PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

EXPLORING MACHINE LEARNING AND DEEP LEARNING APPROACHES FOR DISASTER PREDICTION AND MANAGEMENT: A SURVEY OF DIFFERENT APPROACHES

Hira Farman[1], Noman Islam[2], Affan Alim[3]
[1,2] Faculty of Computing, Iqra University, Karachi Institute of Economics and Technology, Karachi, Pakistan
[3] Department of Computer Science, Mohammad Ali Jinnah University, Karachi, Pakistan

[*]Corresponding author: hira.farman@iqra.edu.pk

Abstract: **Natural disasters affect both human and animal existence everywhere in the world. Not to mention seriously damaging land, wildlife, etc. Thousands of people worldwide lose their lives to natural catastrophes every year, including landslides, cloudbursts, heat waves, storms, tsunamis, floods, earthquakes, and wildfires. On the social networking site Twitter, individuals can exchange thoughts, news, and personal narratives. Real-time data is widely available, and many service agencies routinely use it to identify emergencies, reduce risk, and save lives. However, it has been suggested in numerous studies to provide words in forms that computers can understand and, on the basis of word representations, use machine learning techniques to accurately determine the meaning behind a post. This is because humans are unable to manually filter through the vast number of records and spot hazards in real-time. By posting risks associated with disaster events on social media, the community can keep an eye out for disasters, which has been crucial for emergency preparedness. This study looks at the potential use of the social media site Twitter for disaster-related research. It focuses on the most recent methods for disaster prediction, deep learning, and machine learning. Another objective of the effort is to have a comprehensive understanding of the various data types and their sources in connection to a range of jobs and crisis management scenarios. Furthermore, the study intends to provide a thorough analysis of the different data mining techniques applied to address diverse problems associated with natural disasters in addition to thorough instructions on how to classify tweets as "Related to Catastrophe" or "Not related with Catastrophe" using natural processing techniques.**
Keywords: **Tweets, machine earning, deep learning, big data, natural disasters, NLP**

I. INTRODUCTION

On Earth, natural catastrophes are a significant, harmful, and unbearable phenomenon that frequently occurs [1,2,3]. Disasters, both natural (the power of nature) and man-made (accidental, intentional), have an impact on the lives of millions of individuals, including both humans and other animals. Besides, human lives are lost, with significant implications on infrastructures in terms of property and political stability [4,5]. Exposure to a natural disaster rises in recent months in children under the age of five by 9-18%, increasing their likelihood of getting serious illnesses such as diarrhea, temperature, and extreme respiratory disease. This is in addition to the immediate effect seen. The scope and pattern of these effects are directly related to the socioeconomic condition of the households. The disasters also have significant effects on business establishments as already discussed. Disaster management and monitoring effectively is a global challenge. Both natural and man-made tragedies can affect any community. A disaster is described as an unplanned and frequently sudden event that results in significant damage, destruction, death, and suffering and calls for external assistance at the national or worldwide level. Flood, wildfires, hurricanes, storms, chemical and oil spills, terrorist attacks, nuclear accidents, and other natural and man-made calamities can all result in disasters [6 ,7]. Damage to infrastructure, crops, and housing are just a few examples of how a disaster can affect the local economy directly. Other indirect effects include revenue loss, job loss, and financial distress. Catastrophes continue to take place and the death rate has remained high over the past few years despite various efforts by safety professionals and governmental organizations [8]. The fundamental issues facing force protection, police, public health, fire departments and many other government entities in control of disaster regulation may involve the gathering and processing of vast amounts of information about disasters. In order to respond and cooperate effectively, these entities must process current catastrophe information as rapidly as feasible [9,10]. Twitter has grown to be one of the most widely used microblogging platforms on the internet. More than 400 million tweets are posted daily around the world, and this idea has proven to be very popular. Twitter is becoming a valuable source of freely and publicly provided geographic information. While a large portion of this data is noise made up of meaningless chit-chat and dialogue, other tweets try to engage people in citizen journalism by describing and reporting on their surroundings. Social media communication is done, among other things, to spread awareness, show sympathy, show damages using comments or pictures, talk about different causes, and offer or request help. Another important, but inevitably understudied, category of tweets generated after natural disasters are those that are truly urgent and signal life-threatening situations [11,12,13]. Twitter is an essential tool for exchanging information in times of disaster is social media, which allows for real-time feedback and identification. The benefit of social media over conventional news medium is the capability to gather immediate input from individuals who are affected. This novel bidirectional communication channel can help relief agencies update the public and learn from situational updates provided by impacted individuals. In the context of this, obtaining crisis information through social media posting (such as tweets) can greatly improve situational awareness and lead to quicker reactions [10]. Because of the widespread availability of streaming information, numerous rescue agencies regularly examine these data to identify emergencies, lower the risk, and save lives. Humans, however, are unable to systematically sort through the massive volume of data and spot hazards at the moment. This can be accomplished by providing words in computer-understandable formats and using machine learning techniques to accurately detect the precise meaning and sentiment of text posts. Most of the disaster-related tweets utilized in machine learning classification tasks have been used to determine whether or not the tweets are connected to catastrophes [14, 15]. This information is begun to be used as a base for classifying, monitoring, and examining the characteristics of both natural and man-made catastrophes. Currently, the research being done to determine how disaster-related tweets might be found for awareness, help, management and prediction. In the study [16], the major objective is to give a quick evaluation of losses caused by disasters.

Disasters are classified according to their nature and the intensity of their effects [1-40]. Figure 1 illustrates a range of prevalent disaster types frequently encountered across various countries. Disasters typically have both immediate and long-term effects such as property destruction, long-term psychological impacts, plantations, disruption in lifestyle, disturbance of traffic, loss of life and employment for individuals, breakdown of vital utilities such the supply of electricity, water, and gas, Economic destruction on a national scale, communication network disruption ,viruses and illness spread [17,18,19].

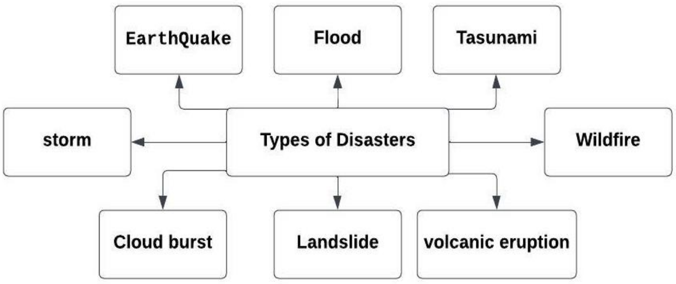

Figure 1: Types of Disaster

An earthquake is a rapid, destructive movement of the earth crust by violently releasing molten beneath the surface of the planet [20]. An abnormally large amount of water spilling over dry land. As a result of inadequate planning and false meteorological predictions Each year, millions of people, animals, and agricultural goods suffer in

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

Pakistan and India specifically [21]. The word "landslide" refers to both the geological formation that results from the downward motion of soil, rock, and other elements caused by gravity. A landslide is a sudden fall of the earth or a mass of rocks from a mountain or cliff caused by movement of the earth surface. Landslides are common in India &Western Ghats and northern sub-Himalayan region [22,23].A massive amount of water is thrown into the air or the ocean by an earthquake, volcano eruptions, or even other underwater explosions [24].Severe weather, like rain or snow that occurs because of strong winds or air currents and is driven on by sudden changes in air pressure over the earth surface [25,26].There is an abrupt, powerful eruption from the surface of the earth of lava, molten rock, ash, steam, fumes, or vapors that is ejected to great altitudes and spreads for numerous kilometers [27,28]. This is a very severe type of unexpected rainstorm that includes hail, thunder, and excessive rainfall that passes quickly. As a terrible result, there was a flash flood in Northern India during 2013 that killed tens of thousands of people and animals [29,30,31]. Burning bushland is referred to as a "bushfire ". They are a type of wildfire that ravages uncontrolled areas of the environment, including bushel, grasses, bushland, and woodlands. These flames are uncontrolled and challenging to put out [32,33]. Millions of users worldwide utilize the online social network (OSN) known as Twitter. It makes it possible for people to keep in touch with their friends, family, and coworkers. With the development of technology, utilizing mobile devices like iPhones and iPads to access Twitter has gotten simpler. It allows users to submit tweets, or messages of 140 characters or fewer, on its platform. The act of reposting a tweet that has previously been sent by another user is known as retweeting. This is comparable to email forwarding. Due to its numerous characteristics, Twitter has established itself as a valuable resource in the field of data mining. Using hashtags, Twitter users can also search for or filter out posts that interest them. Additionally, it provides user security by letting them choose whether to post tweets openly or privately. Most of the time, people talk about common personal experiences, but on occasion, they send messages that can be mined for useful information. These events could include politics, blockage, violence, fires, disasters, thunderstorms, and other disasters etc. [34-37]. However, the fact that information is disseminated in real time over the Twitter network is the most crucial aspect of this study. When 80% of consumers access content on a mobile device, which can provide us with precise geolocation and more recent information, it becomes even more helpful. When conventional methods of communication became ineffective in 2011, after a disaster in Japan. Twitter was used as a means of communication. Twitter and other social media are becoming more popular tools for disaster situational awareness. Using tweets, it is simple to determine the different sorts of disasters. Figure 2 illustrates various categories of Twitter tweets that contribute to analytics and their effective management.

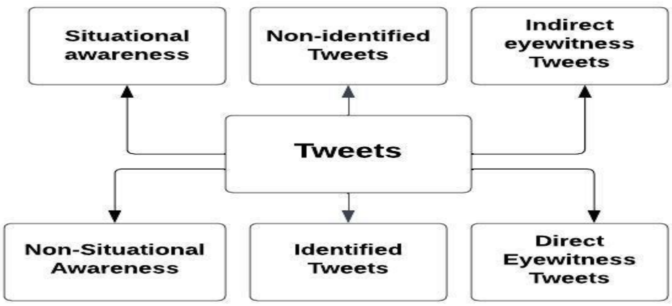

Figure 2: Classification of Tweets

Table 1 shows a few situational knowledge and non-situational knowledge tweets. A number of research articles have been published recently on the general subject of deducing situational & noninstitutional awareness from a post-disaster Twitter account. Twitter and other social media sites make it easier to be informed of time-sensitive situations by disseminating news about casualties and damage, rescue operations, and alerts, as well as by providing relevant information to the population most affected by disasters and first responders who may find it useful for instance, in order to better understand the speed and impact of an earthquake, geotagged tweets have been used in particular [38].

