# **AN EXPLORATORY CASE STUDY ON INFRASTRUCTURE RESILIENCE IN KERALA** **USING SOCIAL MEDIA ANALYSIS**

**Dinesh Pugalenthi1, K. S. Anandh2, K. Prasanna2, Ruggiero Lovreglio3 and**

**Yuvaraj Dhanasekar4 and Rajkumar R5**

Planning Engineer1, Utracon Structural System Pvt Ltd, Chennai, India

Assistant professor2, Department of Civil Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Chennai 603203, India

Research Scholar4, Department of Civil Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Chennai 603203, India

Data Scientist3, Indium Software Ltd, Chennai, India.

Assistant professor5, Department of Data Science and Business System, College of Engineering and Technology, SRM Institute of Science and Technology, Chennai 603203, India

**ABSTRACT:** Disaster resilience is a combination of both infrastructure resilience and community resilience. Technological advancements and its tools, such as the internet and social media can play a key role in community resilience. In major emergency situations, social media can offer wealth of information for capturing details on human behavior, responses, opinions and sentiments during any catastrophic events such as natural disasters. During natural calamities, social media dispenses a plethora of facts and figures which includes information about the nature of the disaster, people’s sentiments and relief efforts.

This study is a new attempt to assess how Twitter data can be used to investigate major disasters. As such, we propose a novel solution to analyze a disaster exclusively by making use of tweets posted by Twitter users during a past event. We applied the proposed solution for the major flood in Kerala in 2018. In this work, diverse analytical techniques are proposed and applied to the collected tweets and generate outputs through comprehensive graphical analysis. The results of our study provide insights into users’ requirements during a disaster, the frequency distribution of tweet timings and the geographical distribution of the flood. This work shows that analyzing data from social media provides a feasible, economical, guileless and instantaneous alternative to traditional methods for analyzing a disaster and the cognizance of the affected population towards a natural calamity.

**KEY WORDS:** Disaster management, Twitter, Text Mining, Kerala Floods, Text Classification

1. **INTRODUCTION**

Many countries compile trillions of infrastructural assets to accelerate their global socio-economic growth [1]; also, by 2050, the population of Megacities is anticipated to increase by 25% [2]. While the community is directly proportional to the existing infrastructure, fragile investments are pumped into protection and resilience techniques. Disaster is a condition in which the targeted population is unable to comply with and rebound from a threat without significant assistance [3]. Disaster Resilience is the capacity of a group exposed to hazards to adapt, resist and adjust to achieve and sustain an appropriate standard of functioning and structure [4]. The post-disaster recovery process is complex, time and cost-consuming [5]. In large-scale disasters, recognizing the extent of events is crucial for effective reconstruction and recovery of safety and necessary services [6]. The need for appropriate coordination in the event of a catastrophic disaster is the nucleus here as we continue to witness its impact [7] [8].

Social media networking dramatically increases the amount of information shared between affected persons and emergency personnel, thereby leading to a higher degree of situational awareness during a disaster, resulting in quicker and more informed choices and actions. Disaster management through Twitter is one of the most popular online microblogging sites, which is recently being studied in the context of humanitarian crises and natural disasters and has become highly popular due to individuals expressing their opinions in real-time, which are visible across the globe [9] [10]. First-hand information from responders often conveys appropriate and credible information that is of greater benefit to official bodies to respond to emergencies appropriately. Social networking tools are ideal for extracting real-world data [11]. Geo-location, the function of determining the position of a social media post – may serve a range of downstream applications, such as advertisement, optimization, event detection, trend analysis and epidemic monitoring [12]. In times of mass emergencies, people turn to Twitter to collect and spread accurate, critical information [13]. Real-time local events can be detected from geo-tagged tweet streams [14]. On the other end, this data has been availed for analysis through Twitter’s Application Programming Interface (API). Apart from mainstream social media, there are various microblogging sites like Tumblr, DailyBooth, Pinterest, etc. The salient aspect of accumulating data through this form is they are mostly uncensored and provide valid information.

A tweet is a 140-character-long status update that addresses the question, “What’s happening?” This message update is associated with the username, date, time of posting, geographic location, and other metadata [15]. People usually post and receive messages related to actual real-world events when they happen as quickly as possible for circulation. Twitter helps users to post status updates or tweets to a network of users using various communication services. Twitter is being updated hundreds of millions of times a day by users all over the world. Its content varies greatly depending on user preferences and actions [16]. The benefit of understanding incidents is the safety of life and assets, the survival of emergency conditions, the support of people and the communication with the police [17]. To date, around 500 million tweets are created every day; among those, interpreting geo-located tweets now seems realistic and relevant [15][18]. Generally, data mining can be used to recognize influential patterns, perceptions or viewpoints, forecast demands, and in specific research to increase awareness and effectively communicate [19].

The major inspiration behind this work is the fact that people’s needs and opinions during and aftermath of a natural disaster are not being heeded to a large extent. In spite of data being available in abundance from various sources, there is still a huge reliance on interviews, manual surveys and offline information. This task can be automated by utilizing a data already available data, such as data posted on Twitter. However, making detailed information accessible from social media in emergencies can be very complicated [20].

In this paper, we aim to assess how public data available on social media can help in investigating disasters. To achieve this goal, we propose a new solution to assess the effects of floods on infrastructure and people through users’ posts on Twitter. We applied the proposed solution to the major flood that occurred in Kerala in 2018. The proposed solution has the potential to store gleaned data from Twitter and provide a comprehensive graphical analysis that covers user sentiments, emotions, and the impact of floods on infrastructure and the geographical distribution of disasters over different regions.

1. **LITERATURE REVIEW**

***2.1 Social media in disaster management:***

In recent years, social media has become an indispensable tool for assisting with disaster risk reduction efforts across the spectrum of preparedness, response, and recovery [21]. Social media dismantles the conventional sender/receiver model in contrast to the one-way communication of traditional mass media. Individuals can gather first-hand information and disseminate it through social media in real-time rather than waiting for professional news reporters to arrive on the scene to report the situation [22]. During a disaster, social media platforms such as Twitter, Facebook, and WhatsApp have enabled affected individuals to connect and share critical information about the disaster. Researchers have utilized social media data to comprehend the impact of natural disasters, assess the damage, and aid disaster response efforts. In times of emergency, social media can be used as a vital tool for communicating with others, sharing maps and updates, coordinating relief efforts, keeping in touch with family members, and soliciting financial contributions. Social media platforms were instrumental in disseminating information, coordinating aid efforts, and providing emotional support to affected communities during the 2010 Haiti earthquake, the 2011 Tohoku earthquake and tsunami, the 2012 Sandy hurricane, the 2015 Nepal earthquake, the 2017 Harvey hurricane, etc., [23][24][25][26][27]. According to the study by Gunessee et al. (2017) on the 2015 Chennai floods, social media facilitates online volunteering. The study discussed the locals' obstacles, enablers, inspirations, and spontaneous contributions to relief efforts [28].

***2.2 Limitations in using social media for disaster management:***

Even though social media platforms play a vital role in information during a disaster and post-disaster reconstruction, there are certain limitations in using the data available from social media platforms. One of the major barriers is the trustworthiness of the data generated. Generally, decisions can be made only from trusted sources. But in the case of social media platforms, anybody can post anything, including false information that may create panic among the people or may be a threat to the safety of the people [29]. The barriers also include the lack of tools for processing a large number of available data and the lack of expert members to handle the tools [30]. The studies conducted on social media were the representation of only the users who posted their views on various platforms. It could not be considered from the perspective of the entire affected population. Therefore a chance of bias might be created on the situational awareness developed using those data [31]. Service interruptions are one of the other drawbacks of using social media for disaster response. It is possible that the internet won't be accessible right away or even for a while after a disaster. It delays the victims in sharing their current situation and creates difficulty in organizing the relief efforts by the relief organizations [32].

