Taylor & Francis
Taylor & Francis Group

International Journal of Digital Earth

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjde20

Detecting crisis events from unstructured text data using signal words as crisis determinants

Hansi Senaratne, Martin Mühlbauer, Stephan Götzer, Torsten Riedlinger & Hannes Taubenböck

To cite this article: Hansi Senaratne, Martin Mühlbauer, Stephan Götzer, Torsten Riedlinger & Hannes Taubenböck (2023) Detecting crisis events from unstructured text data using signal words as crisis determinants, International Journal of Digital Earth, 16:2, 4601-4620, DOI: 10.1080/17538947.2023.2278714

To link to this article: https://doi.org/10.1080/17538947.2023.2278714

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 12 Nov 2023.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjde20

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2023, VOL. 16, NO. 2, 4601–4620
https://doi.org/10.1080/17538947.2023.2278714

OPEN ACCESS Check for updates

Detecting crisis events from unstructured text data using signal words as crisis determinants

Hansi Senaratne[a], Martin Mühlbauer[b], Stephan Götzer[c], Torsten Riedlinger[b] and Hannes Taubenböck[b,d]

[a]TWT GmbH Science & Innovation, Stuttgart, Germany; [b]Earth Observation Center, The German Aerospace Center, Wessling, Germany; [c]Department of Aerospace and Geodesy, Technical University of Munich, Ottobrunn, Germany; [d]Institute of Geography and Geology, University of Würzburg, Würzburg, Germany

ABSTRACT

Earth observation data provides valuable information and support along the disaster management cycle. However, information from satellite remote sensing is often not available in the first hours a crisis occurs, due to several reasons, e.g. pre-defined acquisition times, cloud coverage, downlink capacities. To fill this time gap and add value to the incoming results from remote sensing data, ancillary datasets such as Twitter data become useful to enrich data and get insights into events by leveraging their spatio-temporal and thematic references. However, the main disadvantage of using Twitter data is the noise that is introduced into analyses by these data. Among other reasons, this is mainly caused by the use of insignificant search criteria that are used to harvest the data, that often result in irrelevant, noisy data (e.g. using insignificant keywords or incorrect geotags to filter data). This paper presents a method to identify crisis-event specific signal words, that are then used together with Part Of Speech (POS) tagging to filter the Twitter streams, and gather crisis-event specific data. These data are then used to estimate the location hotspots of the crisis events. The developed methods are applied as a proof-of-concept to determine flood events in May of 2022 .

ARTICLE HISTORY
Received 1 August 2023
Accepted 29 October 2023

KEYWORDS
Crisis-event detection; information retrieval; signal words; VGI

1. Introduction

According to the United Nations Office for Disaster Risk Reduction (UNDRR), a crisis event can be defined as *a situation or occurrence that presents a critical threat to the health, safety, security, or well-being of people, assets, or the environment, and requires urgent and extraordinary response measures to mitigate or resolve the situation* (UNDRR 2017). Crisis events that occur naturally or are made anthropogenically, typically result in loss of life and damage to infrastructure. The knowledge about such events is often fragmented and limited, that reconstructing these events in detail becomes difficult, and finding answers to what, how, when, where becomes an increasingly strenuous process (Albeverio, Jentsch, and Kantz 2006). Earth Observation has proven the capability to support along the disaster management cycle: from the stages of risk assessment and preparedness to response, mitigation, and recovery (e.g. Taubenböck et al. 2008). However, high resolution

CONTACT Hansi Senaratne hansiv.senaratne@gmail.com, hansi.senaratne@twt-gmbh.de TWT GmbH Science & Innovation, Ernsthaldenstr. 17, 70565 Stuttgart, Germany

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

remote sensing data have limited availability to assess and analyse these crisis events, e.g. due to their periodic observation cycles or processing durations, especially for near-real time applications. With the ample availability of web data and Volunteered Geographic Information (VGI), reducing data gaps and thereby improving the detection of such crisis events have become possible (see e.g. Senaratne et al. 2023). For example, Twitter produces over 500 million Tweets a day. Contributors on VGI platforms such as Twitter add thematic, spatial, and temporal references (e.g. through hashtags, geotags, and timestamps) to their content explicitly (e.g. tagging the observed location of an event) and implicitly (e.g. by mentioning the location and other attributes of an event in a microblog on Twitter). These references give insights into the geographical area and populations affected, the time duration of impact, and other associated phenomena such as cascading effects of a given crisis event (Musaev, Wang, and Pu 2014). These insights are valuable input to rapid mapping of affected areas, exposure modelling, risk analysis, and to create pathways for adaptation and resilience building in affected communities. However, a few obstacles still hinder these data from being valuable input to such computations. Some example obstacles of Twitter data are: these microblogs are unstructured, and the useful thematic or spatial references are mostly implicit in the main text body, making it a challenge to gather relevant data. The 1% feed that was freely available through the Twitter API version 1.0 (Twitter API v1) was not randomly sampled for a given search criteria, and the biases in the harvested data are not known (Thomson et al. 2018). Furthermore, as with any type of data, Twitter data are inherent with uncertainties (due to fake news, incorrect references, trolling and malice-driven participation on social media etc.). To gather Twitter data that are relevant for crisis events, for the purpose of training machine learning models (from existing archives, or from the Twitter API v2), particular attention needs to be paid while specifying the filter criteria. Thematic references help to search the crisis events (the what) on Twitter based on textual labels/tags, hashtags, specific keywords mentioned in Tweets etc. To filter a data stream based on a thematic reference, it is important to identify one (e.g. one or more keywords) that significantly represents the crisis event in a pool of data, and not choose a reference only for its likeness to the event. This ensures that as many relevant Tweets as possible related to the event are included in the data for a more comprehensive analysis.

Similarly, determining the most relevant spatial reference on Twitter is important for identifying/approximating where the crisis event took place. The event detection approaches in the literature (briefly reviewed in Section 2 below) use one or more of the available spatial references on Twitter: Tweet location (available only with a GPS-enabled device) – where Tweeters geotag their content on the fly with a specific latitude/longitude point coordinate or a Twitter Place polygon set of 4 latitude/longitude coordinates (both at a user-preferred granularity), mentioned location – where Tweeters mention a place name in their content, profile location – the registered location of the Tweeter account. If an unsuitable spatial reference is used for approximating the location of an event, the results may be misleading. E.g. for identifying the epicentre of an earthquake, parsing the Tweet message for the mentioned location would be more useful than parsing the profile location (as demonstrated also in Figure 1). In contrast, for identifying where people are mostly reacting from for the said event the Tweet location would be more useful [although only 0.8% of the full feed is geotagged (Sloan and Morgan 2015)]. Context always matters when deciding on which spatial reference to use, be it while filtering the data stream or while analysing the filtered data.

In this paper we discuss and demonstrate how to detect selected crisis events by effectively filtering the Twitter data stream based on these references, using established machine learning methods. As example crisis events, floods and earthquakes are considered in this study based on the availability of data. The main contribution of this paper is the application of established machine learning methods to better address the challenge of retrieving relevant data for the detection of floods and earthquakes. To achieve this, a 3-step approach that includes a novel concept of a dictionary of signal words for crisis events is followed, to effectively gather event-related data, and determine the location hotspots of these events. In step 1, a dictionary of signal words is established for Floods

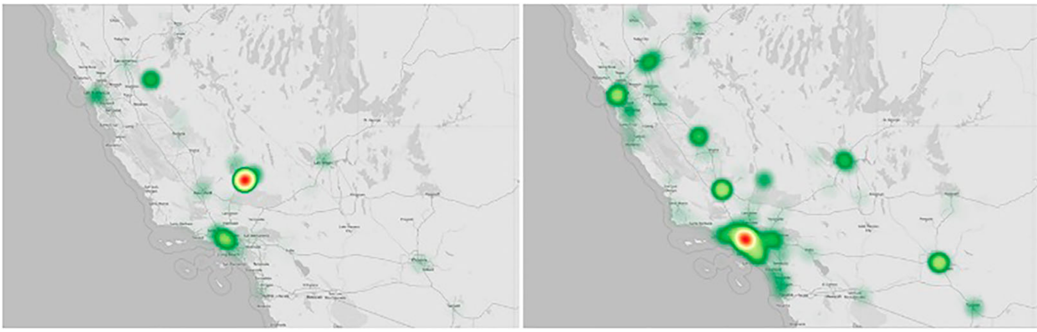

Figure 1. An example to showcase two different ways of localising hotspots for an earthquake event in Ridgecrest, California during 6–8 July, 2019, using Twitter data. Left: mentioned locations. Right: Tweet locations as specified by the Tweeter.

and Earthquakes based on a Logistic Regression model using pre-labelled data specific for Floods and Earthquakes. These signal words are then later used to filter the Twitter live stream for identifying specific events in near-real time. In step 2 the gathered data is tagged based on their Parts-Of-Speech (POS) to further narrow down and classify the dataset to contain only semantically similar Tweets for the crisis events. In step 3 the events are localised by extracting the mentioned locations and the Tweet locations in the Tweets and generating their location hotspots to approximate the locations affected by the crisis events. This workflow is shown in Figure 2. In the following Section 2, an overview of related work is presented, and Section 3 presents the methodology. Section 4 presents the findings, and the limitations and future research perspectives are discussed in Section 5. The paper will conclude the findings in Section 6.

