International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.177

Unleashing the Power of Social Media: A Comprehensive Review of Event Detection Methods

R. Neelakandan,
Research Scholar
Department of Computing Technology
SRM Institute of Science and Technology
Tamil Nadu, India
neelakar@srmist.edu.in

Dr. V. V. Ramalingam
Associate Professor
Department of Computing Technology
SRM Institute of Science and Technology
Tamil Nadu, India
ramalinv@srmist.edu.in

Abstract—Social media platforms have become a vital platform for individuals to share their thoughts, opinions, and experiences, making them an important source of information. The emergence of social media has led to a growing interest in using social media data for event detection - the process of identifying and analyzing significant events or incidents in a given area. Social media has already been utilized successfully in monitoring and predicting epidemics caused by viruses like Zika, Dengue, MERS, and Ebola. More recently, it has been used to track and predict the spread of the COVID-19 pandemic. To evaluate the publicly available crowdsourcing data frequently required to stream, aggregate, filter, and process original data, information retrieval techniques are essential. However, academic literature reveals limited understanding of the crowd power that can be leveraged to foresee or respond to events. This study focuses on analyzing the literature on event detection using social media data, exploring the various techniques and methods utilized to identify and classify events, as well as the challenges and limitations of utilizing social media data for event detection.

Keywords—rowdsource, streaming data, big data, disaster management, unstructured data, social media, outbreaks

I. INTRODUCTION

Identifying events is an important activity in many areas, including crisis management, news analysis, and social science research. The growing use of social networking sites has made it a vital resource for event detection. Social networking sites have become principal outlets for individuals to express their viewpoints, sentiments, and encounters. This has resulted in an increasing interest in utilizing information for event identification. The platform is also a common pastime for billions of people worldwide, resulting in a massive amount of data production. For instance, Facebook (Menlo Park, CA) generates four million posts every minute. Minority opinions and private information are given a platform that are not covered by other sources. According to Bazarova et al. (2015), social media can create a feeling of isolation that enables individuals, especially young people and when discussing intimate topics, to express themselves without filters, more so than in conventional meet and talk. ShengLu, Mingqing Ma, and Qiangqiang (2022) suggest that social media provides an extra, unauthorized source of data that can be used to uncover health content that hasn't been shared with healthcare professionals or health organizations, as well as to express opinions on sensitive health-related topics. Middle East Respiratory Syndrome (MERS), Ebola, Zika, and Dengue disease monitoring efforts may be augmented and improved by using data from networking platform interactions, according to research. (Househ M 2015, Odlum M, and Yoon S 2015; McGough SF et al., 2016; Marques-Toledo CA et al., 2017; Shin SY, Seo DW, An J, et al., 2016).

Researchers can use data processing techniques to automatically collect and analyze social media activity for important insights. This research can be applied to disease surveillance, targeting vulnerable populations, and addressing unsolvable health concerns like adolescent drug and alcohol use. Charles-Smith et al. (2015) emphasized the effectiveness of these tools in monitoring disasters, disease outbreaks, and other events. In this article, we discuss the background, approach, and benefits of utilizing social media for event monitoring. In this article, we provide an overview of the background, methodology, and advantages of utilizing social media to monitor various events, including disasters, outbreaks, and other incidents.

II. BACKGROUND

2.1 *Social media essentials*

Kamruzzaman et al (2019) have suggested that the emergence of social networking sites has drastically transformed the methods through which individuals collaborate with each other and disseminate the details globally.

- These platforms offer exchange of contents in any format, allowing users to express themselves in various ways.
- Furthermore, social media platforms encourage cross-platform communication through social sharing, email, and RSS.
- Users can participate in the platform to varying degrees, posting content and leaving comments. The objective of social platforms is that it enables the rapid and broad distribution of knowledge.
- Additionally, social media platforms allow for one-to-one, one-to-many, and many-to-many communication and can be used synchronously or asynchronously, making communication possible across time.
- Social media websites offer a variety of features that allow users to interact with each other and the platform itself. These includes
- Social bookmarking, where users can tag and browse websites that others have bookmarked

DOI: rp-9788770040723.177

- Social news that enables commenting and voting on articles
- Social networking where users can join groups, add friends, and comment on profiles
- Social photo and video sharing where users can comment on submitted images and videos, and
- Wikis such as Wikipedia where users can create or modify entries.

Beyond these examples, any website that encourages user interaction falls under the definition of social media.

2.2 Social Media in Natural Disasters

Disasters occur frequently, which is difficult for government agencies and emergency services. Individuals and local communities, however, can contribute to the crowd's knowledge to forecast disasters and perform emergency response actions when they occur. Social media makes it possible for the public to participate in disaster relief efforts and allows information to be promptly transmitted from catastrophe-affected areas to those who need it most.

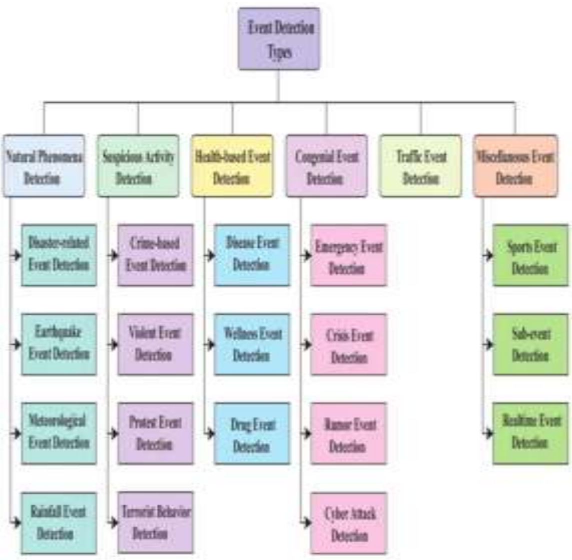

Fig. 1. Detecting a range of events through analysis of crowd sourced data.

Crowdsourcing offers a chance to facilitate the flow of knowledge, resources, money, and labor to anticipate and respond to a crisis in the era of the internet. Natural disasters can seriously harm people and property and are frequently unanticipated. Technology and societal advancements help us respond to disasters more effectively (Tuladhar et al., 2015). The response is faster and more accurate when a modern, powerful, and complicated decision-making system is used. Disasters are happening more frequently, which is making it difficult for government agencies and emergency organizations to respond. Relief efforts may be delayed as a result of unpredictable events, frightened reactions, and incomplete information about affected areas (Kankanamge et al. 2019).