***2.3 Twitter in disaster management:***

For official institutions dealing with disasters, social media like Twitter is regarded as an information and education tool. Onal et al. (2022) analyzed the tweets generated during 2021 by the Disaster and Emergency Management Authority (AFAD) of Turkey to identify tweets related to various types of disasters. They preferred Twitter as a fast and effective platform for the use of social media in disaster management and in critical situations. The efficient use of Twitter during a crisis situation ensures coordination between the general public and governmental agencies, allowing for the quick and secure continuation of intervention and aid processes [33].

Phengsuwan et al. (2021) reported on the use of social media data in disaster management. The researchers reviewed research publications to investigate the contributions of social media data and the techniques for data management and analysis in disaster management. According to their findings, several publications have proposed using social media data for disaster management, with Twitter being one of the most important sources used for disaster management. The temporal and spatial information extracted from Twitter is critical for supporting decision-making in disaster management. Machine learning and information retrieval algorithms were widely used to collect, classify, and extract essential information from social media [34].

Karami et al. (2019) created an analytical framework called Twitter Situational Awareness (TwiSA) to improve situational awareness for disaster preparedness, response, and recovery. Text mining methods like sentimental analysis and topic modeling was used by the framework. The framework was successfully applied to monitor public complaints regarding the 2015 flood in South Carolina [35].

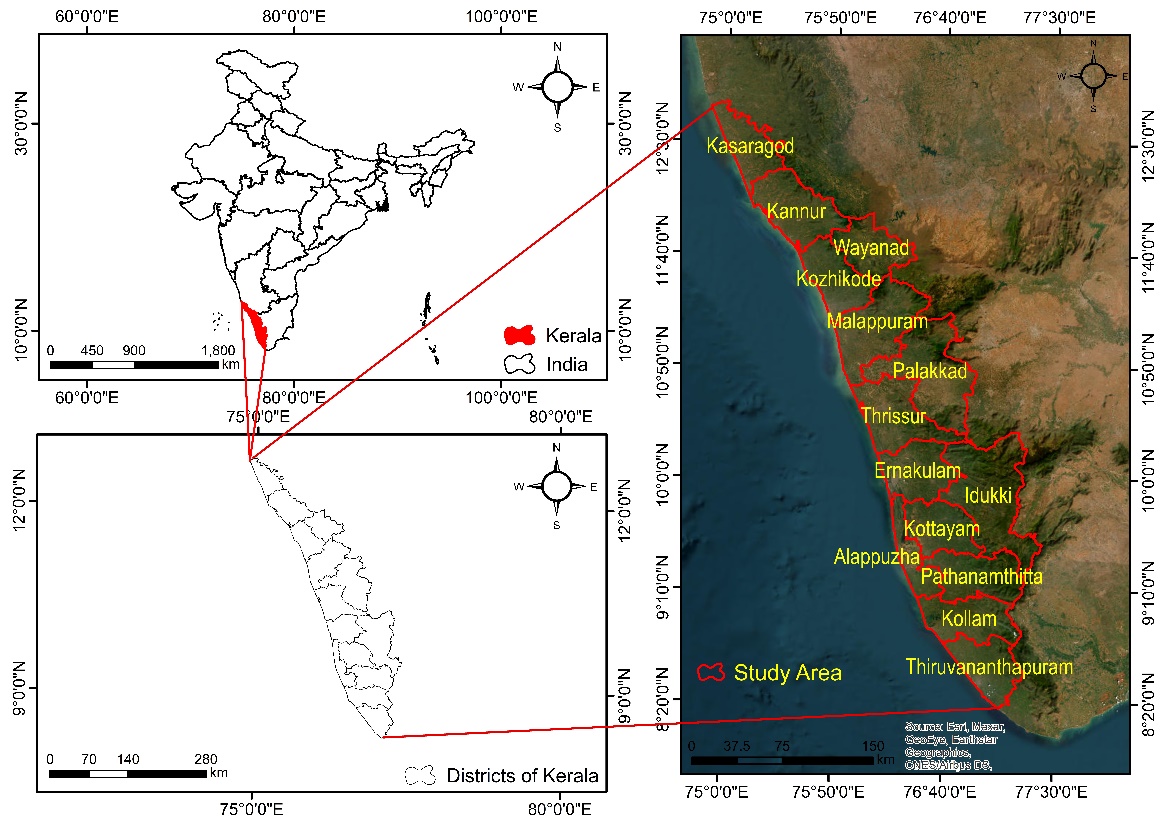
Twitter messages posted by United Kingdom (UK) officials during the heavy snow of December 2010 and the riots of August 2011 were analyzed by Panagiotopoulus et al. (2016). The authors examined the impact of Twitter as a tool for communicating risks to the public. They found both the informational and inferential components present in the Twitter messages. The authors claim that Twitter was used in various ways to communicate and manage associated risks, including messages that offered official updates, prompted protective behaviour, raised awareness, and directed people's attention to preventative measures [36].

Xiao et al. (2015) developed a model to account for the number of tweets based on mass, material, access, and motivation (MMAM). An empirical analysis of the tweets about Hurricane Sandy in New York City was performed, which affirmed the model. According to them, individuals can gather first-hand information about a disaster and disseminate it through social media in real-time rather than waiting for professional news reporters to arrive on the scene to report the situation. But the researchers may not accurately reveal a situation's true picture because of disparate access to social media and heterogeneous motivations for using it [22].

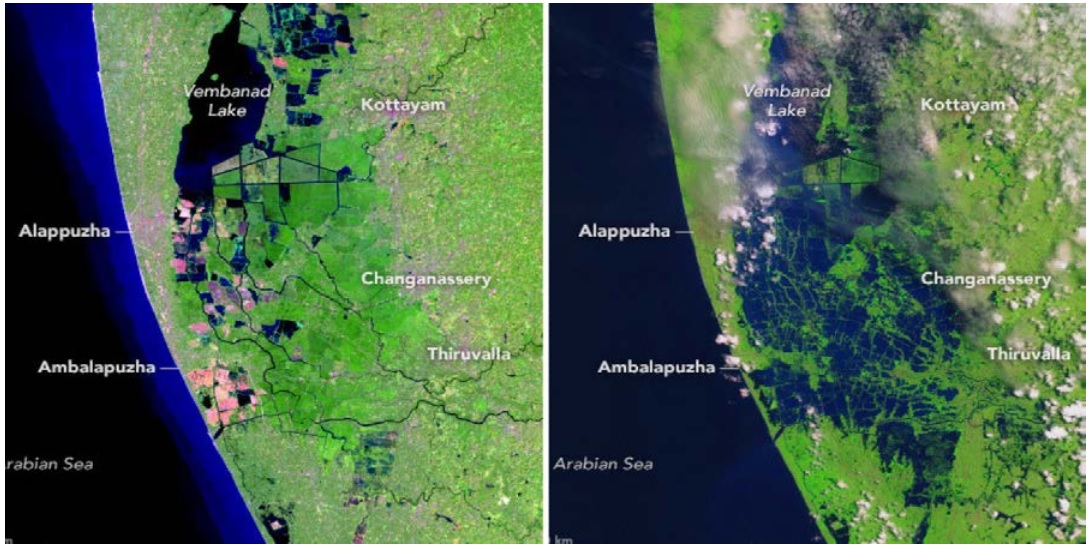
1. **BACKGROUND: THE KERALA FLOODS**

All parts of the Southern States of Kerala, including Mangalore, receive most of its rainfall from the southwest monsoon. The time period from June to September is referred to as the ‘Southwest Monsoon’ period, as the onset of the season starts in June and prolongs till early September. On the whole, the Indian subcontinent receives its primary rainfall from the southwest monsoon. But in 2018, the amount of rainfall received by Kerala was 116 percent more than the regular monsoon season [37]. This was the most devastating flood since the great flood of 99, which was encountered in the year 1924 by Kerala. In the year 1924, rainfall received by Kerala stood at 3368 mm, while in the year 2018, 2086 mm of rain was received by them. There are a total of 14 districts in Kerala and almost all of them were on red alert during this calamity. Particularly in the hilly districts of Idukki and Wayanad, where the normal amount of rainfall is around 290 mm, the continuous rainfall from August 8 to August 18 was excessive and unprecedented; instead, these areas received about 700 mm of rain, which served as the catalyst for the flooding. This disaster caused severe damage and this was amplified due to illegal urban development, construction in hilly areas, mismanagement and improper construction of dams. It was estimated that over 483 people were dead, 140 were missing, and 20 million people were affected, as per local media. The floods inflicted severe losses to the state, costing them around ₹40000 Crore (400 billion) [38].

