NOVA IMS
Information Management School

MGI

Master Degree Program in
Information Management

The Impact of Natural Language Processing in Disaster Management

A Systematic Literature Review

Matilde Martins Lencastre Godinho

Master Thesis

presented as partial requirement for obtaining the Master Degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

The Impact of Natural Language Processing in Disaster Management

A Systematic Literature Review

by

Matilde Martins Lencastre Godinho

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Information Systems and Technologies Management

Supervised by

Pedro da Costa Brito Cabral, PhD

February, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 29th of February 2024

DEDICATION

I dedicate this thesis to my dad, who I know would be extremely proud to see me graduating.

ABSTRACT

In an era where undeniably large amounts of data are produced at an unprecedented rate by a broad variety of sources, artificial intelligence and, more specifically, natural language processing present interesting solutions to today's complex challenges. When properly implemented, natural language processing may play an essential role in all phases of the disaster management cycle by delivering timely and reliable information, optimizing resources and logistics, and assisting in reducing the effect that both natural and man-made disasters have on society. This paper presents a systematic literature review that aims to (1) identify the range of natural language processing methods used in disaster management, (2) explore the main challenges and opportunities in disaster management when using natural language processing, and (3) determine how this field impacts the four phases of disaster as well as the nature of such impact. The data was gathered using Scopus database and, with the help of Zotero citation manager tool, filtered by full-text peer-reviewed quartile 1 journal articles written in English and published between 2014 and 2023. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines, this review is based on 107 articles selected out of 548 search results obtained in February 2024. The findings indicate that social media and sentiment analysis have had an exponential growth in applications and importance within the field of disaster management. We analyze the current academic literature on the link between natural language processing and disaster management, identifying research gaps and potential for future research.

KEYWORDS

Natural language processing; Disaster management; Impact; Systematic literature review; PRISMA

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
BTM	Biterm Topic Model
CA	Conversational Agent
CNN	Convolutional Neural Network
DL	Deep Learning
DM	Disaster Management
GIS	Geographic Information Systems
GPT	Generative Pre-trained Transformer
GSDMM	Gibbs Sampling Algorithm for a Dirichlet Mixture Model
IM	Information Management
KNN	K-Nearest Neighbors
LDA	Latent Dirichlet Allocation
LSTM	Long Short-Term Memory
ML	Machine Learning
MT	Machine Translation
NB	Naïve Bayes
NER	Named Entity Recognition
NGO	Non-Governmental Organization
NLG	Natural Language Generation
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NLU	Natural Language Understanding
POS	Part-of-Speech

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Q1	Quartile 1
QA	Question Answering
RNN	Recurrent Neural Network
RS	Remote Sensing
SJR	Scimago Journal Rank
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
VADER	Valence Aware Dictionary for Sentiment Reasoner

1. INTRODUCTION

Every year hundreds of natural and man-made disasters negatively impact the world. With them, thousands of lives are lost, and many more displaced (Centre for Research on the Epidemiology of Disasters, 2023). Economic devastation and health implications are felt, and extensive environmental damages registered (Akter & Wamba, 2019). Among these, the COVID-19 pandemic has shown how vulnerable the world is to a global catastrophe, proving the extreme importance of having cooperative and preventive measures in place, to reduce the potential risk and impact of hazards turning into disastrous events (IFRC, 2020). On the other hand, whenever a disaster occurs, it is also imperative to assure immediate assistance to all victims as well as a rapid and effective recovery of the affected communities until all systems return to their original or improved state (Wei et al., 2020).

As industry interest, investments, and research on artificial intelligence increase at a fast pace, natural language processing (NLP) is starting to show potential when applied in disaster risk reduction (DRR) and overall humanitarian initiatives (Prior & Williamson, 2023). Correspondingly, a simple search on Scopus database on 17 October 2023 showed that the number of studies mentioning the concepts “natural language processing” and “disaster management” (DM) exponentially increased from 14 in 2012 to 73 in 2018 and to 254 in 2022. Despite the growth on literature surrounding these two topics, there is still a limited number of systematic literature reviews mentioning the terms “natural language processing” and “disaster management” on the title, abstract and/or keywords. Nunavath and Goodwin (2019) applied a systematic literature approach to report on the application of artificial intelligence (AI), machine learning (ML), and deep learning (DL) in disaster management, between 2009 and 2019, without getting into much dept about the NLP subfield. Similarly, literature reviews primarily focused on the impact of AI in disaster management, and which used the term “natural language processing” as a keyword on their search query were identified (Ogie et al., 2018; Stoykova & Shakev, 2023). Cai (2021) also highlighted the potential of NLP to support natural disaster responses while the primarily focus of the study was on natural language processing applied to urban research. More recently, Tounsi and Temimi (2023) published a systematic literature review on the intersection of NLP and DM. However, their study did not investigate the entire spectrum of natural and man-made disasters, only focusing on hydrometeorological hazards (e.g., hurricanes, typhoons, flooding). By filling this study gap, both researchers and practitioners can get a sense of what has been published within these two fields and which areas need further research.

As in any other area, the application of NLP presents both challenges and opportunities when set in the humanitarian context (Rocca et al., 2023). Thus, to identify the impact of such projects, this dissertation aims to answer the following research questions:

RQ1: What are the most common Natural Language Processing applications, pre-processing techniques, and models used in the context of Disaster Management?

The first step into understanding which kind of impact NLP projects have when applied in the humanitarian sector is to detect the distribution of high-quality research within the vast number of existing NLP applications, methods, and techniques. In addition, this exercise supports the identification of knowledge gaps and opportunities for future research.

RQ2: Which challenges derive from applying Natural Language Processing in Disaster Management?

The complexity and fast-growing profile of NLP results in substantial obstacles (Khurana et al., 2023) which, when combined with the equally intricate context of disaster management, may obscure its high potential (Al-Garadi et al., 2022). Thus, it is important to understand the challenges inherent to such initiatives, to ensure that future humanitarian NLP projects have a positive impact.

RQ3: What are the opportunities of using Natural Language Processing in Disaster Management?

To end in a positive note, the last research question aims to identify the benefits of applying NLP in emergency support. Distinguishing successful use cases and what works best in the vulnerable context of disaster management is critical to assure that humanitarian actors take advantage of the best solutions once a disaster strikes (Rocca et al., 2023).

Theoretical in nature, this research makes significant contributions to both theory and practice. For what it concerns theory, the current study fills a gap by integrating interdisciplinary concepts from different research areas including artificial intelligence and disaster management to understand the impact of NLP in a humanitarian disaster environment. As for practice, the elaboration of a systematic review provides a concrete analysis of relevant literature packed with consolidated solutions on how to better apply NLP into current and future humanitarian projects. Thus, this study provides an excellent foundation of concepts and practices to support the conduction and improvement of valuable Natural Language Processing projects which, applied in real-life disasters, may consequently save and protect human lives.

The remaining of this paper is structured as follows: Section 2 provides an overview of the current literature surrounding the concepts of natural language processing and disaster management. In Section 3, using the PRISMA guidelines, we then proceed to lay the foundation for a comprehensive systematic literature review by listing the eligibility criteria and search strategy of this dissertation. Within Section 4, the findings draw attention to social media and, more specifically, to microblogging platforms such as Twitter and Weibo as the most common data sources used in NLP projects applied to DM. In Section 5 the paper includes one table with a summary of NLP methods and pre-processing techniques applied to disaster management and makes reference to another featuring challenges and advantages inherent of using NLP in a humanitarian context. Finally, we conclude by highlighting study limitations and future research recommendations.

2. THEORETICAL BACKGROUND

2.1. NATURAL LANGUAGE PROCESSING

Natural language processing uses techniques from various domains, such as AI, computer science, linguistics, and data science, with the aim of empowering machines in interpreting human language and supporting human-computer interactions (Kusal et al., 2023). NLP can be divided into two components: Natural Language Understanding (NLU) and Natural Language Generation (NLG). While NLU focuses on how machines analyze and extract concepts from text and speech, NLG enables the creation of human language text answers based on a given data input (Khurana et al., 2023). Table 1 compiles the definition of important concepts directly related to NLP for further clarification.

Table 1 - Definition of important concepts surrounding NLP

Concept	Definition
Artificial Intelligence	"(…) a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation (…)" (Kaplan & Haenlein, 2019)
Machine Learning	"Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming." (Hurwitz & Kirsch, 2018)
Deep Learning	"Deep learning is a specific method of machine learning that incorporates neural networks in successive layers in order to learn from data in an iterative manner." (Hurwitz & Kirsch, 2018)

The availability of vast volumes of textual data and the demand for more complex and authentic computer-human communication have led to the exponential increase of NLP in recent years (Patwardhan et al., 2023). Aligned with its rapid growth, NLP has now been applied into a variety of fields (Sharma et al., 2022). Through enabling more effective and efficient communication and information management (IM), NLP has the potential to change numerous industries, including healthcare, education, and customer service (Patwardhan et al., 2023). In Healthcare, natural language processing has been proved to be a reliable solution when it comes to monitoring medical infections (Tvardik et al., 2018). Additionally, several studies have also investigated how data mining and NLP can support doctors convert the extensive healthcare data from Electronic Health Records into useful knowledge (Hossain et al., 2023). In education, NLP has been used to analyze writing skills profiles and identify learning needs of university students (Castillo et al., 2023) as well as to generate automated essays or grade open-ended questions (Kardam et al., 2023). Finally, when it comes to finance and accounting, NLP has been used to automatically detect and classify financial fraud (Chang et al., 2022) as well as to generate an automated bookkeeping system to convert unstructured data into structured accounting entries (Chen et al., 2023).

2.1.1. Natural Language Processing Applications

As seen, NLP is a complex field which deals with the communication between humans and computers. Similar to human-human relationships, human-computer interactions can happen in a variety of forms and embody different purposes. Table 2 features some of the most frequently used NLP applications referenced in the highlighted peer-reviewed papers.

Table 2 - Main NLP applications and linked references

Paper	Dialogue System	Information Extraction	Machine Translation	Question Answering	Sentiment Analysis	Text Summarization
(Khurana et al., 2023)	X	X	X			X
(Kusal et al., 2023)	X		X		X	X
(Liu et al., 2023)	X	X		X	X	X
(Sawicki et al., 2023)	X	X	X	X	X	X
(Sharma et al., 2022)	X	X	X	X		X

2.1.1.1. Dialogue System

Various terminologies are used to describe dialogue systems, including chatbots, conversational agents (CAs), virtual assistants, virtual agents, smart bots, interactive agents, and digital assistants (Kusal et al., 2023). Dialogue systems are currently one of the most popular NLP subjects (Ni et al., 2023), with applications ranging from delivering support to completing specific actions in both businesses and everyday life (Khurana et al., 2023).

CAs communicate with people using natural language (speech and/or text) on both input and output channels (Wołk et al., 2022) and can be divided into two categories: task-oriented dialogue agents and chatbots (Jurafsky & Martin, 2023). As explained by Jurafsky & Martin (2023), task-oriented dialogue agents employ conversations with users to assist them perform tasks, whereas chatbots are developed for extended unstructured conversations, characteristic of human-human contact, primarily for entertainment.

Additionally, there are three major chatbot architectures: rule-based systems, information retrieval systems, and encoder-decoder generators (Jurafsky & Martin, 2023). Rule-based systems are simple to implement and respond intuitively. However, the dialogue flows of these systems are predetermined, which limits their usage to specific scenarios (Ni et al., 2023). On the other hand, information retrieval systems and encoder-decoder generators differ on the process of response generation, either by collecting information to obtain a response from a corpus that is appropriate for the discussion context, or by generating the response using a language model or encoder-decoder, respectively (Jurafsky & Martin, 2023).

Dialogue systems are growing in popularity fueled by the fast advances in AI which translate in an evolving number of applications that now range from food ordering apps to virtual assistants such as Apple's Siri, Microsoft's Cortana, and Amazon's Alexa (Ni et al., 2023).

2.1.1.2. Information Extraction

Information extraction or retrieval consists of automatically identifying and collecting data based on user needs (Jurafsky & Martin, 2023). Accordingly, its main purpose is to extract and convert unstructured text into information that is easier to analyze, search, and visualize (Sharma et al., 2022). The collected information can then be used in different contexts, such as producing a summary, creating databases, identifying keywords, and categorizing text based on predetermined categories (Khurana et al., 2023).

Information retrieval subtasks include named entity recognition (NER), relation extraction, and template filling which can be tackled using techniques based on heuristics or machine learning (Reichenpfader et al., 2023). Similarly, entity linking, the task of connecting entity references in text to their corresponding entities in a knowledge base, is often used to clarify information extracted through automated methods (Shen et al., 2015).

In contrast with the 1990s, during the last decades, "relentless optimization of information retrieval effectiveness has driven web search engines to new quality levels at which (…) web search has become a standard and often preferred source of information finding" (Manning et al., 2008, p. XV). Nowadays, information extraction is applied to a wide range of domains, from healthcare (Ahmad et al., 2023; Reichenpfader et al., 2023) to manufacturing (Kima et al., 2021; Berdyugina & Cavallucci, 2023).

2.1.1.3. Machine Translation

Khurana et al. (2023) define machine translation (MT) as "the task of [automatically] converting the text of one natural language into another language while keeping the sense of the input" (p. 3731).

As most information is predominantly available in English and other languages used in developed countries, machine translation is widely used for information access, helping reducing the digital divide, with Google Translate translating billions of words daily across 100+ languages (Jurafsky & Martin, 2023). According to Jurafsky and Martin (2023), two other common uses of machine translation are assisting human translators and in-the-moment human communication needs.

Even though this technology has been around for centuries, obtaining high quality machine translation still remains a challenge these days (Sawicki et al., 2023). This is because accurate translation requires understanding the precise meanings within a specific context and not solely translating across languages (Sharma et al., 2022). Thus, as highlighted by Khurana et al. (2023), a number of techniques such as position-independent word error rate (Tillmann et al., 1997), generation string accuracy (Bangalore et al., 2000), multi-reference word error rate (Nießen et al., 2000), BLEU score (Papineni et al., 2002), and NIST score (Doddington, 2002) have been recently published for automatically assessing the quality of machine translation.

2.1.1.4. Question Answering

Question answering (QA) systems are intended to meet human information demands during scenarios such as speaking with a virtual assistant, engaging with a search engine, or querying a database (Jurafsky & Martin, 2023).

In recent years, the need to quickly browse and obtain information online has increased, propelling the development of Question Answering in NLP (Hirschman & Gaizauskas, 2001). These systems use advanced NLP techniques such as information retrieval, text classification, and machine learning to evaluate queries, locate relevant material, and generate answers (Ibrahim, 2023).

According to Jurafsky and Martin (2023), there are two major approaches within question answering: information retrieval-based QA and knowledge-based QA. While retrieval-based question answering systems can answer questions from any domain by applying information extraction methods on relevant texts from a vast number of sources, knowledge-based question answering systems convert unstructured or semi-structured information to structured databases which are then queried based on the semantically analyzed user's questions (Hu et al., 2023a). More recently, generative question answering systems which have the capability to produce answers from their own understanding on the subject, have emerged due to advances in large language models and transformers, such as GPT-4 and ChatGLM (Hu et al., 2023a). At last, as exemplified by Liu et al. (2019c), there are also hybrid QA systems which use any of the above approaches simultaneously.

2.1.1.5. Sentiment Analysis

Sentiment analysis is a common text classification task which interprets opinions and captures sentiment, the positive or negative disposition that a user conveys toward an item (Jurafsky & Martin, 2023). Directly related with sentiment analysis lays emotion detection, designed to extract and evaluate fine-grained emotions like happiness, anger, and sadness (Kusal et al., 2023).

Sentiment analysis and emotion detection have grown significantly in recent years (Kusal et al., 2023) as extracting consumer or public sentiment have become increasingly useful in a variety of sectors, from marketing to politics (Jurafsky & Martin, 2023). Devi et al. (2016) and Krishna et al. (2017) each proposed a method for determining the consumer opinion of product features. Whereas, Bagheri et al. (2023) evaluated the performance of several sentiment analysis approaches applied to agricultural knowledge discovery via social media.

Lastly, it is important to highlight Twitter as one of the most popular data sources used in sentiment analysis (Sawicki et al., 2023). Salas-Zárate et al. (2022) reviewed 34 papers on depression identification using social media where 33% of which utilized Twitter as their main data source. A few years earlier, Drus and Khalid (2019) identified that out of 24 selected articles, 85% used twitter to collect information for sentiment analysis.

2.1.1.6. Text Summarization

Within NLP, text summarization can be defined as the automatic summary of text documents (Sharma et al., 2022). One of the most important aspects of automatic summarization is making sure the meanining is kept intact, not only because it allows the understanding of the most important details of an extensive amount of data, but also because it supports the comprehension and consequent transfer of deeper emotional connotations to the summarized text (Khurana et al., 2023).

There are two approaches to text summarization: abstractive and extractive (Sawicki et al., 2023). In addition, automatic summarization can be categorized based on the number of text documents used (single or multi-document), based on the summary type (generic or query-focused summary) and, lastely, based on the summarization task (supervised or unsupervised) (Khurana et al., 2023).

Similarly to other Natural Language Processing applications, text summarization often acts as a subtask, supporting larger NLP tasks (Khurana et al., 2023). For instance, Zin et al. (2023) proposed an unsupervised approach to extract information from lengthy legal contracts which uses a query-based summarization model to generate a coherent summary that is subsequently fed to GPT-3.5 for accurate information extraction.

2.1.1.7. Other Natural Language Processing Applications

Apart from dialogue systems, information extraction, machine translation, question answering, sentiment analysis, and text summarization, NLP can be applied to a variety of scenarios using many other relevant approaches. Below, we highlight additional natural language processing applications that have been discussed in recent literature.

Text categorization, also known as topic classification or text tagging, is "the task of assigning a label or category to an entire text or document" (Jurafsky & Martin, 2023). Text classification is commonly used in conjunction with other NLP tasks such as sentiment analysis or spam detection and filtering (Khurana et al., 2023; Adamson, 2023). For example, Xia et al. (2023) introduced a hybrid architecture partially composed by a hidden Markov model used for short text classification and spam filtering. Taking one step further, Srinivasarao and Sharaf (2023) proposed a method which employed text classification and sentiment analysis for SMS spam detection. In short, spam detection is a binary classification problem that assigns emails to one of two categories: spam or not-spam (Jurafsky & Martin, 2023).

Furthermore, in the same way as machines can automatically recognize and interpret spoken natural language and convert it to text, computers can also transform text into speech. Text-to-speech is used in CAs, to assist those with neurological disorders, as well as in devices and tools that read aloud, such as the one available in Microsoft's Immersive Reader (Jurafsky & Martin, 2023).

Caption generation is another powerful task that makes use of NLP and computer vison to automatically generate natural language descriptions based on the content of images (Kulkarni et al., 2022). Likewise, grammar checker, also known as auto-correct and grammatical error correcting, uses NLP to improve the quality of written text in emails and documents (Mahmoud et al., 2023).

Several other NLP applications can be used together to achieve a stronger and more precise outcome. For instance, topic modeling, topic labeling, and topic extraction are intertwined concepts that are frequently used in conjunction with keywords or information extraction and text summarization methods to identify the most relevant information from large amounts of text and reduce it to a shortened form by easily removing irrelevant information (Sajid et al., 2017; Albalawi et al., 2020). Additionally, event detection, that is used to identify and classify events, is known as an essential step of event extraction, which, in turn, is considered an information extraction subtask (Wan et al., 2023).

Lastly, NLP can also be applied to different sectors such as healthcare, hiring and recruitment, and education (Sharma et al., 2022). Within healthcare, besides lowering costs and improving service quality, NLP techniques can capture unstructured healthcare information, analyze its grammatical structure and meaning, and translate it so that electronic healthcare systems can understand it (Al Kuwaiti et al., 2023). In human resources, NLP is used to support interpretation and screening of candidate curriculums and to improve the efficiency of workers, while contributing to a smooth remote work transition (Mer & Virdi, 2023). Furthermore, in education, NLP is making a positive impact by supporting the development of adaptive learning systems, by providing targeted feedback, and by automating grading and assessment (Alqahtani et al., 2023).

2.1.2. Preprocessing Techniques in Natural Language Processing

Machine learning and natural language processing are strongly reliant on the quality of the data collected (de Silva, 2023). Thus, preprocessing, also often called text normalization or data preparation, is an essential preliminary step of data analysis to assure the model's accuracy (Guarda et al., 2022).

Throughout careful research, Kusal et al. (2023) highlighted some of the most recognized preprocessing techniques. These include data cleaning steps such as lowercasing, stop word removal, and noise removal (punctuation marks, white spaces, hashtags, URLs, emoticons and emojis), as well as other techniques such as stemming, lemmatization, and tokenization. In addition, Sharma et al. (2022) called attention to further NLP core tasks and features, namely part-of-speech (POS) tagging, parsing, and NER.

2.2. Disaster Management

Disaster management is, as its designation entails, the organization and administration of resources and responsibilities for dealing with all humanitarian dimensions of emergencies in order to reduce the impact of both natural and man-made disasters (International Federation of Red Cross and Red Crescent Societies, N/D). Earthquakes, volcanoes, storms, floods, and fires are all examples of natural disasters. On the other hand, war, pollution, nuclear blasts, arsons, hazardous materials spills, explosions, and transportation mishaps are examples of man-made catastrophes (Zibulewsky, 2001). Above all, regardless of the various definitions and classifications, all disasters share certain characteristics, such as the presence of a triggering factor (hazard), the element of surprise, the overwhelming nature, the large-scale impact on society and the surrounding environment, the difficulty of managing it, the intensive resource requirements, and the long-term consequences (Sawalha, 2022).

As seen, DM is a fast-paced, complex, and comprehensive process which asks for immediate coordinated action and donations from a wide variety of actors, including governments, international organizations, non-governmental organizations (NGOs), humanitarian groups, charities, philanthropists, businesses, and impacted local communities (Hesselman & Lane, 2017). Nevertheless, given the potential for profound consequences on a nation, its governance, and its citizens, the government of the impacted country bears the main responsibility for disaster response while optimizing the utilization of existing resources (Carter, 2008). International actors only respond to crises with the approval of the affected state, typically through a request or acceptance of an offer of aid (Office for the Coordination of Humanitarian Affairs, United Nations, 2018).

Governments worldwide are concerned about the potential economic, social, and environmental impacts of natural disasters (Parsons et al., 2021). These hazards are unpredictable and can cause significant damage, casualties, and loss of property (Farazmehr & Wu, 2023). According to the Emergency Events Database (Centre for Research on the Epidemiology of Disasters, 2023) between 2018 and 2022, 2,795 disasters were registered globally, with more than 600 million people affected. These were predominantly natural disasters (73%) and within this group, floods and storms were the most common types, corresponding to 33% and 19%, respectively, of all listed disasters.

In parallel with the rise of disaster occurrences and their higher intensity, literature in prominent journals on the topics of crisis and disaster management has seen a steady increase in publications (Wolbers et al., 2021). With the aim of developing a synthesized outlook of important emerging subjects in the research field of natural hazards and DRR, Zuccaro et al. (2020) proposed five future priority research areas, namely: Improved risk and impact assessment; Better data for a resilient future; Risk governance and partnership; Overcoming the implementation gap in DRR and climate change adaptation; Human behavior and disaster risk. These and other guidelines are essential to support the development of new research and spotlight the need for higher collaboration in all phases of the DM cycle.

2.2.1. Disaster Management Cycle

The disaster management cycle is a continuous process that includes several stages of planning and supporting, to reduce the impact of disasters (Yu et al., 2018). The cycle is famous for its capacity to place concepts and actions in a temporal sequence in connection to one another (Alexander, 2019). Due to the unique nature of each disaster, several frameworks have been proposed and adapted throughout the years.

The idea of creating a theoretical framework for emergency management is believed to date back to the beginning of the twentieth century with Carr's (1932) "sequence-pattern concept" proposal (Alexander, 2019). In 1998, Tufekci and Wallace (1998) highlighted the importance of dividing emergency management efforts in two stages: pre-event and post-event. In line with these two phases, several authors mentioned and applied what is now considered the traditional DM cycle (Figure 1), comprised by four main phases: mitigation, preparedness, response, and recovery (Alexander, 2002; Altay & Green, 2006; Sawalha I. H., 2020; Yu et al., 2018).

Overall, mitigation and preparedness take place before an emergency incidence, whereas response and recovery are considered post-disaster activities (Anaya-Arenas et al., 2014). Each phase is related to a number of procedures and action plans, which may vary amongst organizations, disaster agencies, and academic sources (Sawalha I. H., 2020).

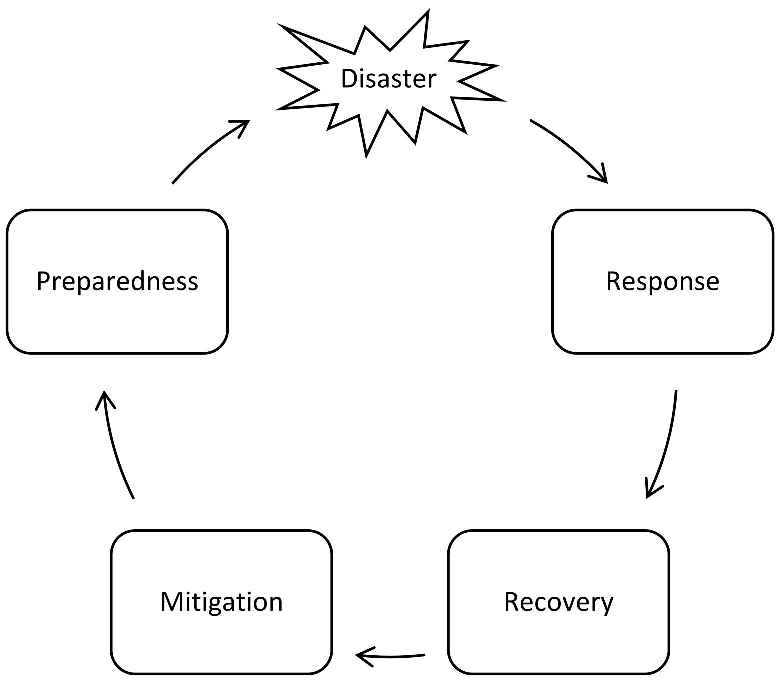

Figure 1 - Traditional disaster management cycle

Mitigation aims to reduce the effects of disasters such as through public education and property inspection, allocation and critical supply routing, and early warning systems. Preparedness focuses specifically on response planning considering, for example, location of aid facilities, goods pre-positioning, resource allocation and arrangement of transport and distribution strategies (Yáñez-Sandivari et al., 2021). On the other hand, response occurs in the immediate aftermath of a disaster and aims to minimize its consequences through the alert and evacuation of affected communities, inventory management, and network revival, while recovery consists of the support and rebuilt of the affected communities after a disaster. Altay and Green (2006) provide a more comprehensive list of typical activities that are usually considered under each phase of the disaster management cycle (p. 481).