To ensure the successful resolution of a crisis, it is important to first determine the effectiveness of today's disaster relief and rescue services. It's important to spread awareness and understanding about various systems, such as those related to geography, search and rescue procedures, and adapting to extreme circumstances. Sharing knowledge about these systems can play a crucial role in disaster preparedness. Studies in information systems have demonstrated that advancements in technology, particularly the use of internet-based platforms that encourage public participation, can help individuals and communities become more resilient in the face of disasters (Nan and Lu 2014).

One approach to improving disaster response systems is to use crowdsourcing to understand the connection between environmental disasters and human actions. By leveraging crowdsourcing, we can gather more data, scale potential solutions, and respond more rapidly to emergencies. Despite the recognition by scholars from various fields on the significance of mass media, community participation, and civic activism in disaster scenarios, these phenomena are not often considered as forms of crowdsourcing (Elbanna et al. 2019, Guo et al. 2021).

Poblet et al (2014) identified two technological approaches for managing disasters in literature: (i) sourcing and (ii) collaboration. The sourcing approach focuses on gathering and analyzing text from social networking sites to produce the timely alerts. Conversely, the collaboration approach aims to improve human and disaster management technology interaction to enhance overall communication.

Applying crowdsourcing to disaster management makes it easier to tap into collective wisdom and mobilize social forces to overcome a variety of challenges, including delayed situational awareness, a shortage of personnel to process and analyze data, and a lack of personnel to conduct rescue operations. Crowdsourcing can help make disaster aid more effective and affordable in this way (Callaghan 2016).

Crowdsourcing involves people from all around the world in disaster management and emergency response to natural disasters. Researchers have examined several types of natural disasters, which include floods (Schumann et al., 2010; Zeman, 2006; Kuhlicke et al., 2015; EMS, 2018; Trambauer et al., 2016), earthquakes (DE A, 2014; Wang J, 2019; TL and J, 2017), and wildfires (Thom, 2016; Quapp et al., 2017).

2.3 Social media in outbreak detection

Before the era of vaccinations, infectious diseases typically followed seasonal patterns and progressed through various phases. It is critical to identify a seasonal outbreak promptly to respond more efficiently and effectively to medical professionals. Failure to do so can lead to fatal diseases like influenza or influenza-like illnesses (ILI) in a region where an epidemic has broken out epidemiologically (Alessa and M. Faezipour, (2018), E. Lau and P. Wark, (2019)).

To lower the risk of illnesses and stop epidemics, early discovery is crucial, but current outbreak detection methods have a lengthy delay between diagnosis and public action (Tao et al, 2020). Additionally, presumptive techniques such as QMRA might not be effective in rapidly identifying

DOI: rp-9788770040723.177

outbreaks as they rely on structured data from pre-planned field trials (Tao et al, 2020).

Online social media (OSM) can be utilized to detect and contain epidemic outbreaks in health surveillance systems (S. Amin, 2021), forecast public illness outbreaks, and spread accurate information (C. Feng et al, 2020, K. Yu, 2021). Rapid disease diagnosis can lessen the impact of seasonal epidemics on population safety (Maharana, A. et al, 2019). Combining machine learning techniques with crowdsourcing provides novel ways to conduct risk assessment and communication for food safety. Social platform monitoring is also an effective tool to disseminate information for disease monitoring and engage with the public to prevent epidemic breakouts (X. Zhang, et al, 2021).

2.4 Understanding Events in Social Media

According to Zhao, Qiankun, and PrasenjitMitra (2007), an event refers to a collection of social player relations on a specific topic within a certain timeframe. According to Hearst and Marti (2009), incidents in social media analysis can be characterized as observable occurrences that involve multiple entities. The process of monitoring the incidents utilizes cognitive processes to accurately detect incidents by analyzing data. Events are noteworthy occurrences at specific times and locations that can be regarded as a single chapter of a larger narrative (Sayyadi, Hassan, Matthew Hurst, and Alexey Maykov, 2009; Panagiotou N et al, 2016). In social media analysis, it's necessary to examine the cumulative trend changes in the data stream to comprehend events. According to KarthikSubbianet all (2012), streaming data analysis involves recognizing significant incidents as they happen. The analysis of social networking sites data can be beneficial for detecting various incidents, whether they are limited to a particular location or widespread, such as traffic congestion on a street or floods in Florida, as explained by McCreadie et al (2013).

The platform has been used for multiple purposes,such as identifying activism, supporting emergency responders, tracking disease spread, analyzing user functions and behaviors, measuring media coverage, providing tourist information, detecting traffic, exposing people to diverse viewpoints, and facilitating political participation. It has also been used to study migration patterns, food preferences, and happiness levels in real-time. Some people still significantly rely on networking platforms for governmental updates and data, despite the fact that the majority of its content has nothing to do with news or public problems (Gil de Ziga et al., 2012). For these research aims, studies have used several billions of tweets or messages (Xu et al., 2014; Avvenuti et al., 2016, 2018; Lampos and Cristianini, 2012; Martinez Teutle, 2010; Cresci et al., 2020; Mazza et al., 2019; Prieto Curiel et al., 2019; Silvestri et al., 2015; D'Andrea et al., 2015; Himelboim et al., 2013; Ausserhofer and Maire, 2013; Coletto, 2017; Amato, 2017; Dodds et al., 2011).

2.5 The challenges of crowdsourcing in crisis / limitations of social media

Crowdsourcing in crisis situations refers to the process of collecting and utilizing information from a large group of individuals, often via social media platforms, to aid in crisis management and response efforts. However, there are several challenges and limitations associated with crowdsourcing in crisis situations.

Due to the use of terms with numerous meanings in social platform streams, ambiguity is a significant issue with human feeds or text (Carbezudo MAS and Pardo TAS, 2017, Gutierrez-Vazquez et al 2016) Although humans can use context and common sense to disambiguate text, computer programs struggle to do so (Alkhatlan A, Kalita J and Alhaddad A, 2018). Moreover, the presence of informal language, shortened words, and initialisms in social networking sites makes it even more challenging to understand their meaning..Unfortunately, most of the techniques used to predict or identify the current events are not equipped to handle this issue. Making it difficult to accurately analyze social platform content.(Katragadda S, Benton R and Raghavan V 2017).