The study area used for work is shown in Figure 1 and the satellite images of the state before and after the disaster is shown in Figure 2.



**Fig. 1: Satellite image showing the study area**

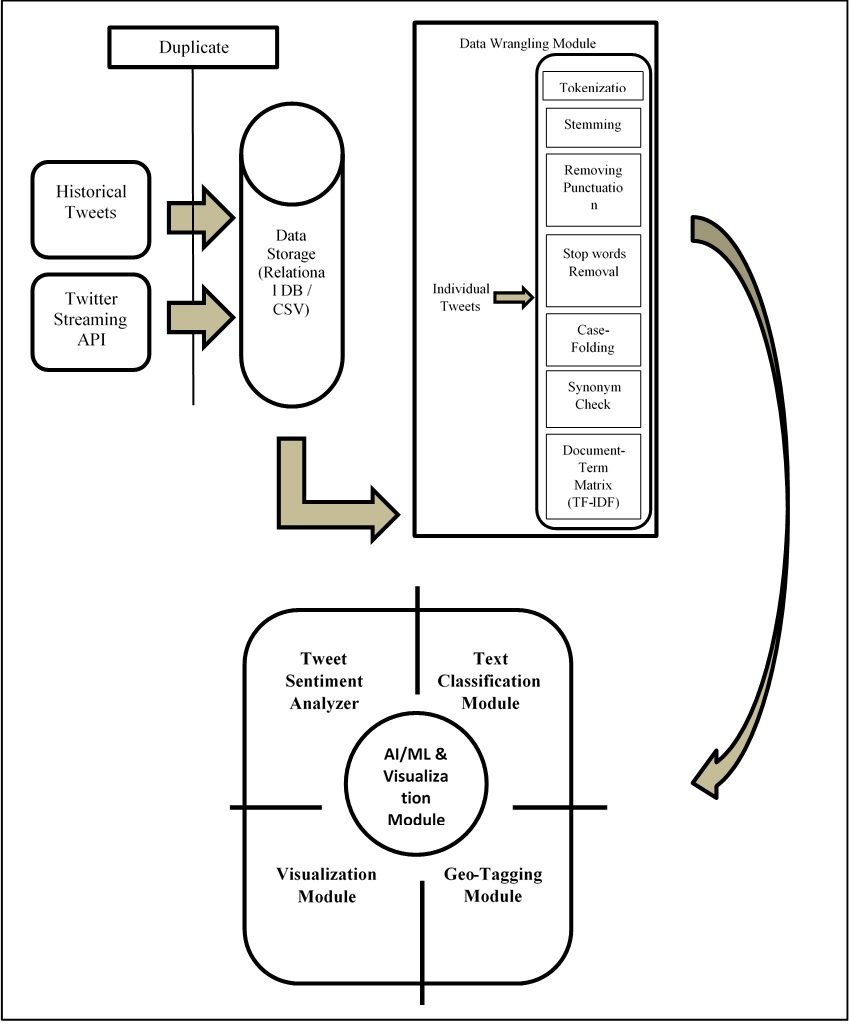


**Fig. 2: Satellite image of the study area before and after the flood**

**Source:** https://earthobservatory.nasa.gov/images/92669/before-and-after-the-kerala-floods

1. **FRAMEWORK**

The proposed solution works upon the preliminary idea that there is always a constant supply of streaming tweets in the cyber universe containing a myriad of information about various domains. Whenever there is an occurrence of any catastrophic event, such as a disaster or crisis, there is usually a spur in the generation of tweets from the region in question. The data collected is filtered and assorted before sending it to the modeling phase. Figure 3 gives out the high-level architecture diagram of the application. The different modules of the system are detailed as follows. In section 4.1, we describe how the data is extracted.

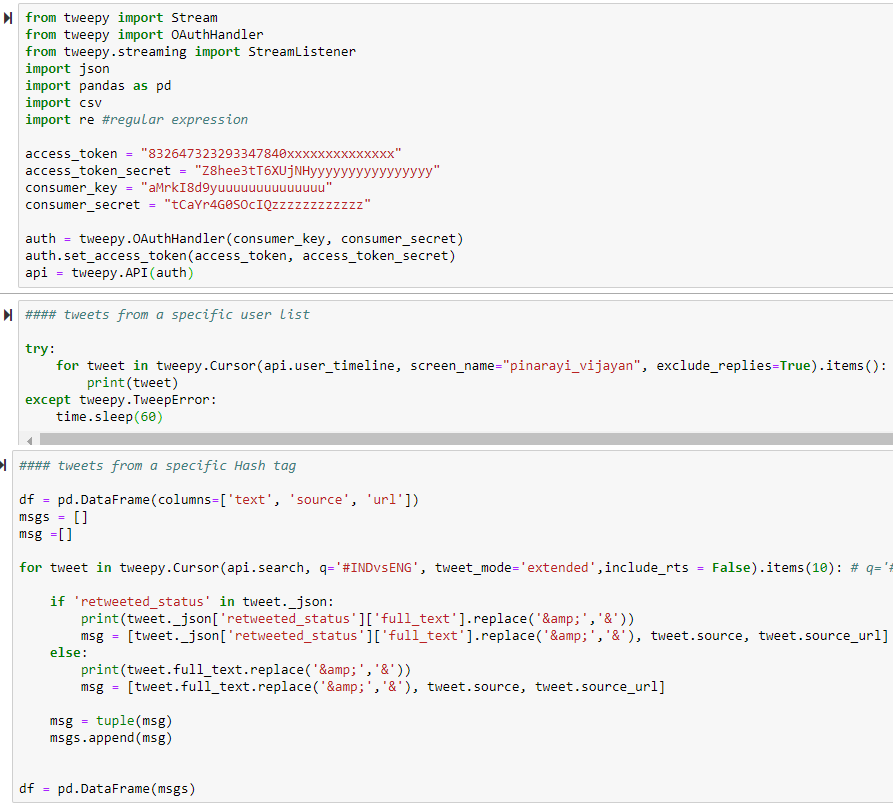
****

**Fig. 3: System architecture for analyzing geo-tagged tweets generated during the natural crisis**

* 1. ***Data Extraction:***

Building an environment to extract/crawl data from Twitter through the ‘Twitter API’ or ‘GetOldTweets3’ python library.

Python has multiple libraries to scrap data from Twitter. ‘Tweepy’ is a library to retrieve tweet content through Twitter API. As a first step, the user has to register with Twitter to get an authentic credential. This registration provides access to Twitter API from any third-party application. The authentication is compressed with a token and a secret key. These two entities are used in ‘Tweepy’ package methods to establish an authentic connection with the Twitter API. Figure 4 shows the implementation of the “Tweepy” library.



**Fig. 4: Connecting to Twitter API through Tweepy**

Twitter’s official API has both time and performance constraints. The extraction of tweets older than a week is restricted. Hence, the ‘Tweepy’ library does not come in handy in gathering historical tweets. ‘GetOldTweets3’ is one such python library that provides an interface to extract historical tweets. Through this, it is possible to extract even the deepest, oldest tweets. Figure 5 displays the implementation of the “GetOldTweets3” library.



**Fig. 5: Extracting Tweet content through GetOldTweets**

* 1. ***Removing Duplication***

It is possible for an automated program to capture the same tweet more than once due to re-tweets by the user. Hence, data gathered from secondary sources are prone to duplication. Performing a redundancy check over the extracted tweets and removing the duplicate accounts for a better analysis.

The redundancy in tweets is avoided either by string matching algorithm or regular expression series method based on the set threshold. A measure of similarity between two string sequences is achieved through the Levenshtein distance string metric. Each incoming tweet will be subjected to the Levenshtein distance algorithm and the distance metric is calculated between the existing corpus of tweets and the new content. If the distance metric crosses the threshold, which means the incoming tweet is similar in information, it is rejected. On the contrary, it is added to the corpus if the distance metric falls below the threshold.

***4.3 Storing Data***

The data which passes the redundancy check is finally stored in a relational database in the form of Structured Query Language (MySQL). It is an open-source relational database available under the terms of the GNU General Public License. MySQL makes storing and fetching data seamless. It is installed on a server/computer and the data operations like insertion, deletion, and retrieval etc., have been performed. The database contains the actual tweet, the username of the person who tweeted, the timestamp, and geolocation. The storage mechanism can store even more information about the posted tweet. The same data has also been stored in Comma Separated Values (CSV’s) for convenient accessibility.

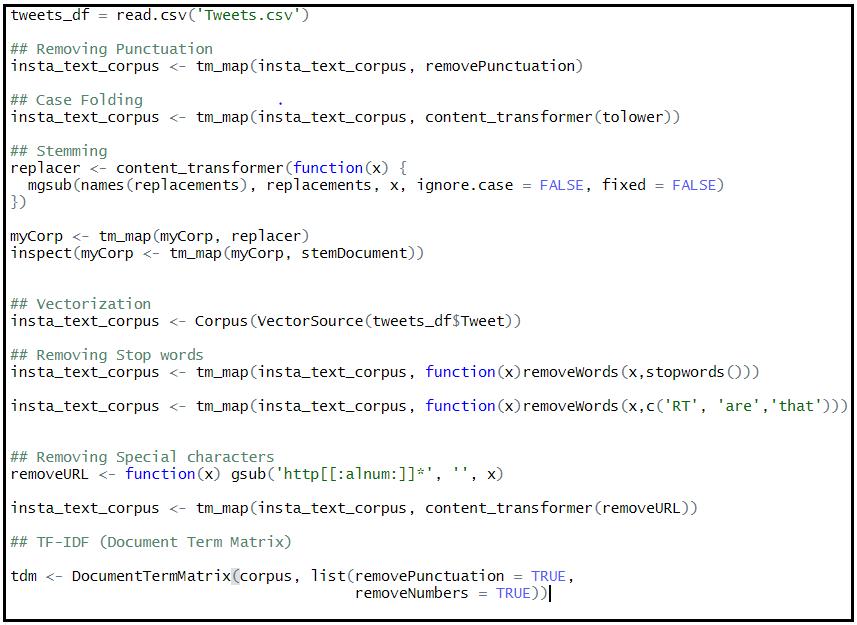
***4.4 Data Wrangling***

The stored tweets are in the raw format. Before processing any textual data into a machine learning pipeline, the non-numerical data must be gleaned carefully to enable it to produce accurate results. The data manipulation procedures vary from numerical data to text data. The ‘tm’ library in R provides powerful functions to perform data wrangling on non-numerical data. It has been used to execute exploratory analyses over a text dataset and detect high-level patterns. Figure 6 shows the implementation of data wrangling in R. The following six pre-processing modules are applied to the stored content:

1. ***Tokenization:***It is a fundamental process of chopping a sequence into multiple pieces. It is divided into sentence and word tokenization. Sentence tokenization, also known as sentence segmentation, is the procedure to divide a string into its component sentences. For instance, in the English language, sentences are tokenized by splitting them whenever a punctuation mark is encountered. Whereas word tokenization, also known as word segmentation, is the procedure to divide a string into its component words.
2. ***Stemming & Lemmatization:***Most of the words in English language are derived from its root word. These words which have the same root have identical meanings. They are called derivationally related words belonging to same families. The main objective of both lemmatization and stemming is to truncate inflectional forms and sometimes derivationally related forms of a word to its common base form. For instance,

*democrat, democratic & democratization -> democracy*

1. ***Removing Punctuation:*** Handling punctuation is one of the vital steps in text processing. Since they do not contain any absolute meaning, while performing an analysis, punctuations do no benefit to the outcome neither in terms of accuracy nor in terms of performance. In turn, they slow down the performance of the model during training. Hence it is essential to remove them before proceeding to the modeling stage. Figure 5 shows the codes used for removing the punctuations.



**Fig. 6: Data Wrangling using ‘tm’ library in R**

1. ***Removing Stop Words:***Every language has its own repository of stop words. They are nothing but most commonly used words in any language. When applying machine learning to each post, stop words have the ability to add a lot of noise to the analysis. In the world of analytics, they are called irrelevant words. Based on the analysis, we have modified the list of stop words. Some of the frequently occurring stop words are ‘and’, ‘an’, ‘the’, ‘is’, ‘a’ etc.
2. ***Case-Folding & Synonym Check:***A simple strategy to execute case-folding is to reduce all the letters to lower case. This is assumed to be a good idea since the absolute starting of any sentence begins with an upper case letter. Through this, tokens have been converted into lower case and also removing the duplicate tokens thereby easing the model building module. In linguistics, a single word can have multiple variants. They are termed as synonyms. For instance, two different words projecting the same would do no good to the analysis. Hence for effective duplicate removal, synonym check has to be done to eliminate redundancy.
3. ***TF-IDF:*** It stands for term frequency-inverse document frequency which is often used in information retrieval and text mining. The tf-idf weight is a statistical measure which denotes how much importance a word holds to a document in a collection of corpus. This importance is directly proportional to the number of occurrences of the word in the corpus. The tf-idf weight is computed by two terms: firstly, the normalized Term Frequency (TF) which is equal to the number of times a word appears in the document, divided by the entire corpus of words in the document. Secondly, the Inverse Document Frequency (IDF) which is equal to the logarithm of the number of documents in the corpus divided by the number of documents where the specific term emerges.

*TF(t) = (No. of times term t appears in a document) / (Total number of words in a document)*

*IDF(t) = log*e*​(Total number of documents / No. of documents with term t in it)*

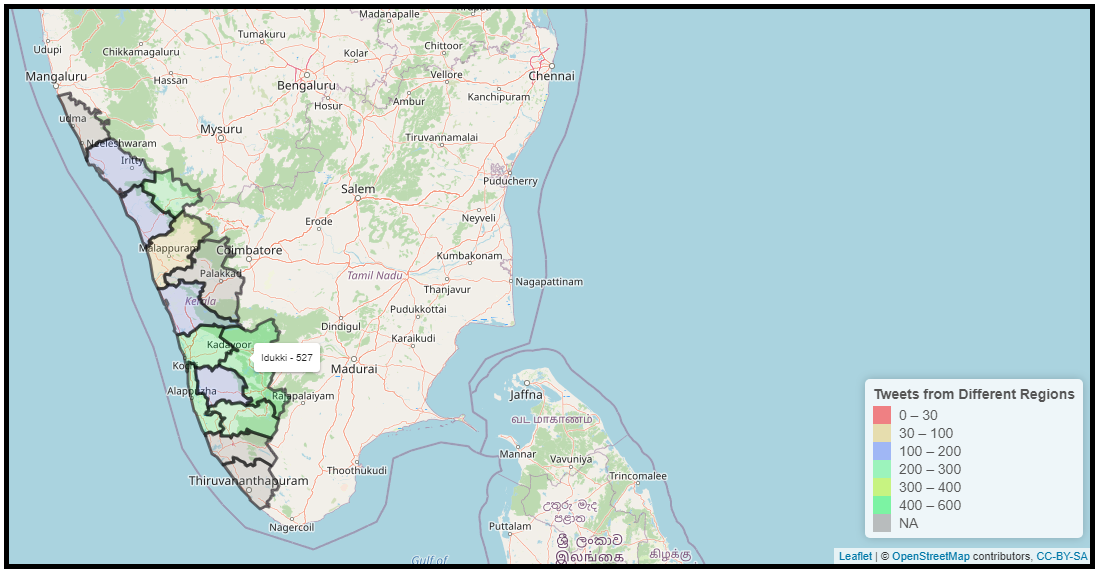
1. **RESULTS**

The processed data is then passed onto to the final phase for analysis and interpretation. This module is divided into four distinct sub-modules in which data is checked for Visualization, Geo-Tagging, Text Classification and Sentiment Analysis.

1. ***Geo-Tagging:*** The whole matter of interest in any disaster analysis is the location of a natural disaster. For extracting locational information from the tweets, a geo-ﬁlter tag is applied on each Twitter post to determine their point of origin. However, in many cases, tweets actually do not originate from the affected region. During such instances, people tweeting about the disaster will include the location of the disaster within the tweet, either through hashtags or comments. The location mentioned inside the tweet has been determined with the help of Google Maps API, wherein all the Twitter posts are passed through a method one by one which chops each tweet into words and tries to ﬁnd the coordinates of that word through the Google Maps API. The technique, in R or Python, returns the geographical coordinates such as latitude & longitude if that particular word exists anywhere in Google Maps, else returns an error. Thus, all the potential locations mentioned in each tweet are here found and labeled along with them.

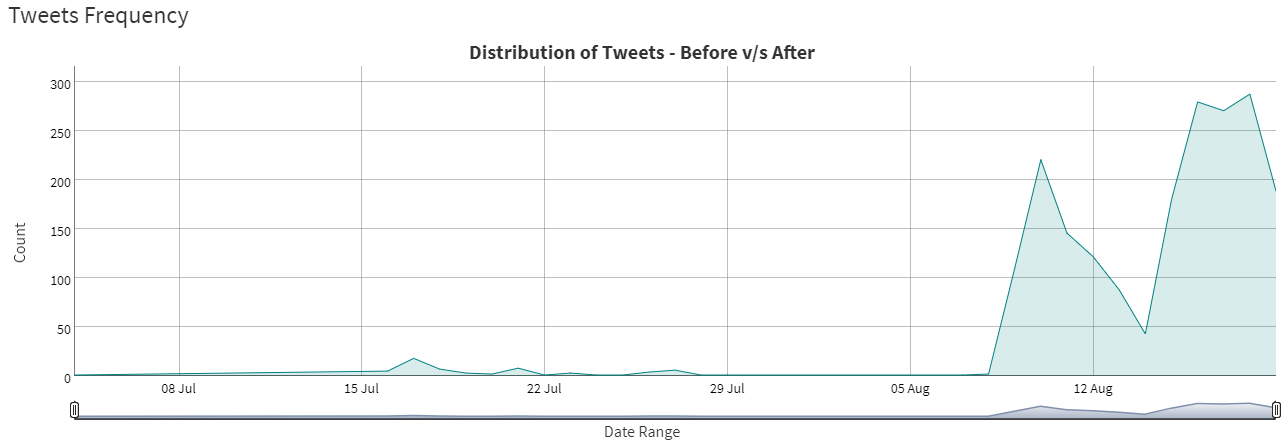
Now that the geographical coordinates for all the tweets have been obtained, the tweets which do not have a location are discarded. Hence the present data has raw tweets along with the parsed location details. Through these coordinates, each tweet can be plotted on a geographical heat map. These heat maps are an interactive way to identify the regions, where something has happened and demonstrate areas of low and high density. These densities are attributed to the point of discussion in the data. Here, since the current use cause is about dealing with data on Kerala Floods, the density shown in the Geographical heat map portrays the areas affected. Wherever the density is high, those areas are highly affected. On the contrary, in the areas where there is no damage, there is low density. The heat map has been implemented with the help of the “leaflet” library present in R. The shape files for drawing the district boundaries are extracted from the census data. The Government of India devises this data on a periodic cycle.

Figure 7 displays the implementation of a geographical heat map for the Kerala floods. It is to be noticed that there are varied densities that show the scale of damage for various regions. During the time of floods, Idukki was the most affected district, according to the PDNA report [39]. This gives an important confirmation that information gathered from social media is in sync with official information.



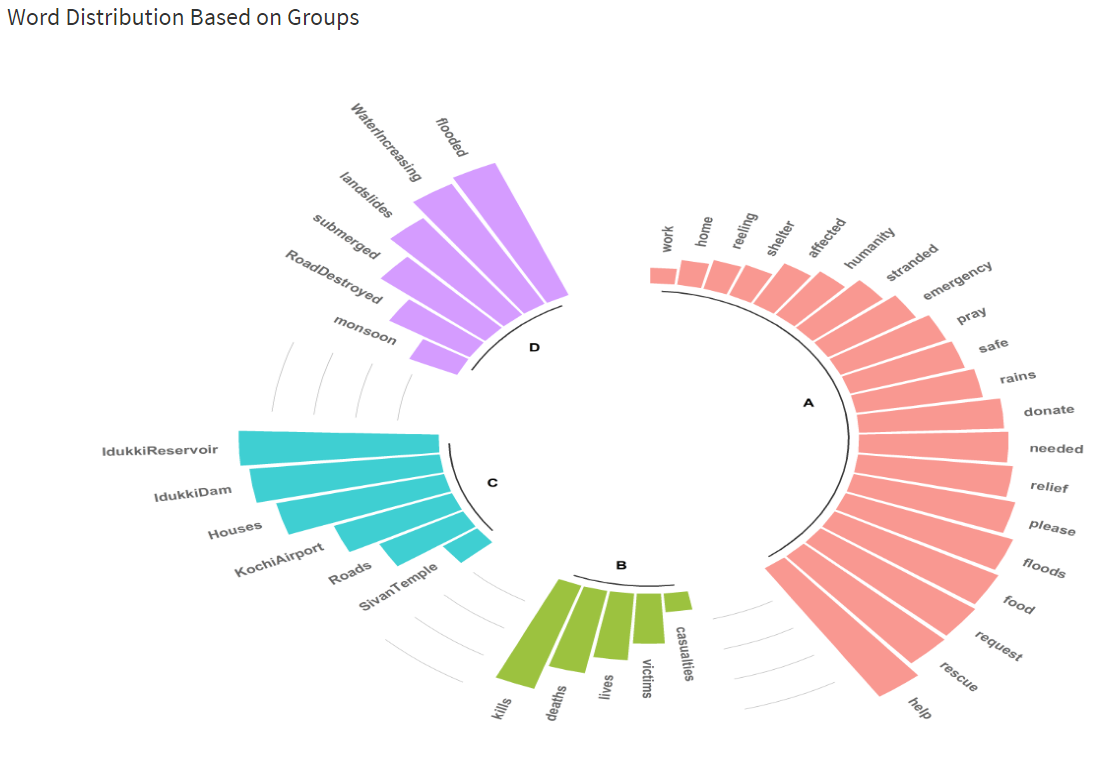
**Fig. 7: Geographical Heat map for Kerala Floods**

1. ***Disaster Distribution Analysis:***This module pertains to performing descriptive analytics on the historical text data through different visualizations to extract insights and patterns from it. These outcomes can be furnished through visualization. Visualization is the graphical representation of data through visual elements like charts and graphs. Data visualization tools provide an easy and convenient way to see and observe trends, outliers, and patterns in data. Figure 8 shows the frequency distribution of tweets based on timing.The tweet count is spread across the time axis. Based on the rainfall and disaster information, they have been classified as ‘Before the Disaster’ & ‘After/During the Disaster.’ As the floods started wreaking havoc, the quantum of tweets shot up exponentially. But on further noticing the small peak generated before the floods started, it projects that it could have been utilized to prepare for the disaster beforehand. Hence, the descriptive analytics section of this tool has been used for both pre-disaster and post-disaster analysis.