Other authors have presented various replacement options to the traditional disaster lifecycle. In his book, Carter (2008) shared a slightly more complex standard cycle comprised by six phases (response, recovery, development, prevention, mitigation, and preparedness) followed by an alternative format with the response, recovery and preparedness phases further divided into several activities (e.g., emergency phase, reconstruction, prevention, preparedness, threat). In the context of the L'Aquila earthquake disaster, Alexander (2019) applied a disaster cycle composed by five phases: mitigation, preparedness, emergency intervention, recovery, and reconstruction. In 2020, Sawalha proposed "a contemporary view to the traditional DM cycle in which recent concepts of management are used to better cope with the uniqueness of the different major incidents." (p. 469).

Independently of the framework applied, it is important to consider that not all disasters follow a cyclical pattern, that not all phases happen in a strict order, and that their duration may vary depending on the severity of the hazard (Alexander, 2019).

2.2.2. Technology and Data in Disaster Management

To effectively combat disasters, emergency managers must be able to foresee and assess possible threats, as well as devise action plans to mitigate impact magnitude and respond to those in need (Tufekci & Wallace, 1998). In addition, successful DM requires a comprehensive technology plan and data standards to facilitate information exchange among systems and stakeholders (Sakurai & Murayama, 2019). Hence, there is general agreement in the humanitarian sector that improved communication, data collecting, and IM will lead to better risk assessment, focused prevention, and more effective preparatory operations in humanitarian action (Altay & Labonte, 2014).

Already in 1998, Beroggi and Wallace argued that developments in communication and computer technology had provided emergency managers with extra abilities to mitigate the potential effects of disasters (Beroggi & Wallace, 1998). Since then, much has change, including groundbreaking discoveries and improvements in the areas of technology and data management (Hillyer, 2020). Below, we highlight four different areas linked with technology and data which have vastly impacted the field of disaster management in recent years.

2.2.2.1. Artificial Intelligence and Big Data Analytics

With the recent advancements in artificial intelligence, many researchers have begun to investigate AI, ML, and DL approaches for big data analytics in disaster management (Nunavath & Goodwin, 2019). Velev and Zlateva (2023) discussed how, when implemented successfully, AI can play an essential role in all phases of the DM cycle by delivering timely and reliable information, optimizing resources and logistics, and assisting in the reduction of disaster consequences on impacted communities.

AI and big data in DM studies have been employed in a variety of fields and data sources. Within studies focused on the intersection between AI and disaster risk communication, Ogie et al. (2018) identified two main research areas: prediction and monitoring for early warning, as well as information extraction and categorization for situational awareness. In a similar context, social media, satellites, and sensors were identified as the most common data sources used for classification and prediction in disaster management (Nunavath & Goodwin, 2019).

Ultimately, several methods such as Support Vector Machines (SVMs), Naïve Bayes (NB), and other common machine learning algorithms and techniques, highlighted by Nunavath and Goodwin (2019), are now frequently used in DM studies. Through a systematic literature review on AI-based methods applied to disaster management between the period of 2012 and 2022, Farazmehr and Wu (2023) concluded that in recent years, deep learning-based approaches have been widely applied to DM, outperforming several classic statistical algorithms like as K-Nearest Neighbors (KNN) and SVM in a variety of prediction tasks on this subject.

2.2.2.2. Geographic Information Systems and Remote Sensing

Geographic Information Systems (GIS) and remote sensing (RS) can be combined with both established and cutting-edge scientific techniques to effectively manage disasters at every stage of the DM cycle (Nefros et al., 2022). While GIS connects databases and maps to produce interactive displays of disaster-related figures and imagery, RS is the art and science of measuring the Earth with aircraft or satellite sensors (Orimoloye et al., 2021).

As explained by Orimoloye et al. (2021), remote sensed imagery is often incorporated into GIS to map disasters. Moreover, remote sensing provides high-resolution, multi-dimensional imagery for disaster assessment which, in turn, provides information about the changing land use system, water bodies, direction, and damaged objects in the impacted area, allowing for more effective decision-making regarding rescue methods (Sarker et al., 2020). Similarly, through a literature review, Wang and Xie (2018) emphasize how RS provides crucial data for mapping water resources, quantifying hydrological fluxes, and monitoring drought and flooding inundation, whereas GIS offers the most effective tools for managing water resources, drought, and flood risk.

3. METHODOLOGY

The present study is theoretical in nature and stands on recent research to investigate the usage of NLP before, during and after natural and man-made disasters. In this regard, we conducted a systematic literature review following the PRISMA 2020 guidelines. The PRISMA methodology, originally published in 2009 and updated in 2020, provides a comprehensive checklist and flow diagram to ensure that all recommended information is accurately identified and reviewed (Page et al., 2021). Furthermore, this method, primarily developed to be used in physical and medical sciences, is now applied into a variety of fields to ensure users produce replicable, comprehensive, and objective research (Booth et al., 2020).

We used one of the biggest bibliographic databases that cover scholarly literature from a multitude of disciplines, Scopus, to identify papers related to Natural Language Processing and Disaster Management as well as related topics, all highlighted below:

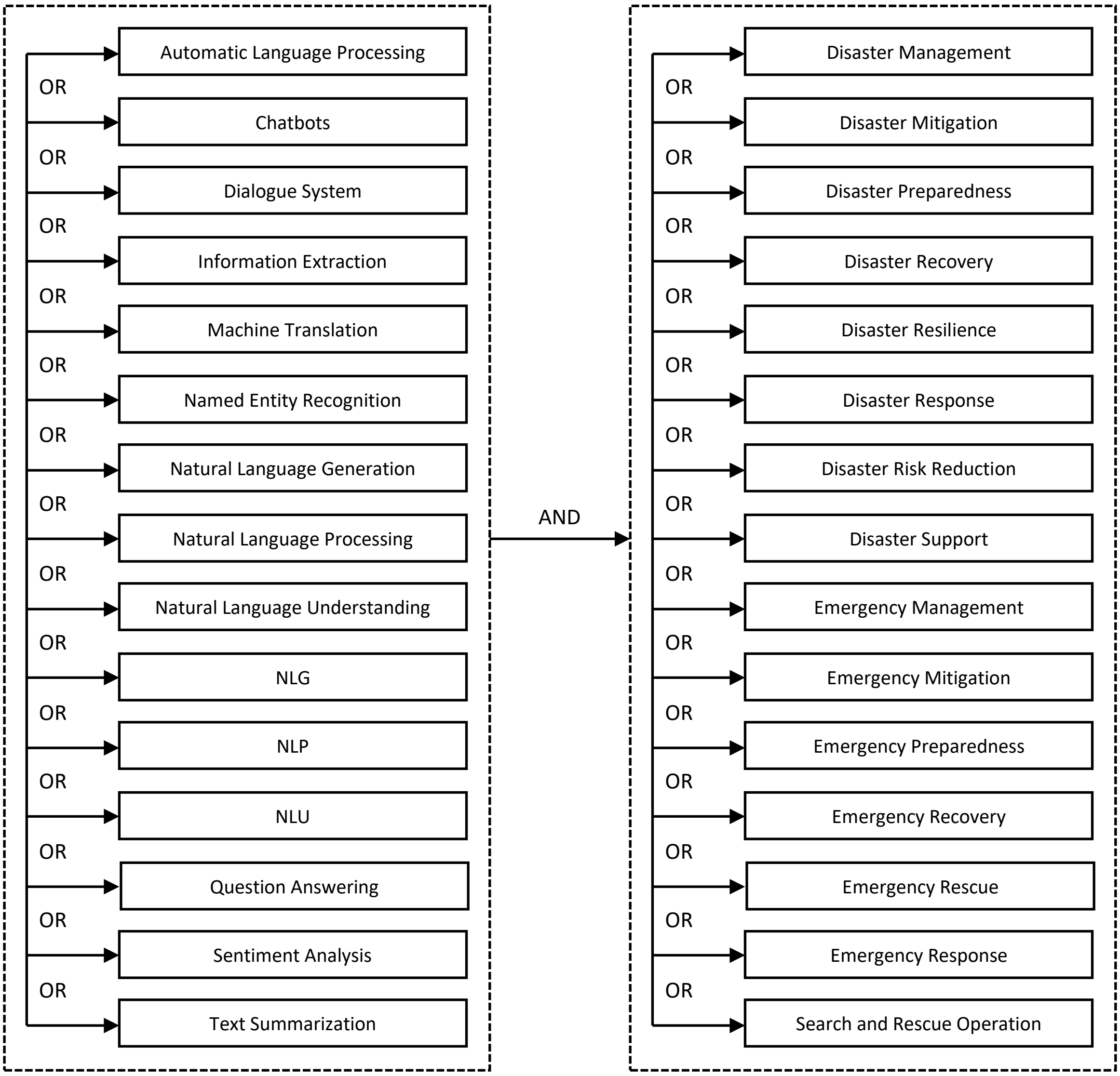

Figure 2 - Conceptual framework of keywords used in the search query

The keywords combined with AND/OR operators, as represented in Figure 2, were selected based on the identification of the most prominent concepts when constructing the theoretical background section. To achieve an accurate outcome, these terms were only searched within the title, abstract and keywords of each document. In addition to the above-mentioned concepts, the search query contained a restriction by the year. Thus, only documents published between 2014 and 2023 were included from the start. Table 3 summarizes the inclusion and exclusion criteria for this study.

Considering February 22nd, 2024, as the last consulted date, this resulted in a total of 548 publications that were then stored and filtered using Zotero. Further screening was performed. After going through the title, abstract and keywords, all documents written in a language other than English were excluded. Moreover, to guarantee the quality of this dissertation, only peer-reviewed quartile 1 journal articles were considered eligible. The assessment of the journal quartile was done using information available in Scimago Journal & Country Rank and Scopus. When uncertain about the correct journal quartile, we considered an average value per topic as well as the publication year of each paper. Finally, full text articles were screened for relevance. During this final step, we excluded all papers that were not directly related with the main topics of research: Natural Language Processing and Disaster Management. In the context of this study, we consider ''disaster management'' as the preparation and support to disastrous events (natural and man-made), and do not include daily response of ambulances, police, or fire departments to normal emergency calls.

Table 3 - Inclusion and exclusion criteria

Criteria	**Inclusion**	**Exclusion**
Document Type	Peer-reviewed Q1 journal articles	Book chapters, books, working papers, or other types of non-peer reviewed and non-Q1 publications
Full Text	Full text is available for download	Entire paper is not available for download
Language	Full documents in English	Non-English papers
Period	Published between 2014 and 2023	Published before January 2014 and after December 2023
Topic	Directly related with NLP and DM	Not directly related with NLP and DM

Several instances of full-text screening were performed to minimize selection bias and ensure the quality of this work. The flow diagram in Figure 3Figure 3 illustrates a detailed representation of the process used to identify the studies included in this review. In the next section, we present the systematic review of the main topics and discussions that emerged through the study of the selected papers.

4. RESULTS

As Figure 3 highlights, after going through the different study identification phases, this research yielded a total of 107 peer-reviewed articles from the initial 548 results.

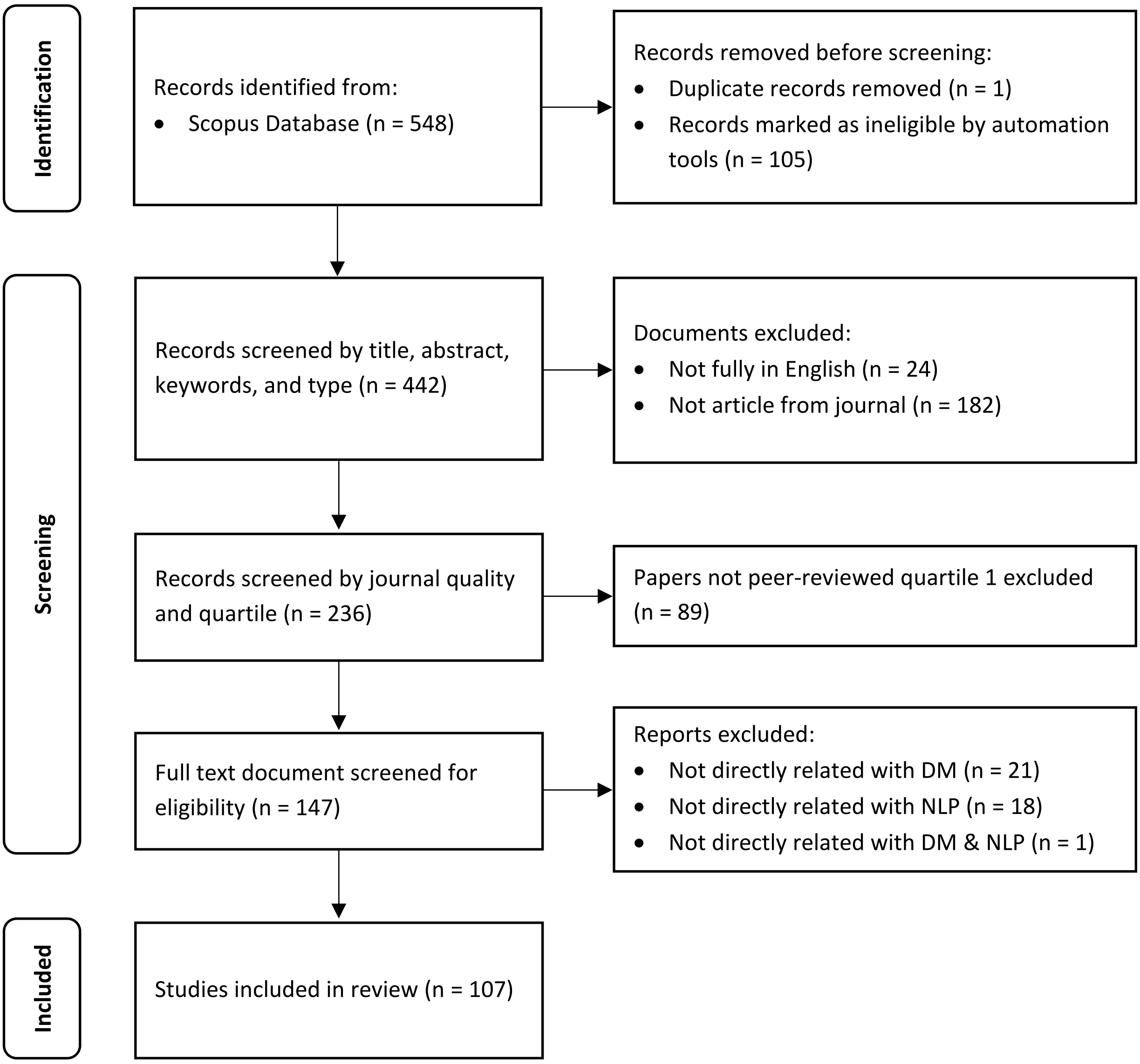

Figure 3 - Document selection based on PRISMA 2020 guidelines (Page et al., 2021)

After importing the RIS file containing the information of the 548 documents obtained through Scopus, Zotero promptly identified two duplicate records. While one of them consisted, in fact, of two distinct parts of the same conference proceedings book (Coman & Vasilache, 2023a; Coman & Vasilache, 2023b), the second set of records seemed correctly identified and the versions were consequently merged (Preum et al., 2020), leaving us with 547 papers.

Subsequently, the “Find available PDFs” functionality within Zotero was used, resulting in the attachment of 158 PDF documents. Each paper that was not automatically identified and attached by Zotero was additionally sought on Google search engine by title and linked to the respective record in Zotero in case it was available for download on another website. As represented in Figure 3, this resulted in a total of 442 papers that proceeded to the next phase.

In the screening phase, we initially took a glance at each one of the 442 papers to manually identify their language and document type. All records that were not fully in English and that were not published in a scientific journal were promptly excluded, leading to a total of 236 articles to be filtered by journal quality and quartile. Later, we individually searched each article’s journal name on the Scimago Journal & Country Rank website. The quartile filtering was carried out based on the publication date, apart from all papers published in 2023, which were filtered based on the ranking for the year 2022 due to a lack of more recent data. All 147 papers that proceeded to the next phase were then fully screened for eligibility and filtered out if not related with any of the main topics of our research: disaster management and natural language processing. Out of the 21 papers that were removed because they did not directly relate to DM, about 43% focused on the employment of NLP techniques in emergency medical services and hospital interventions instead. Kurvers et al. (2023) and Turner et al. (2019) represent two examples of the previous statement.

Taking advantage of the article’s keywords automatically imported into Zotero, we created a network visualization using VOSviewer software. After rectifying and merging a couple of important keywords directly in Zotero, we exported the data linked to the 107 articles and imported it directly into VOSviewer. While setting the minimum number of occurrences of a keyword to 25, we obtained a network visualization, represented in Figure 4, composed by 36 terms distributed between three clusters.

In addition, Table 4 provides a deeper understanding of the data behind Figure 4 as well as a detailed representation of the three mentioned clusters. Through both data representations it is possible to conclude that “Disasters”, “Social Networking”, “Social Media”, and “NLP” are the most used keywords in the included papers.

When taking a closer look at Figure 4, it is possible to identify the keywords with higher link strength between them by looking at the size of each connecting line. These include “Disasters” and “Disaster Prevention” with a total link strength of 130, followed by “Social Media” and “Social Networking” and by “NLP” and “NLP Systems”, exhibiting a total link strength of 105 and 91, respectively. The results present no surprise, as each pair of concepts contains extremely similar and highly interrelated keywords. Therefore, it is expected that the mentioned tags are simultaneously mentioned in the same papers. However, we can already deduce that social media will be a frequently used data source among the included and analyzed articles.

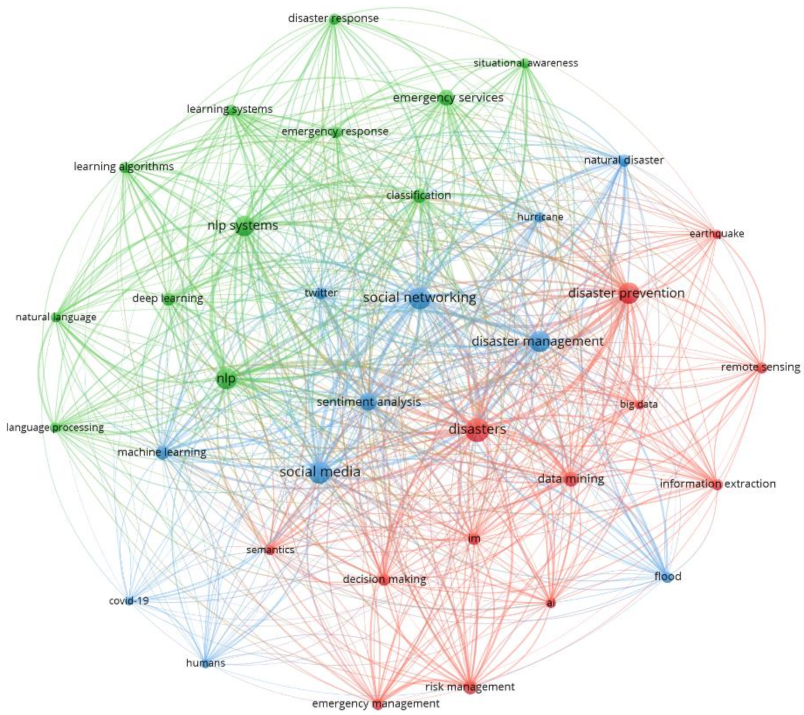

Figure 4 - Network visualization of keywords co-occurrence of the included articles

Table 4 - Clusters classification and number of occurrences

Keywords	**Occurrences**	**Total Link Strength**
Cluster 1 - Disasters & Data		
AI	36	239
Big Data	33	207
Data Mining	78	500
Decision Making	54	367
Disaster Prevention	155	1017
Disasters	200	1275
Earthquake	30	168
Emergency Management	44	268
IM	49	352
Information Extraction	47	237
Remote Sensing	47	171
Risk Management	66	401
Semantics	39	237

Keywords	Occurrences	Total Link Strength
Cluster 2 - Disasters & NLP		
Classification	71	473
Deep Learning	56	330
Disaster Response	46	255
Emergency Response	42	175
Emergency Services	87	468
Language Processing	34	278
Learning Algorithms	42	337
Learning Systems	42	325
Natural Language	36	290
NLP	161	902
NLP Systems	147	893
Situational Awareness	37	231
Cluster 3 - Disasters & Social Media		
COVID-19	26	108
Disaster Management	148	875
Flood	41	255
Humans	29	136
Hurricane	36	243
Machine Learning	69	460
Natural Disaster	53	336
Sentiment Analysis	106	580
Social Media	164	915
Social Networking	172	1187
Twitter	54	319

In the following subsections, we start by representing broader data such as the year of publication and journal distribution. Subsequently, we reveal the findings connected to more specific topics, such as used data sources, disaster types, and disaster management phases, which resulted in the discovery of different forms of challenges, possibilities, and applications of NLP in DM.

4.1. DISTRIBUTION OF PUBLICATIONS BY YEAR AND JOURNAL

Figure 5 depicts the distribution of the screened quartile 1 (Q1) peer-reviewed articles within the timeframe addressed in this study regarding the publication year of the selected research. Although our inclusion criteria comprised studies published between 2014 and 2023, articles from the years 2014 and 2015 initially identified through our search query were excluded as they were not directly related either with natural language processing (Römer et al., 2014; Wu et al., 2014) or with disaster management (Murthy, 2015). Thus, studies meeting the defined inclusion criteria only emerged in 2016, with 2022 and 2023 being the most productive years in the topic of study.

It is interesting to notice the exponential growth of published papers throughout the years. The average number of publications is around 13 studies per year over the past 8 years.

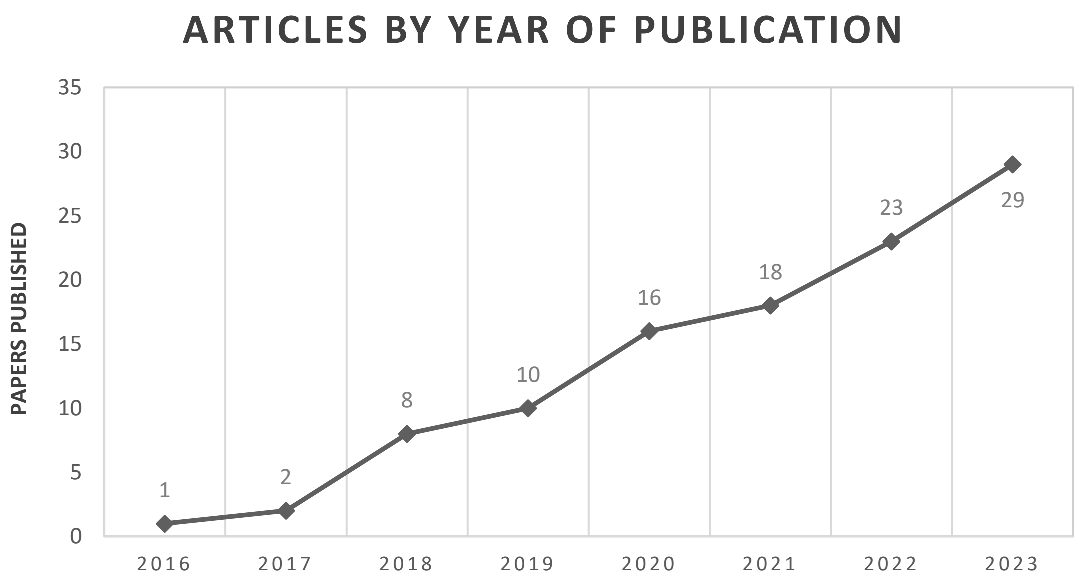

Figure 5 - Distribution of included articles by year of publication

Additionally, we explored the distribution of included studies based on the journal where they were published, which resulted in a total of 63 distinct peer-reviewed Q1 journals. Within this total, 30% are represented in Figure 6 with two or more papers published. As represented in Figure 6, the International Journal of Disaster Risk Reduction has the biggest number of Q1 articles published in the interface of NLP and DM, followed by IEEE Access.

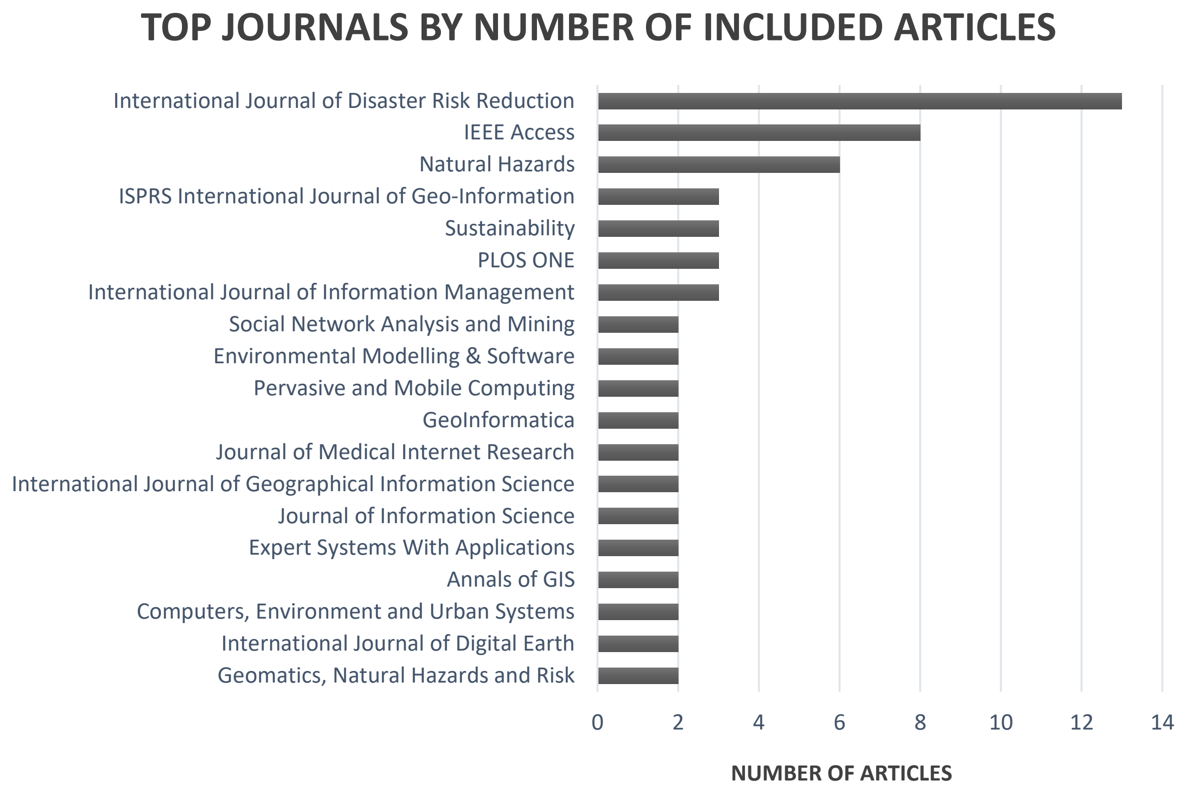

Figure 6 - Distribution of included articles by journal name

The top 15 included journals based on the highest Scimago Journal Rank (SJR) values are given in Table 5. As clarified on the official website, “the SCImago Journal Rank indicator is a measure of journal’s impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years” (Scimago Journal & Country Rank, 2023). Thus, the higher the value, the higher the impact and prestige of a journal.