Social networking sites data gathering and monitoring present many opportunities for multiple entities. However, there are several challenges associated with this field,

1. Data cleansing - One of the key challenges in this field is the need to effectively preprocess the textual data, particularly the high volume of on the fly data, in order to ensure its quality and reliability.
2. Scraping - There is limited academic access of content from the platforms, However, news services such as Thomson Reuters and Bloomberg charge a fee for access to their data, while Twitter recently introduced the Twitter Data Grants program, which provides academics with access to Twitter's historical data and public tweets to enable them to obtain fresh insights from its vast data collection (Twitter has over 500 million tweets each day).
3. Verification and reliability: One of the main challenges of crowdsourcing in crisis situations is verifying the accuracy and reliability of the information that is collected. Social media platforms can be a source of misinformation and rumors, which can lead to confusion and impede crisis response efforts.
4. Privacy concerns: Crowdsourcing in crisis situations can raise privacy concerns, as personal information can be shared and used without consent.
5. Data overload: Social media platforms can generate a large volume of data, which can make it difficult for organizations to effectively process and utilize the information (Agarwal, Liu, Tang, & Yu, 2008).
6. Bias: Crowdsourcing can also be affected by bias, as certain groups of people might be more likely to share information than others, leading to a lack of representation and a narrow perspective of the crisis.
7. Limited data quality: Social media data, although useful, can be unstructured, incomplete and unreliable. This can make it difficult to extract useful information and make accurate predictions.

DOI: rp-9788770040723.177

8. Lack of context: Social media data can lack context and background information, making it difficult to understand the information and to make sense of the events.

9. Streaming data - The real-time handling of crises can be challenging due to the vast amount and speed of social media data. Identifying and categorizing events in real-time can be difficult, which can hinder efforts to respond to crises.

10. Relevance in Context - Social media information is contextual and varies from user to user. For example, some Twitter users may appreciate updates on daily activities while others find it bothersome.

11. Use of Slang Terms and Intentional Spelling Errors - Informal language is prevalent on social media, with people using colloquialisms and intentionally misspelling words for emphasis, such as "so coooooool..." This form of language conveys sentiments, phrases, and feelings and adds a new dimension to written texts. These aspects of communication should not be disregarded as mere typos, as they can provide valuable information. Online dictionaries like UrbanDictionary can be used to decode the meaning of slang, acronyms, and other informal language used by people. However, the abundance of off-topic discussions and noise can distort the analysis.

12. Recency of data - Platforms allow for fast conversations and responses, and people's interests tend to be diverse and changeable over time, as observed in the blogosphere (Hayes &Avesani, 2007). This results in a shift in both people's interests and the social media environment. Additionally, the freshness of the information on social media is constantly evolving.

In conclusion, crowdsourcing in crisis situations can provide valuable information for crisis management and response efforts. However, there are several challenges and limitations associated with crowdsourcing in crisis situations, such as verification and reliability, privacy concerns, data overload, bias, limited data quality, lack of context, and inability to handle crisis in real-time. Thus, organizations should be aware of these challenges and limitations and develop strategies to address them.

III. Methods and Techniques

Two different approaches used to identify incidents from social networking sites data.

The first one is, based on the rules. This approach relies on predefined patterns, policies and guidelines, such as keyword-based and location-based techniques. Keyword-based methods use specific keywords or hashtags to detect events, while location-based methods rely on geographic information to identify events in a certain area.

On the other hand, machine learning-based methods involve the use of algorithms to learn patterns and classify events. These methods can be supervised or unsupervised, where supervised methods use labeled data to train a model, while unsupervised methods use unlabeled data to detect patterns.

IV. Big Data framework for social media analytics

4.1 *Big Data*

A big data framework is a collection of technologies and tools used to handle, examine, and visualize the terabytes of data. To facilitate social media analytics, such a framework can be employed to acquire, retain, process, and analyze voluminous data generated by platforms like Twitter, Facebook, and Instagram.

A data management system typically consists of several essential components, including data collection, storage, processing, analysis, visualization, and governance.

1. The data collection component involves acquiring data from different sources, such as social media platforms, APIs, and web scraping, in various formats, such as text, images, and videos.

2. The data storage component includes preserving the collected data using various storage technologies, like Hadoop Distributed File System (HDFS), NoSQL databases, and cloud storage.

3. The data processing component involves cleaning and processing the collected data using various technologies, such as Apache Spark and Apache Storm, capable of handling large data volumes in real-time.

4. The data analysis component includes analyzing the processed data with the help of big data technologies, which provide machine learning and data mining functionality.

5. The data visualization component involves presenting the analyzed data through various technologie to create interactive visualizations.

6. Lastly, the data governance, security, and compliance component involves ensuring compliance with legal and regulatory requirements and ensuring data security.

The big data framework for social media analytics should also include an orchestration and scheduling layer to schedule and monitor the data pipeline, such as Apache Airflow, Oozie, or Apache NiFi.

Newman et al. (2016) explains that the term "Big Data" implies a vast amount of data, denoting its magnitude or quantity. However, contemporary literature highlights a set of characteristics that distinguish big data and underscore their value. These characteristics, known as the three Vs of big data - volume, variety, and veracity - are widely accepted among scholars. The concept of the three Vs was first introduced by Laney and has been elaborated upon by subsequent authors (Uddin et al., 2014). Volume refers to the size of the data generated, which can range from terabytes to exabytes and even beyond. Big data is continuously produced from various sources such as social media, cloud-based services (such as Amazon), business-related data, and Internet of Things (IoT)-related data (Khan et al., 2014; Storey and Song, 2017). Radicati and Hoang (2011) predicted that the global number of email accounts generated would increase to over 4.3 billion by late 2016, up from 3.3 billion in 2012. According to a survey conducted by IBM in mid-2012, data amounts exceeding one terabyte