2. Related work – filtering tweets for crisis events

Many works have demonstrated the usefulness of Twitter data for gaining insights into crisis events by leveraging their spatio-temporal and thematic information. Depending on which spatial reference has been used, analyses at different geographic scales have been conducted and insights derived. In related work, Pezanowski et al. (2018) developed a tool to support situational awareness during crisis events using Twitter microblogs. Their map-based web application essentially incorporates a visual interface that enables the user to understand place, time, and thematic components of emerging situations. The Tweet location, mentioned location, and the profile location are extracted from all the Tweets in their archive, allowing the users to explore the crisis events based on a selected location. This search by-query application relies on keyword inputs by the user (requiring the user to possess prior knowledge of the situation they want to explore), to then visualise a list of Tweets pertaining to the keywords. Their text content analysis, although limited to the keyword frequency, provides an overview of how often the selected keyword occurs in the dataset in a given time frame. Another similar tool called ScatterBlogs2 (Bosch et al. 2013) focuses mainly on the filtering accuracy where the users are enabled to interactively build task-tailored message filters. At the filter creation stage the user is able to visually create classifiers and filters to train and test them on archived event-based messages. Further, the user

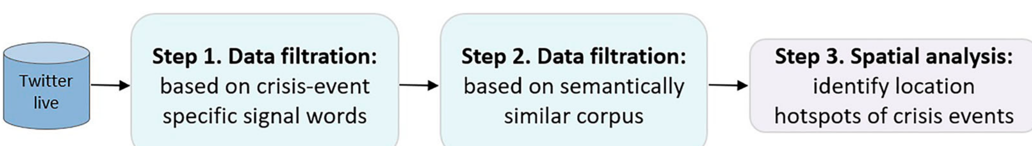

Figure 2. The workflow of the proposed methodology.

has to define keywords that describe the event-relevant messages in order to maximise the search query. In their approach they rely on the Tweet location of the microblogs. A survey of similar toolkits can be found in Ngamassi et al. (2017).

Bouadjenek and Sanner (2019) presents a method, referred to as Viz-TIR, that is designed for real-time clustering and visualisation of Tweets based on their relevance to a query. A Greedy Relevance-Driven Clustering algorithm is used for choosing keywords by making locally optimal decisions at each step. The method is designed for real-time clustering, making it suitable for dynamic information retrieval scenarios, and can be useful for quick and computationally efficient keyword selection. Overall, Viz-TIR is a method designed to create relevance-driven clusters of information elements, offering a balanced approach to precision and recall. It is suitable for real-time applications, however requires careful parameter tuning and relevance score estimation.

Marujo et al. (2015) proposed a methodology that combines traditional approaches with unsupervised feature extraction and predictive modelling to improve keyword extraction from Twitter data. Their use of the MAUI toolkit as a baseline leverages hand-engineered features and decision tree modelling, which can provide robust keyword extraction results. The incorporation of unsupervised methods such as the Brown Clustering and Continuous Word Vectors to address issues related to lexical variants, enhances the model's generalisation. Furthermore, their approach to predicting the number of keywords in a Tweet allows for flexibility and adaptation to different Tweet lengths, addressing the high variance in keyword count. While it addresses specific challenges, such as lexical variants and keyword count variation, it also highlights the importance of domain adaptation and the limitations of certain baselines for Twitter data.

The L-CrisMa framework by Interdonato, Guillaume, and Doucet (2019) offers a flexible and unsupervised approach to extract and rank crisis-related information from social media. The framework starts by collecting microblog posts related to a specific crisis event, and these posts are then preprocessed, including filtering out non-textual content, language filtering, and removing redundancy. The preprocessed Tweets are clustered using techniques such as Latent Dirichlet Allocation (LDA), Nonnegative Matrix Factorization (NMF), or k-means clustering. The goal is to group similar Tweets into clusters. Clusters are then ranked based on their semantic relatedness to a crisis lexicon. Clusters with higher relatedness are considered more informative about the crisis event. Within the top-ranked clusters, individual Tweets are ranked based on their semantic relatedness to the crisis lexicon. The accuracy of the framework heavily depends on the quality of semantic relatedness measures and the crisis lexicon. If these are not well-tailored to the specific crisis, the results may suffer.

Hagras, Hassan, and Farag (2017) presents a method that involves collecting and processing a dataset of Tweets related to the tsunami crisis in Japan in 2011, with the goal of detecting relevant Tweets using Latent Dirichlet Allocation (LDA) as a text clustering algorithm. To collect their data they harvest Tweets containing the word 'tsunami' or the hashtag '#tsunami' from the Twitter stream during the 2011 tsunami crisis. Irrelevant Tweets are filtered based on criteria like relevance to tsunami, reporting of tsunami-related events, or reporting of loss or damages. Latent Dirichlet Allocation (LDA) is applied to the preprocessed dataset to identify topics within the Tweets. LDA is trained with different values of K (number of topics) to find the optimal value. Topics are inferred, and a probability distribution over terms is generated for each topic. A threshold value is used to classify Tweets into topics and detect relevant documents based on their topic distribution. Setting the threshold for relevance detection may be challenging and could affect the accuracy of document classification. The use of the Spark machine learning framework makes the method scalable to handle large datasets.

Gu, Qian, and Chen (2016) presents a methodology that aims to translate Tweets into traffic incident information. Twitter's REST APIs and streaming APIs are used to collect Tweets related to traffic incidents. REST APIs are suitable for historical data retrieval and allow querying by keywords. The goal is to obtain as many traffic incident (TI) Tweets as possible. An initial set of keywords is used to collect Tweets related to traffic incidents (seed words). This adaptive data acquisition process then iteratively expands the dictionary of relevant words based on the frequency of words in acquired TI Tweets. New queries are generated using word combinations, and the process continues until convergence. A Semi-Naive-Bayes (SNB) classifier is employed to classify Tweets as either TI or non-traffic incident (NTI) Tweets. This classification helps filter out irrelevant Tweets. For TI Tweets, a Supervised Latent Dirichlet Allocation (sLDA) classifier is used to identify the category of the traffic incident, such as accidents, road work, hazards, events, or obstacle vehicles. The adaptive data acquisition process allows the methodology to evolve over time to capture new and emerging trends in Twitter language related to traffic incidents. The methodology however relies on manually labelled TI and NTI Tweets, which may introduce labelling biases and limits the availability of a ground truth dataset. Furthermore, the effectiveness of the methodology depends on the quality of the seed dictionary and the relevance of the initial set of chosen keywords.

In A similar use-case, Dabiri and Heaslip (2019) proposed a methodology to detect traffic-related Tweets by leveraging unsupervised deep learning for Tweet modelling and supervised deep learning for classification. Over 50,000 Tweets were collected and labelled into three classes: non-related to traffic, traffic incidents, and traffic condition and information. The methodology uses deep learning techniques, including word-embedding models, Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN), to model Tweets as numerical feature vectors. RNN captures long-term dependencies between Tweet words, while CNN captures local word correlations. These models are employed to discriminate traffic-related Tweets from non-traffic Tweets. The authors showcase the superiority of their approach in detecting traffic-related Tweets from Twitter data against the state-of-the-art. Overall, the proposed methodology offers a robust approach to detecting traffic-related Tweets using deep learning techniques and contributes a valuable labelled dataset to the field. However, it requires careful consideration of data availability, computational resources, and external validation for practical implementation.

In yet another use-case for traffic-incident detection, Alomari, Mehmood, and Katib (2019) describes a methodology focused on automatic traffic event detection from Arabic Tweets using supervised machine learning (ML) algorithms and Apache Spark. Tweets are collected via the Twitter API, focusing on geolocation filtering to obtain Tweets from Saudi Arabia. Arabic Tweets related to specific city events are also collected. The collected data is stored in MongoDB, a NoSQL database, with JSON objects representing Tweets. Term Frequency-Inverse Document Frequency (TF-IDF) is used to represent the importance of terms in Tweets. The TF-IDF vectors are generated using the Spark ML package. A classifier is built to filter out irrelevant Tweets. Three machine learning algorithms (Naïve Bayes, SVM, logistic regression) are used, and model performance is evaluated using precision, accuracy, recall, and F-score. After filtering, relevant Tweets are further classified into eight event categories, such as Traffic Condition, Roadwork, and Fire. Multi-label classification is applied, and binary classifiers are trained for each event type. Random undersampling is used to address class imbalance. The methodology validates the event detection results by extracting information such as time and location from Tweets. The top frequent terms are used to search official news or newspaper websites to confirm event occurrences. Overall, the methodology provides a robust framework for extracting traffic-related events from Arabic Tweets, with a strong focus on data preparation, feature extraction, and classification. However, it relies on the quality and availability of Twitter data and may require careful handling of language-specific challenges.

de Bruijn et al. (2019) presents a methodology to extract information about flood events from Twitter data. This involves filtering relevant Tweets, identifying and verifying flood-related events, and determining their temporal and spatial characteristics. Keywords related to floods in various languages are used to filter relevant Tweets. A machine learning model based on BERT (Bidirectional Encoder Representations from Transformers) is employed to classify Tweets into two categories: relevant to ongoing flood events and irrelevant. A labelled dataset of Tweets, including their spatial and temporal information, is used to train and fine-tune the model. (Near-) duplicate Tweets are discarded to ensure data reliability. A burst detection algorithm called SEQAVG is applied to identify enhanced Twitter activity, potentially indicating flood events. The algorithm considers the time intervals between consecutive Tweets in each region to detect bursts, and regions are categorised as 'flooded' or 'normal' based on Twitter activity. Reliance on keywords for Tweet filtering may result in false positives or missed events if the keywords are ambiguous or context-dependent. The methodology assumes that bursts of Tweets indicate actual events, which may not always be the case.