The SJR is also directly related with the quartile ranking. Thus, taking into account that our study solely contemplates Q1 journals, it isn’t surprising to notice the high SJR values within the list of included journals, ranging from 4.91 to 0.56. Nonetheless, it is important to consider that the values shown in Table 5 are referent to 2022 and based on Scopus data as of April 2023.

By comparing Figure 6 with Table 5, it is easy to conclude that the top 5 journals by number of included articles are not represented in the top 15 journals by SJR values. While the International Journal of Disaster Risk Reduction has a SJR of 1.13, IEEE Access and Natural Hazards scores are 0.93 and 0.75, respectively. Still, as mentioned above, all are peer-reviewed Q1 journals with high admittance standards and several revision phases.

Table 5 - Top 15 included journals by SJR values

Journal Name	SJR 2022
International Journal of Information Management	4.91
IEEE Internet of Things Journal	3.75
Computer-Aided Civil and Infrastructure Engineering	2.96
Journal of Big Data	2.71
Big Data Mining and Analytics	2.53
Automation in Construction	2.44
IEEE Transactions on Geoscience and Remote Sensing	2.40
Sustainable Cities and Society	2.31
Information Processing and Management	2.11
Decision Support Systems	2.08
New Media & Society	2.08
Transportation Research Part D	2.02
Journal of Medical Internet Research	1.99
Journal of Cleaner Production	1.98
Telematics and Informatics	1.88

4.2. DATA TYPES AND SOURCES

In this section, we aim to review the main data types, data sources, and data sub-sources utilized in prominent articles focused on the intersection between NLP and DM.

Within the 107 selected articles, six are literature reviews (Tounsi & Temimi, 2023; Shi et al., 2023; Vongkusolkit & Huang, 2021; Su et al., 2021; Imran et al., 2020; Hou et al., 2020) which are excluded from our quantitative analysis from this point on and solely used for a qualitative comparison of results. This is because literature reviews cover a large range of articles often focused on different areas of study and which use distinct sources of data, making it challenging to pinpoint the specific information required for this analysis. Additionally, considering the conclusions shared in the literature reviews could result in duplicate data analysis as some of the articles studied in those reviews were also selected within our study.

The distribution of reviewed studies with respect to the applied data types is shown in Figure 7. Out of the 101 defined papers, 93,1% uses at least one source of unstructured data, while 6,9% of authors mention using semi-structured data in their analysis.

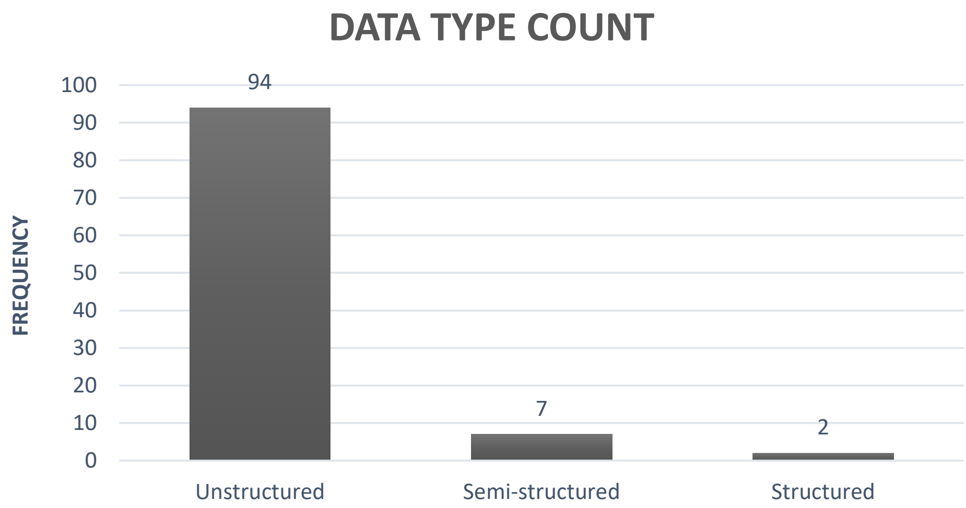

Figure 7 - Distribution of data types used in the shortlisted literature

Apart from the identification and segmentation of studies by data type, we highlight the different data sources and sub-sources applied in the 101 selected articles. For a better understanding of the concepts used in upcoming analysis, Figure 8 illustrates the classification of data sources and sub-sources defined for this study. Data types were not included in the below representation as they may vary between data sources and sub-sources. Thus, while all sub-sources within aerial imagery, social media, and other offline sources were defined as unstructured, other data sources varied by data type. A clear example lies within text documents, where data sub-sources were mostly classified as unstructured except for one research which clearly stated the usage of semi-structured emergency plans (Ni et al., 2023).

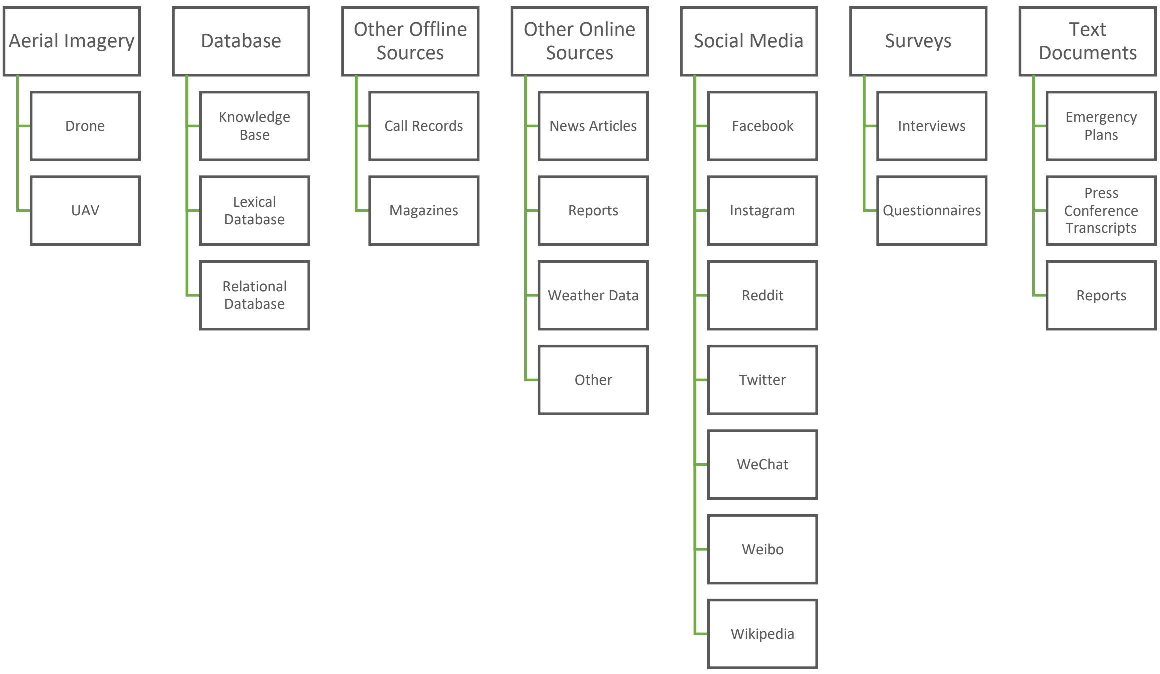

Figure 8 - Proposed data sources and data sub-sources classification scheme

Furthermore, it is important to notice that several authors used more than one data sub-source and even multiple data sources within the same research. For example, in addition to Twitter data, Wang et al. “collected hurricane damage data from FEMA, as well as disaster-related geographical and socioeconomic data from various sources” (2023, p. 5). This and other similar cases are contemplated in Figures Figure 7Figure 9, and Figure 10 and are the reason why the data frequency total doesn’t add up to 101, the total number of selected articles.

It is apparent from Figure 9, that 85 out of the 101 selected studies referred using social media as a data source, followed by 19 articles mentioning other online sources, such as news articles, weather data, etc. Additionally, it is interesting to underline that 15,8% of articles combined both above mentioned data sources in their research to achieve stronger results. Likewise, while most studies used only one social media sub-source as an input to the proposed approaches, four papers applied a variety of social media platforms to enrich their

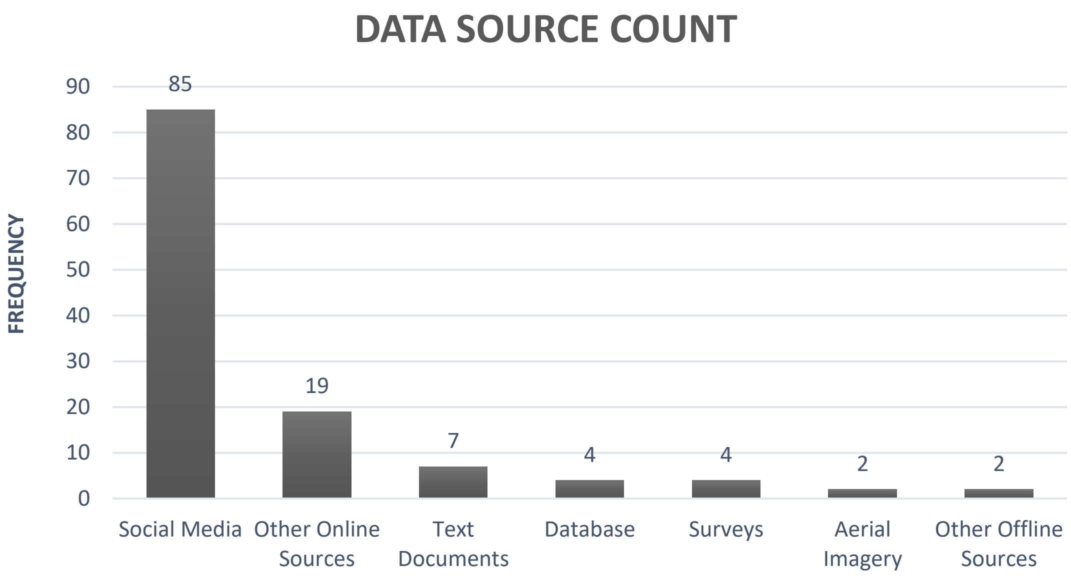

Figure 9 - Distribution of data sources used in the shortlisted literature analysis (Kellner et al., 2023; Lian et al., 2020; Raisio et al., 2022; Shibuya & Tanaka, 2019).

Lastly, our literature review reveals that microblogging platforms play an important role when it comes to data sourcing in NLP projects applied to disaster management. Specifically, as illustrated in Figure 10, Twitter is the most mentioned data sub-source within the 101 selected studies (66,3%), followed by Weibo (14,9%), often called the “Chinese Twitter” as both tools present similar characteristics. However, these results are anything but unexpected as both Vongkusolkit and Huang (2021) and Tounsi and Temimi (2023), authors of two of the literature revisions excluded from the quantitative analysis but included in the 107 shortlisted studies, shared similar values (63,6% and 57,1% for Twitter, respectively). Because of the nature of these microblogging platforms, users post messages with a restricted amount of characters, making it possible to share information in a fast and easy way (Schwarz et al., 2023).

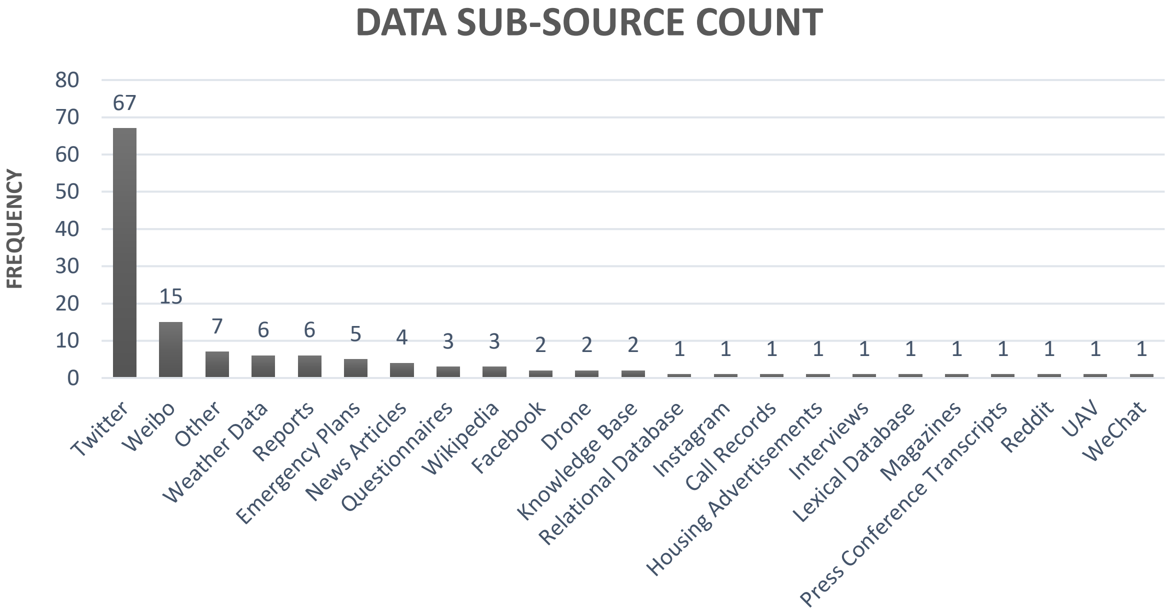

Figure 10 - Distribution of data sub-sources used in the shortlisted literature

4.3. Disaster Types and Disaster Management Phases

Most articles apply the proposed models on a case study referent to a specific disaster occurrence, generally used “as an example to evaluate and analyze the effectiveness of the proposed method” (Wu et al., 2023, p. 1). In our research, we identify that out of the 101 shortlisted articles (107 minus the 6 literature reviews), 81,2% included a case study or used data specific to one or more disasters. Based on those 82 papers we then perform a categorization by disaster group, disaster type, disaster subtype, and country where the mentioned disasters occurred. Additionally, we group the 101 identified papers based on the four phases of disaster management cycle.

Considering the classification scheme proposed in Figure 11, articles are split between “natural” or “man-made”, depending on the disaster group addressed in the included case study. Additionally, these groups are further divided into different types and sub-types. While natural disaster types and sub-types are classified based on the international disaster database (Centre for Research on the Epidemiology of Disasters), man-made disasters are, in turn, broken down into “armed conflict”, “pollution”, “industrial accident” and “terrorist attack” disaster types, with the latter being subdivided based on the classification provided by the New York City Police Department (New York City Police Department, n.d.).

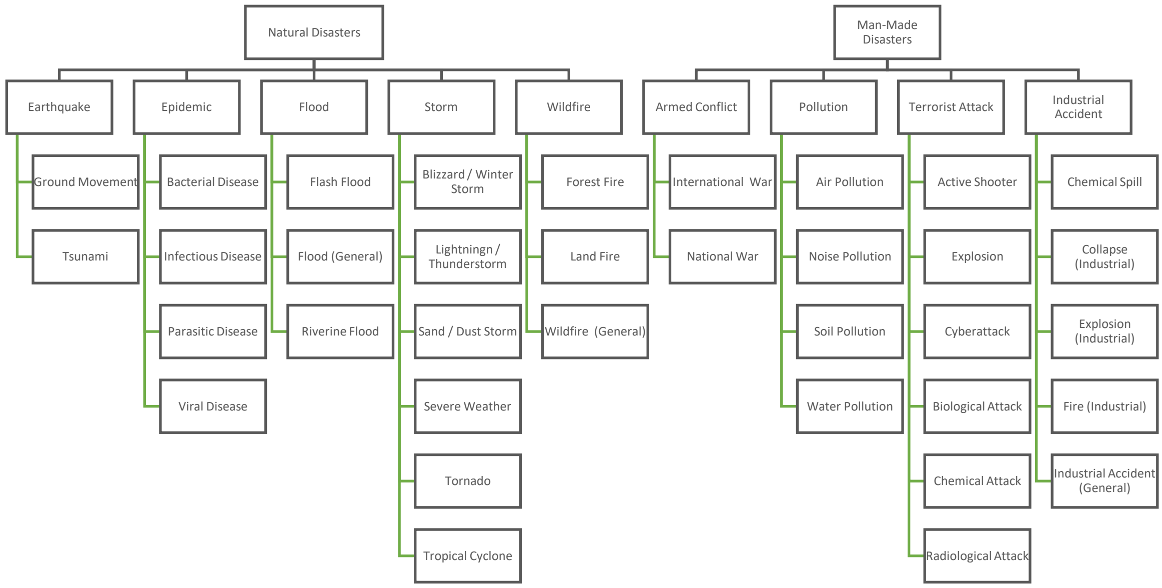

Figure 11 - Proposed disaster classification scheme

In Figure 12, it can be seen that around 87,8% of all articles including a case study address solely natural disasters, while 3,7% focus on both natural and man-made disasters. Considering both the proposed classification scheme in Figure 11 as well as the disaster type distribution illustrated in Figure 13, we conclude that “storm”, “epidemic”, and “flood” are the top 3 disaster types most mentioned in the 82 considered researches. These are all natural disasters.

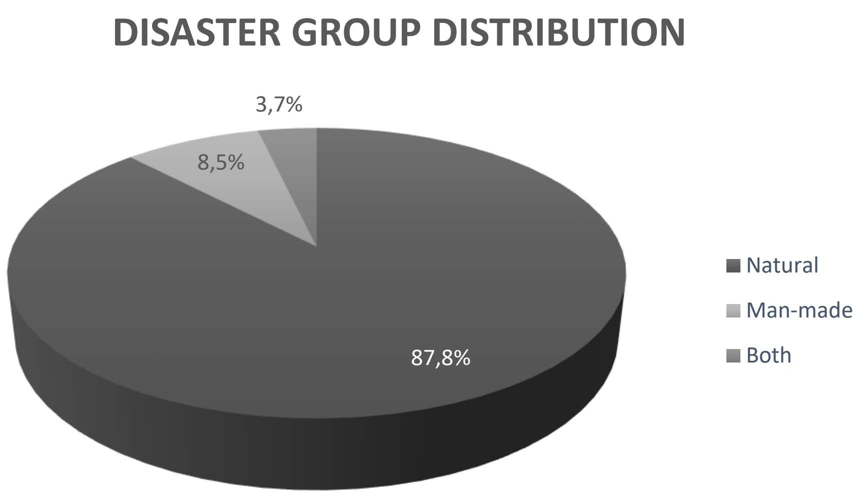

Figure 12 - Distribution of articles based on the disaster group

Moreover, through Figure 13, we can also conclude that 9 out of the 82 papers including a case study analyzed or used data from more than one disaster type. Such is the case of Wahid et al. (2022) research, which applied two different datasets, one composed by COVID-19 twitter data and other related to 7 distinct crisis (2012 Sandy hurricane, 2013 Alberta floods, 2013 Boston bombings, 2013 Oklahoma tornado, 2013 Queensland floods, 2013 Texas explosion and 2011 Joplin tornado), to train their proposed model.

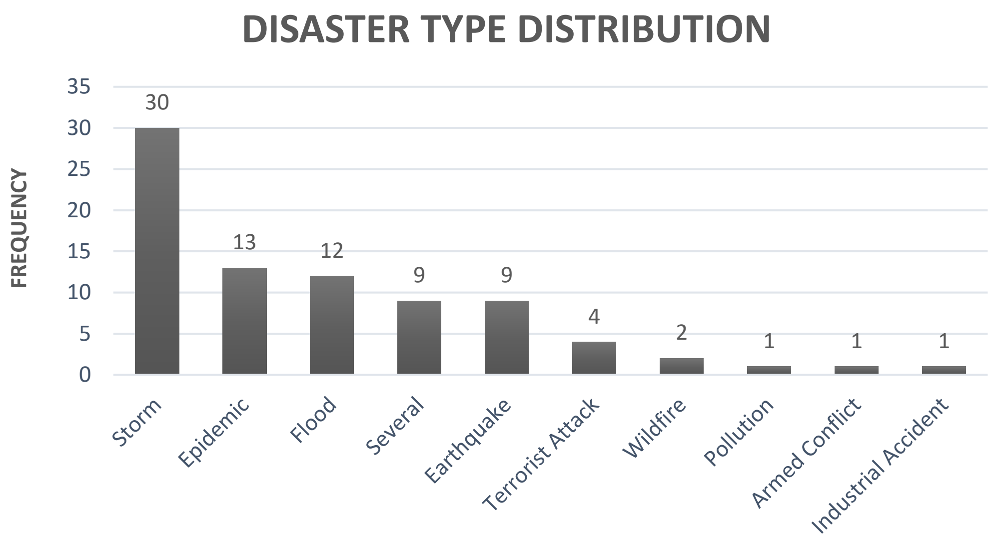

Figure 13 - Distribution of articles based on the disaster type

We have further deconstructed the disaster type data, specifying the disaster subtypes mentioned within the selected papers. It is apparent from Figure 14 that "tropical cyclone" and "viral disease" are the most researched disaster sub-types within the selected papers. Besides, while comparing Figures Figure 13 and Figure 14, we note that from the 30 articles that focus on "storms", 29 are related with "tropical cyclones".

When it comes to the disaster subtype "viral disease", 11 out of the 12 papers (91,7%) underlined in Figure 14 are specifically related with the COVID-19 pandemic. Similarly, most research focused on "tropical cyclone" essentially investigates the 2017 Hurricane Harvey and Hurricane Irma.

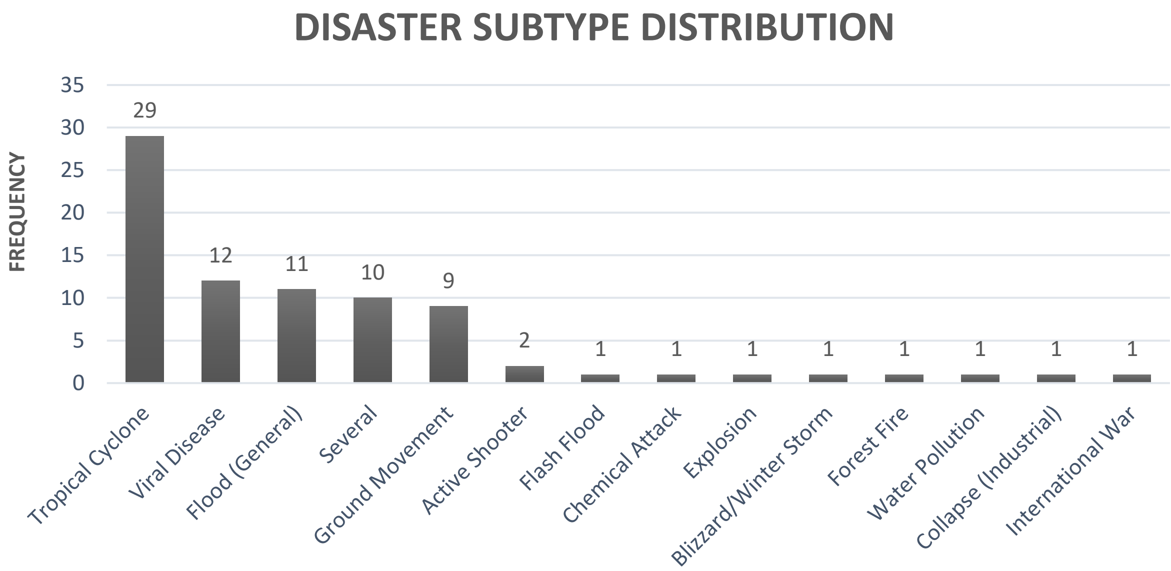

Figure 14 - Distribution of articles based on the disaster subtype

The disasters presented as case studies in the selected articles occurred in different countries. Figure 15 shows how authors mostly used data from disasters which occurred in the United States of America, while other countries such as Brazil and Ukraine were less prominent.

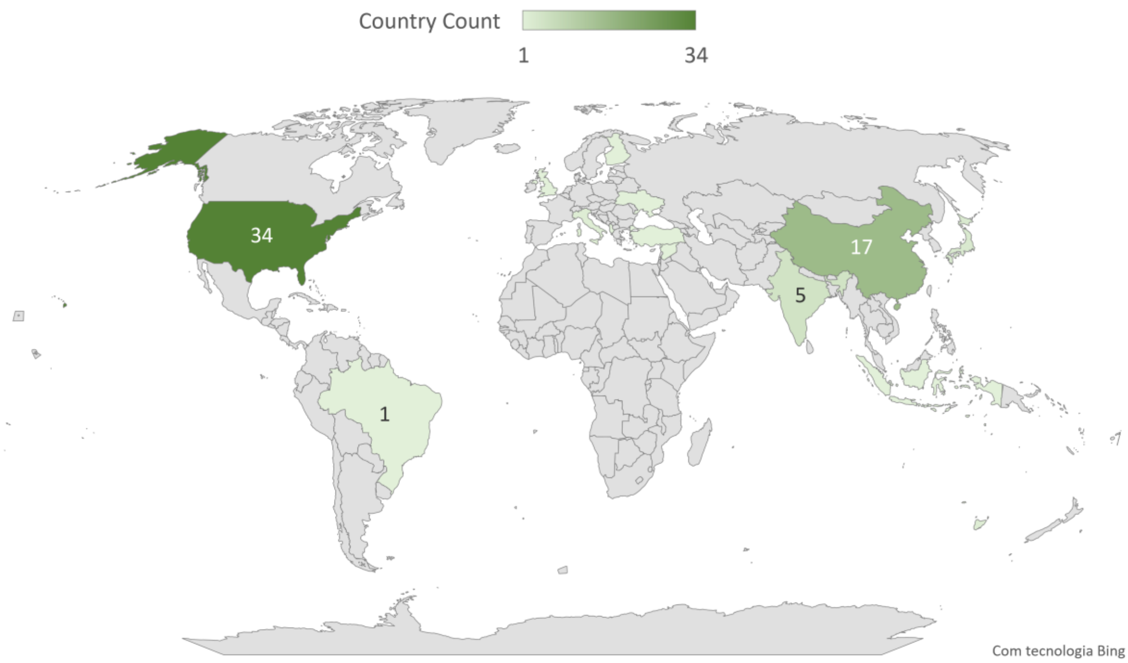

Figure 15 - Distribution of studies by disaster country

In main literature, emergency response operations are divided into four stages, representing the disaster management lifecycle. These are mitigation, preparedness, response, and recovery. We have classified the 101 shortlisted articles based on the above-mentioned disaster phases (see Table 6). It clearly appears that most publications (84,2%) focus on the response phase, while 29,7% may refer to the recovery phase, 21,8% to mitigation and finally 18,8% focus also or only in the preparedness phase of DM.