DOI: rp-9788770040723.177

are considered to be big data (Schroeck et al., 2012). However, this threshold is relative, as the quantification of data volume depends on other factors, such as time and data type. With advances in storage capabilities, larger datasets can be managed over time. According to Chen and Zhang (2014), the concept of Big Data is not only about the volume of data, but also encompasses various dimensions, starting with the letter V and ending with the "Vs" of Big Data. It is evident that a terabyte of textual data is not always equivalent to a terabyte of video data, highlighting the importance of the data-type component. The data sources and types are diverse, and different formats are used to describe data coming from various sources. For instance, structured data can be distinguished from frequently administered Structured Query Language (SQL), which is a programming language designed for maintaining and accessing data within Relational Database Management Systems (RDBMS) (Hashem et al. 2015). Structured data is simple to enter, search for, and save, while unstructured data, such as multimedia (videos, photos, and audios), which has a fixed format, is the primary format that defines Big Data (Gandomi and Haider 2015). Managing unstructured data poses a significant challenge for data scientists. Semi-structured data, such as Extensible Markup Language (XML) and JavaScript Object Notation (JSON) data, are also generated. The incoming and outgoing data are characterized by their velocity, which is assessed in terms of batch size, near real time, and real-time to achieve streaming (Chen and Zhang 2014).

Yaqoob and colleagues (2016) argue that the proliferation of internet-connected mobile devices and sensor-equipped devices has a significant impact on data velocity. Fan and Bifet (2013) suggest that delivering timely responses and updates is critical for evaluating an application's effectiveness. In addition, interpreting and organizing streaming data is a challenge that requires relevant technologies and techniques (Orgaz et al., 2016). Big Data is characterized by additional features, such as veracity, which refers to the reliability and messiness of data, according to IBM and Microsoft (Gandomi and Haider, 2015). Storey and Song (2017) maintain that veracity poses problems with data quality, such as accuracy, completeness, timeliness, consistency, and accessibility. Value, the fourth V of Big Data, was introduced by McKinsey and Oracle to highlight the value of hidden insights in Big Data (Chen and Zhang, 2014). Wang et al. (2017) expand on this concept by proposing five V dimensions for Big Data that consider both value and veracity. Uddin et al. (2014), on the other hand, identify seven Vs of Big Data, taking into account both the validity and volatility dimensions. They define volatility as the retention policy for structured data typically used in organizations, while validity refers to data accuracy and correctness regarding its intended purpose. The many Big Data dimensions and their associated names are shown in Figure 1.

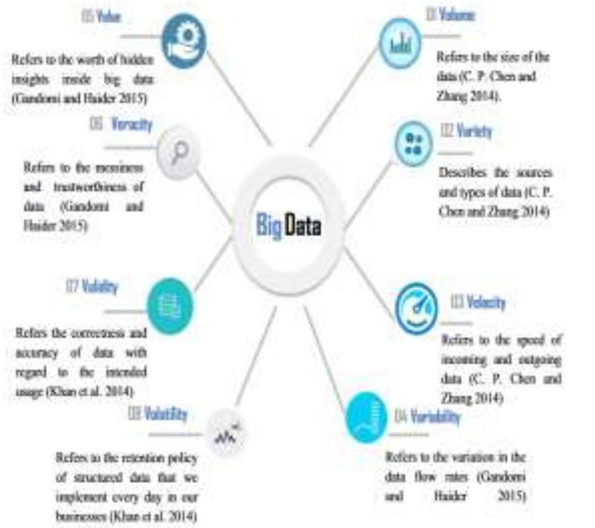

Fig. 2: Big Data dimensions

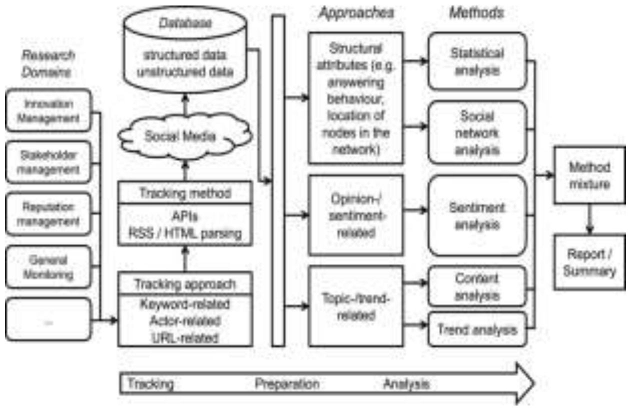

Fig. 3. The Social Media Analytics Framework (Stieglitz et al., 2014, Stieglitz and Dang-Xuan, 2013).

To effectively analyze social media data, organizations can employ a big data framework that involves the use of various technologies and tools. This framework enables the collection, storage, processing, evaluation, and visualization from social media platforms. With this approach, businesses can gain insightful information and use social platform data to guide decisions.

4.2 Natural language processing

Utilizing natural language processing (NLP) is crucial in enhancing event detection through evaluating the online networking platforms data. By employing NLP techniques to extract and scrutinize text data from social media platforms like Twitter, Facebook, and Instagram, events can be identified and classified effectively. One important NLP technique that can be used for event detection is text classification. This technique can be used to classify social media posts into different categories, such as events or non-events. Text classification algorithms can be trained on labeled data to learn the characteristics of events and non-events, and then applied to new data to classify it. For instance, a text classification algorithm can be trained to classify posts that mention a keyword or hashtag as events.

DOI: rp-9788770040723.177

Another NLP technique that can be used for event detection is named entity recognition (NER). This technique can be used to identify and extract entities. NER can be used to extract relevant information from social media posts, such as the location of an event or the people involved in an incident. In addition to text classification and NER, Sentiment analysis is also an important NLP technique that can be used for event detection.

Depending on the polarity of posts, the NLP model can be used to categorize the sentiment.Sentiment analysis is used to determine how the public feels about a specific occurrence or event, which can be useful for crisis management and news analysis. Finally, NLP techniques, such as geotagging and geocoding, can extract geographic information from text data to facilitate geolocation, enabling the identification of the location of an event or incident. This information can be used to identify events in specific areas and to understand the spatial distribution of events.

The authors (McCreadie R et al. 2013) investigated scalable distributed event detection using Twitter feeds and lexical key partitioning technique. The research suggested a framework for automatic distributed real-time event detection that employed Storm topology and utilized Locality Sensitive Hashing (LSH) to handle noisy, temporal, and slang-filled Twitter streams. Authors (Kaleel SB, Almeshary M, Abhari A, 2013) proposed using LSH, which was used for the incident identification. The proposed model compared with leading social platforms Facebook, Twitter.