Behl et al. (2021) classifies Tweets related to disaster relief into two categories: Tweets from victims in need of assistance and Tweets from resource providers (government and NGOs). This classification is achieved using a Multi-Layer Perceptron (MLP) deep learning model. The MLP consists of multiple layers of perceptrons, and training it involves finding the best set of weights and biases for accurate predictions. The model is designed for text classification, specifically for Tweets related to disaster relief. Disaster-specific Word2Vec embeddings are used to represent words as low-dimensional vectors, capturing both semantics and syntax. These embeddings are based on a substantial dataset of crisis-related Tweets. The pre-processed text data is fed into the MLP architecture, which includes an embedding layer, dropout layers, fully connected layers with Rectified Linear Unit (ReLU) activation, and an output layer with Softmax activation for multi-class classification. The model is trained using the Adam optimiser with a specified learning rate. The loss function used is Sparse Categorical Crossentropy, which assesses the model's predictions against the true labels. To address class imbalance, over-sampling and under-sampling techniques are employed.

Kejriwal and Zhou (2019) presents SAVIZ (Situational Awareness Visualization) for visualising information extracted from Twitter during crisis events. SAVIZ ingests a Twitter corpus collected after a crisis event. Methods like CrisisLex and pipeline approaches are used to collect relevant Tweets. The system performs situational awareness analysis, often using systems like ELISA, to categorise Tweets into different labels related to the crisis. SAVIZ employs the fastText word embedding package to convert words and sentences into dense, continuous, and low-dimensional vectors (100-dimensional). All relevant information, including categorical labels assigned during analysis, is compiled into a NoSQL file. The accuracy of situational labels assigned during analysis (e.g. by ELISA) may vary, potentially leading to noisy data visualisations. In some cases, ground truth data for validating the accuracy of situational labelling may not be available.

Rossi et al. (2018) extracts information from social media, particularly Twitter, during extreme weather events. The primary goal is to monitor and analyse Twitter content related to these events, detect emergency situations, and classify Tweets as 'informative' or 'not informative'. Twitter's Streaming API is used to retrieve real-time social media content, with a focus on the identified hazards and languages. Filtering parameters like language and track phrases (keywords and hashtags) are applied to access relevant content effectively. The classification of Tweets relies on vector representations of Tweet text using the fasttext tool, considering sub-word information. The system analyses aggregated volumes of social media content in predetermined time-frames, detecting anomalies or emergency events. Seasonality in Twitter activity is considered, and the EDM (Event Detection Module) triggers alerts in near real-time when outliers are detected. Performance metrics like precision, recall, and f-1 score are used to evaluate the classifier's ability to identify 'informative' Tweets. The definition of 'informative' Tweets can be subjective, and classification accuracy may vary. Also, their dependency on external tools such as fasttext, may produce false positives or miss relevant events, impacting alerting accuracy.

The work of Zhang et al. (2018) focuses on entity extraction and linking from text documents, particularly across low-resource languages. The key components include entity extraction, entity linking, and localisation. The method employs bi-directional long short-term memory (LSTM) networks with a Conditional Random Fields (CRFs) layer for entity extraction, and uses creative methods to derive training data from sources like Wikipedia. After extracting entity mentions, it links them to knowledge bases (KBs), such as GeoNames and Wikipedia, employing name translation and collective inference techniques. The methodology also provides a set of RESTful APIs for running pre-trained models and training customised models for different languages. The

Proposed Methodology addresses the challenge of entity extraction and linking in low-resource languages, making it applicable to a wide range of linguistic contexts. The approach assumes collective inference and salience measures for entity linking, which may not always produce highly precise results. Furthermore, customised model training may require substantial labelled data for low-resource languages, which might be scarce or time-consuming to obtain. The methodology's effectiveness depends on the quality of the available knowledge bases and their coverage of entity mentions.

Senaratne et al. (2014) and Senaratne, Lehle, and Schreck (2022) introduced a methodology that is tailored for crisis event analysis on Twitter. The method provides a comprehensive characterisation of conversation trajectories, considering geospatial (e.g. distance between clusters, speed of cluster propagation) and content-based characteristics (e.g. sentiment, credibility) to provide a multifaceted view of the data. Their approach is designed to analyse Twitter conversations during crisis events, making it valuable for situational awareness and monitoring. The use of density-based sequential DBSCAN clustering allows for adaptive clustering without specifying the number of clusters.

As can be seen in these related works, some approaches are led by preconceived knowledge of events to choose keywords that are then used to filter data streams. Other approaches rely on statistical approaches to identify keywords that significantly represent data pertaining to a given event. Works on flood-event detection based on Twitter data, rely on keywords that are chosen for their likeness to flooding events, and the statistical significance of the dataset to the said event is assessed after the data has already been filtered. This approach however disregards many Tweets during the information extraction phase, that are potentially relevant to the flooding event (but do not contain the chosen keyword). To fill this gap, we are inspired by the established methods to demonstrate an application to identify flooding events around the world, by using a *dictionary of statistically significant signal words to extract relevant Tweets*. This paper further demonstrates through a proof-of-concept how the inclusion of the dictionary of signal words harvests Tweets that are relevant to flooding events, as opposed to using singular keywords such as *flood(s)* or *flooding*, that failed to harvest Tweets for known flooding events in certain regions in the world.

In addition to the reviewed works in this section, other specific related works on context analysis and Natural Language Processing are presented under the respective sections below.

3. Methodology

To improve the information retrieval process by including more relevant Tweets for a given event, we have created a dictionary of signal words for two event scenarios, a flood and an earthquake. These dictionaries of signal words act as crisis determinants by identifying a list of keywords that signals a given crisis event with probability, as opposed to a single randomly chosen keyword, that can be used to filter the data stream. The filtered data are then classified based on part-of-speech tagging to give semantic meaning to the words with respect to the rest of the sentences in the Tweets, and further filter the corpus for semantically similar Tweets. These two steps are described in detail below. To demonstrate the methodology of these two steps, labelled datasets from Imran, Mitra, and Castillo (2016) were used for the considered events.

3.1. Dictionary of signal words

The popular keyword-based filtering method used in many state-of-the-art works for event detection (see Section 2) rely on a single keyword to filter the data stream based on its likeness relating to an event (e.g. using the keyword *earthquake* to identify an earthquake event). However, such keywords can be used in various contexts (e.g. *Clippers vs. Lakers in the playoffs would cause a massive earthquake in LA*), and therefore the information retrieval process will return in addition to the relevant data, also irrelevant data, that may result in inaccurate and inconclusive results, if not properly filtered out. To mitigate such inaccuracies, we use a dictionary of signal words chosen based on a Logistic Regression (LR). LR is a statistical model that measures the statistical significance of each independent variable based on probability (Alam, Joty, and Imran 2018). Works such as Shah et al. (2020), Prabhat and Khullar (2017), Yen et al. (2011), Liu, Zhang, and Wu (2014) have used LR in text classification approaches. Two events are chosen for this study based on the availability of already labelled data from the CrisisNLP repository (Imran, Mitra, and Castillo 2016) – these events are the Nepal earthquake that took place in April, 2015 and the Queensland floods that took place in January, 2013. These data have been collected from the Twitter streaming API v1 by using event-specific keywords and hashtags to filter the stream. A random sample of this data (comprising of 11,670 Tweets for the Nepal earthquake and 10,033 Tweets for the Queensland floods) has been manually labelled through the CrowdFlower crowdsourcing platform as relevant or non-relevant to the events (Alam et al. 2021).

For feature representation in the LR model, we used a bag-of-words method (HaCohen-Kerner, Miller, and Yigal 2020). The frequency of each word is used here as feature weights to train the model, based on the assumption that it is a good indicator to find Tweets related to a given event. This allowed us to extract the top features of the model. These top features, each with a coefficient value, are the signal words that constitute the dictionaries for the two considered events. If a Tweet contains one or more of the signal words in the dictionary, it is classified as a Tweet related to an event. For each type of event a different dictionary has to be created.

As seen in Figure 3, the top 10 features with a coefficient >0 included in the dictionary of signal words for the 2015 Nepal earthquake, as determined by the LR model are: *rt*, *earthquake*, *thoughts*, *rescue*, *pray*, *aid*, *donate*, *relief*, *quake*, *team* (rt, which stands for 'retweet' would be ideally removed with stemming). Figure 4 shows the top 10 features included in the dictionary of signal words for the 2013 Queensland floods. These are: *flood*, *queensland*, *australia*, *flooding*, *floods*, *water*, *power*, *bigwet*, *storm*, *flooded*. These dictionaries of signal words for the different events can be further improved with more accurately labelled datasets collected across different geographic regions and demographics, to ensure the inclusion of more generic keywords with higher coefficients, and thereby omit too specific keywords such as those conferring to colloquial speech (e.g. the word *bigwet* used in Australia to refer to the floods in Queensland), or references specific to one location.

3.2. Part-of-speech tagging and classification

One major drawback of using the keywords-based approach as described in the sections above, as well as in many other NLP-based approaches in the literature, is that it would also consider the Tweet '*…Showtyme_33 took it to the house & Lambeau erupted a la the Earthquake game in Seattle…*' as relevant to an actual earthquake event, although the word *earthquake* in this Tweet is functioning as an adjective modifying the noun 'game', and further helping to provide additional context or reference to a well-known incident that took place in Seattle. Not only do we need relevant and statistically significant keywords to filter data streams, but also methods to semantically enrich the harvested data, and thereby enhance the information retrieval process with the appropriate context. For example, Hidden Markov Models (HMM) (Chakrabarti and Punera 2011), help to disambiguate parts of speech and predict the probability of sequences of language. Some works that go in this direction are Romero and Becker (2019) that refer to external knowledge bases such as DBpedia or Linked Open Data to extract conceptual features of *what*, *who*, *where* in Tweets, Yu et al. (2016) that use Deep Neural Networks (DNN) based on semantic hashing of Tweets, or Khare, Fernandez, and Alani (2017) that use semantic annotation, expansion, and filtering to extract the semantics in Tweets. Future possible research directions on this topic are further elaborated in the Discussion section in Section 5.