Table 1: Examples of the situational and non-situational tweet fragments

Situational Knowledge	
Tweet No	
1.	Oh my God, what is happening?
2.	Oh my Heavenly Father, a tragedy at a public school resulted in the deaths of 19 children and 19 adults.
3. Non- Situational Knowledge	
4.	My thoughts and prayers are with all of the families impacted by the New England public school.
5.	I'm coming to support the victims of the Hyderabad bomb and their families, according to the Hindustan Times.

Table 2 shows some identified and unidentified tweets during a disaster. Twitter is a free social networking site where users may post short articles known as tweets [20]. These tweets could contain links, text, videos, or images. Identified tweets are those tweets easy identification to events not containing sentiments and meaningless information.

Table 2: Examples of the identified and unidentified tweets

Identified Tweets	
Tweet No	
1.	Flood Disaster at the peak in Pakistan in the year 2022
2.	Need Help Flood
Non-Identified Tweets	
3.	# Ooooooo, Ahhhh
4.	I am happy but :(

Table 3 shows some direct and indirect eyewitness tweets provided by eyewitness and non-eyewitness reporters, particularly when a crisis event is still continuing. Most first reports originate from eyewitnesses and observers, i.e. or those who directly view an event [2]. Eyewitness accounts are preferred over other sources of information from the viewpoint of the information seeker (an impacted person or institutional response agency) (e.g. Individuals outside the catastrophe area, for instance) such as non-eyewitness. Indirect eyewitness accounts were also shown to be either regarding feelings of relief or fear. The destruction, safety, or missing people/property were other topics included in indirect eyewitness reports [39].

A. Purpose of the Study

- This research aims to explore the applications of tweets in the context of disaster-related studies.
- It seeks to investigate the algorithms and techniques currently employed for predicting disasters.
- Furthermore, it aims to provide a comprehensive understanding of the different tweets data types and their sources in relation to various tasks and crisis management scenarios.
- Additionally, the research endeavors to present a detailed summary of the diverse data mining techniques utilized for addressing different challenges associated with natural disasters as well how to make tweets clean by using natural processing technique.

This work's remaining portions are arranged as follows. There are two portions in the literature. The sources and features material used for the disaster assessment are included in the first section, and a thorough literature analysis is included in the second. In Section 3, the suggested methodology is covered. The findings and discussion of comparing several deep learning and machine learning algorithms for catastrophe earlier prediction based on tweets are presented. In section 4, the forms of twitter data and a detailed description of disaster management and avoidance mitigation are provided. A conclusion and suggestions for the future are given in Section 5.

Table 3: Instances of fragmented tweets in the form of direct and indirect eyewitness accounts.

Direct Eyewitness attributes		
Tweet No	**Attributes**	**Examples**
1.	Adjectives and pronouns of the first person	we, me, I,huhh
2.	Individualized locations	My workplace, the neighborhood
3.	Words that indicate time [2]	#at this time, now
4.	Words that express the senses of perception	listening, perceiving, and touching
5.	Mentioning the sites of disasters	route, region name, and street name
Indirect Eyewitness attributes		
1.	Statement of places or individuals the author is familiar with	parents and hometown
2.	Manifestation of feelings	ideas, anxiety, and relief
3.	Reporting lost, damaged, and safe	safe, missing

II. LITERATURE REVIEW

The literature contains two sections. The first section contains the sources and features information used for the assessment of disaster and the other section contains a detailed comprehensive review.

B. Source Table

Table 4 highlights 40 of the many useful and practical research papers that were read and reviewed in order to discover the different kinds of disasters and their effects using the Twitter dataset for categorizing disasters using machine learning and deep learning algorithms.

C. Feature Table

In an effort to improve the quality of the research, many research articles that provided a list of factors to be considered for catastrophe prediction were analyzed [41-44]. The characteristics in Table 5 were obtained from a previous literature review and the resulting synthesis. It perfectly exemplifies a number of important contributions [47-50].

D. Related Work

In this section, review of prior research and performed an accurate and algorithmic analysis of a number of catastrophes. In this study [45], the author examined four test datasets, especially those related to the Maria hurricane, the California fire, the Philippines floods, and. The authors presented a new disaster-related dataset of tweets that were hash tagged using a semi-automated lexicon-based method. This study [2,46] introduces the Word2Vec and Lexicon-Based Extraction of features Method's Role in Identifying Direct Eye - witness Disaster-Related Tweets. The feature extraction approach is the main focus of this study's efforts to address large dimension data problems. To create additional features, the author suggests a hybrid strategy that combines lexicon- and word2vec-based feature extraction. The experiment's findings demonstrate that the suggested approach, which uses 150 characteristics and an average AUC rating around 0.84, enhances classification performance. The authors of this research [3] develop a machine learning model that can determine whether or not a person or region is in danger. Dataset from the NLP with Disaster Tweets competition, tracked by Kaggle. The dataset includes 3,263 data for the test set and 7,613 data for the trained set. All of the samples are categorized as either without disaster (without disaster (0)) or with disaster (with disaster (1)).

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

Table 4: Source Table

S#	Source Reference
S1	[1]" On Identifying Hashtags in Disaster TwitterData",2020.
S2	[2]"Automatic identification of eyewitness messages on twitter during disasters",2020.
S3	[3]"Disaster Tweets Classification using BERT-Based Language Model",2022.
S4	[4]"Event Detection using Twitter: A Spatio-Temporal Approach",2014.
S5	[5]"An AI/ML-Based Strategy for Disaster Response and Evacuation of Victims in Aged Care Facilities in the Hawkesbury-Nepean Valley: A Perspective",2022.
S6	[6]"Human Sentiment and Activity Recognition in Disaster Situations Using Social Media Images Based on Deep Learning",2020.
S7	[7]"An Innovative Flood Prediction System Using Improved Machine Learning Approach",2019.
S8	[8]"Similarity-based emergency event detection in social media",2021.
S9	[9]"Machine Learning in Disaster Management",2022.
S10	[10]"A stacked convolutional neural network for detecting the resource tweets during a disaster",2021
S11	[11]"On Identifying Disaster-Related Tweets:Matching-based or Learning-based",2017.
S12	[12]"Social Media Data and Post-Disaster Recovery ",*2019*.
S13	[13]"Machine-learning methods for identifying social media-based requests for urgent help during hurricanes",*2020*
S14	[14]"Automated Machine Learning Approaches forEmergency Response and Coordination via Social Media in the Aftermath of a Disaster: A Review",2021.
S15	[15]"Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data",2022.
S16	[16]"A novel method for identifying the damage assessment tweets during disaster",2021.
S17	[17]"Detecting novelty in social media messages during emerging crisis events",2021.
S18	[18] "Analysis on an Auto Increment Detection System of Chinese Disaster Weibo Text.",2021.
S19	[19]"Automatically Identifying Fake News in Popular Twitter Threads",2017.
S20	[20]"Early Detection of Heterogeneous Disaster Events Using Social Media",2019.
S21	[21]"A Review on Applications of Big Data for Disaster Management",2017.

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

S22	[22]"Social media use in emergency response to natural disasters: a systematic review with a publi health perspective",2020.
S23	[23]" How engaging are disaster management related social media channels? The case of Australia state emergency organizations",2020.
S24	[24]"Disaster and Pandemic Management Using Machine Learning: A Survey",2021.
S25	[25]"Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions",*2020.*
S26	[26]"Analysis and early detection of rumors in a post disaster scenario",2018.
S27	[27]"Rweet Miner: Automatic identification and categorization of help requests on twitter during disasters",2021
S28	[28] "A hybrid machine learning pipeline for automated mapping of events and locations from soci media in disasters",2020.
S29	[29]"AI-based risk assessment for construction site disaster preparedness through deep learning-based digital twinning",2022.
S30	[30]"Flood disaster risk assessment based on random forest algorithm",2021.
S31	[31]"The real-time monitoring system of social big data for disaster management,"2015
S32	[32]"Quantitative analysis of social media sensitivity to natural disasters", 2019
S33	[33]"Detection of areas prone to flood risk using state-of-the-art machine learning models",2021
S34	[34]"Natural Disaster Application on Big Data and Machine Learning: A Review", 2019
S35	[35]"System for Monitoring Natural Disasters using Natural Language Processing in the Social Network Twitter",2016
S36	[36]"A Machine learning approach for rapid disaster response based on multimodal data." ,2021
S37	[37]"A Review on Application of Data Mining Techniques to Combat Natural Disasters",2018
S38	[38]"Artificial Intelligence for Disaster Risk Reduction: Opportunities, challenges, and prospects", 2022
S39	[39"]Machine Learning based Classification of Online News Data for Disaster Management", 2020
S40	[40]"Deep Neural Networks versus Naïve Bayes Classifiers for Identifying Informative Tweets during Disasters", 2018

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930
Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

Table 5: Features table

No	Feature Name	Feature Summary
F1	Data mining	Are data mining techniques (e.g. classification, regression, association, clustering) applied in the study process for disaster prediction? [S1, S37]
F2	Machine learning	Whether the publication's study method utilizes and applies any machine learning algorithm [S13, S24, S34].
F3	Deep Learning	Whether the publication's study method utilizes and applies any deep learning algorithm [S18, S15].
F4	Quantitative Data	Whether a research work on quantitative data.
F5	Qualitative Data	Whether a research work on qualitative data.
F6	Dataset	Whether a publication's study used a dataset for disaster event prediction [S1, S8].
F7	Post disaster	A study that examined post-disaster reaction [S26].
F8	During disaster	Whether the publication's study monitoring resource-related tweets during a crisis and used for disaster analysis [S10, S11, S26].
F9	Real time disaster detection	Whether a publication study works on real time detection of disaster using social networking services [S3].
F10	Risk management	Are evacuation plans and catastrophe risk management discussed in the research papers? [5,30,34]
F11	Disaster prediction	Whether a publication's study work on disaster prediction tasks. [S15, S18, S30]
F12	Disaster management strategy	The extent to which a scientist discusses disaster management techniques in a paper. The importance of social media platforms for disaster management is additionally addressed [S9, S23, S31, S34, S39].
F13	NLTK tool	Using natural language processing (NLTK tool kit) on the social network Twitter, a study may be able to identify tweets about natural catastrophes [S35].
F14	Locations	Does the study technique consider the effects of disasters in various locations and identify locations of disaster?
F15	Social media image	Research papers that particularly address the use of social media picture OR text data to identify disasters [S6].
F16	EDA	Whether the publication uses Exploratory Data Analysis as a research procedure [S4].
F17	Human Sentiment recognition	Whether a publication's study invokes how text about extreme weather events expresses the sentiments and perceptions of people[S6].
F18	space-time scan statistics (STSS), Latent Dirichlet Allocation (LDA)	Is the research methodology used STSS for disaster events detection using Twitter data OR Is the research methodology used Latent Dirichlet Allocation (LDA) method for classifying twitter text for disaster detection[S4]