Table 6 - Distribution of articles based on the disaster phase

Disaster Phase	**Publications**
Mitigation	(Chan & Tsai, 2019), (da Costa et al., 2022), (Guo et al., 2021), (Huang et al., 2022), (Joloudari et al., 2023), (Karimiziarani & Moradkhani, 2023), (Karimiziarani et al., 2022), (Lei et al., 2023), (Ma et al., 2023), (Matsuoka & Rocha, 2021), (Mughal et al., 2021), (Rashid & Wang, 2022), (Rossi et al., 2018), (Shahbazi & Byun, 2022), (Stieglitz et al., 2022), (Sugino et al., 2023), (Terracciano & Han, 2023), (Tian et al., 2023), (Wang et al., 2023), (Zhang et al., 2022), (Zhou et al., 2022)
Preparedness	(He et al., 2023), (Hu et al., 2019), (Huo et al., 2020), (Karami et al., 2020), (Karimiziarani & Moradkhani, 2023), (Karimiziarani et al., 2022), (Karimiziarani et al., 2023), (Kellner et al., 2023), (Matsuoka & Rocha, 2021), (Ni et al., 2023), (Ryan et al., 2021), (Schwarz et al., 2023), (Sermet & Demir, 2018), (Terracciano & Han, 2023), (Valinejad & Mili, 2023), (Wang et al., 2020), (Wang et al., 2021), (Wang et al., 2023), (Zhou et al., 2022)
Response	(Afyouni et al., 2022), (Afyouni et al., 2023), (Alam et al., 2020), (Alkhatib et al., 2019), (An et al., 2022), (Bai & Yu, 2016), (Barker et al., 2019), (Bashir et al., 2021), (Behl et al., 2021), (Cerna et al., 2022), (Chen & Lim, 2021), (Contreras et al., 2022a), (da Costa et al., 2022), (Dereli et al., 2021), (Dong et al., 2021), (Eligüzel et al., 2020), (Fan et al., 2020), (Fang et al., 2019), (Farnaghi et al., 2020), (Gulnerman & Karaman, 2020), (Guo et al., 2018), (Guo et al., 2020), (Guo et al., 2021), (Hampton & Shalin, 2017), (He et al., 2023), (Hu et al., 2023), (Huang et al., 2022), (Jain & Kumar, 2018), (Joloudari et al., 2023), (Karami et al., 2020), (Karimiziarani & Moradkhani, 2023), (Karimiziarani et al., 2022), (Karimiziarani et al., 2023), (Karmegam & Mappillairaju, 2021), (Khattar & Quadri, 2022), (Kim et al., 2022), (Lee & Yu, 2020), (Lei et al., 2023), (Li et al., 2023), (Lian et al., 2020), (Liu et al., 2019), (Lu et al., 2023), (Lyu & Lu, 2023), (Ma et al., 2022), (Ma et al., 2023), (Mukherjee et al., 2022), (Neppalli et al., 2017), (Nguyen et al., 2022), (Nimmi & Janet, 2023), (Paradkar et al., 2022), (Perlstein & Verboord, 2021), (Ragini et al., 2018), (Raisio et al., 2022), (Rashid & Wang, 2022), (Resch et al., 2018), (Rexiline Ragini et al., 2018), (Reynard & Shirgaokar, 2019), (Rossi et al., 2018), (Roy et al., 2020), (Rusho et al., 2021), (Ryan et al., 2021), (Sarkar et al., 2023), (Scalia et al., 2022), (Shahbazi & Byun, 2022), (Singh et al., 2019), (Sit et al., 2019), (Terracciano & Han, 2023), (Tian et al., 2023), (Valinejad & Mili, 2023), (Wahid et al., 2022), (Wang et al., 2020), (Wang et al., 2021), (Wang et al., 2023a), (Wu et al., 2022), (Wu et al., 2023), (Xia et al., 2023), (Yao & Wang, 2020), (Yao et al., 2021), (Yuan et al., 2021), (Zander et al., 2023), (Zhang et al., 2022), (Zheng et al., 2021), (Zhou et al., 2022), (Zhou et al., 2023), (Zhuang et al., 2021)

Disaster Phase	Publications
Recovery	(Bai & Yu, 2016), (Chen & Lim, 2021), (Contreras et al., 2022a), (Contreras et al., 2022b), (Fan et al., 2020), (Hampton & Shalin, 2017) (He et al., 2023), (Karami et al., 2020), (Karimiziarani et al., 2022), (Karimiziarani et al., 2023), (Karmegam & Mappillairaju, 2021) (Lu et al., 2023), (Mukherjee et al., 2022), (Nguyen et al., 2022), (Oz et al., 2018), (Ragini et al., 2018), (Raisio et al., 2022), (Rashid & Wang, 2022), (Rexiline Ragini et al., 2018), (Reynard & Shirgaokar, 2019), (Roy et al., 2020), (Sarkar et al., 2023), (Shibuya & Tanaka, 2019), (Sit et al., 2019), (Terracciano & Han, 2023), (Wang et al., 2020), (Wang et al., 2021), (Wang et al., 2023a), (Yan et al., 2020), (Zander et al., 2023)

Down to the same reason as above, we didn't consider the six literature reviews during the quantitative analysis of disaster groups, types, sub-types, and disaster management phases. At last, we shift the focus to address the researcher's conclusions and compare them to our results.

When studying the application of social media data classification during natural disasters, Vongkusolkit and Huang (2021) obtained different analysis results on the percentage of disaster occurrence depending on the classification methods applied. As an example that coincides with our outcomes, the majority of the articles analyzed by Vongkusolkit and Huang (2021) applied temporal classification to hurricanes (tropical cyclones) and floods. The absence of the "viral disease" subtype on their analysis might have to do with the considered timeframe for the analysis (from 2004 to 2019), which ended right before COVID-19 hit worldwide.

Similarly, Tounsi and Temimi (2023), who reviewed NLP applications for the assessment of hydrometeorological hazards (hurricanes, typhoons, and flooding), concluded that more than 48% of the analyzed studies focused on hurricanes.

Other literature reviews did not make related analysis mainly either due to their broad scope or, contrarily, because of their focus on a single disaster type.

In the following section, we further analyze the 107 shortlisted papers and provide important insights related with the proposed research questions.

5. DISCUSSION

In this work, we aim to determine the variety of natural language processing techniques employed in disaster management while identifying the primary challenges and benefits inherent to that application. To this extent, we review and analyze 101 articles and compare them to the additional six literature reviews obtained through the selection criteria.

To the best of our knowledge, no previous systematic reviews or meta-analyses have been conducted on the intersection of natural language processing and disaster management, focusing on all types of natural and man-made disasters and following the PRISMA 2020 guidelines. To fill these research gaps, the current work leverages the knowledge and efforts of others, published between 2014 and 2023 and focused on the above-mentioned areas.

In the following sections we aim to answer three research questions defined at the beginning of the study:

- RQ1: What are the most common Natural Language Processing applications, pre-processing techniques, and models used in the context of Disaster Management?
- RQ2: Which challenges derive from applying Natural Language Processing in Disaster Management?
- RQ3: What are the opportunities of using Natural Language Processing in Disaster Management?

5.1. NATURAL LANGUAGE PROCESSING APPLICATIONS IN DISASTER MANAGEMENT

Prior to analyzing challenges and opportunities it is important to understand in detail which natural language processing applications and techniques are used within DM. Table 7 provides a summary of all NLP applications, pre-processing techniques as well as NLP methods applied and analyzed in the 101 shortlisted articles. Below, we examine the collected and displayed data in an attempt to answer the first research question.

RQ1: What are the most common Natural Language Processing applications, pre-processing techniques, and models used in the context of Disaster Management?

Advances and disruptions in technology development have enabled disaster management actors such as governments, non-governmental organizations, and the impacted communities to share and collect data more effectively.

It is clear from our analysis that social media plays a big role in this exchange of information. Moreover, it is with no surprise that sentiment analysis is the most frequently used NLP application within the studied research, with 52 out of the 101 studies using this approach.

Conteras et al. (2022b), provides a brilliant summary of the characteristics and specifics within the area of Sentiment Analysis highlighting how it can be performed in three distinct classification levels (document level, sentence level, and individual words) and considering two main techniques: machine learning, using classifiers such as SVM, neural networks, and NB, or lexicon-based, when applying dictionary based or corpus-based approaches. Within the analyzed papers and, more specifically, when reverting to Table 7, we can note this distinction. Authors such as Bai and Yu (2016), Behl et al. (2021), Dereli et al. (2021), and Roy et al. (2020), applied machine learning techniques to extract sentiments out of natural language, while Gulnerman and Karaman (2020), Jain and Kumar (2018), and Terracciano and Han (2023), used lexicon-based techniques for the same purpose.

Still within the context of NLP Applications, text classification and information extraction seem to be the second and third most used approaches in disaster management. However, these are often applied in combination with or as an initial step of sentiment analysis.

While some of the studied articles are straightforward with the type of NLP models applied, others aren't as clear on the defined methods or share more complex algorithms tailored to their specific use cases. In particular, Farnaghi et al. (2020) propose an original dynamic spatio-temporal tweet mining method (DSTTM) to extract event information from geotagged tweets in large study areas. This method is based on a modified clustering structure approach (OPTICS) which combines five different techniques for text encoding and vectorization of Twitter messages (count vector, TF-IDF, Word2vec, GloVe, and FastText) with two metrics (Weighted Sum and Multiplication) to detect clusters for the subsequent step of topic extraction.

Concerning natural language processing methods applied in the humanitarian context, Latent Dirichlet Allocation (LDA) is frequently used for topic modeling, with 18,8% of reviewed authors applying it to their research. As an example, Resch et al. (2018), combined LDA for semantic information retrieval with local spatial autocorrelation to assess the footprint and damage produced by natural catastrophes using social media data. Likewise, 16 out of the 101 shortlisted papers used Bidirectional Encoder Representations from Transformers (BERT) and / or Long Short-Term Memory (LSTM) in their analysis. Namely, to identify rescue request tweets, Zhou et al. (2022) developed and compared ten models, three based on milestone NLP algorithms (GloVe; ELMo; XLNet) and seven BERT-based.

Furthermore, Term Frequency-Inverse Document Frequency (TF-IDF), a statistical approach to determine the significance of a word for one or a set of documents in a corpus (Ma et al., 2023), was also commonly applied in the shortlisted papers for text classification and information extraction. For instance, while exploring the potential of social media data to support disaster damage assessment, Li et al. (2023) trained seven candidate classification models, including the combination of TF-IDF with five distinct machine learning techniques and identified TF-IDF + Logistic Regression as the "best-performing" model.

Regarding machine learning techniques, Dereli et al. (2021) implemented and compared results of several algorithms applied to classification following sentiment analysis such as Naïve Bayes, Support Vector Machine, decision tree, random forest, neural network and k-nearest neighbor, and concluded that SVM obtained the highest value of accuracy while results from ACC and F-measure showed sentiment analysis consistency for SVM, NB, random forest and neural network algorithms. On another context, Eligüzel et al. (2020), performed sentiment analysis on Twitter data using the following machine learning algorithms: Multinomial and Gaussian Naïve Bayes, SVM, Decision Tree, Random Forest, Extra Trees, Neural Network, KNN, Stochastic Gradient Descent, and Adaptive Boosting (AdaBoost).

Lastly, pre-processing techniques such as noise and stop words removal are essential to avoid later problems when it comes to running complex NLP models (Gulnerman & Karaman, 2020). Stop words are challenging to interpret since they occur frequently yet have no deeper significance (Schwarz et al., 2023) while noise such as URLs, emojis, punctuation, and user mentions can significantly distort results. Furthermore, pre-processing techniques also need to be performed taking into account each specific problem and framework goal. For instance, Gulnerman and Karaman (2020), consciously chose not to perform stemming in a way to preserve the rich meaning differences with suffixes existent in the Turkish language.

In many cases, the distinction between NLP pre-processing techniques and NLP models is hardly noticeable. For instance, when it comes to Named Entity Recognition, which is an important task in NLP that seeks to retrieve meaningful entities and entity types from a vast quantity of texts (Lei et al., 2023), researchers may use this technique as a way to obtain the required data to perform further NLP tasks. Additionally, authors may simply apply NER as a means to an end without involving any other NLP techniques. Namely, Hu et al. (2019), used a combination of off-the-shelf and retrained NER models to examine the textual content of housing advertisements, and extract place name candidates for further geospatial analysis.

These and other NLP methods and techniques, when applied correctly and successfully are important tools from which the humanitarian context can also benefit.

Table 7 - Summary of included literature and respective NLP approaches

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Afyouni et al., 2022)	Event Detection; Information Extraction; Text Classification	Noise Removal; Stop Words Removal; Tokenization	KeyBERT; CNN; Bidirectional LSTM.
(Afyouni et al., 2023)	Event Detection	Noise Removal; Stemming; Stop Words Removal; Tokenization	Original / Customized Model: E-ware system composed by two main steps: Incremental Clustering of SpatioTemporal Events and Event detection and spatiotemporal scope.
(Alam et al., 2020)	Sentiment Analysis; Text Classification; Topic Modeling	Lowercasing; Named Entity Recognition; Noise Removal; Stop Words Removal	Sentiment Analysis: Stanford sentiment analysis classifier; Text Classification: Bag of Words; Topic Modeling: LDA; Named Entity Recognition: Stanford NER toolkit.
(Alkhatib et al., 2019)	Text Classification	Named Entity Recognition; Noise Removal; Stemming; Stop Words Removal; Tokenization	Polynomial Networks; NB; KNN; Rachio; SVM.
(An et al., 2022)	Sentiment Analysis	Noise Removal	Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis (SKEP)
(Bai & Yu, 2016)	Sentiment Analysis; Text Classification	Not Described	Authors combined 4 ML classifiers (SVM; NB; KNN; Random Forest) with 2 text representation methods (Bag of Words; Distribution Representation of Words).
(Barker et al., 2019)	Information Extraction; Text Classification	Lowercasing; Noise Removal; Tokenization	Authors leveraged NLP capabilities of Gensim and SciKit Learn libraries including a pre-trained Gensim Doc2Vec Model and SciKit Learn Logistic Regression classifier.

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Bashir et al., 2021)	Sentiment Analysis	Noise Removal; Stop Words Removal; Tokenization	Orange Data Mining Software toolkit; VOSviewer
(Behl et al., 2021)	Sentiment Analysis; Text Classification	Lemmatization; Lowercasing; Noise Removal Stop Words Removal	Pre-Processing: Natural Language Toolkit (NLTK) library + WordNetLemmatizer; Word Embedding / Bag of Words: Word2vec; Text Classification / Sentiment Analysis: LR-TF; CNN-W; CNN-WF; MLP-TF; MLP-W
(Cerna et al., 2022)	Information Extraction; Text Classification	Lemmatization; Stop Words Removal; Tokenization	LSTM; Convolutional Neural Network (CNN); FlauBERT; CamemBERT
(Chan & Tsai, 2019)	Dialogue System; Question Answering; Text Classification	Named Entity Recognition	Original / Customized Model
(Chen & Lim, 2021)	Text Classification	Tokenization	BERT
(Contreras et al., 2022a)	Information Extraction; Sentiment Analysis	Noise Removal	Classification algorithm developed by MonkeyLearn
(Contreras et al., 2022b)	Sentiment Analysis	Noise Removal	Grammarly; MonkeyLearn pre-trained Sentiment Analysis model
(da Costa et al., 2022)	Text Classification	Noise Removal; Stop Words Removal; Tokenization	Multimodal Fusion models; BOW; TF–IDF; Word Embeddings of a Word2Vec and FastText type where both are the type of Skip-Gram and Continuous Bag of Words (CBOW)
(Dereli et al., 2021)	Sentiment Analysis; Text Classification; Topic Labeling; Topic Modeling	Lowercasing; Noise Removal; Stemming; Stop Words Removal; Tokenization	Topic Modeling: LDA + Bag of Words Model (Bag of Ngram model); Sentiment Analysis: NB, SVM, decision tree, random forest, neural network and KNN

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Dong et al., 2021)	Sentiment Analysis; Text Classification	Lowercasing; Noise Removal; Stop Words Removal	Text Classification: TF-IDF; Sentiment Analysis: LR, NB, DT, SVM, KNN, RF, Adaboost, and MNN
(Eligüzel et al., 2020)	Sentiment Analysis; Topic Modeling	Noise Removal; Stemming; Stop Words Removal; Tokenization	Topic Modeling: LDA; Sentiment Analysis: Multinomial and Gaussian NB, SVM, Decision Tree, Random Forest, Extra Trees, Neural Network, k Nearest Neighbor (kNN), Stochastic Gradient Descent (SGD), and Adaptive Boosting (AdaBoost) classifications.
(Fan et al., 2020)	Event Detection; Information Extraction; Text Classification	Named Entity Recognition; Noise Removal	Fine-tuned BERT model; Stanford NER tool; Text Classification: TF-IDF
(Fang et al., 2019)	Information Extraction	Not Described	Word Frequency Analysis; PicData
(Farnaghi et al., 2020)	Event Detection; Information Extraction	Lemmatization; Lowercasing; Noise Removal; Stop Words Removal; Tokenization	Original / Customized Dynamic spatio-temporal tweet mining (DSTTM) method based on a modified clustering structure (OPTICS) approach which combines 5 different methods for text encoding and vectorization of Twitter messages – Count vector, TF-IDF, Word2vec, GloVe, and FastText - and 2 different metrics - Weighted Sum and Multiplication - to merge the spatial distance, temporal distance, and text similarity of tweets. Afterwards, the authors apply a new topic extraction algorithm, called Hierarchical Dirichlet Process (HDP).
(Gulnerman & Karaman, 2020)	Sentiment Analysis; Text Classification	Noise Removal; Stop Words Removal; Tokenization	N-Grams Creation; Sentiment Analysis: SentiTurkNet (STN) lexicon + The Lexicon of Ozturk and Ayvaz (LOA); Text Classification / Filtering: NB; Neural Network; SVM
(Guo et al., 2018)	Information Extraction	Parsing; POS Tagging; Tokenization	Original/ Customized Model using: Stanford Parser; SUNDANCE software

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Guo et al., 2020)	Information Extraction	POS Tagging	Bi-LSTM-CRF
(Guo et al., 2021)	Sentiment Analysis; Topic Modeling	Noise Removal	Topic Modeling: LDA; Sentiment Analysis: Original / Customized Model
(Hampton & Shalin, 2017)	Information Extraction; Sentiment Analysis	Not Described	Sentiment Analysis: SentiStrength; Information Extraction: Linguistic Inquiry Word Count software (LIWC).
(He et al., 2023)	Sentiment Analysis; Topic Modeling	Lemmatization; Lowercasing; Noise Removal; Parsing	JockersRinker sentiment lexicon + NRC emotion lexicon; First-order autoregression models
(Hu et al., 2019)	Information Extraction	Named Entity Recognition	4 NER models: spaCy NER, default Stanford NER, case-insensitive Stanford NER, and Twitter-retrained Stanford NER
(Hu et al., 2023)	Question Answering	Not Described	GPT model
(Huang et al., 2022)	Information Extraction; Text Classification	Segmentation; Stop Words Removal	TF-IDF; Authors integrated social media data analysis into an MCDA based framework that combines AHP, EWM (the entropy weight method), and grey TOPSIS to evaluate the relief demands urgency.
(Huo et al., 2020)	Information Extraction; Text Classification	POS Tagging	BERT-BiLSTM-CRF
(Jain & Kumar, 2018)	Sentiment Analysis; Text Classification; Topic Modeling	Lemmatization; Noise Removal; Stemming; Stop Words Removal; Tokenization	Text Classification: NB; SVM Sentiment Analysis: SentiwordNet AFINN lexicon dictionary; Ekman's emotion theory. Topic Modeling: LDA

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Joloudari et al., 2023)	Sentiment Analysis	POS Tagging; Stemming; Tokenization; Other	BERT; Deep CNN
(Karami et al., 2020)	Sentiment Analysis; Topic Modeling	Not Described	<u>Original / Customized Twitter Situational Awareness (TwiSA) Method:</u> 1. <u>Sentiment Analysis:</u> Linguistic Inquiry and Word Count (LIWC) software; 2. <u>Topic Modeling:</u> LDA
(Karimiziarani & Moradkhani, 2023)	Sentiment Analysis; Topic Modeling; Text Classification	Named Entity Recognition; Noise Removal; Stop Words Removal	Valence Aware Dictionary for Sentiment Reasoner (VADER).
(Karimiziarani et al., 2022)	Information Extraction; Text Classification; Topic Modeling	Noise removal; Stop Words Removal	LDA
(Karimiziarani et al., 2023)	Sentiment Analysis; Topic Modeling; Text Classification	Named Entity Recognition; Noise Removal; Stop Words Removal	VADER
(Karmegam & Mappillairaju, 2021)	Sentiment Analysis	Not Described	SentiStrength tool
(Kellner et al., 2023)	Information Extraction	Named Entity Recognition; Tokenization	BERT; Bi-directional LSTM (BD-LSTM) + 2 independent Recurrent Neural Networks (RNNs).
(Khattar & Quadri, 2022)	Text Classification	Lowercasing; Noise Removal; Stop Words Removal	Cross-Attention Multi-Modal (CAMM) deep neural network with the Bi-LSTM model to extract the features for classification.
(Kim et al., 2022)	Information Extraction; Question Answering	Tokenization	SBERT; BERT

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Lee & Yu, 2020)	Sentiment Analysis	Not Described	Linguistic Inquiry and Word Count (LIWC; theoretically developed dictionaries)
(Lei et al., 2023)	Information Extraction	Named Entity Recognition	Original / Customized Model: Semi-supervised geological disasters NER approach (Semi-GDNER) with two stages: 1. Transfer the parameters of the pre-trained BERT-base model to the BERT layer of the BERT-BiLSTM-CRF backbone model; 2. Continue training the backbone model using a self-training strategy with unlabeled geohazards situation reports. The authors also compared the Semi-GDNER model with other NER approaches (BERT-Softmax, RoBERTa-Softmax, BERT-CRF, BiLSTM-CRF, and BiLSTM- Attention-CRF).
(Li et al., 2023)	Text Classification	Lemmatization; Named Entity Recognition; Noise Removal; Tokenization	The author trained seven candidate classification models in order to identify the best-performing model: TF-IDF + DT, TF-IDF + RF, TF-IDF + NB, TF-IDF + SVM, TF323 IDF + LR, GloVe + LSTM, and FastText + LSTM. TF-IDF + LR is selected as the “best-performing” model.
(Lian et al., 2020)	Information Extraction; Sentiment Analysis; Text Classification	POS Tagging; Stop Words Removal	Pre-Processing: Jieba Toolkit; Information Extraction: TF-IDF + density-based K-means algorithm; Sentiment Analysis: ROST Emotion Analysis Tool; Text Classification: manual-sampling-based, dynamic, incremental clustering algorithm (MS-DICA).
(Liu et al., 2019)	Information Extraction	Lowercasing; Noise Removal; Stemming; Stop Words Removal; Tokenization	Bag of Words Model; Cosine similarity function
(Lu et al., 2023)	Sentiment Analysis; Text Classification	Name Entity Recognition; Other	Pre-Processing: Jieba; Text Classification: Initially, the authors compared SVM and NB and selected SVM as it performed better; Sentiment Analysis: DLUT Emotion Ontology

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Lyu & Lu, 2023)	Machine Translation; Sentiment Analysis; Text Classification; Topic Modeling	Lemmatization; Lowercasing; Noise Removal; Stop Words Removal; Tokenization	Machine Translation: Google Translator; BLEU Score; Text Classification: BERT; Sentiment Analysis: TextBlob; VADER; Topic Modeling: GSDMM
(Ma et al., 2022)	Information Extraction	Not Described	BERT-BiLSTM-CRF
(Ma et al., 2023)	Sentiment Analysis; Text Classification	Noise Removal; Tokenization; Other	Sentiment Analysis: Baidu NLP; Text Classification: TF-IDF
(Matsuoka & Rocha, 2021)	Sentiment Analysis	Noise Removal; Stop Words Removal	Original / Customized Model
(Mughal et al., 2021)	Question Answering	Not Described	Original / Customized Model: ORFFM which uses NLP-based API and software tools (QAKIS) to enable the translation of the natural language to SPARQL query.
(Mukherjee et al., 2022)	Information Extraction; Text Classification; Topic Modeling	Stop Words Removal	N-Grams creation; Information Extraction: TF-IDF; Topic Modeling: LDA; Text Classification: NB; Maximum entropy-based classification
(Neppalli et al., 2017)	Sentiment Analysis	Not Described	SentiStrength algorithm; NB; SVM
(Nguyen et al., 2022)	Information Extraction; Text Classification	Lemmatization; Named Entity Recognition; Parsing; POS Tagging; Stop Words Removal	Pre-Processing: Wordnet Library; Information Extraction: 2 LSTM Models; Text Classification: SVM
(Ni et al., 2023)	Information Extraction	POS Tagging; Other	Bi-directional Long Short Term Memory (Bi-LSTM) networks; Conditional Random Fields (CRF); Interval-valued 2-tuple linguistic representation model

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Nimmi & Janet, 2023)	Text Classification	Lowercasing; Noise Removal; Stemming	Pre-Processing: NLTK library + Regular Expressions (Regex). Text Classification: Authors tested several models such as AlBERT; BERT; RoBERTa; MobileBERT; and proposed an Original / Customized Model: PLE-MobileBERT which outperformed the above methods.
(Oz et al., 2018)	Sentiment Analysis	Not Described	NLTK implementation of VADER
(Paradkar et al., 2022)	Event Detection	Not Described	Semiautomated social media analytics approach for social sensing of Disaster Impacts and Societal Considerations (SocialDISC)
(Perlstein & Verboord, 2021)	Sentiment Analysis; Topic Modeling	Noise Removal	Sentiment Analysis: SentiStrength Software; Topic Modeling: ConText Software.
(Ragini et al., 2018)	Sentiment Analysis; Text Classification	Noise Removal; POS Tagging	Bag of Words Model; N-Grams Creation; Text Classification: SVM
(Raisio et al., 2022)	Sentiment Analysis	Not Described	Original / Customized Model
(Rashid & Wang, 2022)	Event Detection; Information Extraction	Noise Removal	Original / Customized SAS framework (SEIS) which consists of: 1. a Social Data Extraction (SDE) module (uses a keyword-driven NLP-based search tool named Quickview); 2. an Event Investigation (EI) module; 3. a Correlation Analytics (CA) module; 4. a UAV Dispatch (UD) module.
(Resch et al., 2018)	Information Extraction; Topic Modeling	Lowercasing; Noise Removal; Stemming; Stop Words Removal; Tokenization; Other	LDA