In the second step of our approach, we further narrow down the dataset to include Tweets that are similar by definition and context. In natural language processing, part-of-speech (POS) tagging

Tweets	6899
Relevant	3293
Correctly labelled	2141
Falsely labelled	1608
Percentage	65.02

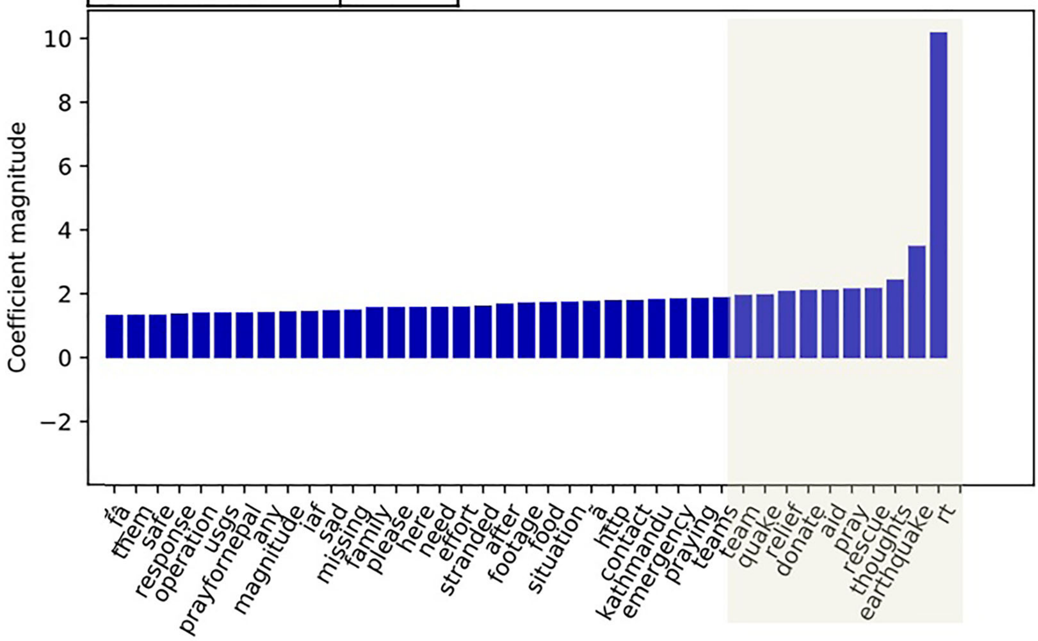

Figure 3. Results of the Logistic Regression for the 2015 Nepal earthquake dataset. The top 10 features are highlighted.

is the process of classifying, and thereby tagging words in a text corpus into their grammatical parts of speech (e.g. verb or noun), based on the definition of the word and its context. In contrast to conventional corpuses, the challenges of text analysis on Twitter data are mainly due to the conversational nature of the text, non-conventional orthography, and the limited character set for each Tweet (Gimpel et al. 2010). Many works such as Gimpel et al. (2010), März, Trautmann, and Roth (2019), AlKhwiter and Al-Twairesh (2021), Nelakuditi, Jitta, and Mamidi (2018) have demonstrated the use of POS in English language corpuses as well as in other languages. We use the NLTK (Millstein 2020) and spaCy (Al Omran and Treude 2017) libraries with built-in tokenizers and POS tag-sets to tag the data. Both libraries yielded similar results within the datasets used. Next, as also done in step 1, the bag-of-words method was used to vectorise the corpus, and the frequency of each word is used here as feature weights to train and test the corpus using LR and Support Vector Machine (SVM) classification models. These two models were used with the two POS tagset libraries for the purpose of comparison. The average classification accuracy is calculated by counting how many actual labels are equal to the predicted label, divided by the total corpus, and then multiplying by 100 to achieve a percentage value (Pranckevičius and Marcinkevičius 2016). The accuracy of the LR and SVM with and without POS tagging for both datasets is shown in Tables 1 and 3 for both of the libraries NLTK and spaCy. We ran the models with test sentences for classification. These examples are shown in Tables 2 and 4. The average classification accuracy

Table 1. Accuracy of the classifiers for the 2015 Nepal earthquake dataset.

Classifier	Accuracy (%)
LR not tagged	76
SVM not tagged	76
LR tagged w. NLTK	77
LR tagged w. SpaCy	77
LR tagged w. SpaCy	77
SVM tagged w. SpaCy	77

Table 2. Classifier results with test sentences – 2015 Nepal earthquake dataset.

Test sentences	LR – NLTK	LR – spaCy	SVM – NLTK	SVM – spaCy
There was a very bad flood in Munich	Not relevant	Not relevant	Not relevant	Not relevant
There was a massive earthquakein LA	Relevant	Relevant	Relevant	Relevant
They film something about this earthquake in Munich	Not relevant	Not relevant	Not relevant	Not relevant

Table 3. Accuracy of the classifiers for the 2013 Queensland floods dataset.

Classifier	Accuracy (%)
LR not tagged	96
SVM not tagged	96
LR tagged w. NLTK	96
LR tagged w. SpaCy	96
LR tagged w. SpaCy	96
SVM tagged w. SpaCy	96

is higher for the 2013 Queensland floods dataset. This can be explained by the fewer falsely labelled Tweets found in the results in Figure 3 in comparison to the falsely labelled Tweets found in the results in Figure 4.

3.3. Localisation of crisis events

A Tweet, within the 280 characters allowed, can propagate information regarding a particular event, and are used to implicitly localise the events based on the locations specified in Tweets. For example, a person could be Tweeting from Atlanta, Georgia (Tweet location) about an event taking place in Ridgecrest, California: '*…7.1 magnitude earthquake strikes Ridgecrest area, according to USGS…*'. It is impossible to determine the absolute location of this event based on the Tweet location of this Tweeter. The earthquake event can be localised by extracting the mentioned location in the text with the help of Natural Language Processing (NLP) methods, as also done in many works in the literature. Figure 1 shows an example of localising this earthquake event based on the Tweet locations versus the mentioned locations. Figure 1(left) shows the mentioned location with the biggest hotspot in Ridgecrest, which was also the epicentre of the earthquake. Figure 1(right) shows the Tweet locations with the biggest hotspot in Los Angeles. A lot of the Tweeters seem to have tweeted about the event from the most populated city in California. For the rest of the analyses in the following section, both of these two types of locations will be used to localise the harvested Tweets. We have conducted experiments with a tool based on the open source library Mordecai (Halterman 2017). Mordecai combines two major functionalities: geoparsing (detection of geographic place names in full texts) and geocoding (assignation of geographic coordinates to a place name), with the help of a gazetteer. Based on the programming language Python we implemented an interface to apply Mordecai's approach on the Twitter live data stream. For this experiment we have used processing power of 4 cores with a Processor Base Frequency of 2.6 GHz, and 8 GB RAM on a CentOS 7 operating system. With this constellation, we used the

Table 4. Classifier results with test sentences – 2013 Queensland floods dataset.

Test sentence	LR – NLTK	LR – spaCy	SVM – NLTK	SVM – spaCy
There was a massive earthquake in LA	Not relevant	Not relevant	Not relevant	Not relevant
There was a very bad flood in Munich	Relevant	Relevant	Relevant	Relevant
Floods are in general a bad thing	Relevant	Relevant	Not relevant	Not relevant

Tweets	6019
Relevant	3248
Correctly labelled	2833
Falsely labelled	123
Percentage	87.22

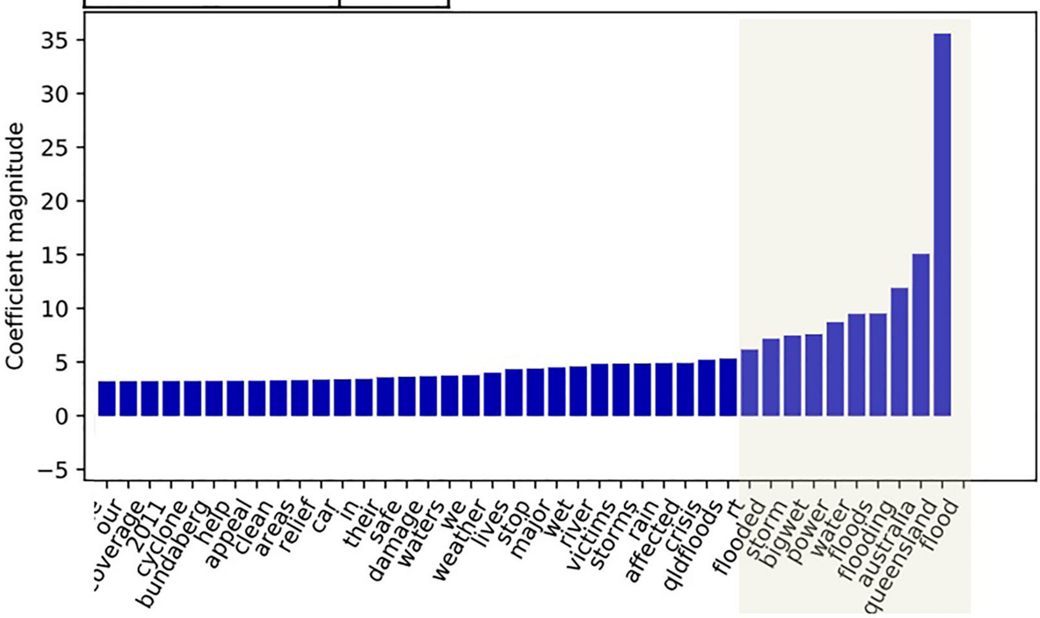

Figure 4. Results of the Logistic Regression for the 2013 Queensland floods dataset. The top 10 features are highlighted.