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
E-ISSN 2224-2333 https://doi.org/10.22555/pjets.v11i1.1004
P-ISSN 2222-9930

F19	Clusters	Whether a publication's study focuses on clustering social media texts for disaster detection and management [S1, S4, S8, S17]
F20	GIS.e ArcGIS10.1	Is the research methodology used ArcGIS 10.1 platform and GIS technology for analyzing disaster risk [30]?
F21	Wiebox text	Whether the paper uses a research approach for analyzing and striving to achieve the Chinese text in order to detect and analyze different types of disasters [18].
F22	English text	Whether the paper uses a research approach for analyzing and striving to achieve the English text in order to detect and analyze different types of disasters
F23	Urdu Text	Whether the paper uses a research approach for analyzing and striving to achieve the Urdu text in order to detect and analyze different types of disasters
F24	UAV and path planning	Is the research methodology used for the safest route to the destination for timely disaster response? [S5]
F25	Matching & Learning method	Whether a study uses learning-based or matching-based methods to detect tweets associated with catastrophes[S11]
F26	TF-IDF	Whether the research methodology uses the information retrieval method TF-IDF (maximum - likelihood document frequency), disaster identification can be done [3].
F27	Hadoop	Whether the methodology processed huge amounts of data using Hadoop or some other framework for research purposes [21].
F28	Mathematical model	Is the research methodology proposing any mathematical model for disaster prediction

In this research [4] Firstly Clusters design of various data in this research and using the STSS algorithm to identify disaster events from tweets or pictures was also recommended, along with the suggestion that the same process might be used for several different types of disasters. Secondly, research should be done to determine whether prospective (STSS) could be used to monitor emergent space time clusters in real-time. Process of Exploratory Data Analysis accomplished. The study in [5] set out to establish a successful way to manage flood risk in order to reduce the disastrous impact of future flood and disaster events in the Hawkesbury Nepean catchment region. In this direction, a novel system for early flood warning, AI and ML-based decision-making about disaster have been advocated. Along with that a novel enhanced disaster prediction, assessment, and response system using UAVs along with path planning have been proposed. The author of the paper [6] introduces a deep learning technique that might be enhanced to distinguish between false and actual disasters using a mixed dataset and to read pictures and express views. Using NLP, ML, and DL, this work may also be done with text datasets and images. Positive 518, Neutral 480, and Negative 2002 are the number of images with sentiment tags utilized in this dataset of 3500 total images. This research [7] main objective is to develop a model that can successfully map water level as a function of climate and predict flood levels at a certain location in the future. The data came from the historical weather data sources Canadian Center for Climate Services, New Brunswick River Watch, and Environment and Climate Change Canada's Weather Statistics. In this work [8-9], the author introduces correlation emergency event identification on social media. This research has conducted a review study rather than using an algorithm to examine how DL and ML approaches have been used in various disaster management operations and improve their effectiveness. According to the authors, tracking and studying flood zones using this technology will help to improve management and planning for such flood-prone areas. The author of this report conducted a survey. None of the dataset and algorithms utilized in this study. In this research [10] the author suggests I-AID, a multimodal technique to automatically classify tweets into different types of information and filter important information from the massive amount of social media data. In this study [11] Identifying Disaster-Related Tweets of earthquake, wildfire, or flood-related or unrelated tweets. Using a Matching based or learning

Volume 11. Issue. 1, PP. 45-73, November 2023
E-ISSN 2224-2333
https://doi.org/10.22555/pjets.v11i1.1004
P-ISSN 2222-9930

based approach. Following algorithm used for this work Logistic regression, PCA Latent Semantic Indexing (LSI), Performance Metrics e.g. Recall and the given methodology is designed for two approaches. Matching Based & Learning Based this research aims to create sentiment maps of the affected areas using the suggested architecture to provide situational awareness in real time during disasters. This study [12] key contribution is to address the existing gap by creating a model that can take the complexities of post-disaster recovery into consideration. In this study [13], the author identified social media-based demands for emergency assistance during hurricanes using machine learning techniques. In order to generalize and recognize urgent tweets across hurricanes and afterwards across any floods, disasters, etc. intend to develop classifiers. GNIP, a data reseller for Twitter, obtained a subset of 2,072,715 unique tweets for this study directly from Twitter. The use of social media for disaster response and coordination using automated machine learning techniques is examined in this research [14]. This study [15] compares the BERT embeddings against more conventional context-free word embedding techniques in order to determine how well they predict disaster using Twitter data. In this research [16] both quantitative and qualitative outcomes are present and this is the Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data. Finding damage assessment tweets during a disaster is beneficial for both victims and relief organizations, according to this research. The majority of earlier studies that have identified tweets during a crisis have focused on situational information, resource availability or demand, infrastructure damage, etc. Only a few studies have been done specifically to identify damage identification tweets. This research proposes a novel technique for locating damage assessment tweets during a crisis. In this research [17], the author presents a technique for locating distinctive subjects in social media posts that surface during a developing crisis. This can be used by emergency responders, observers, and affected parties to track changes in the situation and be made aware of any emerging occurrences as soon as they happen. In the future, the author hopes to be able to incorporate previously published material more effectively, perhaps from sources other than Twitter, such news media. This research [18] stated that consequently, advocating the use of Weibo, the most popular social media site in China, to support the disaster warning system. The aim of this paper is Analysis on and design of an Auto Increment Detection System of Chinese Disaster Weibo Text. The quality of information on social media is a growing concern, according to this study [19], but because to web scale data, it is challenging for experts to assess and correct the majority of the erroneous information, or "fake news," that is prevalent on these platforms. Three datasets were utilized in this experiment: BuzzFeed News Fact-Checking Datasets, Credbank, and Pheme Rumor. In order to improve the situational awareness of emergency services, this article [20] discusses the problem of recognizing crisis related communications on social media. Early Recognition of Diverse Disaster Events The main objective of this study is to use social media. The purpose of this study [21] is to offer a thorough survey of the literature on the use of big data in disaster management. There is no experiment carried out in this study. Using exploratory study, this review's [22,23] purpose is to examine the effects of using social media during natural disasters as well as the question of how engaging social media platforms are involved with disaster management. The study [24] utilized five indices: commitment, viral marketing, involvement, popularity, and usage to evaluate the levels of community engagement by various social media channels. Further examines how modern machine learning (ML) approaches can be used to handle pandemics and disasters. The Internet of Things (IOT), object sensing, unmanned aerial vehicles (UAV), 5G and cellular networks, smartphone-based systems, and satellite-based systems are a few of the technologies that have been used to date to handle disasters and pandemics. But there are limitations and possible research trajectories in this survey. Understanding certain behavioral changes can be useful for understanding how catastrophes affect society and for guiding disaster planning and response. In this study [25], Twitter is used to examine how individuals in Puerto Rico adjusted to prolonged telecommunications outages following Hurricane Maria's destruction in the month of September 2017.This research study [26] proposes an algorithm. For an experiment, tweets on the severe flooding in Chennai, India, between November and December 2015 were gathered. Between December 1 and December 10, 2015, a total of 452,544 tweets were sent, and the data was collected using the Twitter Search Service. The following search terms were used to gather the tweets. The terms ‘Chennai Rains’, ‘Chennai Flood’ and ‘Chennai’. First and foremost, this study [27] properly defines request tweets (also known as rweets) in the context of social networking sites, including their several major categories and subtypes. The detection and classification of rweets are the key contributions of this study. In order to prepare the dataset for rweet identification, two strategies were designed: rule-based and logistic regression. SVM, DT, RF, and AdaBoost were utilized as classifiers for classification. This

Volume 11. Issue. 1, PP. 45-73, November 2023
E-ISSN 2224-2333
https://doi.org/10.22555/pjets.v11i1.1004
P-ISSN 2222-9930

study's [28] goal is to develop and test a hybrid machine learning pipeline to analyze social media posts made during disasters to identify the disaster events that are developing in various regions and correspond to them. To conduct coarse-grained event identification and evaluate the geographic location data from geotagged social media data, previous studies have used machine learning techniques. This research [29] introduces a revolutionary vision-based digital twinning and hazard evaluation paradigm for systematic disaster preparedness in construction sites. Deep learning architectures were used to encode the context of disaster risk in order to identify and examine the traits and effects of potential wind-borne debris in digital twin models of construction sites. This study [30] claims that the increase in natural disasters has enhanced the significance of early flood warnings. The major point of discussion in this study is flood disaster risk analysis using RF algorithms. This research presents a technique of flood catastrophe risk analysis based on GIS (Geographic Information System) for data collection, management, and assessment. The Network Big Board is a real-time monitoring system for social big data for emergency preparedness, according to the author of the article [31]. This approach looks for social big data, specifically tweets concerning disasters on Twitter, analyzes those tweets in real time, and then plots issues and patterns related to those concerns on a map. The theoretical model of emergency decision-making discussed in this article [32] focuses on the fundamental questions that warnings or other danger signs raise for the decision-maker. The theoretical framework offers a structure for combining diverse psychological research results on the consequences of warnings and encounters with danger. The severe and frequently catastrophic repercussions of disasters have been maximized in this work [33] to better deal with contemporary developments in artificial intelligence (AI), particularly in machine learning (ML) and deep learning (DL). The goal of this study [34] is to conclude the review process and appreciate how big data and machine learning are used in the field of disaster management and natural disasters. The outcome of this study provides knowledge and applications of big data, ML, and DL in six domains of disaster management. In this study [35], community organizations and humanitarian organizations are the main subjects of an examination of social media usage in crisis management. Natural disasters are devastating occurrences that shock the entire planet and leave humanity feeling terribly sad and unhappy. The author of this paper [36] discusses the causes, types, and effects of disasters. The author of this study [37] focuses on examining how data mining and analytical approaches have been utilized so far for the creation of disaster management strategies that rely on disaster data collecting, disaster prediction, and catastrophe detection. The report focuses particular emphasis to India, which is among the top five nations in terms of the overall number of lives lost. This article [38] suggests a methodology for predicting the direction that a disaster like a hurricane would adopt by mining real time disaster data collected from Twitter. The author [39] creates a sentiment classifier using streaming tweets about disasters in order to describe consumers feelings during disasters based on their various levels of fear. Finding the situational tweets is a challenging task that is very helpful for both the victims and the assistance organizations. There is a likelihood that a tweet will include both situational and non-situational information. In this research [40] attempts to solve the challenge of detecting the situational tweets using deep learning algorithm. In this study [41] compare the performance of deep neural networks to that of conventional classifiers to better understand the usefulness of deep neural networks This study [42] examines the use of social media in crisis management with a focus on volunteer groups and humanitarian organization. The authors of this [43] research initially propose a deep learning model that combines Attention-based Bidirectional Long Short-Term Memory (BLSTM) and Convolutional Neural Networks (CNN) in order to categorize tweets into several groups. aiming to extract semantic meaning from tweets utilizing global vectors for word representation (Glove) and word vectors. A sensor that uses Twitter to track natural disasters is suggested in this research [44]. The technique uses recurrent Neural Networks and its variations to identify toponyms (identified places mentioned inside the text) in tweets containing information about an event, such as the location of a collapsed building on a specific street or where a person was last seen. In this work [50] applied four machine learning methods (LR, RF, NV, SVM), with as well as without the use of the geo-distance framework, to estimate the location of a tweet based on its lexical content. In this study [51], the novel Advanced Contextual Embedding technique using bidirectional encoder representations Transformers (BERT) was used to predict the disaster related tweets. Text and images from a user-generated tweet are specially used as input modalities in the research [52]and suggested deep learning system. The suggested system's use images and text models and are based on transformers. In one of the researchers, authors focused on tweet classification in emergency response systems. For this purpose, they have employed on a novel transformer based bidirectional attention model. It is obvious