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Rexiline Ragini et al., 2018)	Sentiment Analysis; Text Classification	Named Entity Recognition; Noise Removal; POS Tagging	Bag of Words Model; N-Grams Creation; NB; Decision tree; SVM
(Reynard & Shirgaokar, 2019)	Sentiment Analysis	Not Described	Multinomial logit model; TextBlob
(Rossi et al., 2018)	Event Detection; Text Classification	Lemmatization; Named Entity Recognition	<u>Event Detection:</u> Original / Customized volume-based Event Detection Module which builds upon the generalized Extreme Studentized Deviate test (ESD); <u>Text Classification:</u> fasttext tool developed in the Facebook AI Research group including N-Grams Creation.
(Roy et al., 2020)	Sentiment Analysis; Text Classification	Lemmatization; Named Entity Recognition; Noise Removal; Stemming; Stop Words Removal; Tokenization	<u>Text Classification:</u> TF-IDF with N-Grams Creation; <u>Sentiment Analysis:</u> 8 models: Multinomial NB, logistic regression (LR), KNN, SVC, random forest (RF), DT, multilayer perceptron (MLP), and deep neural network (DNN) applied on 3 approaches: Binary Relevance (BR); Label Power set (LP); Random K-Labelsets (RAKEL). <u>Named Entity Recognition:</u> NLTK
(Rusho et al., 2021)	Topic Modeling	Noise Removal; Stop Words Removal	N-grams Creation; <u>Topic Modeling:</u> LDA + GSDMM
(Ryan et al., 2021)	Sentiment Analysis; Text Classification	Not Described	Stanford CoreNLP sentiment annotator; Valence Aware Dictionary; VADER; Multinomial Logistic Regression
(Sarkar et al., 2023)	Question Answering	Not Described	Supervised attention-based visual question answering framework (SAM-VQA)

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Scalia et al., 2022)	Information Extraction	Named Entity Recognition	CIME geolocation algorithm (CoreNLP NER + Gazetteer)
(Schwarz et al., 2023)	Event Detection; Text Summarization	Lemmatization; Lowercasing; Noise Removal; Stop Words Removal	Event Detection: TF-IDF; CrisisLex; Locality-sensitive hashing method named SimHash; Text Summarization: SpaCy; NLTK; Gensim libraries
(Sermet & Demir, 2018)	Dialogue System	Lemmatization; Parsing	Original / Customized Model
(Shahbazi & Byun, 2022)	Event Detection; Sentiment Analysis; Text Classification	Not Described	Original / Customized architecture with four main components: event identification, automatic reasoning, incident monitoring and blockchain. Mentioned Models: Bag of Words Model; LDA Authors also compared the following algorithms: NB, KNN, SVM, Logistic Regression, XGBoost, and DL.
(Shibuya & Tanaka, 2019)	Sentiment Analysis; Text Classification	Not Described	Text Classification: NB for Facebook; SVM for Twitter. Sentiment Analysis: Japanese open sentiment polarity dictionaries
(Singh et al., 2019)	Event Detection; Sentiment Analysis; Topic Modeling	Lowercasing; Noise Removal; Stop Words Removal	Topic Modeling: LDA + GSDMM; Sentiment Analysis: NRC Emotion Lexicon + various word embedding models.
(Sit et al., 2019)	Event Detection; Information Extraction; Topic Modeling	Lowercasing; Noise Removal; Tokenization	Authors perform an experiment using a series of classification methods (Logistic Regression; Linear SVM; Ridge; CNN; LSTM Networks) and end up employing LSTM networks and LDA
(Stieglitz et al., 2022)	Dialogue System	Not Described	Not Described

Citation	**NLP Applications**	**NLP Pre-Processing Techniques**	**NLP Method**
(Sugino et al., 2023)	Topic Modeling	POS Tagging	Bag of Words Model; Biterm Topic Model (BTM)
(Terracciano & Han, 2023)	Sentiment Analysis; Topic Modeling	Not Described	Bing lexicon dictionary; Eigenvector centrality algorithm
(Tian et al., 2023)	Sentiment Analysis; Topic Modeling	Lowercasing; Noise Removal	Topic Modeling: BERT; Sentiment Analysis: VADER
(Valinejad & Mili, 2023)	Information Extraction; Sentiment Analysis	Not Described	Sentiment Analysis: Linguistic Inquiry and Word Count (LIWC) program; Information Extraction: Web Scraper; Original / Customized Model: Multiagent Cyber–Physical–Social Model with N-Grams Creation
(Wahid et al., 2022)	Information Extraction; Topic Modeling; Text Classification.	Lemmatization; Lowercasing; Noise Removal; Stop Words Removal; Tokenization	N-Grams Creation; Original / Customized Topic2labels (T2L) framework: 1. Used LDA topic model to label data; 2. Applied BERT model to extract features to be used in classification algorithms. 3. The authors utilized DL algorithms suited for text classification such as CNN, LSTM and ANN.
(Wang et al., 2020)	Information Extraction	Named Entity Recognition	NeuroNER (LSTM based)
(Wang et al., 2021)	Sentiment Analysis	Not Described	VADER
(Wang et al., 2023a)	Sentiment Analysis	Parsing	VADER
(Wang et al., 2023b)	Information Extraction	Named Entity Recognition; Other	Pre-trained model multi-feature graph convolutional network (PTM-MFGCN)
(Wu et al., 2022)	Sentiment Analysis; Topic Modeling	Not Described	Topic Modeling: LDA; Sentiment Analysis: SnowNLP library

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Wu et al., 2023)	Information Extraction; Text Classification	Named Entity Recognition; Noise Removal; Parsing	Pre-Processing: JioNLP; Information Extraction: Method of Identifying Spatiotemporal Information of Social Media Multimodal Data (MIST-SMMD) with included LoFTRSeg Geo-Localization Method (LSGL)
(Xia et al., 2023)	Sentiment Analysis	Not Described	BERT; Multilingual Model; Multiperiod difference-in-differences (DID) regression model; Twitter sentiment geographical index (TSGI)
(Yan et al., 2020)	Sentiment Analysis; Topic Modeling	Lemmatization; Noise Removal; Stop Words Removal; Tokenization.	Sentiment Analysis: TextBlob library; Topic Modeling: LDA; Stop Words Removal: NLTK
(Yao & Wang, 2020)	Information Extraction; Sentiment Analysis	Noise Removal	Domain specific sentiment analysis approach specifically for tweets posted during hurricanes (DSSA-H): 1. Information Extraction: NB, SVM, Random Forest (RF), and fully connected neural networks; 2. Sentiment Analysis: Domain-adversarial neural network (DANN) method.
(Yao et al., 2021)	Information Extraction	Noise Removal; Parsing	Original / Customized Model
(Yuan et al., 2021)	Sentiment Analysis; Topic Modeling	Noise Removal; Stop Words Removal; Tokenization	Sentiment Analysis: VADER Topic Modeling: LDA
(Zander et al., 2023)	Sentiment Analysis; Topic Modeling	Lowercasing; Lemmatization; Stop Words Removal; Noise Removal; Tokenization	N-Grams Creation; Bag of Words Model; GSDMM

Citation	NLP Applications	NLP Pre-Processing Techniques	NLP Method
(Zhang et al., 2022)	Sentiment Analysis	Parsing; POS Tagging; Tokenization	CNN-LSTM
(Zheng et al., 2021)	Sentiment Analysis	Not Described	ROST Content Mining System with Bag of Words Model and N-Grams Creation; Emotion Vocabulary of Dalian University of Technology
(Zhou et al., 2022)	Information Extraction; Text Classification	Tokenization; Other	GloveTransformers; ELMoTransformers; BERT-Linear; RoBERTa-Linear; DistilBERT-Linear; ALBERT-Linear; XLNet-Linear; BERT-Nonlinear; BERT-LSTM; BERT-CNN
(Zhou et al., 2023)	Information Extraction; Text Summarization; Topic Modeling	Lowercasing; POS Tagging; Lemmatization; Stop Words Removal; Noise Removal; Tokenization	LDA; TF-IDF
(Zhuang et al., 2021)	Information Extraction; Sentiment Analysis; Topic Modeling	Noise Removal; Stop Words Removal	Pre-Processing: LTP word segmentation tool; LDA-ARMA hybrid model

5.2. NATURAL LANGUAGE PROCESSING IN DISASTER MANAGEMENT CHALLENGES

While the use of NLP helps in predicting, assessing, and supporting disaster management, there are a few challenges that must be addressed before the existing benefits fade away. Table A 1 provides an overview of those issues to support an answer to research question number two.

RQ2: Which challenges derive from applying Natural Language Processing in Disaster Management?

As mentioned, through Table A 1 we can identify an extensive list of challenges derived from applying NLP to disaster management. Within the consolidated list of issues, two main areas can be identified: challenges related with used data sources, and challenges associated with applied NLP methods.

Taking into consideration the challenges related to the applied data sources, difficulties in working with social media data is a common topic. As mentioned by Wang et al. (2020), although social media data can be collected and stored in real-time, extracting this type of information is difficult, requiring a speedy and automatic processing tool to help deal with the huge volume, fast velocity, and significant heterogeneity. Additional concerns involve the ethical and privacy considerations of other data sources besides Twitter (Zander et al., 2023), the existence of misinformation (Lian et al., 2020; Lyu & Lu, 2023; Ahmad et al., 2023; Wang et al., 2020; Zander et al., 2023), the presence of noise, heterogeneity, and sparsity associated with publicly generated data (Fan et al., 2020; Hampton & Shalin, 2017; Lu et al., 2023; Schwarz et al., 2023; Sit et al., 2019; Wu et al., 2023), as well as the bias within the demographic composition of users (Fang et al., 2019; Huang et al., 2022; Li et al., 2023; Xia et al., 2023). Issues with other data sources are also mentioned such as the limitations in assessing public opinions and beliefs from traditional survey methods (Tian et al., 2023) and the incomplete information gathering and high costs of other traditional data sources, such as official reports (Lu et al., 2023).

On the other hand, referring to NLP applications and methods, sentiment analysis is the most mentioned topic, as would be expected. While some authors mention the difficulty of NLP models to classify sarcastic sentences (Contreras et al., 2022a; Contreras et al., 2022b; Lyu & Lu, 2023), others highlight how methods within sentiment analysis are less accurate when applied to non-English languages (Bai & Yu, 2016; Perlstein & Verboord, 2021). Also, simply classifying sentiments as positive, neutral and negative is not “good enough to understand the detailed emotions of the public during the crisis event” (Singh et al., 2019).

Due to the low maturity of NLP methods, analysts and researchers are often led to manually perform certain steps of the analysis process. Chan and Tsai (2019) constitute a clear example of this concern, as they segmented records by hand since the existing tools, Jieba and Stanford Core NLP, could not handle the high level of professionality required for disaster management.

In regards to machine learning techniques, Behl et al. (2021), mention that while supervised learning isn't the best strategy during emergencies because it requires the tagging of large social media datasets, which is costly and time-consuming, the absence of rule-based structural grouping also places limitations on unsupervised or semi-supervised learning.

5.3. NATURAL LANGUAGE PROCESSING IN DISASTER MANAGEMENT OPPORTUNITIES

As seen, NLP in disaster management refers to the application of techniques to extract and evaluate important information throughout all disaster management phases using different sources of data. This allows humanitarian actors to improve disaster aid and resilience. To continue the excellent efforts that have been done and to overcome the challenges mentioned above, it is important to investigate how the automatic processing of natural language can positively support the humanitarian context. In that sense, Table A 1 also provides a summary of the advantages of NLP in disaster management mentioned in the shortlisted articles. Below, we will highlight the most common benefits with the goal of answering the third and last research question.

RQ3: What are the opportunities of using Natural Language Processing in Disaster Management?

During a disaster, the safety of those affected is always the top priority of decision-makers, such as governments. The timely discovery of the locations of victims, as well as the dispatch of rescue personnel, can help to avert further casualties. Faster access to the number and distribution of victims can help decision-makers determine the scope of the disaster and carry out disaster relief activities more efficiently.

In this sense, most authors highlight the importance and possibilities of applying sentiment analysis in disaster management using social media data. On one hand, despite its already mentioned challenges, social media data is largely available and for free at any time, allowing anyone to promptly collect vast amounts of disaster-related information to analyze and resolve several important issues. As mentioned by Karimiziarani and Moradkhani (2023), information from social media can help states and local governments, law enforcement, fire departments, and other authorities prepare for emergencies and handle them more successfully. Similarly, Bashir et al. (2021) discuss how " Twitter is used as a tool to express deep emotions of sadness over the loss of innocent lives, sending out grief and comforting messages, and is beneficial for providing practical support to the people who suffer during the human-made crisis." (p. 6). On the other hand, sentiment analysis techniques used in disaster management aid in improving crisis management and recovery operations planning (Karmegam & Mappillairaju, 2021).

Some authors also emphasize the benefits of combining different NLP approaches with each other or even with other methods to obtain superior results. For example, Terracciano and Han (2023) mention how "combining sentiment analysis with semantic network analysis also addresses the issue of interpretability within traditional sentiment analysis" (p. 4). As Raisio et al. (2022) underline, complementary qualitative sentiment analysis can add context and help analysts comprehend the subtleties of distinct sentiments in order to help increase situational awareness at various stages of disasters.

Regarding NLP methods, Resch et al. (2018) states that LDA has a significant benefit in terms of transferability to various text corpora and languages due to the quick adaption of the unsupervised learning technique to another text corpus, while Zhuang et al. (2021) concludes that the LDA model with the Gaussian function performed better when it came to topic clustering. In addition, other articles draw attention to the benefits of BERT, including requiring cheaper and less time-consuming fine-tuning when compared to other models (Tian et al., 2023).

Several authors, including Wang et al. (2020), also highlight the need for and importance of having a combination of automatic and manual processes set in place in order to achieve better quality results. However, as mentioned before, it seems like the application of manual actions within NLP frameworks, is also a resulting force of the current state and capacity of NLP models (Huo et al., 2020). Once existing frameworks and new models are tested and generated in complex and time-critical applications, the need to use manual processes could reach a minimal state.

In conclusion, we can see that, while challenges in NLP may be amplified when used within the complex contexts of disaster management, there are various benefits that might have a significant impact on these types of projects, and which require further investigation.

5.4. CONTRIBUTION TO LITERATURE

In terms of contribution to literature, this study fills a theoretical research gap by merging interdisciplinary concepts from many research domains, such as artificial intelligence and disaster management, to better comprehend the impact of NLP in a disaster management context. Equally important, the development of a systematic review offers an evident analysis of relevant research, as well as unified ideas for how to successfully incorporate NLP into existing and future humanitarian efforts. Thus, this study provides an ideal foundation of concepts and methods to enable the development and refinement of significant Natural Language Processing projects, which, when utilized in real-world disasters, may save and preserve human lives.

5.5. LIMITATIONS

Our study is not without its limitations. Firstly, although we strived to perform a comprehensive data analysis, certain eligibility criteria influenced the obtained results. For instance, a different choice of keywords could have led to a distinct set of initial documents obtained from Scopus. In addition, only studies in English were considered, reducing the scope of our work. Secondly, the documents available for download available at the time of export may change based on date and access. Lastly, the literature revision and selection were done by only one author which may increase the risk of bias. This challenge was partially overcome by undertaking several reading iterations and reviewing and adjusting all collected data multiple times.

5.6. FUTURE RESEARCH DIRECTIONS

In this paper we presented an analysis of natural language processing methods applied to different disaster contexts and further explored the challenges and advantages of such projects.

More studies should be applied to the phases of mitigation, preparedness, and recovery as well as to less common NLP applications within academic research such as machine learning, question answering and dialogue systems. Additionally, considering the challenges and opportunities explored in sections 5.2 and 5.3, respectively, it would be beneficial to optimize existing models and create new, more efficient, and scalable methods that can perform better in the context of disaster management. Finally, in the future, as disasters can happen anywhere and often affect vulnerable communities the most, researchers may also apply and analyze the results of sentiment analysis to less common languages perhaps even with the help of MT tools.

6. CONCLUSION

In this paper, we discussed the challenges and potential of applying NLP to disaster management. In doing so, we analyzed 101 peer-reviewed articles and six literature reviews and identified important trends within those papers. With that, we aimed to (1) identify the range of natural language processing methods used in disaster management, (2) explore the main challenges and opportunities in disaster management when using natural language processing, and (3) determine how this field impacts the four phases of disaster as well as the nature of such impact. To the best of our knowledge, this is the first systematic review or meta-analysis following the PRISMA 2020 guidelines that has been conducted on the intersection of natural language processing and disaster management, focusing on all types of natural and man-made disasters.

The findings indicate that that social media and sentiment analysis have had an exponential growth in applications and importance within the field of disaster management. Identifying a balance between the challenges and opportunities that social media delivers can be difficult, especially in a context such as disaster response. Thus, it is essential to continue research in this field so that it can be safely applied when required.

Furthermore, we note that the limitations found in this study, as well as the recommended future objectives, will assist the research community in identifying and pursuing prospective research topics to better assist humanitarian organizations during disasters.

BIBLIOGRAPHICAL REFERENCES

Adamson, D. (2023, February 27). Understanding Text Classification In Natural Language Processing - A Beginners' Guide. Retrieved from https://www.linkedin.com/pulse/understanding-text-classification-natural-language-david-adamson-mbcs/

Afyouni, I., Khan, A., & Al Aghbari, Z. (2022). Deep-Eware: spatio-temporal social event detection using a hybrid learning model. *Journal of Big Data, 9*(86), 1-21. doi:https://doi.org/10.1186/s40537-022-00636-w

Afyouni, I., Khan, A., & Al Aghbari, Z. (2023). E-ware: a big data system for the incremental discovery of spatio-temporal events from microblogs. *Journal of Ambient Intelligence and Humanized Computing, 14*, 13949-13968. doi:https://doi.org/10.1007/s12652-022-04104-4

Ahmad, P., Liu, Y., Khan, K., Jiang, T., & Burhan, U. (2023). BIR: Biomedical Information Retrieval System for Cancer Treatment in Electronic Health Record Using Transformers. *Sensors, 23*(9355). doi:https://doi.org/10.3390/s23239355

Akter, S., & Wamba, S. (2019). Big data and disaster management: a systematic review and agenda for future research. *Annals of Operations Research, 283*, 939-959. doi:10.1007/s10479-017-2584-2

Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., . . . Al-Muhanna, F. A. (2023). A Review of the Role of Artificial Intelligence in Healthcare. *Journal of Personalized Medicine, 13*(951). doi:https://doi.org/10.3390/jpm13060951

Alam, F., Ofli, F., & Imran, M. (2020). Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria. *Behaviour & Information Technology, 39*(3), 288-318. doi:https://doi.org/10.1080/0144929X.2019.1610908

Albalawi, R., Yeap, T., & Benyoucef, M. (2020). Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. *Frontiers in Artificial Intelligence, 3*(42), 1-14. doi:https://doi.org/10.3389/frai.2020.00042

Alexander, D. E. (2002). *Principles of Emergency Planning and Management.* Oxford University Press.

Alexander, D. E. (2019). L'Aquila, central Italy, and the "disaster cycle", 2009-2017. *Disaster Prevention and Management, 28*(4), 419-433. doi:https://doi.org/10.1108/DPM-01-2018-0022

Al-Garadi, M. A., Yang, Y.-C., & Sarker, A. (2022). The Role of Natural Language Processing during the COVID-19 Pandemic: Health Applications, Opportunities, and Challenges. *Healthcare, 10*(11). Retrieved from https://doi.org/10.3390/healthcare

Alkhatib, M., El Barachi, M., & Shaalan, K. (2019). An Arabic social media based framework for incidents and events monitoring in smart cities. *Journal of Cleaner Production, 220*, 771-785. doi:https://doi.org/10.1016/j.jclepro.2019.02.063

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., Bin Saleh, K., . . . Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy, 19*, 1236-1242. doi:https://doi.org/10.1016/j.sapharm.2023.05.016

Altay, N., & Green, W. G. (2006). OR/MS research in disaster operations management. *European Journal of Operational Research, 175*(1), 475-493. doi:https://doi.org/10.1016/j.ejor.2005.05.016

Altay, N., & Labonte, M. (2014). Challenges in humanitarian information management and exchange: evidence from Haiti. *Disasters, 38*(1), S50-S72. doi:https://doi.org/10.1111/disa.12052

An, L., An, N., Li, G., & Yu, C. (2022). Research on the dynamic mechanism of group emotional expression in the crisis. *Telematics and Informatics, 71*(101829), 1-15. doi:https://doi.org/10.1016/j.tele.2022.101829

Anaya-Arenas, A. M., Renaud, J., & Ruiz, A. (2014). Relief distribution networks: a systematic review. *Annals of Operations Research, 223*, 53-79. doi:https://doi.org/10.1007/s10479-014-1581-y

Bagheri, A., Taghvaeian, S., & Delen, D. (2023). A text analytics model for agricultural knowledge discovery and sustainable food production: A case study from Oklahoma Panhandle. *Decision Analytics Journal, 9*(100350). doi:https://doi.org/10.1016/j.dajour.2023.100350

Bai, H., & Yu, G. (2016). A Weibo-based approach to disaster informatics: incidents monitor in post-disaster situation via Weibo text negative sentiment analysis. *Natural Hazards, 83*, 1177-1196. doi:https://doi.org/10.1007/s11069-016-2370-5

Bangalore, S., Rambow, O., & Whittaker, S. (2000). Evaluation Metrics for Generation. *INLG'2000 Proceedings of the First International Conference on Natural Language Generation.* Mitzpe Ramon, Israel: Michael Elhadad. doi:10.3115/1118253.1118255

Bashir, S., Bano, S., Shueb, S., Gul, S., Mir, A., Ashraf, R., . . . Noor, N. (2021). Twitter chirps for Syrian people: Sentiment analysis of tweets related to Syria Chemical Attack.

International Journal of Disaster Risk Reduction, 62, 1-10. doi:https://doi.org/10.1016/j.ijdrr.2021.102397

Behl, S., Rao, A., Aggarwal, S., Chadha, S., & Pannu, H. (2021). Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises. *International Journal of Disaster Risk Reduction, 55*(102101), 1-13. doi:https://doi.org/10.1016/j.ijdrr.2021.102101

Berdyugina, D., & Cavallucci, D. (2023). Automatic extraction of inventive information out of patent texts in support of manufacturing design studies using Natural Languages Processing. *Journal of Intelligent Manufacturing, 34*, 2495-2509. doi:https://doi.org/10.1007/s10845-022-01943-y

Beroggi, G., & Wallace, W. (1998). *Operational Risk Management: The Integration of Decision, Communications, and Multimedia Technologies.* Springer.

Booth, P., Chaperon, S. A., Kennell, J. S., & Morrison, A. M. (2020). Entrepreneurship in island contexts: A systematic review of the tourism and hospitality literature. *International Journal of Hospitality Management, 85*, 102438. Retrieved from https://doi.org/10.1016/j.ijhm.2019.102438

Cai, M. (2021). Natural language processing for urban research: A systematic review. *Heliyon, 7*(3). doi:https://doi.org/10.1016/j.heliyon.2021.e06322

Carr, L. J. (1932). Disaster and the Sequence-Pattern Concept of Social Change. *American Journal of Sociology, 38*(2), 207-218. Retrieved from https://www.jstor.org/stable/2766454

Carter, W. N. (2008). *Disaster Management: A Disaster Manager's Handbook.* Asian Development Bank. Retrieved from https://www.adb.org/sites/default/files/publication/27890/disaster-management-handbook.pdf

Castillo, P. P., Soto, C., Asún, R. A., & Gutiérrez, F. (2023). Profiling support in literacy development: Use of natural language processing to identify learning needs in higher education. *Assessing Writing, 58*(100787). doi:10.1016/j.asw.2023.100787

Centre for Research on the Epidemiology of Disasters. (2023, October 25). *EM-DAT | The international disasters database*. Retrieved November 02, 2023, from https://www.emdat.be/

Centre for Research on the Epidemiology of Disasters. (n.d.). *Public EM-DAT.* Retrieved from EM-DAT - The International Disaster Database: https://public.emdat.be/

Chan, H.-Y., & Tsai, M.-H. (2019). Question-answering dialogue system for emergency operations. *International Journal of Disaster Risk Reduction, 41*(101313), 1-13. doi:https://doi.org/10.1016/j.ijdrr.2019.101313

Chang, J.-W., Yen, N., & Hung, J. C. (2022). Design of a NLP-empowered finance fraud awareness model: the anti-fraud chatbot for fraud detection and fraud classifcation as an instance. *Journal of Ambient Intelligence and Humanized Computing, 13*, 4663-4679. Retrieved from https://doi.org/10.1007/s12652-021-03512-2

Chen, Y., Huang, X., & Wu, Z. (2023). From natural language to accounting entries using a natural language processing method. *Accounting and Finance, 00*, 1-15. doi:10.1111/acfi.13067

Coman, A., & Vasilache, S. (2023a). Social Computing and Social Media: 15th International Conference, SCSM 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23–28, 2023, Proceedings, Part I. *HCII: International Conference on Human-Computer Interaction. 14025.* Copenhagen: Springer Cham. doi:https://doi.org/10.1007/978-3-031-35915-6

Coman, A., & Vasilache, S. (2023b). Social Computing and Social Media: 15th International Conference, SCSM 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23–28, 2023, Proceedings, Part II. *HCII: International Conference on Human-Computer Interaction. 14026.* Copenhagen: Springer Cham. doi:https://doi.org/10.1007/978-3-031-35927-9

Contreras, D., Wilkinson, S., Alterman, E., & Hervá, J. (2022a). Accuracy of a pre-trained sentiment analysis (SA) classification model on tweets related to emergency response and early recovery assessment: the case of 2019 Albanian earthquake. *Natural Hazards, 113*, 403-421. doi:https://doi.org/10.1007/s11069-022-05307-w

Contreras, D., Wilkinson, S., Balan, N., & James, P. (2022b). Assessing post-disaster recovery using sentiment analysis: The case of L'Aquila, Italy. *Earthquake Spectra, 38*(1), 1-28. doi:https://doi.org/10.1177/87552930211036486

da Costa, T. A., Meneguette, R. I., & Ueyama, J. (2022). Providing a greater precision of Situational Awareness of urban floods through Multimodal Fusion. *Expert Systems With Applications, 188*(115923), 1-19. doi:https://doi.org/10.1016/j.eswa.2021.115923

de Silva, M. (2023, April 30). Preprocessing Steps for Natural Language Processing (NLP): A Beginner's Guide. Retrieved from https://medium.com/@maleeshadesilva21/preprocessing-steps-for-natural-language-processing-nlp-a-beginners-guide-d6d9bf7689c9

Dereli, T., Eligüzel, N., & Çetinkaya, C. (2021). Content analyses of the international federation of red cross and red crescent societies (ifrc) based on machine learning techniques through twitter. *Natural Hazards, 106*, 2025-2045. doi:https://doi.org/10.1007/s11069-021-04527-w

Devi, D. V., Kumar, C. K., & Prasad, S. (2016). A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine. *2016 IEEE 6th International Conference on Advanced Computing.* Bhimavaram, India: IEEE. doi:10.1109/IACC.2016.11

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. *HLT '02: Proceedings of the second international conference on Human Language Technology Research*, (pp. 138-145). doi:https://dl.acm.org/doi/10.5555/1289189.1289273