**Fig. 8: Frequency Distribution of Tweets Based on Timing**

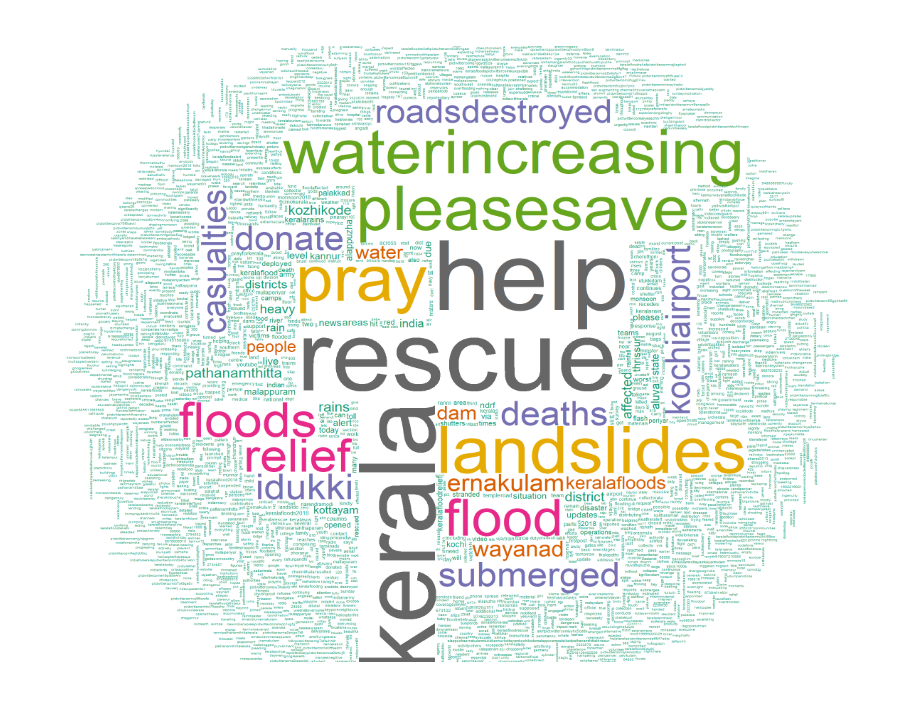
1. ***Word Analytics:***The processed text content is passed through the word analytics engine to obtain granular insights into people’s needs and opinions. The word analytics engine comprises two major text visualization techniques, namely circular bar plot and word cloud. Insights gathered from this module dispense information on the most frequent words used by people. This information has been used to identify the ground-level requirements and needs. Figure 9 shows the outcome of word analytics based on the most frequently used words in the tweets. The distribution of words is split across four categories where A corresponds to people’s requirements, B corresponds to the loss of human life, C corresponds to prominent objects, and D corresponds to the damage caused by the disaster.



**Fig. 9: Circular Bar Plot based on Frequently Used Words**

The separation of the most frequently occurring words into four different categories aids us in having a more granular approach toward the requirements of posted tweets. Having a close look at the category ‘A’, which projects information only about people’s needs, there are a plethora of insights garnered from it. Words like ‘help’, ‘please’, ‘request’ etc. go a long way in explaining the desperate state of the people suffering from the disaster. At the same time, focusing on the other categories, especially ‘D’, tells more about the infrastructure damages reported through Twitter posts. The more the data is passed to this word analytics engine, the better the insights achieved through it.

The findings achieved through the circular bar plot have been validated through a word cloud module. Figure 10 exhibits the outcome of word cloud based on the most frequently used words in the tweets. Word clouds or tag clouds are detailed representations of word frequency that give higher prominence to words that appear more frequently in a document. They display a set of words in the form of a cloud. The more frequent a word appears in the document, the bigger it is projected. Thus, through a visual look at the cloud, identification of the topics on which the documents are commenting on is accomplished. At times, there is also a necessity to perform the word analysis on a comprehensive level without any categories. In contrast to the circular bar plot where the data is divided across different headers, word cloud produces the most frequent words on an overall level. There is a clear similarity between the results obtained through the circular bar plot and the word cloud. Words like ‘landslides’, ‘help’, ‘rescue’, and ‘idukki’ noticed in the circular bar plot are also projected in the word cloud. Since word clouds are both visually appealing and easily understood by all, they are predominantly utilized in word analytics.

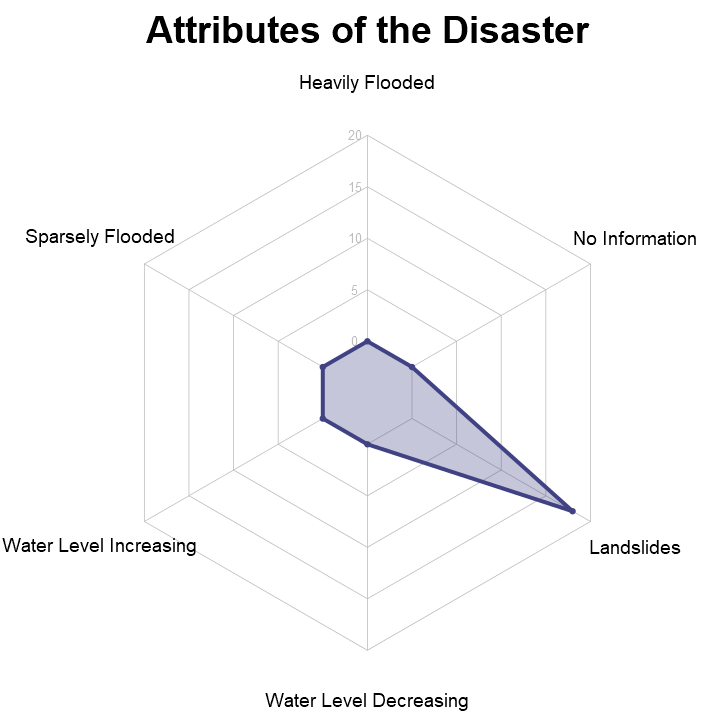
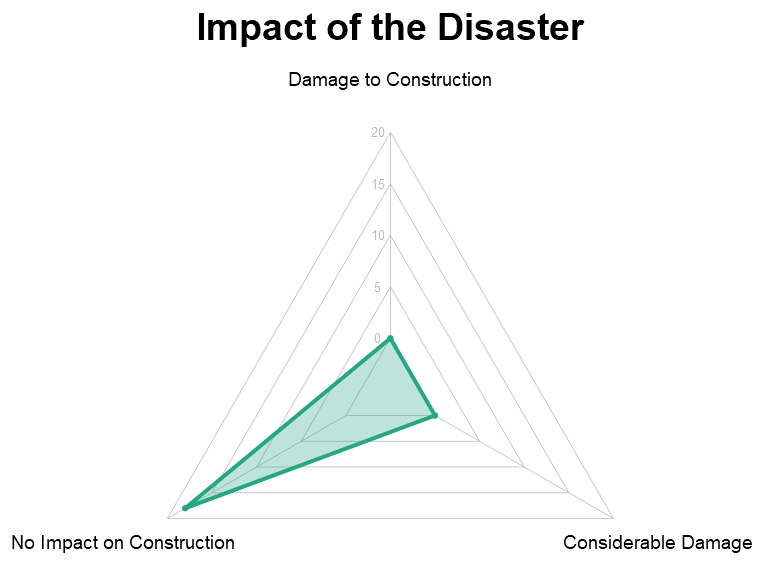
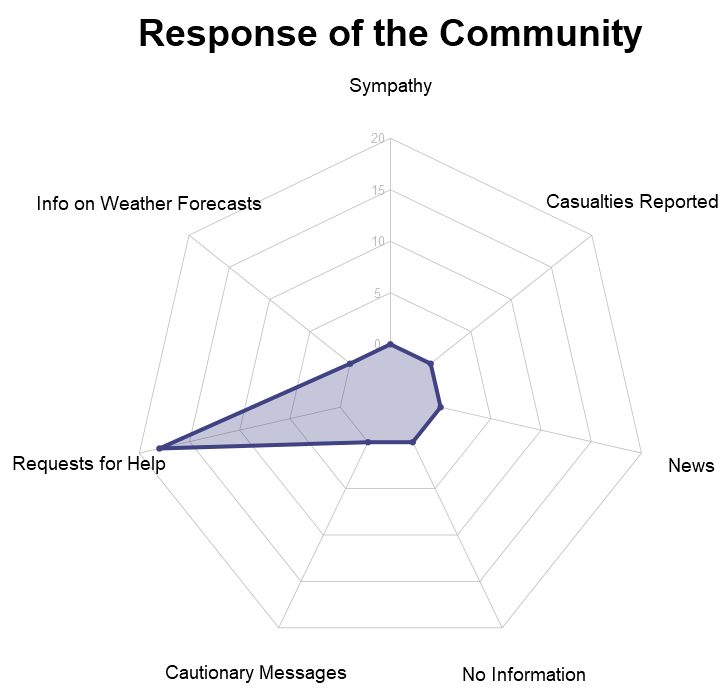


**Fig. 10: Word Cloud Representation Based on Processed Text Data**

1. ***Users’ Emotion Analysis:***Each processed data which is in the form of vector is fed to a Natural Language Processing (NLP) model built out of text classification algorithm. Text classification, also known as text categorization is the process of classifying the text into organized groups. The text classifiers analyze text on its own and then assign a set of pre-defined tags or classes based on its content. Considering the current use case of Kerala Floods, three classifiers have been built with the extracted twitter data. Each of these Machine Learning models would categorize every single incoming tweet based on Attributes of the Disaster, Response of the Community & Impact of the Disaster respectively. The model behind each of these text classifiers is Support Vector Machine (SVM).

SVM is a discriminative classifier explicitly defined through a separating hyperplane. It is used for both classification and regression. In general, it is widely leveraged for classification purposes. In supervised learning, if labeled training data is provided, the algorithm yields an optimal hyperplane which classifies new examples. For a two dimensional space, this hyperplane is a straight line dividing a plane into two parts where in each class lay on either side. Figure 11, shows the classification of tweets based on three essential characteristics. Each attribute contains multiple classes. The text classification outcome is projected in the form of radar chart. Whenever a new incoming tweet is encountered, it is passed through 3 distinct SVN classifier models trained using historical tweet data. The classes for incoming tweets are predicted with an accuracy of around 80%. This accuracy largely depends on the historical data and the amount of pre-processing applied to it. The first model predicts classes for Attributes of the Disaster whereas the second model forecasts categories for Community Resilience and the third model is for predicting classes of impact Caused by Disaster.

The three radar charts shown in figure 11 provide a clear introspection of the predicted outcomes. The most important objective of our analysis is to predict the future outcomes and provide insights in preventing untoward incidents during a disaster. The first radar chart of the lot dispenses the class for the incoming tweet based on disaster attributes. The classes for the first model have been meticulously formulated to identify the nature of the floods which contributes to effective analysis. And the presence of three different radar charts instead of a single one, adds more to the easier interpretation of the results. Similarly, the next two plots play a vital role in dishing out details on how the community is responding to the incident and the impact of the disaster on the construction industry. All of these classes are relatively easier to understand and when integrated with radar chart, makes it suitable for real-time analysis.

**Fig. 11: Classification of Tweets Based on Different Characteristics**

1. **DISCUSSION**

The major contributions of this work are: (1) Analysis of people’s needs and opinions during floods by applying text-classification on the tweet content (2) Illustration of geographical distribution of floods over affected districts within a given time frame merely through tweets (3) Frequency distribution of tweets over a certain period (4) Effects of disaster on infrastructure and community resilience. Also, this involves a multi-disciplinary scientific approach to determine the seriousness of failure tactics with a fusion of probabilistic risk analysis for a potential disaster and vulnerability assessment, with the help of tweets related to the Kerala flood, 2018. The study demonstrated the possibility of assessing the community and infrastructure's resilience using social media data, including texts and geotags. The text of the tweets provides more details about the precise locations of the flood affected regions. The recovery protocols will provide an opportunity to improve the operating of the community so that the probability of future disasters can be overseen.

***v) Advantage of Twitter Data:***

Twitter data can provide a wealth of information for various purposes, including:

Social listening: Twitter data can be used to monitor and analyze public opinion on a particular topic, brand, or event.

Market research: Companies can use Twitter data to gain insight into consumer behavior, preferences, and sentiment towards their products or services.

Real-time events: Twitter data is generated in real-time, providing up-to-the-minute information on breaking news, natural disasters, and other events as they happen.

Sentiment Analysis: Twitter data can be used to analyze the sentiment of tweets, which can be used to understand how people feel about certain topics, brands, and products.

Brand reputation management: Companies can use Twitter data to monitor and respond to mentions of their brand, as well as identify and address any potential reputation issues.

Influencer marketing: Twitter data can be used to identify and engage with influencers in a particular industry or niche, who can then be used to promote a brand or product.

Overall, Twitter data can be used to gain valuable insights into a variety of subjects, making it a valuable resource for businesses, researchers, and individuals.

***vi) Tools available for twitter data analysis***

There are several tools available for analyzing Twitter data, including:

Hootsuite Insights: A social media analytics tool that allows users to track mentions, hashtags, and keywords across multiple social media platforms, including Twitter.

TweetDeck: A social media management tool that allows users to track mentions, hashtags, and keywords on Twitter, as well as schedule and publish tweets.

Brand24: A social listening tool that allows users to track mentions, hashtags, and keywords across multiple social media platforms, including Twitter.

TweetReach: A social media analytics tool that measures the reach and impact of tweets, hashtags, and accounts on Twitter.

Keyhole: A social media analytics tool that allows users to track mentions, hashtags, and keywords on Twitter, as well as monitor campaigns and measure ROI.

Twitter Analytics: Twitter's own analytics tool that provides detailed information on tweets, followers, and audience demographics.

R or Python: Both of these programming languages have a lot of libraries that are capable of scraping twitter data and perform analysis on it.

*Vii) Successful Twitter prediction*

Twitter prediction analysis refers to using data from Twitter to make predictions about future events or trends.

There have been several successful examples of Twitter prediction analysis in various fields, including:

Stock market prediction: Researchers have used Twitter data to predict stock prices, with some studies reporting accuracy rates as high as 86%. This is done by analyzing tweets related to a particular company or industry and using sentiment analysis to determine if the overall sentiment is positive or negative.

Political election prediction: Researchers have used Twitter data to predict the outcome of political elections, with some studies reporting accuracy rates as high as 90%. This is done by analyzing tweets related to the candidates or parties and using sentiment analysis to determine which candidate or party is more popular among Twitter users.

Disaster prediction: Researchers have used Twitter data to predict natural disasters, such as floods and earthquakes, by analyzing tweets related to the disaster and using machine learning algorithms to identify patterns.

Public health prediction: Researchers have used Twitter data to predict the spread of infectious diseases, such as influenza, by analyzing tweets related to the disease and using machine learning algorithms to identify patterns.

These are just a few examples of how Twitter data has been successfully used to make predictions in various fields. However, it's worth noting that while these studies show promising results, they also have some limitations, such as the need for large amounts of data, and the difficulty in handling the noise and bias present in the data.

***vii) Twitter Data Limitation***

Twitter has several limitations on the amount of data that can be retrieved through its API, including:

A limit of 3,200 tweets per user for the user timeline method.

A limit of 7 days of tweets for the search method.

A limit of 450 requests per 15-minute window for the application-only authentication method.

A limit of 900 requests per 15-minute window for the user authentication method.

These limitations are put in place to prevent abuse of the API and ensure that the service remains stable and available for all users. Additionally, some tweets are not available for access due to privacy and security reasons.

**CONCLUSION**

In this paper, a four-step framework has been implemented to analyze the tweets posted during natural calamities to identify the ground requirements and needs of the people from the affected area. The descriptive analytics section assists in performing a post-disaster analysis whereas the AI/ML module is used for real-time data analysis. The descriptive section comprising of charts like Heat map, Dygraph, and bar plots furnish a lot more crucial information about the disaster. These results obtained from post analysis and examination of tweets about the ﬂood could be used for managing and planning for various relief measures. The AI/ML section deals with predicting the classes for the upcoming data through the Machine Learning models. The predictions from the text classifiers have been plotted with the help of radar charts for effective visualization. These outcomes acquired through the radar chart clearly demonstrate the attributes of the disaster and the community’s response in retaliation coupled with infrastructure resilience. further scope of this particular work can be planned to focus on two important aspects. Improving the machine learning model’s accuracy from 75% to near 90% by implementing more sophisticated algorithms like XGBoost and Stochastic Gradient Descent. This will increase the accuracy of the text classifier resulting in better prediction of classes for all the three text classifiers. The next focal point is to deploy the model in production and subjecting it to the streaming data with the help of Big Data technologies like Spark Streaming, Kafka, etc.

**REFERENCES:**

1. J. Ninan, A. Mahalingam., (2017) “Stakeholder Management Strategies in Infrastructure Megaprojects – A Dimensions of Power Perspective,” *Engineering Project Organization conference*.
2. J. C. J. H. Aerts, W. J. W. Botzen., (2014) “Evaluating Flood Resilience Strategies for Coastal Megacities,” *Science*, vol. 344, no. 6183, pp. 473–475.
3. R. Sutton and R. Haigh., (2011) “Private Construction Sector Engagement in Post-Disaster Reconstruction,” *Post-Disaster Reconstruction of the Built Environment*, pp. 192–207.
4. J. C. Matthews., (2016) “Disaster Resilience of Critical Water Infrastructure Systems,” *Journal of Structural Engineering*, vol. 142, no. 8.
5. J. S. Bevington, A. A. Hill., (2011) “Measuring, Monitoring, and Evaluating Post-Disaster Recovery: A Key Element in Understanding Community Resilience,” *Structures Congress.*
6. M. A. Cameron, R. Power., (2012) “Emergency situation awareness from twitter for crisis management,” *Proceedings of the 21st international conference companion on World Wide Web - WWW 12 Companion*.
7. Le Masurier, J. Rotimi., (2006) “A comparison between routine construction and post-disaster reconstruction with case studies from New Zealand*” Proceedings of the 22nd Annual ARCOM Conference, Association of Researchers in Construction Management*, pp. 523-530.
8. F. Mallick, (2013) “Habitat and Infrastructures: A Localized Approach to Resilience,” *Climate Change Adaptation Actions in Bangladesh Disaster Risk Reduction*, pp. 331–340.
9. Karimiziarani, M., Jafarzadegan, K., Abbaszadeh, P., Shao, W., & Moradkhani, H. (2022). Hazard risk awareness and disaster management: Extracting the information content of twitter data. *Sustainable Cities and Society*, 77, 103577.
10. Soesilo, P. K., & Rahman, F. (2022). The Pillars of Survival in the COVID-19 Pandemic: The Case of Indonesia. In Community, Economy and COVID-19 (pp. 267-289). Springer, Cham.
11. S. Hiruta, T. Yonezawa., (2012) “Detection, classification and visualization of place-triggered geotagged tweets,” *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp 12*.
12. M. Dredze, M. Osborne., (2016) “Geo-location for Twitter: Timing Matters,” *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1064-1069.
13. William J. Corvey, Sarah Vieweg., (2010) “Twitter in Mass Emergency: What NLP Techniques Can Contribute William,” *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, pp. 23-24.
14. C. Zhang, D. Lei. (2018) “Geo Burst: Real-Time Local Event Detection in Geo-Tagged Tweet Streams,” *ACM Transactions on Intelligent Systems and Technology*, vol. 9, no. 3, pp. 1–24.
15. J. Capdevila, J. Cerquides., (2017) “Tweet-SCAN: An event discovery technique for geo-located tweets,” *Pattern Recognition Letters*, vol. 93, pp. 58–68.
16. -l
17. Jack Shepherd., (2023) “22 Essential Twitter Statistics You Need to Know in 2023,” <https://thesocialshepherd.com/blog/twitter-statistics>. (last accessed 23rd January 2023)
18. Sarker, I. H., (2021) “Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective,” *SN Computer Science*, vol. 2, 377.
19. Ogie, R. I., Clarke, R. J., Forehead, H., and Perez, P. (2019). Crowdsourced social media data for disaster management: Lessons from the PetaJakarta.org project. *Computers, Environment and Urban Systems*, *73*, 108–117.

1. Xiao, Y., Huang, Q., & Wu, K. (2015). Understanding social media data for disaster management. *Natural Hazards*, *79*(3), 1663–1679.
2. Yates, D., & Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management*, *31*(1), 6–13.
3. PEARY, B. D. M., SHAW, R., & TAKEUCHI, Y. (2012). Utilization of Social Media in the East Japan Earthquake and Tsunami and its Effectiveness. *Journal of Natural Disaster Science*, *34*(1), 3–18.
4. Sadri, A. M., Hasan, S., Ukkusuri, S. v., & Cebrian, M. (2018). Crisis Communication Patterns in Social Media during Hurricane Sandy. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(1), 125–137.
5. Kumar, P. (2022). Social Media, Disasters, and Cultural Heritage: An Analysis of Twitter Images of the 2015 Nepal Earthquake. *Visual Communication Quarterly*, *29*(1), 34–46.
6. Page-Tan, C. (2021). The Role of Social Media in Disaster Recovery Following Hurricane Harvey. *Journal of Homeland Security and Emergency Management*, *18*(1), 93–123.
7. Gunessee, S., Subramanian, N., Roscoe, S., & Ramanathan, J. (2018). The social preferences of local citizens and spontaneous volunteerism during disaster relief operations. *International Journal of Production Research*, *56*(21), 6793–6808.
8. Plotnick, L., & Hiltz, S. R. (2016). Barriers to Use of Social Media by Emergency Managers. *Journal of Homeland Security and Emergency Management*, *13*(2), 247–277.
9. Grace, R. (2021). Overcoming barriers to social media use through multisensor integration in emergency management systems. *International Journal of Disaster Risk Reduction*, *66*, 102636.
10. Spialek, M. L., & Houston, J. B. (2019). The influence of citizen disaster communication on perceptions of neighborhood belonging and community resilience. *Journal of Applied Communication Research*, *47*(1), 1–23.
11. Young, C. E., Kuligowski, E. D., & Pradhan, A. (2020). *NIST Technical Note 2086 A Review of Social Media Use During Disaster Response and Recovery Phases NIST Technical Note 2086 A Review of Social Media Use During Disaster Response and Recovery Phases* (Issue April).
12. Inal Onal, E., Tekeli-Yeşil, S., & Okay, N. (2022). The Use of Twitter by Official Institutions in Disaster Risk Communication and Resilience. *Journal of Emergency Management and Disaster Communications*, *03*(01), 25–40. https://doi.org/10.1142/S2689980922500087
13. Phengsuwan, J., Shah, T., Thekkummal, N. B., Wen, Z., Sun, R., Pullarkatt, D., Thirugnanam, H., Ramesh, M. V., Morgan, G., James, P., & Ranjan, R. (2021). Use of Social Media Data in Disaster Management: A Survey. *Future Internet*, *13*(2), 46. <https://doi.org/10.3390/fi13020046>.
14. Karami, A., Shah, V., Vaezi, R., & Bansal, A. (2020). Twitter speaks: A case of national disaster situational awareness. *Journal of Information Science*, *46*(3), 313–324. https://doi.org/10.1177/0165551519828620
15. Panagiotopoulos, P., Barnett, J., Bigdeli, A. Z., & Sams, S. (2016). Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technological Forecasting and Social Change*, *111*, 86–96. <https://doi.org/10.1016/J.TECHFORE.2016.06.010>
16. Lal, P., Prakash, A., Kumar, A., Srivastava, P. K., Saikia, P., Pandey, A. C., Srivastava, P., & Khan, M. L. (2020). Evaluating the 2018 extreme flood hazard events in Kerala, India. *Remote Sensing Letters*, *11*(5), 436–445. https://doi.org/10.1080/2150704X.2020.1730468
17. Rajiv Gandhi Institute. (2018). *A Report on Kerala Flood 2018 – Disaster of the Century, Thiruvananthapuram, Kerala*. https://sdma.kerala.gov.in/wp-content/uploads/2020/08/Rajeev-Gandhi-Centre-Kerala-flood-2018-The-disaster-of-the-century.pdf
18. “Floods and Landslides, August 2018,” *Kerala, Post Disaster Needs Assessment,* pp. 1-440.