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004
E-ISSN 2224-2333
P-ISSN 2222-9930

that in post-disaster situation, emergency communication lines become overloaded. This leads people to resort to social media to acquire help. In another study [53], authors focus on tweets that are important to first responders during hurricanes are electronically found, categorized, and discovered. The major goal of the research effort is to detect Twitter requests for help using deep learning and artificial intelligence. In order to quickly analyze a catastrophe, this study [54] combines social media data with traditional disaster assessment criteria, disaster bearing carriers (such as exposure and susceptibility), and other elements. Convolutional neural networks were used to automatically categorize text in social media data and train a text classifier. This study's objective is to give a quick evaluation of losses caused by disasters.

III. COMPARATIVE ANALYSIS

Here, comparison of literature evaluation with cutting-edge studies in a table-format.

A. Reference and Features Table

Table 6(a, b) presents the results of the association between the published papers and the relevant features in tabular form. Table (6a) contains 20 research papers and 28 feature connections as well as Table (6b) containing the remaining 20 research papers and 28 same features. It contrasts this survey with various aspects of several study articles to demonstrate similarities, processes, and the connections between the findings of numerous researches. The following table summarizes the investigation to examine several algorithmic and data type parameters for identifying historical disasters.

Table 6(a): Sources and Features

Ft	S1	S2	S3	S4	S5	S6	S7	S8	S9	S 10	S 11	S 12	S 13	S 14	S 15	S 16	S 17	S 18	S 19	S 20
F1	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
F2	N	N	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	Y	Y
F3	Y	N	Y	N	Y	Y	N	Y	Y	Y	N	N	Y	N	Y	N	N	N	N	N
F4	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y
F5	N	N	N	N	Y	N	N	Y	Y	N	N	N	N	Y	N	N	Y	N	N	N
F6	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
F7	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	N	N	Y	N	N
F8	N	Y	N	N	N	Y	N	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	N
F9	N	N	Y	N	N	N	N	N	N	Y	N	N	Y	N	N	Y	N	Y	N	N
F10	N	N	N	N	Y	N	N	Y	Y	N	N	N	N	Y	N	N	N	N	N	N
F11	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y
F12	N	N	N	N	Y	N	N	Y	Y	N	N	N	N	Y	N	N	N	N	N	N
F13	Y	Y	Y	Y	N	Y	N	N	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y
F14	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N
F15	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F16	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
F17	Y	Y	Y	Y	N	N	Y	N	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	Y
F18	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F20	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F22	N	N	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	Y	N	N	N
F23	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F26	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N
F27	Y	Y	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y
F28	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Using deep learning algorithms, pertinent information can be extracted from unstructured data sources of information, particularly when it comes to natural language understanding. This involves determining the salient features of a possible disaster, such as its location, kind, and intensity. In light of tables 6a and 6b Natural language processing (NLP)-based deep learning algorithms are very helpful in certain areas of catastrophe management and prediction. Despite the fact that catastrophe prediction has been done in the past using conventional methods, the complexity and variety of natural catastrophes make it a challenging task.

B. ML and DL Algorithms utilized to identify the type of Disaster from tweets.

The names of the several Machine Learning and Deep Learning algorithms investigated to identify, detect, classify, and predict various types of disasters from tweets dataset are listed in Table 7. Investigating the methods and algorithms currently used for catastrophe prediction is the aim of the research. In the table below, results from comparison accuracy statistics are also provided.

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

Table 6(b): Sources and Features

Ft	S 21	S 22	S 23	S 24	S 25	S 26	S 27	S 28	S 29	S 30	S 31	S 32	S 33	S 34	S 35	S 36	S 37	S 38	S 39	S 40
F1	Y	N	N	N	Y	Y	Y	N	N	N	N	N	N	N	N	Y	Y	N	Y	Y
F2	Y	N	Y	N	N	N	Y	N	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
F3	N	N	N	N	N	N	N	Y	N	N	N	N	Y	N	N	Y	N	N	N	Y
F4	N	N	N	N	N	N	Y	N	N	Y	Y	N	N	Y	N	Y	N	N	Y	Y
F5	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	N	Y	N	Y	Y	N	N
F6	N	Y	N	N	N	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	N	N	Y	Y
F7	N	T	N	Y	Y	N	N	Y	Y	N	N	Y	Y	N	Y	N	Y	N	Y	N
F8	N	N	N	N	N	Y	Y	N	N	N	N	N	N	N	Y	Y	N	N	N	Y
F9	N	N	N	N	N	N	Y	N	N	N	Y	N	N	Y	N	Y	N	Y	N	N
F10	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	N	N	N	Y	N	Y	Y	N
F11	N	N	N	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	N	N	N	N	Y	N
F12	Y	Y	Y	Y	Y	N	N	N	Y	N	Y	N	N	N	N	Y	Y	Y	Y	N
F13	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	Y	Y
F14	N	N	N	N	N	N	N	Y	N	N	Y	Y	Y	Y	N	Y	N	N	N	N
F15	N	N	N	N	N	N	Y	Y	N	Y	Y	Y	N	N	Y	Y	Y	N	Y	N
F16	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	Y	N	N	N	Y	Y
F17	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	Y
F18	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F19	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	Y	N	N	N	N
F20	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
F21	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F22	N	N	N	N	N	Y	Y	Y	N	Y	Y	Y	N	N	Y	N	Y	N	Y	Y
F23	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F24	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F25	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N
F26	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	Y	Y
F27	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N
F28	N	N	N	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	N	N	N	N	N	N

Table 7: ML and DL Algorithms Identification used for Disaster prediction

Ref no	Algorithm	Disaster type	Country	Year	Accuracy	Dataset
[1]	LSTM	Hurricane, Fire, Floods,	Maria, California, Philippines	2020	A maximum F1 score is 92.22%	Compiled publicly accessible datasets that were crawled while disasters were occurring and contained tweets on a broad range of (37 disasters total)
[2]	Random forest	Floods, - Hurricane, Earthquakes Wildfires	Not specific	2020	1-Floods: Eye=0.567, Non-eye=0.8 Don't know=0.74 2-Hurricane: Eye=0.806, Non-eye=0.84Don't know=0.810 3-Earthquakes Eye=0.927, Non-eye=0.857, Don't know=0.7194. Wildfires: Eye=0.795 Non-eye=0.74, Don't know=0.90	The Crisis NLP repository provides the dataset: URL link: https://crisisnlp.qcri.org/.
[3]	BERT	Any disaster	Not specific	2022	F1 = 0.8867:	Kaggle Datasets:3,263 test set and 7,613 training set,

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

[4]	-STSS(Space time scan statistic), -LDA(Late Dirichlet Allocation) -STPM (space-time permutation model)	Any emerging disaster/event from twitter (e.g. London helicopter crash).	London	2014	Result P value London Helicopter Crash:4.0610215 Train Delay A 1.961024	Data is obtained via the Twitter streaming application programming interface (API),
[6]	Deep Learning Model Yolo based sentiment and activity detector	disaster	Not specific	2020	F1 rating Positive: 95.81; Unfavorable: 94.98 positive:95.26	Sentiment Tags Number of Images: 3500 Expressions for human activity use dataset statistics.
[7]	Linear Regression Model	Flood	Not specific	2019	Adjusted R^2=0.625	Data was gathered from historical, Weather Station, Canadian Center for Climate Services, New Brunswick River Observe.
[10]	1-Bert encoder, 2-(textgat)text-graph neural network.	Any disaster	Nepal and Italy earthquakes in 2015 and 2016.	2021	F1W. AVG (BERT=0.5 TEXTGAT=0.2, I-AID=0.69)Jaccard Index(BERT=0.34,TEXTGAT=0.28, I-AID=0.43) Hamming Loss(BERT=0.11, TEXTGAT=0.24, I-AID=0.07)	Datasets: TREC-IS, Train=27,467, Valid=6,867 Test=8,584, Classes=25 COVID-19 Tweets Train=4,844, Valid=1,211, Test=1,514, Classes=12
S[13]	CNN, SVM,MLP	Hurricane Harvey	Houston, U.S	2020	F1 scores on CNN=0.87,	not defined
S[15]	DT, RF, Logistic regression, LSTM, BERT	Multiple disasters	Not specific	2022	Glove=0.79% Bert=0.83%	Twitter dataset from a recent Kaggle competition (NLP)10,876 total tweets
S[18]	SVM, Random Forests, KNN Naive Bayes	Earthquake	Leshan	2021	Classifier are: (i)SVM A:0.912 (ii)RFA:0.899 (iii)KNN A:0.857 (iv)NB :0.874 ,	1500 Weibo messages were gathered both before and after the Leshan Earthquake.,1211 unnecessary negative samples and 289 positive samples

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930
Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

S[28]	(i) BERT for classifying posts with humanitarian categories (ii)graph-based clustering to identify situat ional information	Hybrid machine learning pipelines automatically map out events and locations.	----	2020	Model BiGRU Validation Accuracy (74.475), CNN LSTM Validation Accuracy (71.14%) Test Accuracy (58.09%), DP CNN Validation Accuracy (73.15%) Test Accuracy (63.74%), BERT model's test accuracy (75.37%),	The data set gathered from Twitter during the 2017 Hurricane Harvey in Houston is used to illustrate how the study is applied.
S[30]	Random forest & GIS technology i	Flood	Not specific	2021	According to the results, it is simple to evaluate the internal controls of the risk of flooding when the random forest approach is combined with GIS technology.	Data from remote sensing, geographically fundamental data, and statistical data were all utilized during this investigation.
S[53]	BERT and XLNet	Hurricane Harvey	Not specific	2023	BERT=0.70% XLNet=0.71% CNN=0.53%	The College of Texas at North Texas (UNT) dataset used in this study includes 7,041,866 total tweets
S[54]	Convolutional Neural Network (ANN)	Typhoon “Mangkhut” and Lekima”	China Guangdong and Zhejiang Province	2023	Pre-disaster prevention F1=0.99 Disaster notification F1 =0.90 Reminders and advice F1=0.91	Real time data collected Mangkhut=21,901 (tweets) Lekima=(25,967) (tweets)

C. Description and Categories of tweets on social media.

The study of tweet actions for catastrophe behavior prediction and interpretation is a key challenge from social media. The descriptions that proceed with align to the tweet actions listed in table 8. The tweet behavioural patterns are described in the table 8.