Drus, Z., & Khalid, H. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. *Procedia Computer Science, 161*, 707-714. doi:10.1016/j.procs.2019.11.174

Eligüzel, N., Çetinkaya, C., & Dereli, T. (2020). Comparison of different machine learning techniques on location extraction by utilizing geo-tagged tweets: A case study. *Advanced Engineering Informatics, 46*(101151), 1-19. doi:https://doi.org/10.1016/j.aei.2020.101151

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. doi:10.1109/ACCESS.2020.2965550

Fang, J., Hu, J., Shi, X., & Zhao, L. (2019). Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm. *International Journal of Disaster Risk Reduction, 34*, 275-282. doi:https://doi.org/10.1016/j.ijdrr.2018.11.027

Farazmehr, S., & Wu, Y. (2023). Locating and deploying essential goods and equipment in disasters using AI-enabled approaches: A systematic literature review. *Progress in Disaster Science, 19*(100292), 1-12. doi:https://doi.org/10.1016/j.pdisas.2023.100292

Farnaghi, M., Ghaemi, Z., & Mansourian, A. (2020). Dynamic Spatio-Temporal Tweet Mining for Event Detection: A Case Study of Hurricane Florence. *International Journal of Disaster Risk Science, 11*, 378-393. doi:https://doi.org/10.1007/s13753-020-00280-z

Guarda, T., Portela, F., & Augusto, M. (2022). *Advanced Research in Technologies, Information, Innovation and Sustainability.* Springer. doi:https://doi.org/10.1007/978-3-031-20319-0

Gulnerman, A., & Karaman, H. (2020). Spatial Reliability Assessment of Social Media Mining Techniques with Regard to Disaster Domain-Based Filtering. *ISPRS International Journal of Geo-Information, 9*(245), 1-34. doi:https://doi.org/10.3390/ijgi9040245

Guo, D., Zhao, Q., Chen, Q., Wu, J., Li, L., & Gao, H. (2021). Comparison between sentiments of people from affected and non-affected regions after the flood. *Geomatics, Natural Hazards and Risk, 12*(1), 3346-3357. doi:https://doi.org/10.1080/19475705.2021.2012530

Hampton, A., & Shalin, V. (2017). Sentinels of Breach: Lexical Choice as a Measure of Urgency in Social Media. *Human Factors, 59*(4), 505-519. doi:https://doi.org/10.1177/0018720817691612

He, S., Li, D., Liu, C.-H., Xiong, Y., Liu, D., Feng, J., & Wen, J. (2023). Crisis communication in the WHO COVID-19 press conferences: Aretrospective analysis. *PLoS ONE, 18*(3), 1-12. doi:https://doi.org/10.1371/journal.pone.0282855

Hesselman, M., & Lane, L. (2017). Disasters and non-state actors – human rights-based approaches. *Disaster Prevention and Management, 26*(5), 526-539. doi:10.1108/DPM-07-2017-0174

Hillyer, M. (2020, November 18). *How has technology changed - and changed us - in the past 20 years?* Retrieved from World Economic Forum: https://www.weforum.org/agenda/2020/11/heres-how-technology-has-changed-and-changed-us-over-the-past-20-years/

Hirschman, L., & Gaizauskas, R. (2001). Natural language question answering: the view from here. *Natural Language Engineering, 7*(4), 275-300. doi:https://doi.org/10.1017/S1351324901002807

Hossain, E., Rana, R., Higgins, N., Soar, J., Barua, P. D., Pisani, A. R., & Turner, K. (2023). Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. *Computers in Biology and Medicine, 155*, 106649. Retrieved from https://doi.org/10.1016/j.compbiomed.2023.106649

Hou, Q., Han, M., & Cai, Z. (2020). Survey on Data Analysis in Social Media: A Practical Application Aspect. *Big Data Mining and Analytics, 3*(4), 259-279. doi:10.26599/BDMA.2020.9020006

Hu, S., Zhang, H., & Zhang, W. (2023a). Domain Knowledge Graph Question Answering Based on Semantic Analysis and Data Augmentation. *Applied Sciences, 13*(8838). doi:https://doi.org/10.3390/app13158838

Hu, Y., Mao, H., & McKenzie, G. (2019). A natural language processing and geospatial clustering framework for harvesting local place names from geotagged housing advertisements. *International Journal of Geographical Information Science, 33*(4), 714-738. doi:https://doi.org/10.1080/13658816.2018.1458986

Huang, L., Shi, P., & Zhu, H. (2022). An integrated urgency evaluation approach of relief demands for disasters based on social media data. *International Journal of Disaster Risk Reduction, 80*(103208), 1-15. doi:https://doi.org/10.1016/j.ijdrr.2022.103208

Huo, Y., Tang, J., Pan, Y., Zeng, Y., & Cao, L. (2020). Learning a Planning Domain Model From Natural Language Process Manuals. *IEEE Access, 8*, 143219-143232. doi:10.1109/ACCESS.2020.3013237

Hurwitz, J., & Kirsch, D. (2018). *Machine Learning For Dummies.* John Wiley & Sons, Inc. Retrieved from https://www.ibm.com/downloads/cas/GB8ZMQZ3

Ibrahim, M. (2023, March 10). Weights & Biases. *The Answer Key: Unlocking the Potential of Question Answering With NLP*. Retrieved from https://wandb.ai/mostafaibrahim17/ml-articles/reports/The-Answer-Key-Unlocking-the-Potential-of-Question-Answering-With-NLP--VmlldzozNTcxMDE3#:~:text=Question%20Answering%20in%20NLP%2C%20is,and%20present%20an%20accurate%20answer.

IFRC. (2020). *World Disasters Report 2020: Come heat or high water.* Geneva: International Federation of Red Cross and Red Crescent Societies. Obtido de https://media.ifrc.org/ifrc/world-disaster-report-2020

Imran, M., Ofli, F., Caragea, D., & Torralba, A. (2020). Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions. *Information Processing & Management, 57*(102261), 1-9. doi:https://doi.org/10.1016/j.ipm.2020.102261

International Federation of Red Cross and Red Crescent Societies. (N/D). *About disaster management*. Obtido de https://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management/

Jain, V. K., & Kumar, S. (2018). Effective surveillance and predictive mapping of mosquito-borne diseases using social media. *Journal of Computational Science, 25*, 406-415. doi:https://doi.org/10.1016/j.jocs.2017.07.003

Joloudari, J. H., Hussain, S., Nematollahi, M., Bagheri, R., Fazl, F., Alizadehsani, R., . . . Talukder, A. (2023). BERT-Deep CNN: State-of-the-Art for Sentiment Analysis of COVID-19 Tweets. *Social Network Analysis and Mining, 13*(99), 1-20. doi:10.1007/s13278-023-01102-y

Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd Draft ed.). Prentice Hall. Retrieved from https://web.stanford.edu/~jurafsky/slp3/

Kaplan, A., & Haenlein, M. (2019). Siri, siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15-25. doi:https://doi.org/10.1016/j.bushor.2018.08.004

Karami, A., Shah, V., Vaezi, R., & Bansal, A. (2020). Twitter speaks: A case of national disaster situational awareness. *Journal of Information Science, 46*(3), 313-324. doi:10.1177/0165551519828620

Kardam, N., Misra, S., & Wilson, D. (2023). Is Natural Language Processing Effective in Education Research? A case study in student perceptions of TA support. *ASEE Annual Conference and Exposition.* Baltimore: American Society for Engineering Education.

Karimiziarani, M., & Moradkhani, H. (2023). Social response and Disaster management: Insights from twitter data Assimilation on Hurricane Ian. *International Journal of Disaster Risk Reduction, 95*(103865), 1-8. doi:https://doi.org/10.1016/j.ijdrr.2023.103865

Karimiziarani, M., Jafarzadegan, K., Abbaszadeh, P., Shao, W., & Moradkhani, H. (2022). Hazard risk awareness and disaster management: Extracting the information content of twitter data. *Sustainable Cities and Society, 77*(103577), 1-11. doi:https://doi.org/10.1016/j.scs.2021.103577

Karimiziarani, M., Shao, W., Mirzaei, M., & Moradkhani, H. (2023). Toward reduction of detrimental effects of hurricanes using a social media data analytic Approach: How climate change is perceived? *Climate Risk Management, 39*(100480), 1-15. doi:https://doi.org/10.1016/j.crm.2023.100480

Karmegam, D., & Mappillairaju, B. (2021). Information extraction using a mixed method analysis of social media data: A case study of the police shooting during the anti-Sterlite protests at Thoothukudi, India. *Information Development, 39*(4), 1-11. doi:https://doi.org/10.1177/02666669211049153

Kellner, D., Lowin, M., & Hinz, O. (2023). Improved healthcare disaster decision-making utilizing information extraction from complementary social media data during the COVID-19 pandemic. *Decision Support Systems, 172*(113983), 1-12. doi:https://doi.org/10.1016/j.dss.2023.113983

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications, 82*, 3713-3744. Retrieved from https://doi.org/10.1007/s11042-022-13428-4

Kima, L., Yahia, E., Segonds, F., Véron, P., & Mallet, A. (2021). i-Dataquest: A heterogeneous information retrieval tool using data graph for the manufacturing industry. *Computers in Industry, 132*. doi:https://doi.org/10.1016/j.compind.2021.103527

Krishna, M. H., Rahamathulla, K., & Akbar, A. (2017). A Feature Based Approach for Sentiment Analysis using SVM and Coreference Resolution. *International Conference on Inventive Communication and Computational Technologies (ICICCT 2017).* Coimbatore, India: IEEE. doi:10.1109/ICICCT.2017.7975227

Kulkarni, C., Monika, P., B, P., & S, S. (2022). A novel framework for automatic caption and audio generation. *Materials Today: Proceedings, 65*, 3248-3252. doi:https://doi.org/10.1016/j.matpr.2022.05.380

Kurvers, R. H., Nuzzolese, A. G., Russo, A., Barabucci, G., Herzog, S. M., & Trianni, V. (2023). Automating hybrid collective intelligence in open-ended medical diagnostics. *Proceedings of the National Academy of Sciences of the United States of America, 120*(34), 1-7. doi:https://doi.org/10.1073/pnas.2221473120

Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Vora, D., & Pappas , I. (2023). A systematic review of applications of natural language processing and future challenges with special emphasis in text-based emotion detection. *Artificial Intelligence Review, 56*, 15129-15215. Retrieved from https://doi.org/10.1007/s10462-023-10509-0

Lei, X., Song, W., Fan, R., Feng, R., & Wang, L. (2023). Semi-supervised geological disasters named entity recognition using few labeled data. *GeoInformatica, 27*, 263-288. doi:https://doi.org/10.1007/s10707-022-00474-1

Li, L., Bensi, M., & Baecher, G. (2023). Exploring the potential of social media crowdsourcing for post-earthquake damage assessment. *International Journal of Disaster Risk Reduction, 98*(104062). doi:https://doi.org/10.1016/j.ijdrr.2023.104062

Lian, Y., Liu, Y., & Dong, X. (2020). Strategies for controlling false online information during natural disasters: The case of Typhoon Mangkhut in China. *Technology in Society, 62*(101265), 1-10. doi:https://doi.org/10.1016/j.techsoc.2020.101265

Liu, R., Mao, R., Luu, A. T., & Cambria, E. (2023). A brief survey on recent advances in coreference resolution. *Artifcial Intelligence Review, 56*, 14439-14481. Retrieved from https://doi.org/10.1007/s10462-023-10506-3

Liu, Y., Xu, B., Yang, Y., Chung, T., & Zhang, P. (2019c). Constructing a Hybrid Automatic Q&A System Integrating Knowledge Graph and Information Retrieval Technologies. In M. Chang, E. Popescu, Kinshuk, N.-S. Chen, M. Jemni, R. Huang, . . . D. G. Sampson, *Foundations and Trends in Smart Learning: Proceedings of 2019 International Conference on Smart Learning Environments* (pp. 67-76). Springer. doi:http://dx.doi.org/10.1007/978-981-13-6908-7_9

Lu, S., Huang, J., & Wu, J. (2023). Multi-Dimensional Urban Flooding Impact Assessment Leveraging Social Media Data: A Case Study of the 2020 Guangzhou Rainstorm. *Water (Switzerland), 15*(4296), 1-20. doi:https://doi.org/10.3390/w15244296

Lyu, S., & Lu, Z. (2023). Exploring Temporal and Multilingual Dynamics of Post-Disaster Social Media Discourse: A Case of Fukushima Daiichi Nuclear Accident. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), 1-24. doi:https://doi.org/10.1145/3579484

Ma, M., Gao, Q., Xiao, Z., Hou, X., Hu, B., Jia, L., & Song, W. (2023). Analysis of public emotion on ood disasters in southern China in 2020 based on social media data. *Natural Hazards, PREPRINT (Version 1)*. doi:https://doi.org/10.21203/rs.3.rs-2374215/v1

Ma, Z., He, J., Qiu, J., Cao, H., Wang, Y., Sun, Z., . . . Chen, W. (2022). BaGuaLu: targeting brain scale pretrained models with over 37 million cores. *Proceedings of the 27th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, (pp. 192-204). Seoul, Republic of Korea. doi:https://doi.org/10.1145/3503221.3508417

Mahmoud, Z., Li, C., Zappatore, M., Solyman, A., Alfatemi, A., Ibrahim, A. O., & Abdelmaboud, A. (2023). Semi-supervised learning and bidirectional decoding for effective grammar correction in low-resource scenarios. *PeerJ Computer Science, 9*, 1-25. doi:https://doi.org/10.7717/peerj-cs.1639

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511809071

Matsuoka, Y., & Rocha, E. (2021). The role of non-government stakeholders in implementing the Sendai Framework: A view from the voluntary commitments online platform. *Progress in Disaster Science, 9*(100142), 1-20. doi:http://dx.doi.org/10.1016/j.pdisas.2021.100142

Mer, A., & Virdi, A. S. (2023). Navigating the Paradigm Shift in HRM Practices Through the Lens of Artificial Intelligence: A Post-pandemic Perspective. In P. Tyagi, N. Chilamkurti, S. Grima, K. Sood, & B. Balusamy, *The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part A* (pp. 123-154). Emerald Publishing Limited.

Mughal, M., Shaikh, Z., Wagan, A., Khand, Z., & Hassan, S. (2021). ORFFM: An Ontology-Based Semantic Model of River Flow and Flood Mitigation. *IEEE Access, 9*, 44003-44031. doi:10.1109/ACCESS.2021.3066255

Murthy, D. (2015). Twitter and elections: are tweets, predictive, reactive, or a form of buzz? *Information, Communication & Society, 18*(7), 816-831. doi:https://doi.org/10.1080/1369118X.2015.1006659

Nefros, C., Kitsara, G., & Loupasakis, C. (2022). Geographical Information Systems and Remote Sensing Techniques to Reduce the Impact of Natural Disasters in Smart Cities. *IFAC-PapersOnLine, 55*(11), 72-77. doi:https://doi.org/10.1016/j.ifacol.2022.08.051

New York City Police Department. (n.d.). *Types of Terrorist Attacks*. Retrieved from Plan Now NYC: https://plannownyc.cityofnewyork.us/types-of-terrorist-attacks/

Ni, J., Young, T., Pandelea, V., Xue, F., & Cambria, E. (2023). Recent advances in deep learning based dialogue systems: a systematic survey. *Artificial Intelligence Review, 56*, 3055-3155. doi:https://doi.org/10.1007/s10462-022-10248-8

Ni, W., Shen, Q., Liu, T., Zeng, Q., & Xu, L. (2023). Generating textual emergency plans for unconventional emergencies - A natural language processing approach. *Safety Science, 160*(106047), 1-18. doi:https://doi.org/10.1016/j.ssci.2022.106047

Nießen, S., Och, F. J., Leusch, G., & Ney, H. (2000). An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00).* European Language Resources Association (ELRA). Retrieved from http://www.lrec-conf.org/proceedings/lrec2000/pdf/278.pdf

Nunavath, V., & Goodwin, M. (2019). The Use of Artificial Intelligence in Disaster Management - A Systematic Literature Review. *6th International Conference on Information and Communication Technologies for Disaster Management.* doi:10.1109/ICT-DM47966.2019.9032935

Office for the Coordination of Humanitarian Affairs, United Nations. (2018). *UN-CMCoord Field Handbook.* Field HB. Retrieved from https://emergency.unhcr.org/sites/default/files/UN%20Civil%20Military%20Coordination%20Handbook%202018.pdf

Ogie, R. I., Rho, J. C., & Clarke, R. J. (2018). Artificial Intelligence in Disaster Risk Communication: A Systematic Literature Review. *5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, (pp. 1-8). doi:https://doi.org/10.1109/ICT-DM.2018.8636380

Orimoloye, I., Belle, J., & Ololade, O. (2021). Exploring the emerging evolution trends of disaster risk reduction research: a global scenario. *International Journal of Environmental Science and Technology, 18*, 673-690. doi:https://doi.org/10.1007/s13762-020-02847-1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372*(71). doi:https://doi.org/10.1136/bmj.n71

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, (pp. 311-318). Philadelphia. doi:https://doi.org/10.3115/1073083.1073135

Parsons, M., Reeve, I., McGregor, J., Hastings, P., Marshall, G., McNeill, J., . . . Glavac, S. (2021). Disaster resilience in Australia: A geographic assessment using an index of coping and adaptive capacity. *International Journal of Disaster Risk Reduction, 62*(102422), 1-15. doi:https://doi.org/10.1016/j.ijdrr.2021.102422

Patwardhan, N., Marrone, S., & Sansone, C. (2023). Transformers in the Real World: A Survey on NLP Applications. *Information, 14*(4), 242. doi:https://doi.org/10.3390/info14040242

Perlstein, S., & Verboord, M. (2021). Lockdowns, lethality, and laissez-faire politics. Public discourses on political authorities in high-trust countries during the COVID-19 pandemic. *PLOS ONE, 16*(6), 1-23. doi:https://doi.or g/10.1371/journal.pone.0253175

Preum, S., Shu, S., Alemzadeh, H., & Stankovic, J. (2020). EMSContExt: EMS protocol-driven concept extraction for cognitive assistance in emergency response. *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 13350-13355). New York: AAAI Press. doi:https://doi.org/10.1609/aaai.v34i08.7048

Prior, H., & Williamson, D. J. (2023, May 04). Revolutionising disaster response: Natural Language Processing presents game changing opportunity. New Zealand. Retrieved from https://www2.deloitte.com/nz/en/blog/consulting/2023/revolutionising-disaster-response.html

Raisio, H., Puustinen, A., & Lindell, J. (2022). #StrongTogether? Qualitative Sentiment Analysis of Social Media Reactions to Disaster Volunteering during a Forest Fire in Finland. *Sustainability, 14*(3983), 1-18. doi:https://doi.org/10.3390/su14073983

Rashid, M., & Wang, D. (2022). SEIS: A spatiotemporal-aware event investigation framework for social airborne sensing in disaster recovery applications. *Pervasive and Mobile Computing, 87*(101717), 1-25. doi:https://doi.org/10.1016/j.pmcj.2022.101717

Reichenpfader, D., Müller, H., & Denecke, K. (2023). Large language model-based information extraction from free-text radiology reports: a scoping review protocol. *BMJ Open, 13*(12). doi:10.1136/bmjopen-2023-076865

Resch, B., Usländer, F., & Havas, C. (2018). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science, 45*(4), 362-376. doi:https://doi.org/10.1080/15230406.2017.1356242

Rocca, R., Tamagnone, N., Fekih, S., Contla, X., & Rekabsaz, N. (2023). Natural language processing for humanitarian action: Opportunities, challenges, and the path toward humanitarian NLP. *Frontiers in Big Data, 6*. Retrieved from https://doi.org/10.3389/fdata.2023.1082787

Römer, H., Kersten, J., Kiefl, R., Plattner, S., Mager, A., & Voigt, S. (2014). Airborne near-real-time monitoring of assembly and parking areas in case of large-scale public events and natural disasters. *International Journal of Geographical Information Science, 28*(4), 682-699. doi:https://doi.org/10.1080/13658816.2013.866240

Rossi, C., Acerbo, F., Ylinen, K., Juga, I., Nurmi, P., Bosca, A., . . . 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. doi:10.1016/j.ijdrr.2018.03.002

Roy, K., Hasan, S., & Mozumder, P. (2020). A multilabel classification approach to identify hurricane-induced infrastructure disruptions using social media data. *Computer-Aided Civil and Infrastructure Engineering, 35*(12), 1387-1402. doi:https://doi.org/10.1111/mice.12573

Ryan, J., Sellak, H., & Brindal, E. (2021). The Psychosocial Impacts of COVID-19 on a Sample of Australian Adults: Cross-sectional Survey and Sentiment Analysis. *JMIR Public Health and Surveillance, 7*(7), 1-13. doi:http://dx.doi.org/10.2196/29213

Sajid, A., Jan, S., & Shah, I. (2017). Automatic Topic Modeling for Single Document Short Texts. *2017 International Conference on Frontiers of Information Technology* (pp. 70-75). Islamabad, Pakistan: IEEE. doi:10.1109/FIT.2017.00020

Sakurai, M., & Murayama, Y. (2019). Information technologies and disaster management – Benefits and issues. *Progress in Disaster Science, 2*(100012), 1-4. doi:http://dx.doi.org/10.1016/j.pdisas.2019.100012

Salas-Zárate, R., Alor-Hernández, G., Salas-Zárate, M., Paredes-Valverde, M. A., Bustos-López, M., & Sánchez-Cervantes, J. (2022). Detecting Depression Signs on Social Media: A Systematic Literature Review. *Healthcare, 10*(291). doi:https://doi.org/10.3390/healthcare10020291

Sarkar, A., Chowdhury, T., Murphy, R., Gangopadhyay, A., & Rahnemoonfar, M. (2023). SAM-VQA: Supervised Attention-Based Visual Question Answering Model for Post-Disaster Damage Assessment on Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing, 61*(4702716), 1-17. doi:https://doi.org/10.1109/TGRS.2023.3276293

Sarker, M., Peng, Y., Yiran, C., & Shouse, R. (2020). Disaster resilience through big data: Way to environmental sustainability. *International Journal of Disaster Risk Reduction, 51*(101769), 1-9. doi:https://doi.org/10.1016/j.ijdrr.2020.101769

Sawalha, I. H. (2020). A contemporary perspective on the disaster management cycle. *Foresight, 22*(4), 469-482. doi:http://dx.doi.org/10.1108/FS-11-2019-0097

Sawalha, I. H. (2022, March 11). *Warfare as a Type of Disaster*. Retrieved from Disaster Recovery Journal: https://drj.com/journal_main/warfare-as-a-type-of-disaster/

Sawicki, J., Ganzha, M., & Paprzyck, M. (2023). The state of the art of Natural Language Processing – a Systematic Automated Review of NLP Literature Using NLP Techniques. *Data Intelligence, 5*(3), 1-47. Retrieved from https://doi.org/10.1162/dint_a_00213

Scalia, G., Francalanci, C., & Pernici, B. (2022). CIME: Context-aware geolocation of emergency-related posts. *Geoinformatica, 26*, 125-157. doi:https://doi.org/10.1007/s10707-021-00446-x

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology, 3*(4), 552-631. Retrieved from https://doi.org/10.1016/0010-0285(72)90022-9

Schwarz, K., Aranda, D. A., & Hartmann, M. (2023). Towards Automated Situational Awareness Reporting for Disaster Management - A Case Study. *Sustainability, 15*(7968), 1-14. doi:https://doi.org/10.3390/su15107968

Scimago Journal & Country Rank. (2023, April). *Journal Rankings*. Retrieved from Scimago Journal & Country Rank: https://www.scimagojr.com/journalrank.php

Sermet, Y., & Demir, I. (2018). An intelligent system on knowledge generation and communication about flooding. *Environmental Modelling & Software, 108*, 51-60. doi:https://doi.org/10.1016/j.envsoft.2018.06.003

Shahbazi, Z., & Byun, Y.-C. (2022). Blockchain-Based Event Detection and Trust Verification Using Natural Language Processing and Machine Learning. *IEEE Access, 10*, 5790-5800. doi:10.1109/ACCESS.2021.3139586

Sharma, R., Agarwal, P., & Arya , A. (2022). Natural Language Processing and Big Data: A Strapping Combination. In R. Sharma, & D. Sharma, *New Trends and Applications in Internet of Things (IoT) and Big Data Analytics* (Vol. 221, pp. 255-271). Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-030-99329-0_16

Shen, W., Wang, J., & Han, J. (2015). Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions. *IEEE Transactions on Knowledge and Data Engineering, 27*(2), 443-460. doi:10.1109/TKDE.2014.2327028

Shi, K., Peng, X., Lu, H., Zhu, Y., & Niu, Z. (2023). Application of Social Sensors in Natural Disasters Emergency Management: A Review. *IEEE Transactions On Computational Social Systems, 10*(6), 3143-3158. doi:10.1109/TCSS.2022.3211552

Shibuya, Y., & Tanaka, H. (2019). Using Social Media to Detect Socio-Economic Disaster Recovery. *IEEE Intelligent Systems, 34*(3), 29-37. doi:https://doi.org/10.1109/MIS.2019.2918245

Singh, N., Roy, N., & Gangopadhyay, A. (2019). Analyzing The Emotions of Crowd For Improving The Emergency Response Services. *Pervasive and Mobile Computing, 58*(101018), 1-17. doi:https://doi.org/10.1016/j.pmcj.2019.04.009

Sit, M., Koylu, C., & Demir, I. (2019). Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. *International Journal of Digital Earth, 12*(11), 1205-1229. doi:https://doi.org/10.1080/17538947.2018.1563219

Srinivasarao, U., & Sharaff, A. (2023). Machine intelligence based hybrid classifier for spam detection and sentiment analysis of SMS messages. *Multimedia Tools and Applications, 82*(20), 31069-31099. doi:https://doi.org/10.1007/s11042-023-14641-5

Stieglitz, S., Hofeditz, L., Brünker, F., Ehnis, C., Mirbabaie, M., & Ross, B. (2022). Design principles for conversational agents to support Emergency Management Agencies. *International Journal of Information Management, 63*(102469), 1-12. doi:https://doi.org/10.1016/j.ijinfomgt.2021.102469

Stoykova, S., & Shakev, N. (2023). Artificial Intelligence for Management Information Systems: Opportunities, Challenges, and Future Directions. *Algorithms, 16*(357). doi:https://doi.org/10.3390/a16080357

Su, Z., McDonnell, D., Bentley, B., He, J., Shi, F., Cheshmehzangi, A., . . . Jia, P. (2021). Addressing Biodisaster X Threats With Artificial Intelligence and 6G Technologies: Literature Review and Critical Insights. *Journal of Medical Internet Research, 23*(5), 1-16. doi:https://doi.org/10.2196/26109

Sugino, H., Sekiguchi, T., Terada, Y., & Hayashi, N. (2023). “Future Compass”, a Tool That Allows Us to See the Right Horizon—Integration of Topic Modeling and Multiple-Factor Analysis. *Sustainability, 15*(10175), 1-20. doi:https://doi.org/10.3390/su151310175