In 2015, Musaev et al. proposed a system called LITMUS to identify "landslides" related information from social media using a keyword-based approach. The system extracted categorization features from Wikipedia and utilized an enhanced Explicit Semantic Analysis (ESA) algorithm to categorize the data into relevant and irrelevant categories. The location was approximated using semantic clustering that relied on measuring the semantic distance. However, only data that had been tagged with location information were taken into consideration, and the analysis did not take into account the handling of SAB words in the semantic-based approach. The system was later used to identify crossover activities between two social media streams.

TABLE 1 :AN OVERVIEW - ADVANTAGES AND LIMITATIONS OF MODELS IN EVENT DETECTION.

Category	Unsupervised ML	Semi-supervised ML	Supervised ML	Semantic approach
Benefits	Approach can detect events without considering their nature, and can handle a large volume of data.	The method is especially beneficial when it's challenging to identify important characteristics in the data. Additionally, even a small dataset can produce a substantial increase in accuracy.	The outcomes are very precise and reliable.	More accurate and precise results due to its ability to consider the context and meaning of the text being analyzed.
Shortcoming	Working with data streams that have many attributes or features, using this technique may pose a challenge. This technique may not account for the spatial relationships within the data, which can be a limitation in certain applications.	Iteration results are not stable and accuracy is low.	Challenges, including the need for significant time and resources, dealing with large and complex datasets, addressing the issue of concept drift, and the requirement for expertise to properly label input and output variables.	A crucial time and resource commitment, in addition to a detailed understanding of algorithms, are needed to develop semantic-based techniques.

(Romero and Becker, 2019) and (Sun et al., 2021) proposed different approaches for event detection in social media using natural language processing techniques. Romero and Becker used a hybrid approach that combined with multiple traditional approaches NER, TF-IDF and others to classify events in tweets. On the other hand, Sun et al. proposed scoring and word embedding for the same which employed Word2vec with enhanced embeddings. Both studies employed pre-processing steps. However, neither of the approaches addressed the ambiguity related to SAB terminologies in social platforms. Overall, natural language processing techniques have the potential to improve event detection in social media by extracting and analyzing text data using methods such as text classification, named entity recognition, sentiment analysis, and geolocation.

V. CONCLUSION

In conclusion, studies on social networking sites have shown that it might be a useful data resource for incident detection. Various natural language processing techniques have been used to extract and analyze text data from social media platforms, including text classification, named entity recognition, sentiment analysis, and geolocation. While these techniques have shown promising results, there are still challenges to overcome, such as handling ambiguity and variability in language, improving the accuracy and scalability of models, and addressing ethical and privacy concerns. Overall, more investigation into this topic has the potential to improve our capacity for real-time recognition and response to events, with significant implications for sociology, health services, and emergency preparedness.

VI. FUTURE WORK

There are several potential areas of focus for future research that explore the use of online networking platforms for event detection. These include:

1. Ambiguity and variability in language: Social media language is often informal and ambiguous, making it difficult to accurately detect events. Existing natural language processing techniques have been successful in addressing some of these challenges, but more

sophisticated approaches could be explored to improve accuracy.

2. Scalability and efficiency: Social media platforms generate vast amounts of data in real-time, making it difficult to process and analyze the data in a timely manner. Future work could focus on developing scalable and efficient algorithms for event detection.

3. Ethical and privacy concerns: Important ethical and privacy issues are brought up by the use of social media data to identify events, notably in relation to data privacy, informed permission, and potential bias. Future research might concentrate on creating moral standards and best practices for gathering and utilizing social networking websites.

4. Incorporating multimedia data: Social media platforms also contain a wealth of multimedia data such as images and videos, which could provide valuable information for event detection. Future research could explore how to effectively integrate multimedia data into event detection models.

5. Adapting to new social media platforms and evolving language: As new social media platforms emerge and language use evolves over time, it will be important for event detection models to adapt to these changes. Future research could explore how to build models that are flexible and adaptable to new platforms and changing language use.

6. Application to specific domains: The method of event detection from social networks has multiple applications across diverse industries including crisis management, public health, and social sciences. Future work could focus on applying event detection models to specific domains and assessing their effectiveness in real-world settings.

7. Real-time event detection: Social networking platforms produce vast amounts of data in real-time and event detection algorithms need to be able to operate in real-time as well. Future research could explore how to improve the speed and efficiency of event detection algorithms to enable real-time monitoring of events on social media.

8. Handling bias and overfitting: One potential challenge with event detection algorithms is that they may be biased towards certain types of events or may overfit to the data used for training. Future research could explore how to minimize bias and overfitting in event detection models, for example, by using more diverse training data or by incorporating regularization techniques.

Overall, event detection through the social platform is a crucial and also complex task with the potential to improve our understanding of various events and incidents. The forthcoming investigations in this field can concentrate on several significant challenges including the development of methods that can be applied to stream data, improving the ability to handle noise and ambiguity in social media language, integrating other sources of information such as multimedia data, addressing issues related to bias and overfitting, and finding ways to ensure privacy and security concerns are addressed. By addressing these challenges, we can make significant strides towards developing more accurate and effective event detection algorithms that can be applied in diverse domains, from crisis management to marketing and beyond.

References

1. G. Tuladhar, R. Yatabe, R.K. Dahal,,and N.P. Bhandary,"Disaster risk reduction knowledge of local people in Nepal,"GeoenvironDisast.,vol. 2, no. 1, 2015,https://doi.org/10.1186/s40677-014-0011-4
2. Nan, Ning and Yong Lu. "Harnessing the Power of Self-Organization in an Online Community During Organizational Crisis." MIS Q., vol. 38,pp. 1135-1157, 2014.
3. N. Kankanamge, T. Yigitcanlar, A. Goonetilleke, and M. Kamruzzaman, "Can Volunteer Crowdsourcing Reduce Disaster Risk? A Systematic Review of the Literature," International Journal of Disaster Risk Reduction, vol. 35, p. 101097, 2019.
4. A. Elbanna, D. Bunker, L Levine, and A. Sleigh, "Emergency Management in the Changing World of Social Media: Framing the Research Agenda with the Stakeholders through Engaged Scholarship," International Journal of Information Management vol. 47, no. 1, pp. 112–120, 2019.
5. M. Poblet, E. García-Cuesta, and P. Casanovas, "Crowdsourcing Tools for Disaster Management: A Review of Platforms and Methods," In: Casanovas, P., Pagallo, U., Palmirani, M., Sartor, G. (eds) AI Approaches to the Complexity of Legal Systems. AICOL 2013. Lecture Notes in Computer Science(), vol 8929, 2014, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-45960-7_19
6. C. W. Callaghan, "Disaster Management, Crowdsourced R&D and Probabilistic Innovation Theory: Toward Real Time Disaster Response Capability," International Journal of Disaster Risk Reduction,vol. 17, pp. 238–250, 2016.
7. C. Begg,G. Walker,and C.Kuhlicke, "Localism and flood risk management in England: the creation of new inequalities?" Environ Plann C Gov Policy, vol. 33, no. 4, pp. 685–702, 2015, https://doi.org/10.1068/c12216
8. E.M.SCopernicus, European Union, Earth Observation Programme – Copernicus, Annual Fire Reports effisjrceceuropaeu/reports-and-publications/annual-fire-reports/? Accessed 15 May 2018
9. D. Eilander,P.Trambauer,J.Wagemakerand A. van Loenen,"Harvesting social Media for Generation of near real-time flood maps,"Proc Eng., vol. 154, pp. 176–183,2016, https://doi.org/10.1016/j.proeng.2016.07.441
10. B. Hossain,"Role of organizations in preparedness and emergency response to flood disaster in Bangladesh,"GeoenvironDisast., vol. 7, no. 1, p. 33, 2020, https://doi.org/10.1186/s40677-020-00167-7
11. B. Merz, J. Hall, M. Disse, and A. Schumann,"Fluvial flood risk management in a changing world," Nat Hazards Earth Syst Sci., vol. 10, no. 3, pp. 509–527, 2010, https://doi.org/10.5194/nhess-10-509-2010
12. J. Schanze, "Flood risk management – a basic framework," In: Schanze J, Zeman E, Marsalek J (eds) Flood Risk Management: Hazards, Vulnerability and Mitigation Measures. Springer Netherlands, Dordrecht, pp 1–20, 2000.
13. D.E. Alexander,"Social Media in Disaster Risk Reduction and Crisis Management,"SciEng Ethics, vol. 20, no. 3, pp. 717–733, 2014, https://doi.org/10.1007/s11948-013-9502-z
14. X. Han, and J. Wang, "Earthquake information extraction and comparison from different sources based on web text," ISPRS Int J Geo Inf., vol. 8, no. 6, 2019, https://doi.org/10.3390/ijgi8060252
15. J. Xu, and T.L. Nyerges,"A framework for user-generated geographic content acquisition in an age of crowdsourcing,"CartogrGeogrInf Sci.,vol. 44, no. 2, pp. 98–112, 2017, https://doi.org/10.1080/15230406.2016.1163644
16. S. Daly, and J.A. Thom,"Mining and classifying image posts on social media to analyze fires," In: Antunes P, BanulsSilvera VA, Porto de Albuquerque J, Moore KA, Tapia AH (eds) Information Systems for Crisis Response and Management, ISCRAM, 2016.

17. M. Nayebi, R. Quapp, G. Ruhe, M. Marbouti, and F. Maurer,"Crowdsourced exploration of mobile app features: a case study of the fort McMurray wildfire," Institute of Electrical and Electronics Engineers Inc, Bueons Aires, pp 57–66, 2017.

18. D.Tao, D.Zhang,R. Hu, E. Rundensteiner, andH. Feng, "Crowdsourcing and machine learning approaches for extracting entities indicating potential foodborne outbreaks from social media," Scientific Reports, vol. 11, no. 1, pp. 1-12, 2021.

19. S.F. McGough, J.S. Brownstein, J.B. Hawkins, M. Santillana,and "Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data,"PLoS Neglect Trop Dis.,p. 11:e0005295, 2017, [PMC free article] [PubMed] [Google Scholar]

20. C.A. Marques-Toledo, C.M. Degener, and L.Vinhal et al., "Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level,"PLoS Neglect Trop Dis.,vol. 11, p. e0005729, 2017, [PMC free article] [PubMed] [Google Scholar]

21. S.Y. Shin, D.W. Seo, and J. An, et al., "High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea,"Sci Rep.,p. 6:32920, 2016, [PMC free article] [PubMed] [Google Scholar]

22. M. Househ, "Communicating Ebola through social media and electronic news media outlets: a cross-sectional study," Health Informat J.,vol. 22, pp. 470–478, 2016, [PubMed] [Google Scholar]

23. M. Odlum,and S. Yoon,"What can we learn about the Ebola outbreak from tweets?" Am J Infect Control,vol. 43, pp. 563–571,2015, [PMC free article] [PubMed] [Google Scholar].

24. Bin, ShengLu, Mingqing Ma, and Qiangqiang, The Effect of Mobile Social Media Sharing Behavior on the Subjective Well-Being and Mental Health Regulation of Internet Users 7573322, VL - 2022, Mobile Healthcare Social Media (MHSM) https://doi.org/10.1155/2022/7573322, 10.1155/2022/7573322

25. Bazarova, Natalya (Natalie) and Choi, Yoon andSchwandaSosik, Victoria andCosley, Dan and Whitlock, Janis, Social Sharing of Emotions on Facebook,pp. 154-164, 2015, 10.1145/2675133.2675297.

26. L.E. Charles-Smith, T.L. Reynolds, M.A. Cameron, M. Conway, E.H. Lau, J.M. Olsen, J.A. Pavlin, M. Shigematsu, L.C. Streichert, K.J. Suda,and C.D. Corley, "Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review,"PLoS One,vol. 10, no. 10, p. e0139701,2015,doi: 10.1371/journal.pone.0139701. PMID: 26437454; PMCID: PMC4593536.

27. Alessa and M. Faezipour, "A review of influenza detection and prediction through social networking sites," Theoretical Biology and Medical Modelling, vol. 15, no. 1, pp. 1–27, 2018.

28. E. Lau and P. Wark, "Twitter-Based influenza detection after flu peak via tweets with indirect Information : text mining study," Journal of Medical Internet Research, vol. 4, no. 3, pp. 1–27, 2019.

29. S. Amin, M. I. Uddin, M. A. Zeb, A. A. Alarood, M. Mahmoud, and M. H. Alkinani, "Detecting dengue/flu infections based on tweets using LSTM and word embedding," Institute of Electrical and Electronics Engineers Access, vol. 8, pp. 189054–189068, 2020.

30. S. Amin, M. Irfan Uddin, M. Ali Zeb, A. AbdulsalamAlarood, M. Mahmoud, and M. H. Alkinani, "Detecting information on the spread of dengue on twitter using artificial neural networks," Computers, Materials and Continua, vol. 67, no. 1, pp. 1317–1332, 2021.

31. C. Feng, K. Yu, M. Aloqaily, M. Alazab, Z. Lv, and S. Mumtaz, "Attribute-based encryption with parallel outsourced decryption for edge intelligent IoV," Institute of Electrical and Electronics Engineers Transactions on Vehicular Technology, vol. 69, no. 11, pp. 13784–13795, 2020.

32. X. Zhang, L. Yang, Z. Ding, J. Song, Y. Zhai, and D. Zhang, "Sparse vector coding-based multi-carrier NOMA for in-home health networks," Institute of Electrical and Electronics Engineers Journal on Selected Areas in Communications, vol. 39, no. 2, p. 325, 2021.

33. K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastava, and P. Chatterjee, "Efficient and privacy-preserving medical research Support platform against COVID-19: a blockchain-based approach," Institute of Electrical and Electronics Engineers Consumer Electronics Magazine, vol. 10, no. 2, pp. 111–120, 2021.

34. Zhao, Qiankun, and PrasenjitMitra. "Event Detection and Visualization for Social Text Streams." ICWSM. 2007.

35. Hearst, Marti. Search user interfaces. Cambridge University Press, 2009.

36. Sayyadi, Hassan, Matthew Hurst, and Alexey Maykov. "Event detection and tracking in social streams." Third International AAAI Conference on

37. Weblogs and Social Media. 2009.

38. Aggarwal, Charu C., and KarthikSubbian. "Event detection in social streams." Proceedings of the 2012 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, 2012.

39. McCreadie, Richard, et al. "Scalable distributed event detection for twitter." 2013 IEEE international conference on big data. IEEE, 2013

40. Panagiotou N, Katakis I, Gunopulos D. Detecting events in online social networks: definitions, trends and challenges. In: Michaelis S, editor. Solving large scale learning tasks: challenges and algorithms. Cham: Springer; 2016. p. 42–84.

41. W.W.Xu, Y.Sang, S.Blasiola, and H.W.Park,"Predicting opinion leaders in Twitter activism networks: the case of the Wisconsin recall election," Am Behav Sci., vol. 58, pp. 1278–1293,(2014.

42. M. Avvenuti, S.Cresci, A.Marchetti, C.Meletti, and M. Tesconi,"Predictability or early warning: using social media in modern emergency response," IEEE Internet Comput., vol. 20, pp. 4–6, 2016.

43. M. Avvenuti, S. Cresci, F. Del Vigna, T.Fagni, and M. Tesconi,"CrisMap: a big data crisis mapping system based on damage detection and geoparsing,"InfSyst Front., vol. 20, pp. 1–19, 2018.

44. V. Lampos, and N. Cristianini,"Nowcasting events from the social web with statistical learning," ACM T IntellSyst Technol., vol. 3, p. 72, 2012.

45. A.R.MartínezTeutle, "Twitter: network properties analysis," 20th International Conference on Electronics, Communications and Computer (CONIELECOMP) IEEE, pp. 180–186, 2010.

46. S.Cresci, R.Di Pietro, M. Petrocchi, A. Spognardi, M.Tesconi,"Emergent properties, models, and laws of behavioral similarities within groups of Twitter users,"ComputCommun, vol. 150, pp. 47–61, 2020.

47. Mazza M, Cresci S, Avvenuti M, Quattrociocchi W, Tesconi M (2019) RTbust: exploiting temporal patterns for botnet detection on Twitter. In: The 11th International Conference on Web Science (WebSci'19), ACM

48. Prieto Curiel R, Cabrera Arnau C, Torres Pinedo M, González Ramírez H, Bishop SR (2019) Temporal and spatial analysis of the media spotlight. Comput Environ UrbSyst 75:254–263

49. Barchiesi D, Preis T, Bishop SR, Moat HS (2015) Modelling human mobility patterns using photographic data shared online. R Soc Open Sci 2:150046

50. S. Cresci et al.,"Towards a dbpedia of tourism: the case of Tourpedia," In: Proceedings of the 2014 International Semantic Web Conference (ISWC'14), pp. 129–132, 2014.

51. C.I.Muntean, Nardini FM, Silvestri F, Baraglia R,"On learning prediction models for tourists paths", ACM T IntellSyst and Technolvol. 7, p. 8, 2015.

52. D'Andrea E, Ducange P, Lazzerini B, F. Marcelloni,"Real-time detection of traffic from Twitter stream analysis," IEEE T IntellTranspSyst, vol. 16, pp. 2269–2283, 2015.

53. I. Himelboim, S. McCreery, and M. Smith,"Birds of a feather tweet together: integrating network and content analyses to examine cross-ideology exposure on Twitter,"JComput-MediatCommun, vol. 18, pp. 40–60, 2013.

54. I. Himelboim, D.Hansen, and A.Bowser,"Playing in the same Twitter network: political information seeking in the 2010 US gubernatorial elections,"InfCommun Soc., vol. 16, pp. 1373–1396,2013.

55. M. Coletto et al.,"Perception of social phenomena through the multidimensional analysis of online social networks," Online Social Networks and Media vol. 1, pp .14–32, 2017, http://www.sciencedirect.com/science/article/pii/S246869641630009X

56. P.S. Dodds, K.D. Harris, I.M. Kloumann, C.A. Bliss, C.M.Danforth,"Temporal patterns of happiness and information in a

global social network: hedonometrics and Twitter,"PLoS ONE, p. 6:e26752, 2011.