D. Sample of Tweets for Analyzing Multiple Disaster Category

Primary data sources for disaster-related research could be field observations, surveys, interviews, information gathered directly from impacted communities, or information obtained via specialized sensors and equipment. These resources are frequently customized to meet the unique goals and research questions of an investigation. Conversely, Twitter data is typically utilized as an additional or secondary data source. For the sake of helping in understanding the tweet text, Table 9 presents a few tweet examples and a map of their behaviors that correspond with Table 8 tweet actions.

E-ISSN 2224-2333
P-ISSN 2222-9930
Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930
Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

Table 8: Categories of tweets with description

Categories	Tweets action	Description
C1	Asking for support	Tweets from individuals requesting quick assistance or support during or after the tragedy or specific disaster
C2	Negative perception	Tweets from people expressing dissatisfaction, describing difficulties, or talking about tragedies
C3	Positive viewpoint	Tweets from people seeking solutions, anticipating achievement, and sharing their happiness with life
C4	Fear	Tweets posted by those who expressed dread both during and after tragedy
C5	Promoting	Tweets that advertise a company's goods or services or give information about the sector
C6	Informative and useful	Helpful information such as impacted persons, infrastructure damage, availability, and resource requirements are provided through informative tweets.
C7	Non-Informative tweets	On the other hand, ineffective tweets don't offer useful information on victims or humanitarian agencies.
C8	Incident alert	Report of an issue, error defect, mishap or fault
C9	Social Interaction	Awful debate of societal and personal concerns,
C10	Noise	Unreadable tweets or those written in languages other than English or with a high proportion of symbols
C11	Job advertisement	Tweets from individuals that announce the open positions inside the organization
C12	Mention Tweet	A mention is a tweet that contains an @ symbol followed by your username within the body of the tweet; individual or user mentioned someone
C13	Coordinating relief efforts	Tweets include promoting the event and creating awareness of it, accepting donations and giving them, locating ways to assist and listing them, as well as offering information on crisis response.
C14	Transportation	Tweets referencing a method or system for moving people or commodities from one location to another using a ship, plane, or vehicle
C15	Mental Health	Tweets on Disaster mental/behavioral health support: giving and receiving
C16	Crime	Tweets on the crime events
C17	Criticizing the government	Tweets addressing the economic causes, effects, and consequences for events
C18	Delivering good wishes and reminiscing	voicing feelings, worries, well intentions, remembering the losses, and disseminating knowledge on disaster response, healing, and restoration
C19	Examining reasons	covers analyses of the academic, spiritual, and other factors that led to the tweets.
C20	Writing on the matter from a particular, current standpoint	Giving and receiving advice on how to prepare for emergencies as well as specific disaster alerts are included in this usage. This covers specifics regarding one's situation and environment.
C21	Reporting on the state of affairs (second hand reporting)	This element includes providing news coverage of the event, detecting and signaling disasters, documenting and comprehending what is happening in the disasters, and much more.

C22	Other	Other tweets except these categories

Table 9: Sample of tweets and its category

No	Tweets	Category
1.	I enjoy how delightful smoking is.	C3
2.	the foolish experiences I've had over the past 15 days	C2
3.	Rengoku ignites the fire in my heart ♡P.S. I missed this coloring style, so I've included it here:#鬼滅の刃https://t.co/YrUF9g68s0	C10,C22
4.	Before tomorrow, friends, USGS reports that the 7-day aftershock forecast was altered by the latest 5.9 earthquake this morning…	C8
5.	I need help I am stuck in a building near plasma airport# help! please.	C1,C4
6.	Our workshop attendees developed a disaster detection system in pairs by designing and programming it! This team created a countryside fire detection system.	C20
7.	I'm furious because that sounds exactly like you.	C22
8.	Terrorists, War, OH MY GOD, Disaster management: Reach out to management.	C21
9.	Traffic congestion is bad, I guess. A sudden disaster has occurred.	C8,C14,C21
10.	"Boston police: At least 3 died in bombing during Boston Marathon"#bostonmarathon".	C21
11.	"Really saddened by the tragedy: (, may they RIP", #Santamaria.	C18
12.	#OHmY gOD #YAHOO♡ #oH NO.	C10

IV. METHODOLOGY APPROACH

In the today's world, Twitter has been used in emergency situations to coordinate real-time rescue and relief efforts with officials. Machine learning and Deep learning algorithms can be used to predict disasters and assist in disaster management operations. such as figuring out pathways for crowd rescue, examining posts on social media, managing the disaster's consequences and prediction for disaster and disaster related activities. Figure 3 illustrates the methods analyzed in this work for the detection and prognosis of disaster through tweets. This methodology section is divided into four (4) primary sections. Part I discusses the method for predicting catastrophes, Part II discusses data techniques and Part III discusses the stages of text processing. Part IV includes techniques for disaster management.

A. Algorithm Used for Disaster Prediction

Artificial intelligence methods such as machine learning and deep learning are used to predict various types of disasters. Support vector methods (SVM), decision trees (DT), reptree (RF), KNNs (K nearest neighbors), linear and logistic regressions (LRs), K means clustering, and ensembles of linear programming (LPs) are a few examples of machine learning approaches. However, Deep Learning methodologies use Neural Network designs such as Convolutional Neural Networks (CNN), Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory Neural Networks (LSTM), Bidirectional Encoder Representations from Transformers (BERT), Transformers Architecture, and Generative Adversarial Networks (GAN).

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004
E-ISSN 2224-2333
P-ISSN 2222-9930

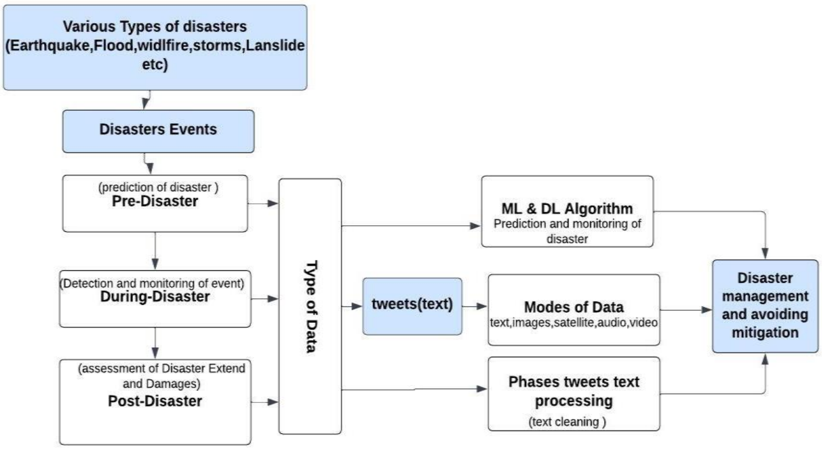

Figure 3: Workflow diagram

Systems that forecast disasters, assist with response and recovery after such events, and lead to the development of valuable decision-support tools can be built using large and complex datasets along with ML and DL. These methods make use of the capability of modifying different data kinds from numerous sources in order to look for patterns that could reveal information that would otherwise be unclear. These are just a few of the numerous big data technologies, which also include wireless sensor networks, satellite imaging, unmanned aerial vehicles (UAVs), social media, crowdsourcing, geographic information systems (GIS), and dataset worldwide websites like Kaggle and the UCI repository. In this study, how to identify disasters using Twitter are already discussed. In this inquiry, a number of algorithms—some of which are listed below—were examined for disaster prediction.

K-Nearest Neighbor (K-NN): The KNN approach is applied to solve classification and regression prediction issues. It belongs to the group of supervised machine learning techniques. The KNN approach predicts the probability of new data points based on feature resemblance which suggests that the new data point's value will rely on how much it resembles the points in the training set. Although there are numerous distance functions, Euclidean is the most widely used one for calculating distance. The equation (i), (ii) and (iii) define distance equations for single and multi-dimensions.

1-dimension:

$$d = \sqrt{[(x2 - x1)^2]} \qquad \ldots(i)$$

2-dimension:

$$d = \sqrt{[(x2 - x1)^2 + (y2 - y1)^2]} \qquad \ldots(ii)$$

For m-dimension:

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \qquad \ldots(iii)$$

Segregate Vector Machine (SVM): Supervised learning techniques include support Vector Machines (SVM).When there are only two classes in the data, it is utilized for binary grouping. SVM can be used to solve specific classification or regression problems. It just draws the hyper plane between 2 classes. The newest data and the class they belong to are predicted using SVC. Take into consideration how far these classes are from the line. The primary method for arriving at a decision is to use linear regression, as demonstrated in Equation (iv). Considering a given sample x,

$$\sum_{i\varepsilon SV} Y_i a_i K(x_i, x) + b \qquad \ldots(iv)$$

Here, the component α_i is referred to as the dual coefficients, and its upper bound is set by C. The independent term b must be calculated. The kernel is $K(x_i, x)$, where x is the input vector.