Terracciano, E., & Han, A. (2023). Twitter communication during winter storm Uri in San Antonio, Texas Implications for climate resiliency planning. *Cities, 139*(104407), 1-10. doi:https://doi.org/10.1016/j.cities.2023.104407

Tian, X., He, W., He, Y., Albert, S., & Howard, M. (2023). Using health belief model and social media analytics to develop insights from hospital-generated twitter messaging and community responses on the COVID-19 pandemic. *Journal of Enterprise Information Management, 36*(6), 1483-1507. doi:https://doi.org/10.1108/JEIM-06-2021-0267

Tillmann, C., Vogel, S., Ney, H., Zubiaga, A., & Sawaf, H. (1997). Accelerated DP based Search for Statistical Translation. *Proc. 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, (pp. 2667-2670). doi:10.21437/Eurospeech.1997-673

Tounsi, A., & Temimi, M. (2023). A systematic review of natural language processing applications for hydrometeorological hazards assessment. *Natural Hazards, 116*, 2819–2870. Retrieved from https://doi.org/10.1007/s11069-023-05842-0

Tufekci, S., & Wallace, W. A. (1998). The emerging area of emergency management and engineering. *IEEE Transactions on Engineering Management, 45*(2), 103-105. doi:https://doi.org/10.1109/TEM.1998.669742

Turner, A. M., Choi, Y. K., Dew, K., Tsai, M.-T., Bosold, A. L., Wu, S., . . . Meischke, H. (2019). Evaluating the Usefulness of Translation Technologies for Emergency Response Communication: A Scenario-Based Study. *JMIR Public Health and Surveillance, 5*(1), 1-14. doi:https://doi.org/10.2196/11171

Tvardik, N., Kergourlay, I., Bittar, A., Segond, F., Darmoni, S., & Metzger, M.-H. (2018). Accuracy of using natural language processing methods for identifying healthcare-associated infections. *International Journal of Medical Informatics, 117*, 96-102. doi:10.1016/j.ijmedinf.2018.06.002

Valinejad, J., & Mili, L. (2023). Cyber–Physical–Social Model of Community Resilience by Considering Critical Infrastructure Interdependencies. *IEEE Internet of Things Journal, 10*(19), 17530-17543. doi:https://doi.org/10.1109%2FJIOT.2023.3277450

Velev, D., & Zlateva, P. (2023). Challenges of Artificial Intelligence Application for Disaster Risk Management. *39th International Symposium on Remote Sensing of Environment (ISRSE-39) "From Human Needs to SDGs"* (pp. 387-394). Antalya, Türkiye: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. doi:https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-387-2023

Vongkusolkit, J., & Huang, Q. (2021). Situational awareness extraction: a comprehensive review of social media data classification during natural hazards. *Annals of GIS, 27*(1), 5-28. doi:https://doi.org/10.1080/19475683.2020.1817146

Wahid, J., Shi, L., Gao, Y., Yang, B., Wei, L., Tao, Y., . . . Yagoub, I. (2022). Topic2Labels: A framework to annotate and classify the social media data through LDA topics and deep learning models for crisis response. *Expert Systems With Applications, 195*(116562), 1-15. doi:https://doi.org/10.1016/j.eswa.2022.116562

Wan, X., Mao, Y., & Qi, R. (2023). Chinese Event Detection without Triggers Based on Dual Attention. *Applied Sciences, 13*(4523), 1-17. doi:https://doi.org/10.3390/app13074523

Wang, K., Lam, N. S., & Mihunov, V. (2023). Correlating Twitter Use with Disaster Resilience at Two Spatial Scales: A Case Study of Hurricane Sandy. *Annals of GIS, 29*(1), 1-20. doi:https://doi.org/10.1080/19475683.2023.2165545

Wang, K., Lam, N. S., Zou, L., & Mihunov, V. (2021). Twitter Use in Hurricane Isaac and Its Implications for Disaster Resilience. *ISPRS International Journal of Geo-Information, 10*(116), 1-20. doi:https:// doi.org/10.3390/ijgi10030116

Wang, L., Liu, X., Liu, Y., Li, H., Liu, J., & Yang, L. (2023). Emergency entity relationship extraction for water diversion project based on pre-trained model and multi-featured graph convolutional network. *PLOS ONE, 18*(10), 1-18. doi:https://doi.org/10.1371/journal. pone.0292004

Wang, R.-Q., Hu, Y., Zhou, Z., & Yang, K. (2020). Tracking Flooding Phase Transitions and Establishing a Passive Hotline With AI-Enabled Social Media Data. *IEEE Access, 8*, 103395-103404. doi:10.1109/ACCESS.2020.2994187

Wang, X., & Xie, H. (2018). A Review on Applications of Remote Sensing and Geographic Information Systems (GIS) in Water Resources and Flood Risk Management. *Water, 10*(608), 1-11. doi:http://dx.doi.org/10.3390/w10050608

Wei, X., Li, L., Tang, C., & Subramaniam, S. (2020). UAVFog-Assisted Data-Driven Disaster Response: Architecture, Use Case, and Challenges. *Web Information Systems Engineering – WISE 2020. 12343*, pp. 591-606. Springer. Retrieved from https://doi.org/10.1007/978-3-030-62008-0_41

Wolbers, J., Kuipers, S., & Boin, A. (2021). A systematic review of 20 years of crisis anddisaster research: Trends and progress. *Risks Hazards Crisis Public Policy, 12*, 374-392. doi:https://onlinelibrary.wiley.com/doi/10.1002/rhc3.12244

Wołk, K., Wołk, A., Wnuk , D., Grzes, T., & Skubis, I. (2022). Survey on dialogue systems including slavic languages. *Neurocomputing, 477*, 62-84. doi:https://doi.org/10.1016/j.neucom.2021.11.076

Wu, H., Cheng, Z., Shi, W., Miao, Z., & Xu, C. (2014). An object-based image analysis for building seismic vulnerability assessment using high-resolution remote sensing imagery. *Natural Hazards, 71*, 151-174. doi:10.1007/s11069-013-0905-6

Wu, Y., Chen, Y., Zhang, R., Cui, Z., Liu, X., Zhang, J., . . . Wu, Y. (2023). A Spatial Information Extraction Method Based on Multi-Modal Social Media Data: A Case Study on Urban Inundation. *ISPRS International Journal of Geo-Information, 12*(368), 1-21. doi:https://doi.org/10.3390/ijgi12090368

Xia, T., Chen, X., Wang, J., & Qiu, F. (2023). A Hybrid Model with New Word Weighting for Fast Filtering Spam Short Texts. *Sensors, 23*(8975). doi:https://doi.org/10.3390/s23218975

Xia, X., Zhang, Y., Jiang, W., & Wu, C. (2023). Staying Home, Tweeting Hope: Mixed Methods Study of Twitter Sentiment Geographical Index During US Stay-At-Home Orders. *Journal of Medical Internet Research, 25*, 1-14. doi:http://dx.doi.org/10.2196/45757

Yáñez-Sandivari, L., Cortés, C. E., & Rey, P. A. (2021). Humanitarian logistics and emergencies management: New perspectives to a sociotechnical problem and its optimization approach management. *International Journal of Disaster Risk Reduction, 52*, 101952. doi:10.1016/j.ijdrr.2020.101952

Yu, M., Yang, C., & Li, Y. (2018). Big Data in Natural Disaster Management: A Review. *Geosciences, 8*(5). doi:http://dx.doi.org/10.3390/geosciences8050165

Zander, K. K., Garnett, S. T., Ogie, R., Alazab, M., & Nguyen, D. (2023). Trends in bushfire related tweets during the Australian 'Black Summer' of 2019/20. *Forest Ecology and Management, 545*(121274), 1-11. doi:https://doi.org/10.1016/j.foreco.2023.121274

Zhang, W., Li, L., Zhu, Y., Yu, P., & Wen, J. (2022). CNN-LSTM neural network model for fine-grained negative emotion computing in emergencies. *Alexandria Engineering Journal, 61*, 6755-6767. doi:https://doi.org/10.1016/j.aej.2021.12.022

Zhou, B., Zou, L., Mostafavi, A., Lin, B., Yang, M., Gharaibeh, N., . . . Mandal, D. (2022). VictimFinder: Harvesting rescue requests in disaster response from social media with BERT. *Computers, Environment and Urban Systems, 95*(101824), 1-12. doi:https://doi.org/10.1016/j.compenvurbsys.2022.101824

Zhuang, M., Li, Y., Tan, X., Xing, L., & Lu, X. (2021). Analysis of public opinion evolution of COVID-19 based on LDA-ARMA hybrid model. *Complex & Intelligent Systems, 7*, 3165-3178. doi:https://doi.org/10.1007/s40747-021-00514-7

Zibulewsky, J. (2001). Defining Disaster: The Emergency Department Perspective. *Baylor University Medical Center Proceedings, 14*(2), 144-149. doi:https://doi.org/10.1080/08998280.2001.11927751

Zin, M., Nguyen, H., Satoh, K., Sugawara, S., & Nishino, F. (2023). Information Extraction from Lengthy Legal Contracts: Leveraging Query-Based Summarization and GPT-3.5. *Frontiers in Artificial Intelligence and Applications, 379*, 177-186. doi:10.3233/FAIA230963

Zuccaro, G., Leone, M., & Martucci, C. (2020). Future research and innovation priorities in the field of natural hazards, disaster risk reduction, disaster risk management and climate change adaptation: a shared vision from the ESPREssO project. *International Journal of Disaster Risk Reduction, 51*(101783), 1-14. doi:https://doi.org/10.1016/j.ijdrr.2020.101783

APPENDIX A

Table A 1 - Challenges and opportunities included in the shortlisted articles

Citation	Challenges	Opportunities
(Afyouni et al., 2022)	1. Traditional data analysis algorithms and techniques do not scale in high computational complexity with large datasets from social media	1. The proposed hybrid approach for spatio-temporal event extraction has a major advantage for real-time spatio-temporal event detection and tracking from social media. 2. DL and NLP algorithms were built on transformer-based word embeddings (KeyBERT) to enable unbiased feature extraction for event detection.
(Afyouni et al., 2023)	1. Many of the tweets that are not geotagged, and which may contain valuable information about events cannot be considered by the proposed system (E-ware).	1. Event and anomaly detection from social media can provide deeper insights about user's and community interactions on a variety of unspecified topics of interest. 2. E-ware integrates a scalable spatio-temporal index for managing dynamic event clusters, thus efficiently accommodating incoming data streams into newly extracted events, or within existing clusters. 3. The continuous and incremental processing of data streams can make the proposed event detection system more accurate and efficient.
(Alam et al., 2020)	1. Making sense of social media data is a challenging task due to limitations of available tools to analyze high-volume and high velocity data streams, and challenges in dealing with information overload. 2. The extraction of named entities from tweets is more challenging due to the noisy structure of social media data.	1. Research studies have revealed the usefulness of the data available on Twitter for several disaster response tasks. 2. A richer set of information useful for crisis managers can be provided when combining findings from different techniques.

Citation	Challenges	Opportunities
(Alkhatib et al., 2019)	1. The accurate determination of geo-tags associated to crowdsourcing data. 2. The validation of information reports. 3. The automatic summarization of reports. 4. Since Arabic Dialects differ from one another phonologically, lexically, morphologically, and linguistically, their classification and automated processing becomes challenging.	1. The ability to obtain information on demand about a certain incident before arriving on the scene could provide an important decision support tool for emergency responders, thus leading to fast and accurate situational assessment and effective response to incidents and emergencies.
(An et al., 2022)	1. Public emotions are complex and changeable.	1. The algorithm of Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis (SKEP), provides a unified and powerful emotional semantic representation for various sentiment analysis tasks, which can be used for the prediction of sentence-level sentiment.
(Bai & Yu, 2016)	1. Sentiment classification using emotion rule-based method performs poorly due the complexity of the Chinese language (homophonic words, dialect, emoticons in social media). 2. One of the difficulties of sentiment analysis in microblogging is the sparsely of the texts.	1. Microblogging services have proven to be a useful source of information for gaining first-hand information about disaster events from the community experiencing the disaster. 2. Compared with the traditional bag-of-words algorithm, the new distribution representation of words (DRW) text representation algorithm may improve each classifier's accuracy by more than 3 %.
(Barker et al., 2019)	1. The volume of data from social media proves difficult to identify relevant information for decision-making. 2. Automatic identification of individual tweets for flooding relevance might be difficult to achieve.	1. Effective data mining of social media to improve crisis response is sought by the humanitarian sector. 2. Social media can help with gathering and disseminating information to accurately identify the localized flood extent.
(Bashir et al., 2021)	Not Discussed	1. Twitter has entered the current news and information distribution landscape, particularly during crises and disasters, and is used for situational awareness and saving lives.

Citation	Challenges	Opportunities
(Behl et al., 2021)	1. Supervised learning isn't the best strategy during emergencies because it requires the tagging of large social media datasets, which is costly and time-consuming. 2. The absence of rule-based structural grouping places limitations on unsupervised or semi-supervised learning.	1. Strong feature extraction techniques should be incorporated with advanced ML techniques for accurate and quick performance. 2. Emotional association using the temporal and spatial features could be deployed for the missing patterns through the social media information for homophily effect.
(Cerna et al., 2022)	1. The rarity of phenomena, such as floods or heating, makes them difficult to predict due to the limited data available. 2. RNN and LSTM models can be long and difficult to train.	1. For large datasets CNN can be an interesting alternative. 2. The main advantages of the transformer model are the following: 2.1. The distance between two tokens is no longer a parameter considered by the model. 2.2. The attention matrix calculation allows to parallelize the process of encoding and then decoding the sequences, thus accelerating the calculations. 2.3. No labeled data are required to pre-train these models.
(Chan & Tsai, 2019)	1. The segmentation tools Jieba and Stanford Core NLP could not handle the high level of professionality required for DM.	1. The current generation and consumption of large amounts of data in a variety of formats, sources, and attributes, has further challenged decision makers to accomplish their mission accurately and efficiently. 2. Developing dialogue systems to handle the massive and frequent transfer of data has become more common.
(Chen & Lim, 2021)	1. The main obstacle in the quantification of textual descriptions is the impossibility of compiling a complete vocabulary and grading each word in it. 2. With a more comprehensive geographic coverage, a better assessment of damage would have been achieved.	1. Emotions manifested on social media could be advantageous evidence to understand damage severity and relief needs.

Citation	Challenges	Opportunities
(Contreras et al., 2022a)	1. Sentiment analysis models have a difficult time classifying sarcastic sentences, often present in social media.	1. Social media provides opportunities for citizens to engage in emergency management by disseminating information to the public. 2. For larger tweet data, unsupervised classification saves a significant amount of time compared to supervised classification and may enhance the consistency of polarity categorization over the full text dataset.
(Contreras et al., 2022b)	1. One of the difficulties with the proposed method is that positive information can be detected as neutral because the algorithm does not know the context of the post-disaster recovery process. 2. The ML algorithm developed by MonkeyLearn does not recognize more than 3-grams which can lead to inaccurate results. 3. The unsupervised algorithm does not recognize sarcasm in the tweets.	1. The extraction of sentiments during a disaster contributes to a vital situational awareness of the disaster zone dynamics. 2. Supervised classification allows to detect ironic tweets and harsh colloquial expressions.
(da Costa et al., 2022)	1. Mining social network opinion is difficult since ML algorithms cannot fully reflect the context of messages. 2. The fact that users on Twitter, write their messages in an informal way, makes it difficult to extract useful knowledge.	1. Contextual data inclusion resulted in a 22.8% increase in situation awareness of urban floods via tweets and contextual data, demonstrating that multimodal approaches are more promising than unimodal strategies.
(Dereli et al., 2021)	1. Decision tree and KNN algorithms are not affective to provide satisfactory results.	1. Random forests are attractive from a computational perspective, because they are relatively quick to predict and train. 2. KNN is widely preferred in classification problems. 3. It is important to use social media effectively to increase people’s awareness about humanitarian activities. 4. The higher value of accuracy is obtained from SVM at 76%.

Citation	Challenges	Opportunities
(Dong et al., 2021)	1. Extracting useful disaster-related information from social media is still challenging.	1. The systematic use of social media can be beneficial for emergency management by receiving victim requests for assistance, monitoring user activities, and updating the public situational awareness. 2. In general, supervised ML techniques perform better than the unsupervised ML methods.
(Eligüzel et al., 2020)	1. Most Twitter users do not share their locations in their tweets making location extraction a challenging task.	1. Geospatial information is an important aspect on Twitter to improve situation awareness in DM and rescue operations. 2. Results demonstrate that the Multinomial NB classification algorithm and Extra Trees provide the best results for the applied three cases.
(Fan et al., 2020)	1. Reliable automated mapping of events across time and space during disasters remains a challenge. 2. Heterogeneous posting behaviors from social media such as the use of emoji and multiple exclamation marks lead to a challenge of removing noise for computational analysis. 3. Social media users tend not to share their locations due to privacy concerns which leads to a challenge related to obtaining the location of events included in posts.	1. The fine-tuned BERT-based classifier performs better than existing deep-learning classifiers in classifying the tweets into different humanitarian categories. 2. The proposed pipeline can be implemented on other sources of data, such as news articles, crowdsourced platforms, and RS data, to extend its applications and gain better situation awareness for disaster-affected areas.
(Fang et al., 2019)	1. Ensure the accuracy of online information. 2. The social and spatial heterogeneity in the usage of social media might lead to a biased analysis of actual disaster impacts.	1. The expansion of social media applications may facilitate better decision-making toward disaster impacts mitigation.

Citation	Challenges	Opportunities
(Farnaghi et al., 2020)	1. Most of the density-based clustering algorithms like DBSCAN and its branches do not account for the spatial heterogeneity of the Twitter data. 2. Traditional frequency-based vectorization methods like count vector, term frequency, and TF-IDF often result in huge vectors for representing the tweets, neglect the effect of synonyms/antonyms, the context, and the semantics of the texts, are not capable of modeling the abbreviations and misspelled words that are frequently used in tweets, and output unfavorably sparse vectors.	1. The opportunities provided by geotagged messages make social networks a potential source of information in different domains, and particularly disaster management. 2. The proposed method provides the ability to address the spatial heterogeneity in Twitter data and sensitivity to the changes in the density of tweets in different locations. 3. The proposed method also could consider spatial and temporal distances along with textual similarity in real-time extraction of spatiotemporal clusters.
(Gulnerman & Karaman, 2020)	1. Apart from the acquisition of the data, scientists are still concerned with the reliability and accuracy of using social media data. 2. NB requires trained datasets for each domain which might be challenging to find as it will not be possible to gather training data until a damaging earthquake occurs.	1. The NB classifier performance over two datasets is clearly better than that of the other four techniques. 2. The social media data generated by billions of human sensors throughout the world and by nearly half of the total population of Turkey are crucially significant as a data source, during and after a disaster.
(Guo et al., 2018)	1. The existence of a lot of noise in Chinese word segmentation makes it difficult to extract process models from natural language text. 2. The performance of Chinese language processing tool is much lower than that of English.	1. The proposed approach can be used to assisted experts to model emergency response process. 2. The application of NLP techniques helps achieve deeper verification and analysis from emergency plans.

Citation	Challenges	Opportunities
(Guo et al., 2020)	1. Because of the flexibility of natural language, emergency response plans normally contain unwanted ambiguities, and it is difficult to check consistency and completeness. 2. It is difficult to establish an appropriate indicator system because of the variety of emergency types and complex response plan systems. 3. The proposed approach cannot be used directly for other languages.	1. The proposed approach can be used to facilitate revisions and improvements of emergency response plans. 2. The proposed automated analysis approach can save evaluation cost and improve evaluation efficiency.
(Guo et al., 2021)	1. Large sample size and time span of collected data.	1. The sentiment dictionary is suitable for low granularity texts, with the advantages of speedy procedures and high accuracy.
(Hampton & Shalin, 2017)	1. Difficulty remains in finding, interpreting, and scaling relevant, actionable signal in a virtual (social media) firehose of noise.	1. There is an opportunity of research in Twitter data analysis for the exchange of information between the public and the authorities.
(He et al., 2023)	1. Only examining the monotonic trend may not be sufficient to account for the dynamic development of sentiments and emotions over the analyzed period.	1. In face of challenges exacerbated by COVID-19, timely and effective communication is needed to promote information transparency, which in turn, helps keep members of the public informed of the pandemic.
(Hu et al., 2019)	1. It is challenging to determine which place instance was referred to by a name in the text. 2. Another challenge lies in the informal format of housing advertisements, especially those posted by individuals on local websites.	1. Authors use a combination of off-the-shelf and retrained NER models. 2. A major advantage of using CRF for detecting named entities is that each word is not treated independently but is considered within a sequence.

Citation	Challenges	Opportunities
(Hu et al., 2023)	1. Many of the complex location descriptions consisting of multiple entities cannot be directly recognized by typical NER tools designed for identifying single entities. 2. Despite their impressive performance in generating human-like text, GPT models have triggered various societal concerns. 3. Malicious actors may take advantage of the disaster victims' submitted messages on the proposed online GPT model.	1. Merging geo-knowledge and a GPT model can help in recognizing complete location descriptions and their categories. 2. One advantage of the proposed method is its flexibility in adjusting the geo-knowledge used to guide the model.
(Huang et al., 2022)	1. The population biases in crowdsourced data may impact the evaluation results.	1. The proposed method can identify specific disaster relief needs, to improve the timeliness and accuracy of emergency rescue. 2. By gathering social media data over time, it is feasible to identify dynamic changes in the urgency of disaster relief requests, allowing relevant departments to adapt rescue plans at any time and transfer relief resources to the most in need areas. 3. To improve the accuracy of evaluation results, more social media sources can be considered.
(Huo et al., 2020)	1. Automating a planning model requires the analysis and understanding of natural language text and a general learning techniques which do not exist in real-world applications. 2. When deploying the framework in a practical application of generating typhoon contingency plans, the user interaction is still necessary in many AI planning systems.	1. Besides generating typhoon contingency plans, the proposed method may also perform well in the domain learning for narrative generation. 2. Introducing grammar parsing into the proposed framework may further reduce the need for manual effort.

Citation	Challenges	Opportunities
(Jain & Kumar, 2018)	1. The process of collecting, integrating, and analyzing disease-related information from diverse web-based channels becomes more difficult and challenging as social media platforms proliferate and the amount of data quickly multiply.	1. The proposed framework focused on alternative methods of analysis and visualization of user's opinions that do not depend upon the assumption of normality and historical data. 2. Future studies may consider other data resources such as blogs, news articles, forums etc.
(Joloudari et al., 2023)	1. As a result of their characteristics, sentiment analysis of texts like tweets remains challenging. 2. In a phenomenon such as the COVID-19 pandemic, fake news or incorrect scientific results can lead to the sending of mass messages on social networks that do not reflect reality.	1. In a pandemic situation, analyzing social media texts to uncover sentimental trends can be very helpful in gaining a better understanding of society's needs and predicting future trends. 2. BERT has recorded state-of-the-art performance in NLP tasks, including sentiment analysis, by substantial margins. 3. The transformer approach outperforms RNNs and CNNs in MT and answering systems thanks to a considerable reduction in training resource requirements. 4. GPT-2 can generate high-quality conditional synthetic text samples outperforming rival methods. This approach has also performed well on tasks such as question answering, reading comprehension, summarizing, and MT.
(Karami et al., 2020)	1. The unpredictable nature of natural disasters behavior makes it hard to have a comprehensive situational awareness to support disaster management.	1. The growth of social media has provided a great opportunity to track public opinion. 2. Big real-time social media data can help disaster managers to develop a better situational awareness during a natural disaster.

Citation	Challenges	Opportunities
(Karimiziarani & Moradkhani, 2023)	1. Information provided on social media platforms can be challenging to comprehend in time-sensitive situations because of the large volume and velocity of such data streams. 2. Twitter users may not accurately reflect the sentiments and experiences of people who are affected by natural disasters but do not use social media.	1. Social media data can be beneficial for several humanitarian objectives including "situational awareness". 2. State and local governments, police, fire departments, and other authorities can all benefit from social media information and use it to plan for disasters and respond to them more effectively.
(Karimiziarani et al., 2022)	1. Only a small portion of tweets contain geographical coordinates. 2. Filtering irrelevant and noisy content, recognizing misleading information, and managing enormous amounts of information from social media data are some of the challenges faced by crisis responders.	1. NLP methodologies may help in improving the quality of retrieved social media data. 2. The valuable and constant stream of content shared by users on Twitter during natural hazards can be monitored and evaluated to benefit disaster responders and crisis managers.
(Karimiziarani et al., 2023)	1. A proper understanding of social media data in time-critical situations is complicated due to the large volume and velocity of social media data streams. 2. Acquiring Twitter data for events in the past is a quite challenging task and depends on availability of a prepared dataset by external data providers. 3. The vast volumes of disaster-related content on Twitter, particularly during an active natural hazard, requires optimized and effective computational power that can function in real-time.	1. Twitter's large user base provides prompt and large volumes of data. 2. Social media data can provide significant insights into the social response to catastrophes and satisfy the information needs of disaster management entities and crisis responders.
(Karmegam & Mappillairaju, 2021)	1. Automated sentiment analysis may not be sufficient to completely understand the social media content. 2. Examining people's thoughts by traditional methods like interviews and group discussions entails a huge number of resources such as money, manpower, and time. 3. Information overload in social media may occur during emergencies which makes qualitative analysis more complicated.	1. Sentiment analysis techniques applied to DM help in planning the response and recovery operations and to handle the crisis far more effectively. 2. Extracting social media messages can be done within a short period of time at almost no cost.

Citation	Challenges	Opportunities
(Kellner et al., 2023)	1. The timeliness becomes an obstacle in situations where the data collection consists of a complex process.	1. Analyzing user-generated content from various social media platforms is beneficial for decision-making during health-related disasters. 2. Microblogging content is widely available, and researchers can utilize it to analyze topics discussed in the general population. 3. Complex models, such as the proposed LSTM model, can extract more useful information from more diverse data. 4. More complex models are suitable for longer forecasting horizons, like one or two weeks.
(Khattar & Quadri, 2022)	1. Filtering informative and actionable messages from a vast pool of noisy Twitter messages.	1. Tweet data may provide crucial information during an emergency. 2. Authors confirm the advantages of using an attention mechanism. 3. The performance of multimodal classification models is better than unimodal text-only and image-only models.
(Kim et al., 2022)	1. CNN-based models do not reflect sequential information and are unsuitable for tasks requiring an understanding of the entire text. 2. RNN-based models have a long-term dependency problem, and performance decreases as the length of the input sequence increases. 3. QA with multiple answers remains a challenging task.	1. CNN-based models can accomplish tasks, such as sentiment analysis and text classification. 2. RNN-based models can reflect sequential information.
(Lee & Yu, 2020)	1. The proposed approach results may differ when applied to other types of natural disaster events.	1. The proposed NLP approach allows to overcome limitations related with the analysis of a large volume of tweet data set.