57. H. Gil de Zúñiga, N. Jung, and S. Valenzuela,"Social media use for news and individuals' social capital, civic engagement and political participation," J Comput-MediatCommun., vol. 17, pp. 319–336, 2012, https://doi.org/10.1111/j.1083-6101.2012.01574.x.
58. Carbezudo MAS, Pardo TAS. Exploring classical and linguistically enriched knowledge-based methods for sense disambiguation of verbs in Brazilian Portuguese news texts. Nat Lang Process. vol. 59, pp. 83–90, 2017.
59. Y. Gutierrez-Vazquez, S. Vazquez S, Montoyo A. A semantic framework for textual data enrichment. Expert Syst Appl., vol. 57, pp. 248–69,2016.
60. A. Alkhatlan, J. Kalita, and A. Alhaddad, "Word sense disambiguation for Arabic exploiting WordNet and word embedding,"ProcediaComput Sci.,vol. 142, pp. 50–60,2018, https://doi.org/10.1016/j.procs.2018.10.460.
61. S. Katragadda, R. Benton, and V. Raghavan, "Framework for real-time event detection using multiple social media sources," Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS). Waikoloa, Hawaii, p. 1716–1725, 2017, https://doi.org/10.24251/HICSS.2017.208
62. N. Agarwal, and Liu, H. Modeling and Data Mining in Blogosphere,vol. 1, 2009, (R. Grossman, Ed.)
63. C. Hayes, andP. Avesani, "Using tags and clustering to identify topic-relevant blogs," International Conference on Weblogs and Social Media (ICWSM), 2007.
64. S. Stieglitz, L. Dang-Xuan, A. Bruns, C. Neuberger, Social Media Analytics – An Interdisciplinary Approach and Its Implications for Information Systems Business and Information Systems Engineering, vol. 6, no. 2, pp. 89-96, 2014,10.1007/s11576-014-0407-5
65. S. Stieglitz, L. Dang-Xuan Social media and political communication: A social media analytics framework Social Network Analysis and Mining, vol. 3, no. 4, pp. 1277-1291, 2013,10.1007/s13278-012-0079-3
66. R. Newman, Chang V, Walters RJ, Wills GB, Web 2.0–the past and the future. Int J InfManag., vol. 36, no. 4, pp. 591–598, 2016.
67. M.F. Uddin, and N. Gupta et al.,"Seven V's of Big Data understanding Big Data to extract value," In: American Society for Engineering Education (ASEE Zone 1), Zone 1 Conference of the IEEE, pp 1–5, 2014.
68. N. Khan, Yaqoob I, Hashem IAT, Inayat Z, Mahmoud Ali WK, Alam M, Shiraz M et al.,"Big data: survey, technologies, opportunities, and challenges,"Sci World J., 1–18, pp. 2014.
69. V.C. Storey, and I.Y Song,Big data technologies and management: what conceptual modeling can do. Data Knowl Eng. Vol. 108, pp. 50–67, 2017.
70. Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., &Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.
71. M. Schroeck, R. Shockley, Smart J, Romero-Morales D, P. Tufano,"Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report," In: IBM Institute for Business Value and Said Business School at the University of Oxford, 2012.
72. A.Gandomi, and M. Haider,"Beyond the hype: big data concepts, methods, and analytics,"Int J InfManag., vol. 35, no. 2, pp. :137–144, 2015.
73. C.P. Chen and C.Y Zhang,"Data-intensive applications, challenges, techniques and technologies: a survey on Big Data,"Inf Sci., vol. 275, pp. 314–347, (2014.
74. Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU. The rise of "big data" on cloud computing: review and open research issues. InfSyst, vol. 47, pp. 98–115, 2015.
75. I. Yaqoob, Hashem IAT, Gani A, Mokhtar S, Ahmed E, Anuar NB, Vasilakos AV,"Big data: from beginning to future,"Int J InfManagvol. 6, no. 6, pp. 1231–1247,2016.
76. Fan W, Bifet A (2013) Mining big data: current status, and forecast to the future. ACM SIGKDD ExplorNewsl 14(2):1–5
77. Orgaz, G.B., J.J. Jung and D. Camacho,"Social big data: recent achievements and new challenges"InfFus, vol. 28, pp. 45–59, 2016.
78. A.F. Haryadi, Hulstijn J, Wahyudi A, Voort H, van der, Janssen M,"Antecedents of big data quality: an empirical examination in financial service organizations," In: IEEE international conference on Big Data (Big Data), IEEE, pp.116–121, 2016.
79. D. Agrawal, P. Bernstein, Bertino E, Davidson S, Dayal U, Franklin M, Gehrke J, Haas L, Halevy A, Han J, Jagadish HV, Labrinidis A, Madden S, Papakonstantinou Y, Patel JM, Ramakrishnan R, Ross K, Shahabi C, Suciu D, Vaithyanathan S, Widom J, 2012, Challenges and opportunities with big data—a community white paper developed by leading researchers across the United States. http://cra.org/ccc/docs/init/bigda tawhitepap er.pdf
80. A. Corbellini, Mateos C, Zunino A, Godoy D, Schiaffino S (2017) Persisting big-data: the NoSQL landscape. InfSyst 63:1–23
81. Rajesh, M., &Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0.Computers and Electrical Engineering, 104, 108481.
82. Sun X, Liu L, Ayorinde A, Pannerselvam J. ED-SWE: event detection based on scoring and word embedding in online social networks for the internet of people. Digital Commun Net. 2021. https://doi.org/10.1016/j.dcan.2021.03.006.
83. McCreadie R, Macdonald C, Ounis I, Osborne M, Petrovic S. Scalable distributed event detection for Twitter. 2013 IEEE International Conference on Big Data. Silicon Valley, CA, USA, 2013. p. 543–549.
84. Kaleel SB, Almeshary M, Abhari A. Event detection and trending in multiple social networking sites. Proceedings of the 16th Communications and Networking Symposium. San Diego, CA, USA, 2013:5.
85. Musaev A, Wang D, Shridhar S, Lai C, Pu C. Toward a real-time service for landslide detection: Augmented explicit semantic analysis and clustering composition approaches. 2015 IEEE International Conference on Web Services New York, NY, USA, 2015. p. 511–518. https://doi.org/10.1109/ICWS.2015.74.
86. S. Romero, K. Becker, "A framework for event classification in tweets based on hybrid semantic enrichment," Expert Syst Appl. vol. 118, pp. 522–38,2019, https://doi.org/10.1016/j.eswa.2018.10.028.