Twitter live streaming API v1 to extract Tweets for the specific signal words found in the sections above. These Tweets were extracted in a matter of seconds after being posted, and then geocoded for both Tweet location and mentioned location in a matter of 0.3 s per Tweet, making data available for analysis in near-real time.

4. Proof of concept – detecting flood events in May, 2022

In this section, approaches developed in Section 3 are combined for a proof of concept. We chose flood events that were prevalent during the time period 18–24, May of 2022 (beginning of this study) as our example crisis event. During this time the Twitter API v1 was tasked to harvest Tweets along with their various metadata from the live stream. For a comparative analysis, the live stream was filtered on the one hand for the singular keyword *flood*, and its derivatives *floods*, *flooding*, and

Table 5. The number of Tweets gathered with the different filter criteria.

Filter criteria	No. of tweets
Singular keywords *flood, floods, flooding*	30,592
Dictionary of signal words established through the LR	653,329

on the other hand for the top features found in the dictionary of signal words established for floods (except for keywords that refer to locations and confer to colloquial language): *flood*, *flooding*, *floods*, *water*, *power*, *storm*, *flooded*. The number of Tweets gathered through these two filter criteria are shown in Table 5. As evident from Table 5, the number of Tweets gathered with the dictionary of signal words are over 20 times more than the data gathered with the singular keywords for the considered time frame. The data gathered with the dictionary of signal words are then classified using LR and SVM with the NLTK and spaCy POS tagging. An excerpt of the results is shown in Table 6. The data gathered through both filter criteria were localised with their mentioned locations extracted from the Tweet text and Tweet locations, and a Kernel Density Estimation (KDE) was performed on these locations to identify the location hotspots. These location hotspots were visualised as heatmap visualisations. Figures 5–10 shows for various regions around the world, a comparative analysis of the Tweets harvested for the dictionary of signal words (blue colour hotspots), the keyword *flood* and its derivatives (green colour hotspots), and ground truth data (red colour symbols) for flood events for the time period 18–24 of May, 2022. As evident from Figures 5 and 6, similar hotspots can be found for both filter criteria, with more intense hotspots for the keywords flood and it's derivatives in Java, Indonesia in Figure 6. However, as evident from Figures 7–10, many more intense hotspots of clusters for the flood events can be seen for the the dictionary of signal words, in comparison to the hotspots derived based on the data for the keyword *flood* and its derivatives. Many of these hotspots are at, or around the verified locations of flood events. For the Shantou region and the Hainan island in China in Figures 9 and 10, no hotspots of Tweets could be found for the data extracted through the keyword flood and its derivatives, although groundtruth data showed flood events prevalent in these regions, as also observed with some of the hotspots derived based on the dictionary of signal words.

Table 6. An excerpt of the classifier results for 18–24 May, 2022 flood events dataset.

Test sentences from Twitter	LR – NLTK	LR – spaCy	SVM – NLTK	SVM – spaCy
Despite the magnitude of Australia's environmental decline, we still have the opportunity and ability to turn things around	Not relevant	Not relevant	Not relevant	Not relevant
Getting awfully close: flood 2 feet above normal high tide tonight in NYC	Relevant	Relevant	Relevant	Relevant
Assam dangerous flood situation land slides collapse destroy everything	Relevant	Relevant	Relevant	Relevant
Free State Premier Sisi Ntombela has expressed her gratitude for the love and solidarity shown by residents after the donation of various items to flood-stricken KwaZulu-Natal	Relevant	Relevant	Relevant	Relevant
This is my son and I the week we brought him home from the hospital. A home should be a sanctuary…a place to find peace and hope. If you experience a flood or fire, call us at Paul Davis of East Michigan. We'll help you Restore Hope quickly	Not relevant	Not relevant	Not relevant	Not relevant
Double tragedy induced by #ClimateCrisis affecting many children in Dasenech Woreda in #Ethiopia. Proud of the work UNICEF teams doing in Health @ Nutrition, Education	Not relevant	Not relevant	Not relevant	Not relevant
With more heavy rain on the way worth reflecting on the almost criminal negligence of the Palaszczuk government that it has done nothing about flood mitigation in 8 years despite the plans being there	Relevant	Relevant	Not relevant	Not relevant
Tree Fact: Trees can defend them self from attacking insects. They flood their leaves with chemicals called phenolics. As well as warning other trees about a future attack	Relevant	Not relevant	Not relevant	Not relevant

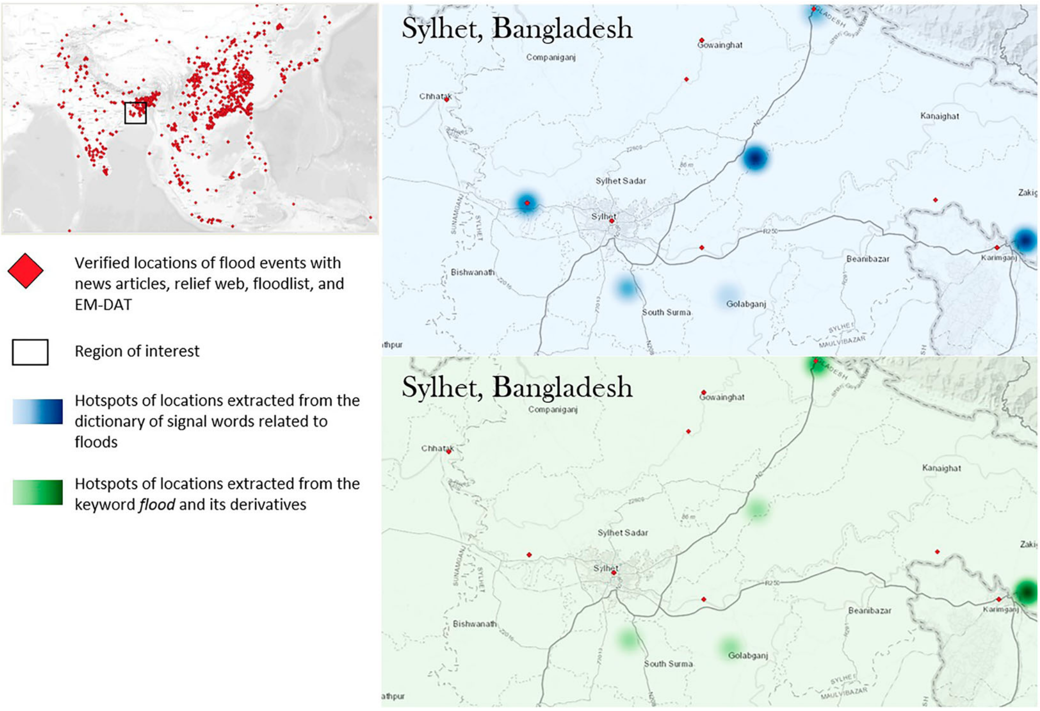

Figure 5. Comparative analysis of the data harvested from the Twitter live stream and groundtruth data, for flood events in the Sylhet region of Bangladesh during 18–24 May, 2022.

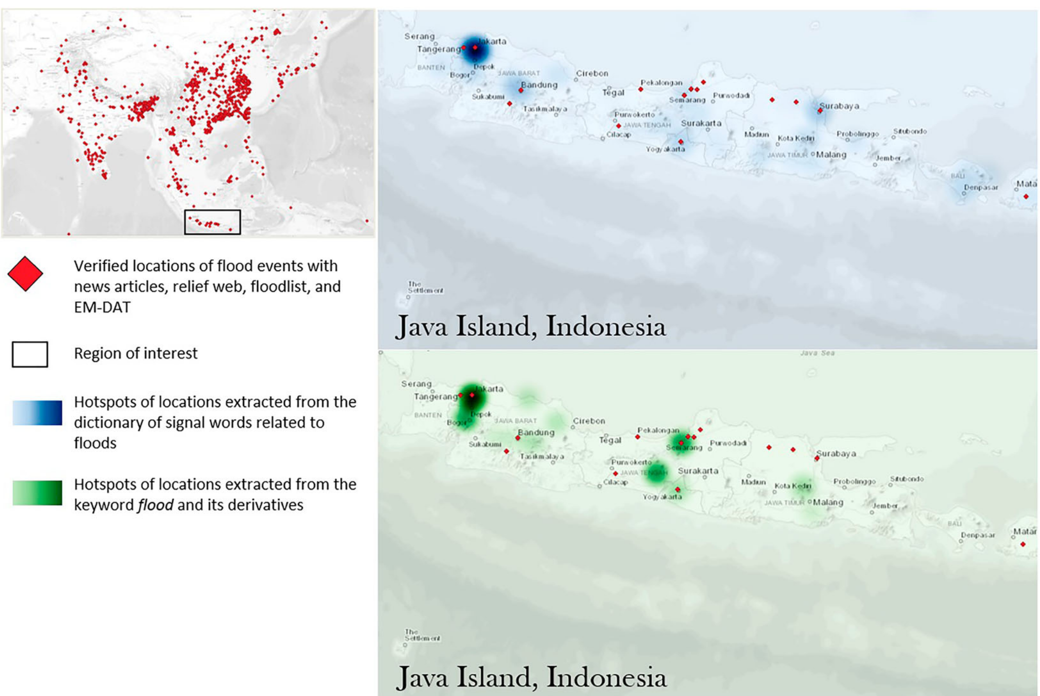

Figure 6. Comparative analysis of the data harvested from the Twitter live stream and groundtruth data, for flood events in Java, Indonesia during 18–24 May, 2022.

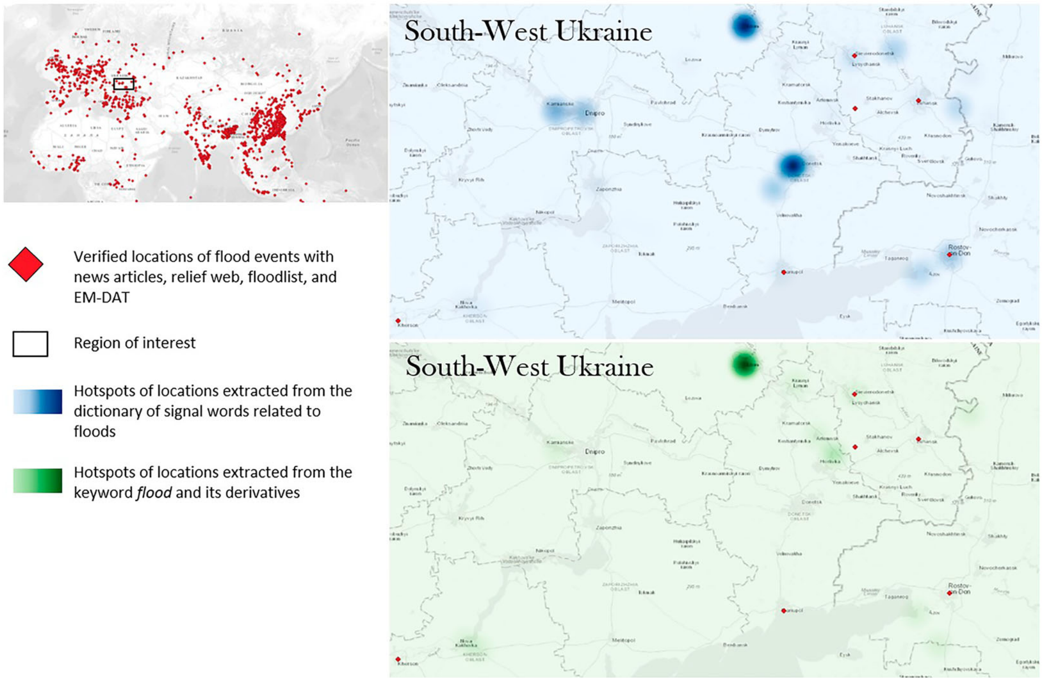

Figure 7. Comparative analysis of the data harvested from the Twitter live stream and groundtruth data, for flood events in the South-West region of Ukraine during 18–24 May, 2022.

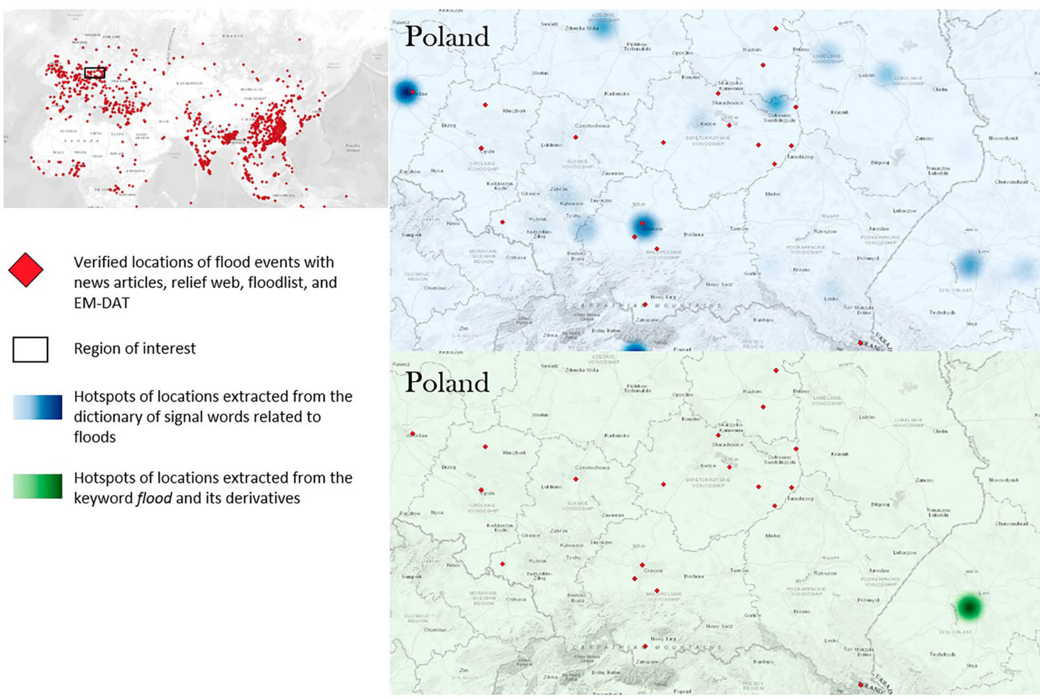

Figure 8. Comparative analysis of the data harvested from the Twitter live stream and groundtruth data, for flood events in Poland during 18–24 May, 2022.

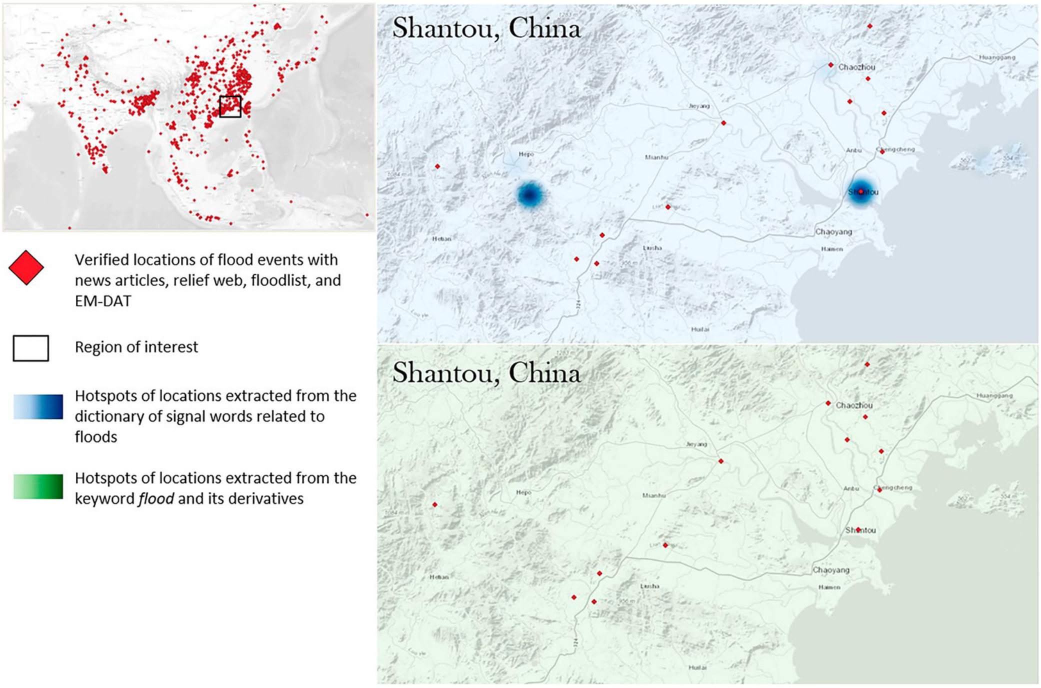

Figure 9. Comparative analysis of the data harvested from the Twitter live stream and groundtruth data, for flood events in Shantou, China during 18–24 May, 2022.

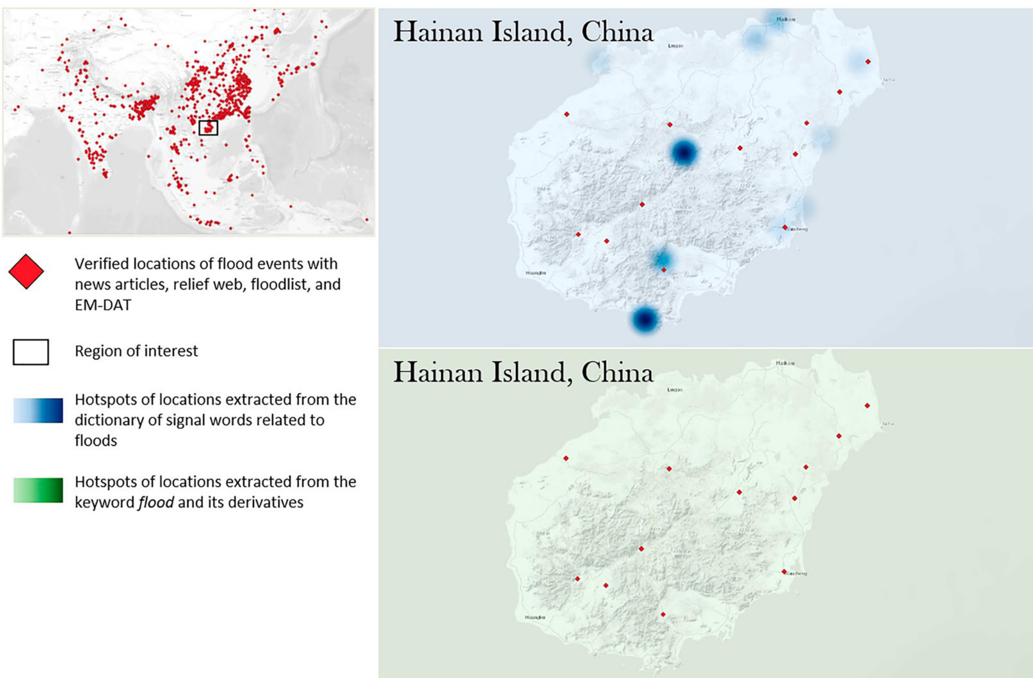

Figure 10. Comparative analysis of the data harvested from the Twitter live stream and groundtruth data, for flood events in the Hainan island of China during 18–24 May, 2022.