Citation	Challenges	Opportunities
(Lei et al., 2023)	1. Without training on large-scale labeled data, current NER methods based on DL models cannot identify specific geological disaster entities from geological disaster situation reports. 2. It is challenging to obtain large-scale annotation training data in the field of geological disasters. 3. The word vector obtained by Word2Vec cannot separate the different semantics of polysemous words as well as words with different information.	1. The proposed model (Semi-GDNER) can effectively extract geological disaster entities when a few manually labeled and unlabeled in-domain data are available. 2. The proposed model outperforms other NER approaches. 3. BERT pre-training model can capture long-distance features and has good parallel computing capabilities.
(Li et al., 2023)	1. Assessments based on social media may be subject to challenges related to population representation. 2. The TF-IDF method may not be ideal in situations where tweet data significantly deviate from the training samples as this approach does not account for the semantic meanings of words within a sentence. 3. Determining the precise location of tweets in non-English speaking regions (spatial bias).	1. social media data can provide useful information in a timely manner as they are available in large quantities and can be collected in near-real-time.
(Lian et al., 2020)	1. Existence of misinformation in social media data.	1. In the domain of emergency response to disasters, the increasingly prominent role of social websites has been confirmed. 2. The researchers propose ways on how to effectively manage the spread of false online information about natural disasters.

Citation	Challenges	Opportunities
(Liu et al., 2019)	1. There is no transparent procedure in place to allow use of social media data for decision making purposes. 2. The dynamic nature of social media, and its abbreviated and informal styled content makes it difficult to craft and disseminate alert messages for public consumption. 3. Social media users tend to have varying perceptions and educational backgrounds, and possess distinct life experiences, which often leads to generation of dubious quality data.	1. Both population and degree of damage contribute positively to increase in tweet volume, while population count appears to have a bigger impact on tweet volume increase.
(Lu et al., 2023)	1. Traditional data sources, such as official reports, RS, and surveys, often suffer from delayed responses to actual disaster events, incomplete information gathering, and high costs which hamper their ability to provide timely disaster impact assessments. 2. Difficulty in finding the impacted area of a disaster due to noise on extracted data. 3. Sentiment lexicon-based methods are well-suited for low granularity texts characterized by shorter lengths and sentences and are more efficient and accurate.	1. Analyzing social media content enables monitoring of the development of disaster situations. 2. A lexicon-based approach allows for tracking specific sentiment categories. 3. Location information derived from social media texts can unveil the geographic distribution of impacted areas. 4. Collecting and analyzing social media data is a reliable and feasible way of conducting rapid impact assessment for disasters. 5. The quality of lexicon-based sentiment analysis results depends on the efficacy of the sentiment dictionaries used. 6. NER can be used to directly extract location information from microblog texts and overcome the issue of data-scarce scenarios for geo-tagged information in social media.
(Lyu & Lu, 2023)	1. Identification of many sarcastic tweets in the imported dataset which may not reflect the actual and intended sentiments and may compromise the analysis. 2. Low availability of geo-tagged tweets. 3. Misinformation within social media data. 4. Translation distortion when using Google Translator.	1. Analyzing crisis-related tweets can help capture insights for public situational awareness development, crisis global response coordination, and post-disaster policymaking. 2. During and after crises, individuals use social media to seek help and exchange important information.

Citation	Challenges	Opportunities
(Ma et al., 2022)	1. Deep learning-based models exhibited powerful geoparsing capabilities mainly when evaluated on a single simple dataset. 2. The Stanford NER tool's applicability is limited to informal natural language texts written by various web users. 3. The proposed method (BERT-BiLSTM-CRF) fails to identify many toponyms when long and complex toponym descriptions exist in location names.	1. The Stanford NER tool is an essential NER technique that labels sequences of words in natural language text and detects person and location names. 2. The results benefit from the inclusion of a BERT layer which models the multiple meanings of words and can reflect the complex characteristics of words, including syntax, semantics, etc.
(Ma et al., 2023)	1. The low computational efficiency of hydrodynamic simulations makes it harder to perform fast and high precision simulations at the urban scale of hundreds of square kilometers.	1. When a natural disaster suddenly occurs, emotion analysis can help crisis response managers understand the state, trend, and abnormal changes in public emotion. 2. In flood-related studies, social media is suitable for acquiring data on a larger scale and in real time. 3. Weibo text can be used to study the flooding situation and public emotion.
(Matsuoka & Rocha, 2021)	Not Discussed	1. Sentiment analysis addresses the problem of language processing and/or mining of texts through an automated opinion discovery and summarization process.
(Mughal et al., 2021)	1. Emerging technologies with IoT and cloud-based storage are complex to extract, process, and comprehend relevant data of complex integrated multiple domains involving many participants. 2. Native data acquisition format for organization-specific needs, regional operational terminologies and procedure, data scarcity, data heterogeneity, poorly defined context, lack of semantics in models and procedures, are limiting factors of the effective disaster coordination. 3. The decentralized environment with different IM structures makes coordination the main challenge among disaster management actors.	1. The proposed knowledge base may contribute towards streamflow optimization and flood mitigation. 2. There is an emerging demand for a common data sharing and representation framework.

Citation	Challenges	Opportunities
(Mukherjee et al., 2022)	1. It is more challenging to gather data from social media and run complex computational methods in developing countries due to data and internet availability.	1. Importance of computational intelligence in predicting the popularity of tweets in case of mass emergencies in developing nations. 2. The NB classifier performs considerably better than other complex algorithms for small dataset sizes.
(Neppalli et al., 2017)	1. The sentiment classification of tweets faces many challenges including dealing with very short texts as well as unstructured text and noisy user input	1. Understanding elements of a message can potentially shed light into the type of information being spread during disasters in large microblogging communities and can better inform emergency managers on how to reach the widest audience in the fastest way.
(Nguyen et al., 2022)	1. Handling social media's data volume and complexity makes impractical to be directly used by analysts within the humanitarian context. 2. RNNs are often hard to train.	1. Social media may contribute to situational awareness during crisis. 2. Social media data may also help in accurately predict disaster-related needs. 3. LSTMs address the training problem of neural networks by introducing a set of gating functions to control the flow of information.
(Ni et al., 2023)	1. It may not be possible to draw a comprehensive evaluation conclusion from Emergency Plans if they only rely on a single notion of preference. 2. An unconventional emergency generally involves evolving emergent situations that may even interact in a complex way with incomplete information, making it hard for an expert to give a precise judgment.	1. It can be beneficial to refine the emergency response knowledge extracted from raw emergency plans and perform empirical evaluation by using more real-world unconventional emergency cases.

Citation	Challenges	Opportunities
(Nimmi & Janet, 2023)	1. Close relationship between certain emergence classes and the scarcity of resources on call transcripts. 2. Imbalance in class and the usage of less common / more complex words in a sentence can lead to misclassification by the PLE-MobileBERT model.	1. The proposed PLE-MobileBERT approach outperforms other models in terms of speed and effectiveness, adding the model's ability to understand phrase-level information. 2. MobileBERT has the same task-agnostic qualities as BERT and can be easily deployed on resource constrained devices due to its lower memory utilization, small size, and high inference speeds.
(Oz et al., 2018)	1. Sentiment analysis was presented as a challenge for the authors due to the length of the text in tweets.	1. One of the major advantages of working with digital traces is that event time or location has almost no effect in collecting disaster related information.
(Paradkar et al., 2022)	1. The language used in Twitter posts is abstract, informal, and full of jargon, making annotation difficult. 2. Location obtained from social media data can be difficult to identify and are open to multiple interpretations.	1. The content of social media posts and the embedded geocoordinates can be leveraged to enhance disaster situational awareness. 2. Authors found a very high consistency level (94%) between the point-type mentioned locations and the geotags in the tweets.
(Perlstein & Verboord, 2021)	1. Disaster events tend to happen suddenly and rapidly, limiting the amount of time researchers must plan out their studies and gather data. 2. NLP methods within sentiment analysis and topic modeling are less accurate when applied to non-English languages.	1. NLP techniques allow researchers to extract information about the thoughts and feelings present in a text corpus, or about social and semantic networks.

Citation	Challenges	Opportunities
(Ragini et al., 2018)	1. Social networks generate high volume of data every second and the major challenges are filtration and analysis of those big data for a specific query. 2. The challenges involved in using social media for disaster response include difficulty in collecting the disaster related data to build a better sentiment model for disaster analysis, lack of standard crisis data set or disaster related lexicon for accurate evaluation of the needs of the people.	1. During disaster situations, such emergency requests need to be mined from the pool of big data for providing timely help.
(Raisio et al., 2022)	1. A reported weakness of computer-based sentiment analysis is its lack of depth.	1. Complementary qualitative sentiment analysis may provide more context and help to understand the nuances of different sentiments. 2. Sentiment analysis can help improve situational awareness during the different stages of disasters.
(Rashid & Wang, 2022)	1. The development of an integrated system may find challenges in handling sparse data, in optimizing the UAV deployment, and in carefully extracting and analyzing latent semantic features embedded in the social media data. 2. Social media and unmanned aerial vehicles have different update frequencies and operational characteristics.	1. The proposed model (SEIS) achieves better event veracity estimation, event urgency inference, and deadline hit rate compared to state-of-the-art baselines.

Citation	Challenges	Opportunities
(Resch et al., 2018)	1. The central challenge in detecting events in online systems in near real time is that no prior knowledge about an event and no previously collected data are analyzed, but data that arrive as continuous streams. 2. Remote-sensing-based methods face the central disadvantage of temporal lags of approximately 48–72h before information layers can be produced, which are relevant to disaster management. 3. Remotely sensed data are limited in terms of their spatial, spectral, and temporal resolution, their usability, and their cost-efficient availability.	1. A central advantage of LDA is its transferability to other text corpora, and languages because of the rapid adaptation of the unsupervised learning approach to another text corpus.
(Rexiline Ragini et al., 2018)	1. The usage of social media data for the search and rescue operations during any crisis is a challenging task due to the high volume of information collected from various sources and the inability to segregate the information that are received from the affected people.	1. Social media data, if leveraged effectively, can help bring the crisis under control and reduce the risks of the disaster affected people or disaster-prone areas.
(Reynard & Shirgaokar, 2019)	1. Tweets may not be representative of the location where they originate, and populations that are unable to access social media networks like Twitter may be completely overlooked.	1. So that disaster managers may benefit from real time risk models to predict demand for services, analysts may need to use the early tweets to train a classification algorithm, and proactively use it as the tweet volume grows during and immediately after the event.
(Rossi et al., 2018)	1. Including data from social media in emergency management processes poses several challenges, including the availability of location information, the truthfulness and accuracy of the shared information, as well as the big volume, velocity, and variety of data.	1. In the emergency domain social media data can become a powerful resource for assessing in near real-time the evolution of a hazardous event, its impact and how it is perceived by the affected population.
(Roy et al., 2020)	1. During a disaster, emergency managers face challenges to monitor the massive volume of social media posts in real time.	1. More complex classification methods such as probabilistic neural networks, dynamic neural networks, and hierarchy-based models provide a better classification accuracy.

Citation	Challenges	Opportunities
(Rusho et al., 2021)	1. Due to limited information regarding the origin of tweets, the study could not capture spatial variations of social media interactions on active shootings.	1. Proposed infographics could serve as a data dictionary in future active shooting scenarios to maximize peer influence. 2. Social media is a readily available widely used tool for alarm dissemination without incurring any time for source authentication. 3. ML techniques and data-driven approaches can help to analyze the intrinsic motives, thoughts, and sentiments retrieved through social media interactions and facilitate more informed decision making.
(Ryan et al., 2021)	1. The sample of the study was biased toward women and relied upon self-report measures, which can be subject to social desirability bias. 2. The demography of social media platforms is biased and tends to overrepresent demographics such as urban residents, men, and people with specific interests in sports and politics.	1. The proposed approach offers the ability to automatically monitor or rapidly analyze unstructured data to identify and comprehend unanticipated or unforeseen health- and medical-related needs in the community. 2. NLP can help in understanding disease transmission and epidemic and pandemic trajectories by shedding light on community attitudes, behaviors, and experiences related to specific diseases. 3. NLP has a key advantage of being able to process large qualitative data sets rapidly and with greater objectivity than manual analysis.

Citation	Challenges	Opportunities
(Sarkar et al., 2023)	1. Difficulty of image assessment when using drone imagery based VQA algorithm: 1.1. Top-down pictorial representations make it very difficult to distinguish between several objects in the case of damage assessment. 1.2. The degree of scene complexity gets much higher due to noises coming from many sources. 2. Attention-based VQA frameworks fail to obtain relevant visual attention weights from RS images (resolved by implementing visual supervision to the model).	1. The VQA method can promptly deliver high-level scene information from images through interaction, which is limited in other computer vision tasks. 2. The direct interaction between a non-domain expert aand a machine makes the presented data-driven approach faster compared to other methods.
(Scalia et al., 2022)	1. The volume of emergency-related posts during disasters is large, but only a small fraction of posts is relevant and provide objective information. 2. It can be difficult to associate the available geolocation information with a location associated with the post. 3. Wrongly recognized surface forms negatively affect the linking/disambiguation phase. 4. Since the proposed method (CIME) does not rely on any prior information about the event, automatically learning event-specific parameters appears to be challenging.	1. Explore the benefits of crawling other sources besides Twitter.

Citation	Challenges	Opportunities
(Schwarz et al., 2023)	1. Short-text summarization and filtering can be improved. 2. Irrelevant data shared on social media is often difficult to filter. 3. When employing a BERT-based multi-task learning approach and labeling real-time capabilities are sacrificed to achieve higher accuracy. 4. Existence of noise, spam, and hate speech among the extracted social media data significantly distorts the results. 5. Minor events and side notes caused problems and could not always be appropriately detected or were largely over-detected at times.	1. People witnessing a disaster often use social media to report, inform, summarize, update, or warn each other faster than traditional news and mainstream media. 2. Automated situational awareness reporting allows for quick filtering, detection, and summarization of important information during disasters. 3. The temporal and spatial information extracted from Twitter is critical for supporting decision-making in DM. 4. The outcomes of the proposed method can help ensure adequate preparation and training in the future after a disaster. 5. Automation tools and methods allow for faster evaluation and summarization of information with the need of fewer manual steps and resources. 6. The proposed method focuses on delivering undistorted information fast to provide decision-makers with rapid situational awareness reports. 7. The SimHash algorithm is very computation time efficient and guarantees that unique semantic properties are not lost during the hashing process.
(Sermet & Demir, 2018)	1. Most environment related datasets are not easily accessible and understand by the public or other non-targeted groups. 2. Complexity of natural language. 3. NLP tools and systems such as Stanford NLP, wit.ai, NlpTools, take extra time to parse questions and provide limited capabilities to personalize the methods to connect ontologies. 4. Accurate assessment of date and time is one of the most challenging issues for the proposed NLP module.	1. Recent breakthroughs in sensor networks and RS technologies allow scientists to gather large-scale high-resolution datasets on the environment, water quality, and weather conditions. 2. Flood AI eliminates the need for traversing through complex information systems and datasets by allowing the user to access the knowledge as if by talking with an actual flood expert.

Citation	Challenges	Opportunities
(Shahbazi & Byun, 2022)	1. Social media content is changing continuously which is not a practical within disaster management.	1. The presented blockchain and ML pipeline in this system gives a significant direction for the future research work.
(Shibuya & Tanaka, 2019)	Not Discussed	1. Using public sentiment on social media can improve situational awareness in recovery operations.
(Singh et al., 2019)	1. Sentiment polarity (positive and negative) is identified in the data, but it is not good enough to understand the detailed emotions of the public during the crisis event.	1. The huge amount of Twitter data inspire the researchers to explore the trending topics, event detection and event tracking which help to postulate the fine-grained details and situation awareness.
(Sit et al., 2019)	1. Due to the noise in social media data, and the locational and contextual uncertainty of shared data, it is challenging to extract useful and actionable information generated by citizen sensors in real-time or in the aftermath of an event.	1. Lexical features help to achieve better precision, while semantic and stylistic features derived from metadata help improve recall. 2. In this study, CNN-based models were significantly better than other conventional ML algorithms.
(Stieglitz et al., 2022)	1. Receiving location-based information raises further challenges for several disaster management actors as this highly sensitive information has to be stored, processed and provided to align to the legal requirements. 2. CAs such as chatbots cannot simply be implemented in the same way in the context of Emergency Management Agencies as in other contexts.	1. Disaster management experts showed receptibility towards the application of Conversation Agents in crisis communication. 2. CAs can provide important information as a basis for decision making.

Citation	Challenges	Opportunities
(Sugino et al., 2023)	1. Surveys based on closed-ended questions may limit respondents' attitudes and choices and lead to biased results. 2. Authors used cross-sectional data from a single point in time through a questionnaire survey, which limits the ability to examine changes over time. 3. Due to the online nature of the survey, authors were unable to include residents aged 70 years and above.	1. The utilization of bottom-up public opinion collection, quantitative analysis, and information visualization methods, may aid towards resolving issues regarding coastal social–ecological systems. 2. The free-association survey method allows respondents to respond without being restricted by the researcher's prior knowledge, preconceptions, or assumptions. 3. BTM can guarantee a better topic quality compared to other methods, even when dealing with short documents. 4. BTM allows topic estimation to calculate topic scores on each response. 5. "Future Compass" demonstrates broader potential for identifying key trends in natural language data.
(Terracciano & Han, 2023)	1. Communities with better socioeconomic conditions have easier access to information via social media and may be more likely to benefit from the response phase of disaster recovery. 2. Twitter data collected does not fully represent the entire affected population. 3. Sentiment analysis may lead to mislabeling and misclassification of concepts.	1. Including elected officials' and utility providers' social media communications could allow to examine whether their responses were timely and adequate. 2. Combining sentiment analysis with semantic network analysis addresses the issue of interpretability within traditional sentiment analysis. 3. Eigenvector centrality may provide a more intuitive interpretation of varying priorities of analyzed keywords.

Citation	Challenges	Opportunities
(Tian et al., 2023)	1. Traditional survey methods have limitations in assessing public opinions and beliefs. 2. Several Tweets were not within any set category and therefore were not classified and analyzed. 3. Analysis during the disaster response period may not result in a complete picture of the disaster. 4. Results referent to a specific country cannot be blindly generalized to other areas and countries.	1. User generated content from social media may help guide researchers, policy makers, hospital administrators and public health authorities to more informed policy and decision-making. 2. Twitter may be used as a tool to dispel opposition to vaccination. 3. Twitter has become an alternate via of communication and a source of information in emergency situations. 4. BERT is a pre-trained model in which fine-tuning is less expensive and requires less time. 5. VADER is more effective than other lexicon-based methods.
(Valinejad & Mili, 2023)	1. Surveys have several disadvantages, such as high cost, limited sample size, and the possibility of response bias.	1. LIWC can detect meaning in a variety of experimental contexts. 2. The proposed method can be extrapolated into other disaster types. 3. The proposed method can be used to simulate a variety of situations that are either prohibitively expensive or impossible to test in the real world.
(Wahid et al., 2022)	1. It is hard to extract useful information from Twitter as posts often lack enough context and contain ambiguous terms. 2. For contextual classification of textual data, supervised learning approaches need well labeled data for training which is challenging to obtain during disasters. 3. Texts in non-English languages are hard to label and their lexicons are hard to find. 4. LDA model selection to determine optimal number of topics was a complex and time-consuming task.	1. The proposed method handles noisy values, sparse data and slang language from datasets in an efficient way, improving the classification performance. 2. The proposed framework performs better than the baseline approaches.

Citation	Challenges	Opportunities
(Wang et al., 2020)	1. High-quality geo-tagged urban flooding data are challenging to collect. 2. Extracting social media information in real-time is challenging, as it requires a rapid and automatic processing tool to help deal with the high volume, fast velocity, and strong heterogeneity of the data. 3. Misinformation is often found in social media stream and has been seen in the context of disasters.	1. Social media is an emerging data source that has a great potential to meet the rising data challenge of urban flooding. 2. Neural network-based models generally outperform traditional models such as CRF in the task of NER. 3. Manual data inspection based on automatically filtered tweets has demonstrated promising results in improving data mining outcomes.
(Wang et al., 2021)	1. Mining social-media data is challenging due to the huge quantity, unbalanced user composition, redundancy, presence of incorrect or imprecise user locations, and false information. 2. The biased demographic composition of users in social media will contribute to bias in the study results.	1. There is a great need to develop efficient data-mining algorithms and frameworks to extract useful information from social-media data for resilience studies.
(Wang et al., 2023a)	1. Social media data often represents a biased demographic composition of users. 2. Social media data may include false information. 3. Answers to the same question may vary when the analysis scale changes. 4. Unable to get / use accurate locational data to / from all users.	1. A fine-grained study can help stakeholders in identifying problem areas and developing targeted strategies more efficiently. 2. Having results conducted at multiple scales translate in a better understanding of the relationships between Twitter use and population conditions during disasters.
(Wang et al., 2023b)	1. Rule-based knowledge extraction methods suffer from drawbacks such as cumbersome rule definition, poor scalability, and low extraction accuracy. 2. The GCN model presents a lower extraction accuracy when faced with long textual data. 3. The entity recognition algorithm may encounter difficulties in accurately identifying certain entities.	1. Extracting information from emergency documents using information technology is essential to enhance the efficiency and capacity of emergency management. 2. Rule-based knowledge extraction methods possess advantages such as strong interpretability and high controllability. 3. DL methods based on RNN can effectively leverage the temporal dependencies in text and better capture the complex relationships within the data. 4. The pre-training phase enables the model to better understand the meaning and usage of terminology in a particular domain.

Citation	Challenges	Opportunities
(Wu et al., 2022)	1. Hard to identify and make sure the geo-location data collected from microblogs is accurate. 2. Multisource data fusion could overcome the data shortcomings and obtain more accurate results. However, this technology is not sufficiently mature.	1. Many studies have shown that social media can be used for rapid damage assessments during typhoon disasters, 2. An advantage of using social media data to extract victim location is that emergency management information can be obtained at the same time.
(Wu et al., 2023)	1. Presence of noise, heterogeneity, and sparsity associated with publicly generated social media data. 2. Microwave RS is limited and cannot monitor or extract surface water information, due to factors such as cloud cover and vegetation canopy.	1. The introduction of other modal data, such as images and videos, resulted in a significant improvement in spatial error. 2. IoT sensors can accurately and swiftly respond to urban inundation issues and facilitate real-time alerts and monitoring.
(Xia et al., 2023)	1. More specific data was partially unavailable. 2. Impossible to avoid sampling bias in web-based data.	1. Twitter can be a valuable tool for studying public opinion because it allows researchers to collect large amounts of data in real time and at a low cost. 2. As Twitter data is considered public, the study did not require research ethics approval.
(Yan et al., 2020)	1. Qualitative approaches such as interviews and focus groups are to some extent costly because they rely on researchers' personal and embodied linkages with the field and are confined to a relatively small proportion of stakeholders.	1. Social media is ubiquitous, rich in content, and big in volume, velocity, and variety. 2. social media is a valuable source for monitoring and assessing the post-disaster recovery of tourism destinations.
(Yao & Wang, 2020)	1. Extracting useful information from social media and interpreting data accurately for disaster management is still challenging.	1. Social networking platforms such as Twitter can provide real-time information for disaster managers and affected populations during large-scale disasters.

Citation	Challenges	Opportunities
(Yao et al., 2021)	1. Once the text length of phrase gets bigger, the complexity of word combination increases dramatically, and may decrease the accuracy when mapping and matching. 2. The social media system is a complex network with numerous nodes and high internal accessibility.	1. Social media can provide useful supplement information, especially in the very early stage of a sudden disaster. 2. Social media data containing massive quantities of information has the advantages of timeliness, efficiency, and multiple spatiotemporal scales, and can supplement existing methods.
(Yuan et al., 2021)	1. The collection of timely data for evaluating social aspects of disaster resilience during disasters remains a critical challenge.	1. Understanding social aspects of disaster resilience through citizens' dynamic interactions with disasters and the built environment can benefit crisis response managers in situation awareness and further support them to design and implement response strategies. 2. A rule-based method for investigating different demographic groups' sentiment expressions is applied.
(Zander et al., 2023)	1. Other data sources, besides Twitter, may lead to ethical and privacy considerations and challenges. 2. Social media data may not represent the entire population. 3. Misinformation spread is often a common problem of social media.	1. An understanding of how people use social media during natural disasters can help detect and track hazard events early as well as tailor policy and practices in Emergency Services. 2. Twitter data is generally considered public and can be easily obtained using an API.
(Zhang et al., 2022)	1. Governments are facing the challenge of analyzing the emotional tendency of online public opinion during emergencies to regulate people's emotions more effectively and maintain social stability. 2. Sparse features, ignoring fine-grained negative emotions, is a common problem when dealing with large-scale short, unordered texts and extracting text features. 3. The existing emotion classification methods focus on only three emotion polarities (positive, negative, and neutral) whose emotional granularity is not fine enough.	1. The results show that the proposed CNN-LSTM model achieved competitive classification performance. The method was rapid, effective, and feasible and could be more suitable for optimizing emotion regulation policies. 2. The CNN model efficiently extracts higher level features using convolutional layers and max-pooling layers. 3. The LSTM model is capable to capture long-term dependencies between word sequences.

Citation	Challenges	Opportunities
(Zheng et al., 2021)	1. Weibo is moderated, and posts can be removed.	1. Policymakers, organizations, corporations, and members of the public can all benefit from a better understanding of how social media is used in crisis circumstances. 2. An approach which considers more than 2 or 3 levels of sentiments permits more complex analysis than dichotomous approaches.
(Zhou et al., 2022)	1. Efficiently leveraging social media in rescue operations remains challenging sue to the lack of tools to identify rescue request messages on social media fast and automatically. 2. People using languages other than English on social media cannot benefit from the research results. 3. Location bias, temporal bias, and reliability issues present in social media posts.	1. People using languages other than English on social media cannot benefit from the research results. 2. Location bias, temporal bias, and present in social media posts.
(Zhou et al., 2023)	1. Evaluating LDA models is notoriously challenging due to its unsupervised training process. 2. Determining the classification schema into which the tweets should be classified. 3. Rapidly and precisely update and extract SA information to government agencies and public citizens. 4. Difficulty in independently analyze the topics solely for one specific event.	1. Identifying emerging topics during different disaster events. 2. Providing multilabel references to the classification schema. 3. The proposed model helps to quickly identify and extract situational awareness info.
(Zhuang et al., 2021)	1. Text may be expressed in a complex context that includes slang and emoticons. 2. The weights of feature words in the original LDA "bag-of-words" models are uniform, resulting in the problem that the topic model favors high-frequency terms.	1. Authors propose an improved LDA-ARMA theme analysis model that can effectively solve the problem of complex public opinion analysis. 2. The LDA model improved using the Gaussian function gave better performance in terms of topic clustering.

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa