Disaster Risk Reduct*, 77, 1–12. https://doi.org/10.1016/j.ijdrr.2022.103070.

Neyshabur, B., H. Sedghi, and C. Zhang. 2020. "What is Being Transferred in Transfer Learning?" In *Proc., 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 1–12. Vancouver, Canada: Neural Information Processing Systems.

Nguyen, P. H., Lines, B. C., Maali, O., Sullivan, K. T., Hurtado, K., & Savicky, J. (2022). Current state of practice in the procurement of information technology solutions: Content analysis of software requests for proposals. *International Journal of Procurement Management, 15*(3), 406-423.

Nigam, K., McCallum, A. K., Thrun, S., & Mitchell, T. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39, 103–134.

NOAA US Department of Commerce. 2020a. "Tropical Storm Zeta Affects North and Central Georgia." Accessed May 5, 2023. https://www.weather.gov/ffc/TropicalStormZeta.

NOAA US Department of Commerce. 2020b. "Hurricane Zeta 2020." *NOAA's National Weather Service.* Accessed May 5, 2023. https://www.weather.gov/tae/zeta2020#:~:text=George%20Island%2031%20MPH%200139,28%2029.79N%2F84.88W%20.

NOAA US Department of Commerce. 2022. "Hurricane Ian's Path of Destruction." *National Environmental Satellite Data and Information Service*. Accessed March 8, 2023. https://www.nesdis.noaa.gov/news/hurricane-ians-path-of-destruction.

Ogie, R. I., & Verstaevel, N. (2020). Disaster Informatics: An Overview. *Progress in Disaster Science, 7*, 1-11. https://doi.org/10.1016/j.pdisas.2020.100111.

Ozbay, K., Yang, H., Xie, K., Li, J., & Satish, V. U. (2016). Modeling disaster operations from an interdisciplinary perspective in the New York–New Jersey area. *Region II University Transportation Research Center Research Project Report*.

Peirelinck, T., Kazmi, H., Mbuwir, B. V., Hermans, C., Spiessens, F., Suykens, J., & Deconinck, G. (2022). Transfer Learning in Demand Response: A Review of Algorithms for Data-Efficient Modelling and Control. *Energy and AI, 7*, 1–14. https://doi.org/10.1016/j.egyai.2021.100126.

Peng, Y., Zhang, Y., Tang, Y., & Li, S. (2011). An Incident Information Management Framework Based on Data Integration, Data Mining, and Multi-Criteria Decision Making. *Decision Support Systems, 51*(2), 316–327. https://doi.org/10.1016/j.dss.2010.11.025.

Pinto, G., Wang, Z., Roy, A., Hong, T., & Capozzoli, A. (2022). Transfer Learning for Smart Buildings: A Critical Review of Algorithms, Applications, and Future Perspectives. *Advances in Applied Energy, 5*, 1–23. https://doi.org/10.1016/j.adapen.2022.100084.

Pollard, J. A., Spencer, T., & Jude, S. (2018). Big Data Approaches for Coastal Flood Risk Assessment and Emergency Response. *WIREs Climate Change, 9*(5), 1–14. https://doi.org/10.1002/wcc.543.

Pollock, N., & Williams, R. (2007). Technology choice and its performance: Towards a sociology of software package procurement. *Information and Organization, 17(3)*, 131-161.

Psyllidis, A., Bozzon, A., Bocconi, S., & Titos Bolivar, C. (2015). A Platform for Urban Analytics and Semantic Data Integration in City Planning. In *Proc., CAAD Futures 2015: Computer-Aided Architectural Design Futures. The Next City - New Technologies and the Future of the Built Environment* (pp. 21–36). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-47386-3_2.

Purohit, H., Kanagasabai, R., & Deshpande, N. (2019). Towards Next Generation Knowledge Graphs for Disaster Management. In *Proc., 2019 IEEE 13th International Conference on Semantic Computing (ICSC)* (pp. 474-477). Newport Beach, CA, USA: IEEE. http://doi.org/10.1109/ICOSC.2019.8665638.

Qi, S., Brown, D. G., Tian, Q., Jiang, L., Zhao, T., & Bergen, K. M. (2009). Inundation Extent and Flood Frequency Mapping Using Landsat Imagery and Digital Elevation Models. *GIScience & Remote Sensing, 46*(1), 101–127. https://doi.org/10.2747/1548-1603.46.1.101.

Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X. (2020). Short text topic modeling techniques, applications, and performance: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(3), 1427–1445.

Ribeiro, M., Grolinger, K., ElYamany, H. F., Higashino, W. A., & Capretz, M. A. M. (2018). Transfer Learning with Seasonal and Trend Adjustment for Cross-Building Energy Forecasting. *Energy and Buildings, 165*, 352–363. https://doi.org/10.1016/j.enbuild.2018.01.034.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399–408).

Rossi, C., Acerbo, F. S., Ylinen, K., Juga, I., Nurmi, P., Bosca, A., Tarasconi, F., Cristoforetti, M., & Alikadic, A. (2018). Early detection and information extraction for weather-induced floods using social media streams. *International Journal of Disaster Risk Reduction, 30*, 145-157.

Roy, K. C., Hasan, S., Mohaimin, A. M., & Cebrian, M. (2020). Understanding the Efficiency of Social Media Based Crisis Communication During Hurricane Sandy. *International Journal of Information Management, 52*, 1–13. https://doi.org/10.1016/j.ijinfomgt.2019.102060.

Sack, R., Foreman, C., Forni, S., Glynn, R., Lehrer, A., Linthicum, A., & Perruzzi, A. (2019). Social media practices in traffic safety (Report No. DOT HS 812 673). *Washington, DC: National Highway Traffic Safety Administration*.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010, April). Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 851-860).

Salas, A., Georgakis, P., & Petalas, Y. (2017). Incident Detection Using Data from Social Media. In *Proceedings of IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 751–755). Yokohama, Japan: IEEE. http://doi.org/10.1109/ITSC.2017.8317967.

Salley, C., Mohammadi, N., & Taylor, J. E. (2022). Semi-supervised Machine Learning Framework for Fusing Georeferenced Data from Social Media and Community-Driven Applications. In *Proceedings of Computing in Civil Engineering 2021* (pp. 114–122). https://doi.org/10.1061/9780784483893.015.

Samuels, R., Mohammadi, N., & Taylor, J. E. (2020). *Social Media-Informed Urban Crisis Detection* (Report No. FHWA-GA-20-1834) https://rosap.ntl.bts.gov/view/dot/53905.

Samuels, R., & Taylor, J. E. (2020). Deepening the Divide: Crises Disproportionately Silence Vulnerable Populations on Social Media. *Journal of Management in Engineering, 36*(6): 04020083. https://doi.org/10.1061/(ASCE)ME.1943-5479.000084.

Schmidtlein, M. C., Finch, C., & Cutter, S. L. (2008). Disaster Declarations and Major Hazard Occurrences in the United States. *The Professional Geographer, 60*(1), 1–14. https://doi.org/10.1080/00330120701715143.

Schubert, E., Weiler, M., & Kriegel, H.-P. (2014). SigniTrend.

Senarath, Y., Nannapaneni, S., Purohit, H., & Dubey, A. (2020, December). Emergency incident detection from crowdsourced Waze data using Bayesian information fusion. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 187-194). *IEEE*.

Sheffi, Y., & Goentzel, J. (2015). *Big Data During Crisis: Lessons from Hurricane Irene* (Report No. MITR24-10). *New England University Transportation Center*.

Sidauruk, A., & Ikmah. (2018). Congestion Correlation and Classification from Twitter and Waze Map Using Artificial Neural Network. In *Proceedings of 3rd International Conference on Information Technology, Information Systems and Electrical Engineering* (pp. 224–229). Yogyakarta, Indonesia: IEEE. http://doi.org/10.1109/ICITISEE.2018.8720995.

Smith, C. (2023). Waze Statistics and User Count (2023). *DMR*. Accessed April 12, 2023. https://expandedramblings.com/index.php/waze-statistics-facts/.

Sontag, D., & Roy, D. (2011). Complexity of inference in latent dirichlet allocation. In *Advances in Neural Information Processing Systems 24*.

Sujon, M., & Dai, F. (2021). Social Media Mining for Understanding Traffic Safety Culture in Washington State Using Twitter Data. *Journal of Computing in Civil Engineering, 35*(1): 04020059. https://doi.org/10.1061/(ASCE)CP.1943-5487.00009.

Sullivan, L., & Schwartz, E. (2018). FEMA Report Acknowledges Failures in Puerto Rico Disaster Response. *NPR*. Accessed April 16, 2023. https://www.npr.org/2018/07/13/628861808/fema-report-acknowledges-failures-in-puerto-rico-disaster-response.

Tankovska, H. (2022). Twitter: Number of Users Worldwide 2019-2024. *Statista*. Accessed April 12, 2023. https://www.statista.com/statistics/303681/twitter-users-worldwide/.

The Eastern Transportation Coalition. (2017). Going My WAZE to Closing Real Time Data Gaps: Crowdsourcing Summit Summary Brief, 1-19. Philadelphia, PA: ETC.

Thales. (n.d.). *What is a Software Maintenance Process? 4 Types of Software Maintenance*. https://cpl.thalesgroup.com/software-monetization/four-types-of-software-maintenance#:~:text=Software%20maintenance%20is%20the%20process,to%20boost%20performance%2C%20and%20more.

Tien, I., Musaev, A., Benas, D., Ghadi, A., Goodman, S., & Pu, C. (2016). Detection of Damage and Failure Events of Critical Public Infrastructure Using Social Sensor Big Data. In *Proceedings of International Conference on Internet of Things and Big Data (IoTBD 2016)* (pp. 435–440). Rome, Italy: IoTBD.

UNDRR. (n.d.). Terminology. *UNDRR*. Accessed March 9, 2023. https://www.undrr.org/terminology.

US National Institute of Health. (2022). *Competencies Proficiency Scale*. Retrieved from https://hr.nih.gov/working-nih/competencies/competencies-proficiency-scale

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, 94, 101582.

Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A., & Anderson, K. (2011). Natural Language Processing to the Rescue? Extracting 'Situational Awareness' Tweets during Mass Emergency. In *Proceedings of International AAAI Conference on Web and Social Media, 5*(1), 385–392. Barcelona, Spain: AAAI. https://doi.org/10.1609/icwsm.v5i1.14119.

Vinh, N. X., Epps, J., & Bailey, J. (2009). Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In *26th Annual International Conference on Machine Learning* (pp. 1073–1080).

Wang, J., Xia, L., Hu, X., & Xiao, Y. (2019). Abnormal event detection with semi-supervised sparse topic model. *Neural Computing and Applications, 31*, 1607–1617.

Wang, Q., & Taylor, J. E. (2014). Quantifying human mobility perturbation and resilience in Hurricane Sandy. *PLoS One, 9*(11), e112608.

Wang, Q., & Taylor, J. E. (2016). Process Map for Urban-Human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms. *Journal of Computing in Civil Engineering, 30*(2), 04015004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000469.

Wang, Y., & Taylor, J. E. (2018). Coupling Sentiment and Human Mobility in Natural Disasters: A Twitter-based Study of the 2014 South Napa Earthquake. *Natural Hazards, 92*, 907–925.

Wang, Y., & Taylor, J. E. (2018). Urban crisis detection technique: A spatial and data-driven approach based on latent Dirichlet allocation (LDA) topic modeling. *In Construction Research Congress 2018* (pp. 250–259).

Wang, Y., Wang, Q., & Taylor, J. E. (2017). Aggregated responses of human mobility to severe winter storms: An empirical study. *PLoS One, 12*(12), e0188734.

Wang, Y., & Taylor, J. E. (2019). DUET: Data-Driven Approach Based on Latent Dirichlet Allocation Topic Modeling. *Journal of Computing in Civil Engineering, 33*(3), 04019023. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000819.

Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic Crime Prediction Using Events Extracted from Twitter Posts. *SBP 12*, 231–238.

Waze. (2017). Waze Traffic-data Specification Document, 1–21.

Wright, B. (2011). Race, Place, and the Environment in the Aftermath of Katrina. *Anthropology of Work Review, 32*(1), 4–8. https://doi.org/10.1111/j.1548-1417.2011.01052.x.

Wu, D., & Cui, Y. (2018). Disaster Early Warning and Damage Assessment Analysis using Social Media Data and Geo-location Information. *Decision Support Systems, 111*, 48–59. https://doi.org/10.1016/j.dss.2018.04.005.

Xie, J., Salley, C. J., Mohammadi, N., & Taylor, J. E. (2023, November). Online Confirmation-Augmented Probabilistic Topic Modeling in Cyber-Physical Social Infrastructure Systems. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 390-397).

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 1445–1456).

Yang, Y., Zhang, C., Fan, C., Mostafavi, A., & Hu, X. (2020). Towards Fairness-aware Disaster Informatics: An Interdisciplinary Perspective. *IEEE Access, 8*, 1–15. http://doi.org/10.1109/ACCESS.2020.3035714.

Yi, X., & Allan, J. (2009). A comparative study of utilizing topic models for information retrieval. In *Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009* (pp. 29–41).

Yin, J., Karimi, S., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2015). Using social media to enhance emergency situation awareness. In *IJCAI International Joint Conference on Artificial Intelligence* (pp. 4234–4239).

Yin, J., & Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 233–242).

Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using Social Media to Enhance Emergency Situation Awareness. *IEEE Intelligent Systems, 27*(6), 52–59.

Yu, J. J. Q. (2019). Online Traffic Speed Estimation for Urban Road Networks with Few Data: A Transfer Learning Approach. In *Proceedings of 2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 4024–4029). Auckland, New Zealand: IEEE. http://doi.org/10.1109/ITSC.2019.8917502.

Yuan, F., & Liu, R. (2018). Integration of social media and Unmanned Aerial Vehicles (UAVs) for Rapid Damage Assessment in Hurricane Matthew. *Construction Research Congress 2018*, 513–523.

Yuan, F., & Liu, R. (2020). Mining social media data for rapid damage assessment during Hurricane Matthew: Feasibility Study. *Journal of Computing in Civil Engineering, 34*(3), 05020001. https://doi.org/10.1061/(ASCE)CP.1943-5487.000087.

Yuen, D. A., Kadlec, B. J., Bollig, E. F., Dzwinel, W., Garbow, Z. A., & Silva, C. R. (2005). Clustering and Visualization of Earthquake Data in a Grid Environment. *Visual Geosciences, 10*(1), 1–12.

Zhang, C., Liu, L., Lei, D., Yuan, Q., Zhuang, H., Hanratty, T., & Han, J. (2017). TrioVecEvent: Embedding-based online local event detection in geo-tagged tweet streams. In *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 595–604).

Zhang, C., Zhou, G., Yuan, Q., Zhuang, H., Zheng, Y., Kaplan, L., Wang, S., & Han, J. (2016). GeoBurst: Real-time local event detection in geo-tagged tweet streams. In *SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Zhang, M., & Pan, Q. (2019). Utilize Crowd-Sourced Data and Machine Learning Technology to Enhance Planning for Transportation Resilience to Flooding. *Transportation Review Board*. In progress research at the time.

Zhou, D., Bousquet, O., Navin Lal, T., Weston, J., & Schölkopf, B. (2004). Learning With Local and Global Consistency. In *Proceedings of Advances in Neural Information Processing Systems, 16*, 1–8.

Zou, L., Lam, N. S., Cai, H., & Qiang, Y. (2018). Mining Twitter Data for Improved Understanding of Disaster Resilience. *Annals of the American Association of Geographers, 108*(5), 1422–1441. https://doi.org/10.1080/24694452.2017.1421897.

Zubiaga, A., Voss, A., Procter, R., Liakata, M., Wang, B., & Tsakalidis, A. (2017). Towards Real-Time, Country-Level Location Classification of Worldwide Tweets. *IEEE Transactions on Knowledge and Data Engineering, 29*(9), 2053–2066. http://doi.org/10.1109/TKDE.2017.2698463.