5. Discussion and future research perspectives

The proposed approach for extreme event detection using Twitter data advances the field of information extraction in several ways. The methodology introduces a novel approach by creating a dictionary of signal words using logistic regression (LR) to filter Twitter data related to specific events (e.g. floods or earthquakes). Unlike traditional methods that rely on single keywords, this approach incorporates a range of relevant keywords, improving the accuracy of data collection. It further mitigates the inclusion of irrelevant data that may be associated with event-related keywords in unrelated contexts. This significantly reduces noise in the collected dataset, improving the overall quality of information extraction. The methodology leverages part-of-speech (POS) tagging to semantically enrich the harvested data. This step helps disambiguate the meaning of keywords within the context of the Tweets, reducing false positives caused by keyword ambiguity. Semantic analysis enhances the precision of event-related information extraction.

In summary, the proposed methodology offers a more nuanced and accurate approach to information extraction for extreme event detection on Twitter. It addresses the challenges of noise reduction, context disambiguation, and location-based event identification. However, there are ongoing challenges related to language variability and the potential for false negatives, which require further research and refinement of the methodology. For instance, the construction of signal word dictionaries relies on labelled datasets and LR. While effective, the accuracy of these dictionaries is contingent on the quality and representativeness of the training data. Improvements could be made by expanding and diversifying the training datasets. Furthermore, Twitter data often includes non-standard language, slang, abbreviations, and misspellings, making it challenging to identify relevant keywords and phrases. Addressing this variability in language use is an ongoing challenge in Twitter-based event detection. While the signal word dictionary approach reduces false positives, it may still result in some false negatives, where relevant Tweets do not contain the predefined signal words. Enhancements to the methodology could involve exploring context-aware machine learning techniques.

To achieve a deeper understanding of crisis-related terms, such as ‘Earthquake game’, and differentiate it from a generic term such as ‘earthquake’, we need to go beyond traditional part-of-speech tagging and use more advanced techniques in Natural Language Processing and semantic analysis. For example, Named Entity Recognition (NER) is a technique that aims to identify and classify named entities, such as specific events, locations, or people, within a text. By applying NER, we can potentially recognise ‘Earthquake game’ as a named entity and distinguish it from a general ‘earthquake’. Lexical semantics involves understanding the meanings and relationships between words. Word Sense Disambiguation (WSD) is a technique used to determine the correct meaning of a word within a given context. By analysing the surrounding words and context, we can potentially disambiguate between the general meaning of ‘earthquake’and the specific sense of ‘Earthquake game’Furthermore, using external resources such as knowledge graphs, ontologies, or domain-specific databases can provide additional context and information about specific events. By linking the term ‘Earthquake game’to relevant entries or entities in these resources, we can enrich our understanding and differentiate it from a generic ‘earthquake’. Also, advanced NLP models, such as contextual embedding models (e.g. BERT, GPT), can capture the contextual information and meaning of words based on their surrounding context. By leveraging such models, we can potentially grasp the specific connotation and significance of ‘Earthquake game’ based on the overall context of the sentence or document. It is important to note that achieving a deeper understanding of specific terms or entities often requires a combination of these techniques, as well as domain-specific knowledge or data sources, depending on the context and the specific task at hand.

In addition to the signal words of crisis events, there are other unique signatures that determine the different crisis events, that may help in future research perspectives to further improve the information retrieval process to gather relevant data. Such signatures to consider are, the burstiness and the spatiotemporal propagation of Tweets (e.g. floods have a longer time duration and a larger geographic extent compared to most earthquakes that are characterised by a shorter time duration and a smaller geographic extent. These characteristics are also inherent in the data that represents these events), language usage (e.g. earthquakes *shaking*, floods *washing up*, storms and cyclones *carrying away*), individual parameters of the hazards (e.g. Richter scale measurements for earthquakes, speed and velocity of storms and hurricanes, inundation levels of floods). One main issue with location-based social media data such as Twitter is the uncertainty in such data. Future work should also develop methods to tackle the different types of uncertainties in these data as found in Senaratne et al. (2017), to further improve the retrieval of relevant information for crisis events.

6. Conclusions

In this paper we have presented an application that leverages established Machine Learning methods to detect and geolocate crisis events in near-real time on Twitter by (1) filtering the data stream based on a dictionary of signal words that were established through a Logistic Regression model, (2) further filtering the gathered data for semantically similar Tweets in the dataset by tagging the dataset for their POS, and (3) localising the filtered data by geoparsing the mentioned locations on the gathered Tweets and the geotagged Tweet locations. For step (1), a Logistic Regression was applied on a set of labelled data for two example crisis events – a flood in Queensland in 2013 and an earthquake in Nepal in 2015. For each of these events a dictionary of signal words was created. For step (2), the NLTK and spaCy libraries were used to tag the Tweets with their grammatical parts of speech based on the definition of the words and their context, and a Logistic Regression and a Support Vector Machine model were trained with the POS libraries to classify the data. These two models with the two POS libraries were compared for accuracy. With this two-step filtration process we were left with Tweets that were relevant to the events. For step (3), the Mordecai library is used to geoparse the mentioned locations in the gathered Tweets, and thereby localise them, in addition to the geotagged Tweet location. A kernal Density Estimation has been performed on these Twitter locations to estimate the hotspots of locations of the crisis events. This approach is demonstrated as a proof of concept for detecting location hotspots of flood events around the world during the time period May 18–24, 2022, using the dictionary of signal words *flood*, *flooding*, *floods*, *flooded*, *water*, *power*, *storm*, as well as singular keyword *flood* and its derivatives *floods*, and *flooding* to filter the Twitter live stream. The comparative analysis shows similar hotspots, however, the data gathered with the dictionary of signal words are 20 times more in quantity than the data gathered with the singular keywords, thereby producing more intense location hotspots. The inclusion of more relevant Tweets in the dataset with the proposed approach paves the way for more comprehensive and conclusive analyses in the future.

Acknowledgments

The corresponding author Hansi Senaratne was affiliated to the Earth Observation Center, The German Aerospace Center, during the initial submission of this manuscript. Her current affiliation at the time of major revision is TWT GmbH Science & Innovation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study has been conducted as part of the Megacities project of the VW Momentum Initiative, and the project Open Search @DLR phase II (internal DLR project).

References

Alam, Firoj, Shafiq Joty, and Muhammad Imran. 2018. "Domain Adaptation with Adversarial Training and Graph Embeddings." arXiv, arXiv:1805.05151, preprint.

Alam, Firoj, Hassan Sajjad, Muhammad Imran, and Ferda Ofli. 2021. "CrisisBench: Benchmarking Crisis-related Social Media Datasets for Humanitarian Information Processing." In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 15, 923–932, PKP Publishing Services Network.

Albeverio, Sergio, Volker Jentsch, and Holger Kantz. 2006. *Extreme Events in Nature and Society*. Springer Science & Business Media.

AlKhwiter, Wasan, and Nora Al-Twairesh. 2021. "Part-of-Speech Tagging for Arabic Tweets Using CRF and Bi-LSTM." *Computer Speech & Language* 65:101–138. https://doi.org/10.1016/j.csl.2020.101138.

Alomari, Ebtesam, Rashid Mehmood, and Iyad Katib. 2019. "Road Traffic Event Detection using Twitter Data, Machine Learning, and Apache Spark." In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 1888–1895. IEEE.

Al Omran, Fouad Nasser A., and Christoph Treude. 2017. "Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments." In *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*, 187–197. IEEE.

Behl, Shivam, Aman Rao, Sahil Aggarwal, Sakshi Chadha, and H. S. Pannu. 2021. "Twitter for Disaster Relief Through Sentiment Analysis for COVID-19 and Natural Hazard Crises." *International Journal of Disaster Risk Reduction* 55:102101. https://doi.org/10.1016/j.ijdrr.2021.102101.

Bosch, Harald, Dennis Thom, Florian Heimerl, Edwin Püttmann, Steffen Koch, Robert Krüger, Michael Wörner, and Thomas Ertl. 2013. "ScatterBlogs2: Real-time Monitoring of Microblog Messages Through User-guided Filtering." *IEEE Transactions on Visualization and Computer Graphics*19 (12): 2022–2031. https://doi.org/10.1109/TVCG.2013.186.

Bouadjenek, Mohamed Reda, and Scott Sanner. 2019. "Relevance-driven Clustering for Visual Information Retrieval on Twitter." In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 349–353. New York: ACM.

Chakrabarti, Deepayan, and Kunal Punera. 2011. "Event Summarization using Tweets." In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 5, 66–73. Barcelona, Catalonia, Spain: PKP Publishing Services Network.

Dabiri, Sina, and Kevin Heaslip. 2019. "Developing a Twitter-based Traffic Event Detection Model Using Deep Learning Architectures." *Expert Systems with Applications* 118:425–439. https://doi.org/10.1016/j.eswa.2018.10.017.

de Bruijn, Jens A., Hans de Moel, Brenden Jongman, Marleen C. de Ruiter, Jurjen Wagemaker, and Jeroen C. J. H. Aerts. 2019. "A Global Database of Historic and Real-time Flood Events Based on Social Media." *Scientific Data* 6 (1): 311. https://doi.org/10.1038/s41597-019-0326-9.

Gimpel, Kevin, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, et al. 2010. *Part-of-speech Tagging for Twitter: Annotation, Features, and Experiments*. Technical Report. Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

Gu, Yiming, Zhen (Sean) Qian, and Feng Chen. 2016. "From Twitter to Detector: Real-time Traffic Incident Detection Using Social Media Data." *Transportation Research Part C: Emerging Technologies*67:321–342. https://doi.org/10.1016/j.trc.2016.02.011.

HaCohen-Kerner, Yaakov, Daniel Miller, and Yair Yigal. 2020. "The Influence of Preprocessing on Text Classification Using a Bag-of-words Representation." *PLoS ONE* 15 (5): e0232525. https://doi.org/10.1371/journal.pone.0232525.

Hagras, Mohamed, Ghada Hassan, and Nadine Farag. 2017. "Towards Natural Disasters Detection from Twitter Using Topic Modelling." In *2017 European Conference on Electrical Engineering and Computer Science (EECS)*, 272–279. Bern, Switzerland: IEEE.

Halterman, Andrew. 2017. "Mordecai: Full Text Geoparsing and Event Geocoding." *Journal of Open Source Software* 2 (9): 91. https://doi.org/10.21105/joss.

Imran, Muhammad, Prasenjit Mitra, and Carlos Castillo. 2016. "Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages." arXiv, arXiv:1605.05894, preprint.

Interdonato, Roberto, Jean-Loup Guillaume, and Antoine Doucet. 2019. "A Lightweight and Multilingual Framework for Crisis Information Extraction From Twitter Data." *Social Network Analysis and Mining* 9 (1): 1–20. https://doi.org/10.1007/s13278-019-0608-4.

Kejriwal, Mayank, and Peilin Zhou. 2019. "SAVIZ: Interactive Exploration and Visualization of Situation Labeling Classifiers Over Crisis Social Media Data." In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 705–708. New York: Association for Computing Machinery.

Khare, Prashant, Miriam Fernandez, and Harith Alani. 2017. "Statistical Semantic Classification of Crisis Information.".

Liu, Huawen, Shichao Zhang, and Xindong Wu. 2014. "MLSLR: Multilabel Learning Via Sparse Logistic Regression." *Information Sciences* 281:310–320. https://doi.org/10.1016/j.ins.2014.05.013.

Marujo, Luis, Wang Ling, Isabel Trancoso, Chris Dyer, Alan W. Black, Anatole Gershman, David Martins de Matos, Joao P. Neto, and Jaime G. Carbonell. 2015. "Automatic Keyword Extraction on Twitter." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 637–643. Cambridge, Massachusetts, United States: MIT Press.

März, Luisa, Dietrich Trautmann, and Benjamin Roth. 2019. "Domain Adaptation for Part-of-Speech Tagging of Noisy User-generated Text." arXiv, arXiv:1905.08920, preprint.

Millstein, Frank. 2020. *Natural Language Processing with Python: Natural Language Processing Using NLTK*. Scotts Valley, California, USA: CreateSpace Independent Publishing Platform.

Musaev, Aibek, De Wang, and Calton Pu. 2014. "LITMUS: A Multi-service Composition System for Landslide Detection." *IEEE Transactions on Services Computing* 8 (5): 715–726. https://doi.org/10.1109/TSC.2014.2376558.

Nelakuditi, Kovida, Divya Sai Jitta, and Radhika Mamidi. 2018. "Part-of-Speech Tagging for Code Mixed English-Telugu Social Media Data." In *Computational Linguistics and Intelligent Text Processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, Revised Selected Papers, Part I 17*, 332–342. Springer.

Ngamassi, Louis, Abish Malik, Jiawei Zhang, and David S. Ebert. 2017. "Social Media Visual Analytic Toolkits for Disaster Management: A Review of the Literature." ISCRAM.

Pezanowski, Scott, Alan M. MacEachren, Alexander Savelyev, and Anthony C. Robinson. 2018. "SensePlace3: A Geovisual Framework to Analyze Place–Time–Attribute Information in Social Media." *Cartography and Geographic Information Science* 45 (5): 420–437. https://doi.org/10.1080/15230406.2017.1370391.

Prabhat, Anjuman, and Vikas Khullar. 2017. "Sentiment Classification on Big Data using Naïve Bayes and Logistic Regression." In *2017 International Conference on Computer Communication and Informatics (ICCCI)*, 1–5. IEEE.

Pranckevičius, Tomas, and Virginijus Marcinkevičius. 2016. "Application of Logistic Regression with Part-of-the-Speech Tagging for Multi-class Text Classification." In *2016 IEEE 4th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, 1–5. IEEE.

Romero, Simone, and Karin Becker. 2019. "A Framework for Event Classification in Tweets Based on Hybrid Semantic Enrichment." *Expert Systems with Applications* 118:522–538. https://doi.org/10.1016/j.eswa.2018.10.028.

Rossi, C., F. S. Acerbo, K. Ylinen, I. Juga, P. Nurmi, A. Bosca, F. Tarasconi, M. Cristoforetti, and A. Alikadic. 2018. "Early Detection and Information Extraction for Weather-induced Floods Using Social Media Streams." *International Journal of Disaster Risk Reduction* 30:145–157. Communicating High Impact Weather: Improving Warnings and Decision Making Processes. https://www.sciencedirect.com/science/article/pii/S2212420918302735. https://doi.org/10.1016/j.ijdrr.2018.03.002.

Senaratne, Hansi, Arne Bröring, Tobias Schreck, and Dominic Lehle. 2014. "Moving on Twitter: Using Episodic Hotspot and Drift Analysis to Detect and Characterise Spatial Trajectories." In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, 23–30. New York: ACM.

Senaratne, Hansi, Dominic Lehle, and Tobias Schreck. 2022. "Characteristics Analysis of Moving Conversations to Detect Events on Twitter." ." In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 476–483. New York: IEEE Press.

Senaratne, Hansi, Amin Mobasheri, Ahmed Loai Ali, Cristina Capineri, and Mordechai Haklay. 2017. "A Review of Volunteered Geographic Information Quality Assessment Methods." *International Journal of Geographical Information Science* 31 (1): 139–167. https://doi.org/10.1080/13658816.2016.1189556.

Senaratne, Hansi, Martin Mühlbauer, Ralph Kiefl, Andrea Cárdenas, Lallu Prathapan, Torsten Riedlinger, Carolin Biewer, and Hannes Taubenböck. 2023. "The Unseen – An Investigative Analysis of Thematic and Spatial Coverage of News on the Ongoing Refugee Crisis in West Africa." *ISPRS International Journal of Geo-Information* 12 (4): 175. https://doi.org/10.3390/ijgi12040175.

Shah, Kanish, Henil Patel, Devanshi Sanghvi, and Manan Shah. 2020. "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification." *Augmented Human Research* 5 (1): 1–16. https://doi.org/10.1007/s41133-019-0017-2.

Sloan, Luke, and Jeffrey Morgan. 2015. "Who Tweets with Their Location? Understanding the Relationship Between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter." *PLoS ONE* 10 (11): e0142209. https://doi.org/10.1371/journal.pone.0142209.

Taubenböck, Hannes, Joachim Post, Achim Roth, Kai Zosseder, Günter Strunz, and Stefan Dech. 2008. "A Conceptual Vulnerability and Risk Framework As Outline to Identify Capabilities of Remote Sensing." *Natural Hazards and Earth System Sciences* 8 (3): 409–420. https://doi.org/10.5194/nhess-8-409-2008.

Thomson, Robert, Christopher Dancy, Ayaz Hyder, and Halil Bisgin. 2018. *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Proceedings*, Washington, DC, USA, July 10–13, 2018. Vol. 10899. Springer.

UNDRR. 2017. "Sendai Framework for Disaster Risk Reduction 2015–2030." Accessed May 24, 2023. https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

Yen, Show-Jane, Yue-Shi Lee, Jia-Ching Ying, and Yu-Chieh Wu. 2011. "A Logistic Regression-based Smoothing Method for Chinese Text Categorization." *Expert Systems with Applications* 38 (9): 11581–11590. https://doi.org/10.1016/j.eswa.2011.03.036.

Yu, Yuhai, Hongfei Lin, Jiana Meng, and Zhehuan Zhao. 2016. "Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks." *Algorithms* 9 (2): 41. https://doi.org/10.3390/a9020041.

Zhang, Boliang, Ying Lin, Xiaoman Pan, Di Lu, Jonathan May, Kevin Knight, and Heng Ji. 2018. "ELISA-EDL: A Cross-lingual Entity Extraction, Linking and Localization System." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, New Orleans, Louisiana, June 41–45. Association for Computational Linguistics. https://aclanthology.org/N18-5009