Decision Tree: The algorithm for supervised learning is the decision tree. It is also incorporated into classification and regression. Using training data, quickly estimation of the class or value of any target variable using decision tree techniques or learning decision rules. Decision trees compute the entropies of groups to determine the purity of such groups. Entropy of a C-class decision tree: The equation (v)can be used to get the entropy:

$$E = -\sum_{i=1}^{N} p_i \log_2 p_i \qquad \text{…(v)}$$

Where: Pi is the likelihood of class. The entropy values range from 0 to 1, where 1 indicates the largest number of impure groups and 0 indicates entire pure groups Information gain, which is a decrease in entropy, is another statistical characteristic of decision trees. Depending on the provided feature values. It establishes the variation in the dataset's entropy between the two splits before and after as in equation (vi).

$$E = -\left(p_r \log_2 p_r + p_p \log_2 p_p + p_y \log_2 p_y\right) \qquad \text{…(vi)}$$

Here, "before" refers to the dataset that was there prior to the split, "k" signifies the number of subsets produced by the split, and "j, after" designates subset j that emerged following the split.

Naive Bayes Classifier: Numerical data can occasionally be classified using the Naive Bayes algorithm. High-dimensional datasets can also be successfully used with the Naive Bayes classifier. By estimating the likelihood of each class based on the feature vector for text classification for continuous large data, the Naive Bayes classifier solves the issues caused by the curse of dimensionality. Bayes theorem is applied in naive Bayes. According to equation (vii), the Bayes theorem, sometimes referred to as Bayes' Rule, must be applied in order to determine the likelihood of a hypothesis given certain prior information. determined using the conditional probability.

$$P(A|B)=P(B|A)P(A)/P(B) \qquad \text{…(vii)}$$

The probability that a particular hypothesis (A) will really happen is expressed as the posterior probability, or P(A|B). P(B|A) stands for Likelihood Probability, which quantifies the likelihood that a given hypothesis is true based on the available data.

Linear Regression: Linear model independent variable is used for estimation of the other reliant variable. But in various regression one or more variables are used for the prediction of the dependent variable. An illustration of a classical linear model in equation (viii).

$$Y = \alpha + \beta X + \varepsilon \qquad \text{...(viii)}$$

where Y is the dependent variable, X is the independent variable, β is slope measure how much Y changes when X changes, ε is the error term.

Random Forest: A decision tree classifier known as a random forest analyses different subsamples of the data sets. It helps forecast outcomes more accurately. An extension of a bagged decision tree is random forest. Equation (ix) defines each decision tree procedure.

$$ni_j = W_j C_j - W_{left(j)} C_{left(j)} - W_{right(j)} C_{right(j)} \qquad \text{…(ix)}$$

Node j's relevance is equal to n_i sub(j), where w sub is the weighted number of samples that reach node j.C sub(j) = node j's impurity value. The leftmost child node of node j splits to the left (j), while the rightest child node of node j separates as from the right (j).The decision tree's features are given weights according to the equation(x).

$$fi_i = \frac{\sum_{j:node\ j\ splitson\ feature i} ni_j}{\sum_{k \varepsilon all\ nodes} ni_k} \qquad \text{...(x)}$$

A characteristic's significance is determined by fi sub(i) and ni sub(j). These can then be normalized to a value between 0 and 1 by dividing by the sum of feature relevance shown in equation(xi):

$$normfi_i = \frac{fi_i}{\sum_{j \varepsilon all\ features} fi_j} \qquad \text{…(xi)}$$

The average of the feature importance across all trees at the Random Forest level determines the final feature significance. The outcome of calculating the importance of each attribute for each tree and dividing the total by the number of trees is given in equation (xii):

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004
E-ISSN 2224-2333
P-ISSN 2222-9930

$$RFfi_i = \frac{\sum_{j \in all trees} normfi_{ij}}{T} \quad \ldots(xii)$$

RFfi$_i$ sub(i) gives the Random Forest model's aggregate tree's determination of feature i relevance. In tree j, feature i's normalized feature importance is sub (ij), T is the total number of trees.

Logistic Regression: A statistical method to machine learning is logistic regression. It is derived from the study of statistics, like many other machine learning approaches, but despite its name, it is not an algorithm for regression issues where the goal is to predict a continuous outcome. Instead, the preferred technique for binary classification is logistic regression. Discrete binary result in the range of 0 and 1 is received. Its result is either one thing or another, to put it another way. An illustration of a classical linear model reference to equation(xiv).The logistic regression model provides the best fit when categorizing or limiting the range of values for the dependent variable. The following equation (xiii) contrasts logistic regression with linear regression:If P is substituted for Y with the assumption that the probability ranges between 0 and 1, then P's odds are taken.

$$\frac{p}{1-p} = \alpha + \beta X \quad \ldots(xiii)$$

The spectrum in this equation is restricted, resulting in fewer data points and, ultimately, a lower correlation. To prevent this, the answer for P is to take the log of the odds and add the exponent to both sides. The given equation(xiv) is the logistic regression model's sigmoid function for predicting any dichotomous dependent variable .

$$P = \frac{1}{1+e^{-(\alpha+\beta X)}} \quad \ldots(xiv)$$

Neural Network: In essence, a neural network functions as a processor and it is shown in Figure 4. It receives inputs, processes them according to layers (input, hidden, and output layer), and then produces output. It looks for patterns and trends that are challenging for both people and computers to understand. It can be represented as in given figure

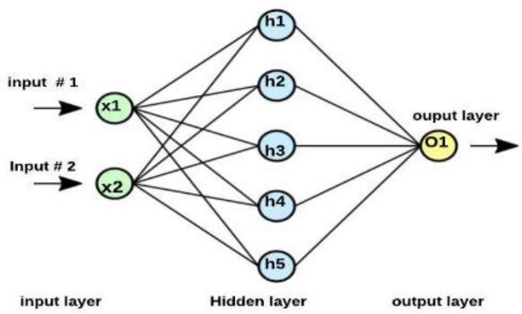

Figure 4: Neural Network [47]

The mathematical equation for calculating the values of a1, a2, and a3 in layer 2 would look like this as a function of the inputs x1, x2, and x3. Additionally, the value of a1 in layer 3 is represented by the values of a1, a2, and a3 in layer 2.Visualizing the output values produced by the three hidden units in the hidden layer will be the first step. According to equation (xva, xvb, xvc), Layer 1 represents the input layer, Layer 2 represents the hidden layer, and Layer 3 represents the output layer.

$$a_1^{(2)} = g(\theta_{10}^{(1)} x_0 + \theta_{11}^{(1)} x_1 + \theta_{12}^{(1)} x_2 + \theta_{13}^{(1)} x_3 \quad \ldots \text{(xv a)}$$

$$a_2^{(2)} = g(\theta_{20}^{(1)} x_0 + \theta_{21}^{(1)} x_1 + \theta_{22}^{(1)} x_2 + \theta_{23}^{(1)} x_3 \quad \ldots \text{(xv b)}$$

$$a_3^{(2)} = g(\theta_{30}^{(1)} x_0 + \theta_{31}^{(1)} x_1 + \theta_{32}^{(1)} x_2 + \theta_{33}^{(1)} x_3 \quad \ldots \text{(xv c)}$$

The output value of the node or unit in the output layer is defined by equation (xvd). A value of x1, x2, and x3 in the input layer may be the value that is represented in the preceding nodes or units as a function of a1, a2, and a3.

$$a_1^{(2)} = g(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a^{(2)}{}_1 + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)} \quad \ldots \text{(xv d)}$$

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004
E-ISSN 2224-2333
P-ISSN 2222-9930

Gradient Boosting Classifier: Boosting is one ensemble strategy in which the predictors are created sequentially rather than independently. Gradient boosting is a machine learning technique that provides a predictive model when applied in an ensemble to solve classification and regression problems. As base learners, it is frequently used with decision trees (most commonly CART trees). Sequential classifier that lowers bias and variance. Any supervised learning algorithm's goal is to define and reduce a loss function. Gradient Boosting is composed of Gradient Descent and Boost. To reduce the empirical risk, the gradient boosting method creates a model F: Xp->R repeatedly. Equation (xvi a) represents the empirical risk as the overall average loss of training dataset D.

Empirical Risk = (1/n) * L(F(xi), yi) [i=1 to n] … (xvi a)

Trying to lower this amount through the boosting process.The model is updated as equation (xvi b) for each iteration.

t: F(t)(x) =F(t-1) (x)+learning_rate*h(t) (x) ...(xvi b)

Here, the 'learning_rate' hyper-parameter controls how much each weak learner adds to the overall ensemble.

Voting Classifier: Voting Predictions are combined using a variety of models, often of different kinds, and basic statistics (such as computing the mean). For classification via a majority or plurality vote, the Ensemble Vote Classifier is a meta-classifier that combines machine learning classifiers that are conceptually similar or distinct. Figure 5 shows the structure of the voting classifier.

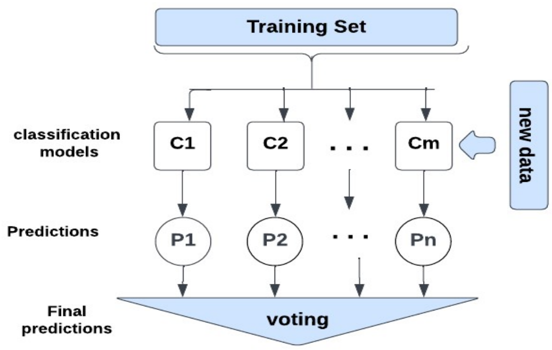

Figure 5: Voting Classifier [48]

Hard voting is the most basic kind of majority voting. In equation (xvii) each classifier's majority vote (plurality). C_j predicted the classification y:

$$\hat{Y} = \operatorname{mod} e\{C_1(x), C_2(x), ..., C_m(x)\} \quad \text{…(xvii)}$$

Assuming that combination of three classifiers in equation (xviii) that categorize a training sample according to the following categories: Classifier 1 for class 0; Classifier 2 for class 0; Classifier 3 for class 1.By a majority decision, classify the sample as "class 0."

$$\hat{Y} = \operatorname{mod} e\{0,0,1\} = 0 \quad \text{…(xviii)}$$

Bagged Decision tree for classification: Constructing numerous models (usually of the same type) using various training dataset subsamples. A straightforward ensemble technique called bagging entails the construction of numerous independent predictors, models, and trainers, which are then combined using model averaging methods. (for instance, the weighted average, a majority vote, or the average). It manages over-fitting as well as variance reduction.it has Parallel classifier function. Bootstrap Using replacement, you take numerous samples from your training dataset, aggregate them, and then train a model for each sample [49]. In the given Figure 6 bagging illustration

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

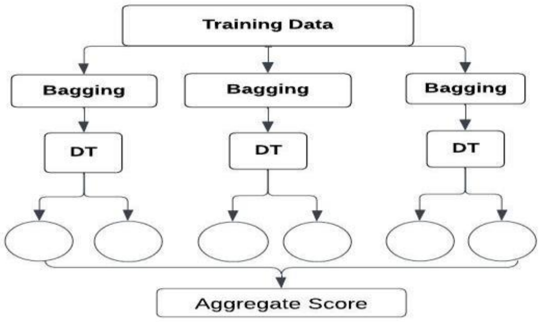

Figure 6: Bagging classifier [49]

Long Short-Term Memory (LSTM): Deep learning and artificial intelligence make use of a Long Short-Term Memory (LSTM) artificial neural network. Unlike traditional feed-forward neural networks, LSTM has feedback connections. As an example, LSTM can be applied to tasks such as speech recognition, robot control, machine translation, prediction, networked, un-segmented handwriting recognition, video games, and medical care. In tasks involving sequence prediction, many Recurrent neural networks (RNN) are able to learn long-term dependencies. The LSTM gates formula are represented in equation (xix):

$$it=\sigma(wi\ [ht-1\ ,\ xt]+ bi)\ ;\ ft= \sigma(wf\ [ht-1\ ,\ xt]+ bf);\ \ ot= \sigma(w_o\ [h_{t-1,}\ xt]+bo) \qquad \ldots(xix)$$

where, it shows an input gate, ft represent the forgotten gate, ot Show an output gate, Ơ sigmoid function is shown,wx The weight of the corresponding gate(x) neurons,(ht-1) Results from the earlier Lstm block (at timestamp t-1), xt Specify the current state stamp, bx the current timestamp's biases

B. Modes of Tweets Data

There are several ways to access data for disaster detection and management. Primary data sources for disaster-related research not text as shown in Figure 7 but also visual content as shown in Figure 8 etc

Disaster Detection through Text Tweets Corpus typically takes the form of documents that can be words, sentences, or even entire paragraphs of available text. Text data can be found in language corpora, library resources, social networks, organizations, information extraction, and audio and video transcriptions. Automatic information extraction from unstructured text input is accomplished using a machine learning technique known as text analysis (TA). This study concentrated on English-language text tweets.

Figure 7 shows facts on text tweet kinds. There are many distinct types of tweets written in many different languages. In this research, investigation of a number of disaster detection techniques utilizing English tweets has been analyzed.

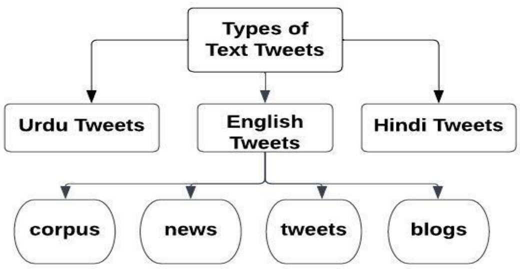

Figure 7: Types of text tweets

Analysis of Disaster from Social Media Using Visual Content: Online content that primarily uses images is known as "visual content. “Visual content can take many different forms, as shown in Figure 8 including images, videos, infographics, call-to-action buttons, questionnaires, interactive tools, drawings, memes, visual quotations, data visualizations, gifs, spreadsheets, and slideshows.

PJETS
E-ISSN 2224-2333
P-ISSN 2222-9930
Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

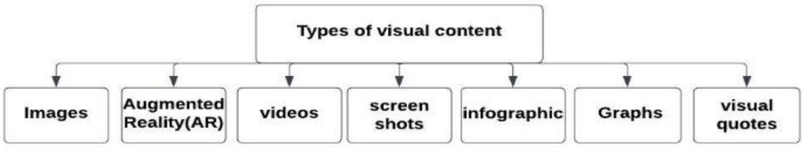

Figure 8: Types of visual content

Disaster Identification through Audio: An auditory medium is a type of media communication that disseminates information using audio or voice recordings as the medium. An innovation that lets persistent gushing of sound records like music, podcasts, voice-overs, addresses, etc. over the web is Sound streaming. Figure 9 depicts audio content, which is essential for evaluating the surroundings before, during, or after a disaster.

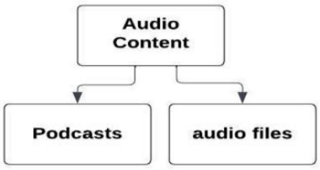

Figure 9: Types of audio content

Disaster Detection through camera: Increasing need for security and safety application of security camera technology to video analytics. A camera is essentially a device used to record moving images as either photos, films, or video transmission. One of the top concerns after a disaster is to save the victims' lives Cameras are useful for finding and identifying disaster victims as well as raising the alert when an incident occurs, as seen in Figure 10. However, one of its limitations is that they have difficulty distinguishing people who are shrouded in darkness.

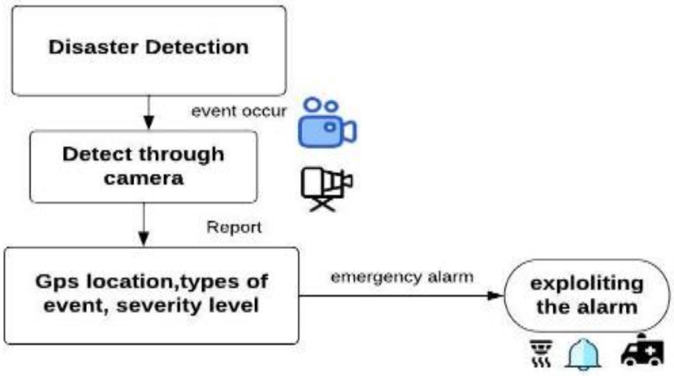

Figure 10: Disaster detection through camera

Disaster Detection through Satellite Imagery: Satellites provide current macro information on a wide range of geographic areas while securely residing above the Earth's atmosphere. This information, together with cutting-edge analytical methods like artificial intelligence and machine learning, provides crucial information about a natural disaster. Figure 11 depicts the numerous forms of satellite data for disaster identification and prediction.

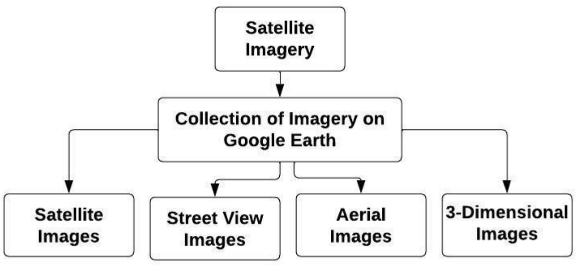

Figure 11: Types of Satellite Imagery

C. Identified Phases for Tweets Text processing for disaster prediction

Natural language Processing (NLP), which uses computational methods, involves a number of stages or phases to process and understand human language. A pipeline that converts unstructured text into data that computers can process often includes several phases. The following modules, which have been explained in the following sections, make up the system [20].

Data gathering: Retrieving every tweet both during and after an event is difficult. Studies gathered tweet data from many websites using the Twitter Stream API in order to combine tweets with additional meta-information, such as user data, location specific, hashtags, categories etc., bin order to predict disasters using tweets.

Cleaning and preprocessing of data: Then preparation is made for disaster prediction and management via tweets. Preprocessing entails eliminating tweets that aren't in English, removing noise and duplication, special characters, stop-words, and jargon. Remove the URL, as it provides no information when trying to evaluate text from words, especially on Twitter where user names are converted to codes to save space. Punctuation, special characters, and digits, as well as character normalizing lowering the enclosure lower and higher. Eliminating noisy data is a part of data cleaning. No unique Frameworks are utilized. Comparatively speaking to data cleaning, data processing is challenging. Data processing is more difficult than data cleaning. Following things, text processing and cleaning are required.

- **Text Cleaning:** For any type of text analysis, including sentiment analysis and, more broadly, text mining, cleaning text data is a need. It doesn't matter whether sentiment analysis (or other textual analysis) is done "manually" or automatically using machine learning techniques.
- **Noise Removal:** The term "noise removal" in the context of data processing refers to the process of deleting unnecessary or worthless data from a dataset. The elimination of noise is necessary to improve the reliability of analysis, modeling, and insight extraction.
- **Vectorization:** text data into numerical vectors that can be used as input for machine learning techniques [4].
- **Tokenization and Part-of-Speech Tagging:** Tokenization aims for dividing a text into token-sized pieces. Despite the individual words being used in most of the cases; complete phrases or even complete sentences can also be considered as tokens. As a pre-processing step, some of the characters along with punctuation marks may be removed during the tokenization process.
- **Text Normalization:** It is the process of transforming text into a single standard/ normal form such that different variations of the same text can be processed in a standard way.
- **Sentence Splitting:** Text is divided into sentences by the process of sentence splitting.
- **POS Tagging:** A key component of many NLP pipelines, including word-sense disambiguation, question answering, and sentiment analysis, is the POS tagging of raw text. In its most basic form, POS tagging is the process of recognizing nouns, verbs, adjectives, adverbs, and other words in a sentence.
- **Labeling:** Annotating a word, phrase, sentence, or document with its sentiment label positive or negative is known as sentiment labeling. There are three types of labeling: interactive, automatic, and manual.
- **Lower casing**: All data should be converted to lowercase as this will aid in preprocessing and later parsing stages of the NLP application.
- **Stemming:** It is a technique for removing the word's suffix and returning it to its base form. The normalization method known as stemming is used in natural language processing to minimize the number of computations needed.
- **Stop Word Removal:** Stop words include often used words like"if","but","we,","he" "she" and "them". The majority of the time (but not always), removing these terms enhances a model's performance without altering the text's semantics.
- **Padding:** Padding is used to ensure that every sequence is the same length while processing sequences in batches and feeding them into models that need constant input sizes.

Feature Extraction: This stage is carried out in several studies for automatic disaster eyewitness message detection and disaster prediction from tweets on the Twitter site [2,30]. There are typically three different approaches to extracting features from text: vector space representation, word-embedding vectors based, and lexicon-based

- **Word embeddings:** Most studies use word embedding representations to represent each key word while maintaining its semantic, syntactic, and linguistic significance [43]. The Word Embedding methods One Hot Encoding, TF-IDF, Word2Vec, and Fast Text are frequently used to convert text to numeric.
- **Vector space representation:** Most of the studies used vector space representation used for calculating bigram unigram n gram for disaster prediction [46].
- **Strategy based on a lexicon:** The lexicon-based approach is one of the methods or methods of semantic analysis.

Model Building: In this stage, the text is analyzed and understood using machine learning or deep learning models used for Detection and prediction of numerous disasters: Machine learning and deep learning techniques were used in most studies to evaluate the weighted accuracy for different disaster prediction, management, automatic recognition of damage assessment tweets during [16] or after a catastrophe, and emotion detection for volunteering and care [22].

Post-processing: After the model's output anticipates post-processing techniques can be used to the model's predictions to enhance the output. Ensuring that the output is formatted correctly, Removing ineffective or incorrect results and integrating predictions from multiple models or resources.

Evaluation Metrics: Utilizing the appropriate assessment criteria, testing the NLP model's performance, and modifying the model based on user feedback and performance. This measure is employed to determine the accuracy of models. When evaluating classification models for balanced datasets, measures like accuracy, precision, and recall are useful; however, when the data is unbalanced, alternative approaches, such ROC/AUC, perform better. The ROC curve represents an association between the classifier's true positive rate and false positive rate .The following mathematical equations(xx),(xxi),(xxii),(xxiii) used as evaluation of model.

$$precision = \frac{TP}{TP+FP} \quad \text{... (xx)}$$

Where Tp is the overall sample count from the true class that has been determined to be true, and the function Fp is the entire sample count from the false class that has been determined to be true.

$$recall = \frac{TP}{TP+FN} \quad \text{…(xxi)}$$

The number of valid class samples that were assigned a false label, as well as Fn.

$$F1 = \frac{2 * precision * recall}{precision + recall} \quad \text{…(xxii)}$$

The F1 Score, which is the average of precision and recall, combines precision and recall into a single, simple indicator. It is mathematically represented by eq.(xxii).

$$accuracy = \frac{TP}{TP+FN+TN+FP} \quad \text{…(xxiii)}$$

Each forecast is made up of all possible positive (P) and negative (N) events. False negatives (TN) and false positives (FP) together make up (FN), whereas FP and TN make up (P).

V. DISASTER MANAGEMENT AND AVOIDANCE / MITIGATION

Disaster management and risk reduction work to decrease the likelihood and potential effects of a disaster going forward. According to the United Nations Office for Disaster Risk Reduction, disaster risk reduction can be defined as the systematic process of using administrative directives, organizations, and operational expertise and abilities. The objective is to implement those strategies, policies, and enhanced functional capacities that can lead to reduction in the adverse impacts of hazards and the possibility of disaster..

For this purpose use the trained model to instantly and automatically categorize incoming tweets. Depending on the size of the activity, either local servers or cloud services can be used for this. When a tweet is flagged as being related to a crisis, integrate the model with an alerting system that automatically notifies emergency response teams or pertinent authorities. This makes it possible to react quickly to possible threats. Urge the public to proactively report dangers on social media. Encourage the use of particular hashtags or keywords that the model is capable of identifying. Educate the public on the value of prompt reporting by using social media and neighborhood engagement. Utilise the knowledge gathered from the tweet analysis to create focused awareness and readiness campaigns. Describe the safety precautions, emergency shelters, and evacuation routes.

PJETS
Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
E-ISSN 2224-2333
https://doi.org/10.22555/pjets.v11i1.1004
P-ISSN 2222-9930

Disasters according to emergency managers, are re-occurring events that go through the following four stages: preparation, response, mitigation, and recovery. The life cycle of catastrophe management is depicted in Figure 12.

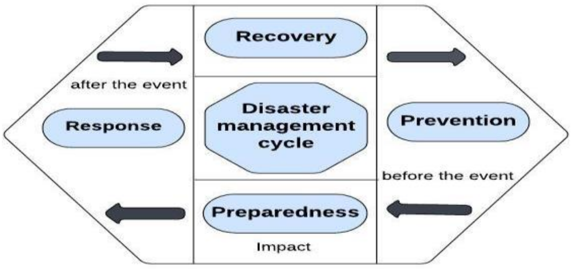

Figure 12: Disaster Management Cycle

A. Response

When a hazard arises, a response is described as the steps done to reduce death and sickness as well as to stop future property damage. Implementing preparedness strategies is the response. Possible response actions include: triage for search and rescue, urgent medical attention, battling fires, providing shelter for victims, and moving medical records

B. Recovery

The steps necessary to get back to normal after an event are known as recovery. e.g. (repairing houses, changing Holmes)

C. Protection Prevention

The term "consistent activities that minimize the risk of a hazard (probability of occurrence) of a hazard, or to limit the possible bad effects incurred by persons and/or assets" is used to describe prevention and mitigation. It may entail doing things like:

- structure relocation or elevation to decrease the effects of floods
- Anchoring shelves and water heaters to the walls in earthquake-prone regions
- use fire-resistant materials while building new structures to lower the risk of fire.

D. Preparedness:

Plans or processes designed to reduce property and human harm when an event happens are examples of preparedness. By engaging in these activities, disaster (emergency) managers can be prepared to offer the best possible response when a disaster strikes. The crucial components are as planning ,educating employees, tablecloth drills for emergencies.

VI. OBSERVATION & RECOMMENDATION

Natural disasters are one of the leading causes of fatalities as well as damage to property and infrastructure. Major property losses and casualties could result from any type of calamity. The literature review covered text analysis and categorization techniques as well as the most effective methods to utilize tweets to predict disasters. The work demonstrated a useful method for the binary machine learning problem of attempting to identify questionable text samples. Determining whether or not a passage of text pertains to an actual catastrophe is the machine's hardest hurdle. People update social media sites like Twitter in real-time when a crisis emerges. Organizations working on disaster relief and response are especially in need of this information because it can immediately make them aware of their obligations. Effective disaster prevention, mitigation, observation, and early warning systems must be constructed in risk areas, and suitable protection plans for the most dangerous landslides must be created. Because disaster recovery activities should be sustainable, research should focus on using ML and DL to enhance mitigation efforts, reduce vulnerabilities, and assess resilience, including of vital infrastructure. Algorithms that can determine whether a text passage pertains to a real disaster or not may ultimately help in the detection of real disasters and eliminating out of fake ones, allowing the concerned team to respond more effectively. The

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004
E-ISSN 2224-2333
P-ISSN 2222-9930

complexity and significance of catastrophic operations necessitate robust and proven ML and DL solutions. In this study, focus on how deep learning and machine learning methods are used to disaster management and disaster prediction has been done, as well as how social media platform data in the form of tweets (English text) that have been converted from unstructured to structured assist in disaster detection has been done. Since Twitter data is unstructured, Natural Language Processing (NLP) is required to divide it into the categories "Related to Catastrophe" and "Not associated with Catastrophe." The issue is that, the algorithm may struggle to determine if a word is being used properly, semantically related to disaster or symbolically.

VII. CONCLUSION

Natural disasters are catastrophic occurrences carried on by the globe's natural processes. They are capable of causing extensive harm to society, the environment, and infrastructure, which frequently results in fatalities and has serious financial repercussions. Natural disasters can take many different forms, and each has unique traits and effects. The development of ML and DL is assisting in managing the complexity of disasters. to support disaster management efforts and enhance their effectiveness. This survey has reviewed ML and DL strategies used in various disaster management and prediction. Fortunately, the diversity of users, ongoing development of social media platforms and free access present academics with a unique opportunity to investigate people's attitudes and beliefs regarding a number of issues. As a result, the extent of the damage caused by disasters can be seen reflected in how people use social media. Despite the fact that prior studies on geotagged Twitter data produced a variety of practical techniques. With a focus on how tweets assist with disaster identification after being converted from unstructured to structured using natural language processing techniques, this study aims to evaluate how tweets can be used in disaster-related investigations. It looks into the methods and algorithms used for predicting disasters at the moment. A comprehensive understanding of the numerous data kinds and their sources in relation to a variety of operations and crisis management scenarios is another goal of the work. Additionally, the study aims to provide a thorough overview of the many data mining approaches applied to various difficulties related with natural catastrophes as well as information on how to clean tweets using natural processing techniques. This work can be used to determine the areas for future research on automated information classification and disaster coordination systems, which will help emergency response teams by using social media platform data (tweets) to make judgements in the case of an extensive catastrophe.

VIII. ABBREVIATIONS

The following abbreviations are used in this manuscript:

- TF-IDF Term Frequency-Inverse Document Frequency.
- CNN Convolutional Neural Network.
- LR Logistic Regression.
- NB Naive Bayes.
- SVM Support Vector Machine.
- RF Random Forest.
- DT Decision Tree
- ML Machine Learning
- DL Deep Learning
- NLP Natural language Processing
- EDA Exploratory Data Analysis

REFERENCES

[1] Jishnu Ray Chowdhury, Cornelia Caragea, Doina Caragea, “On Identifying Hashtags in Disaster TwitterData”, 2020,Vol. 34 No.01:AAAI20 Technical Tracks 1
[2] Kiran Zahraa, Muhammad Imran(b), Frank Ostermannc, “Automatic identification of eyewitness messages on twitter during disasters”,Volume 57, Issue 1, January 2020, 102107
[3] Anh Duc Le, “Disaster Tweets Classification using BERT-Based Language Model”, Arxiv Computing science journal,2022
[4] Tao Cheng and Thomas Wicks,”Event Detection using Twitter: A Spatio-Temporal Approach”,PLOS ONE | www.plosone.org ,June 2014 | Volume 9 |issue 6 | e97807

PJETS
Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
E-ISSN 2224-2333
https://doi.org/10.22555/pjets.v11i1.1004
P-ISSN 2222-9930

[5] Hafiz Suliman Munawar , Mohammad Mojtahedi 1,Ahmed W. A. Hammad 1, Michael J. Ostwald,"An AI/ML-Based Strategy for Disaster Response and Evacuation of Victims in Aged Care Facilities in the Hawkesbury-Nepean Valley: A Perspective", January 2022,Buildings 12(1):80

[6] Amin Muhammad Sadiq , Huynsik Ahn and Young Bok Choi ,"Human Sentiment and Activity Recognition in Disaster Situations Using Social Media Images Based on Deep Learning",Sensors Multidisciplinary Digital Publishing InstituteVolume ,2020, 20(24), 7115

[7] Cynthia Cui and Leonardo Cui,"An Innovative Flood Prediction System Using Improved Machine Learning Approach" ,The canadian Science Fair journal,CSFJ | Volume 2 | Issue 2 © Cui & Cui 2020

[8] LidaHuang,GangLiua,TaoChena,HongyngYuan,PanpanShib,YujiaMiaob,"Similarity-based emergency event detection in social media",2020,Journal of Safety Science and Resilience 2(1)

[9] Mehdi Jamali,Ali Nejat, Saeed Moradi , Souparno Ghosh, Guofeng Cao, Fang Jin,"Machine Learning in Disaster Management,Recent Developments in Methods and Applications", 2022 ,Machine Learning and Knowledge Extraction 4(2):446-473

[10] Zahera, Hamada M; Jalota, Rricha; Sherif, Mohamed Ahmed; Ngomo, Axel-Cyrille Ngonga,"A stacked convolutional neural network for detecting the resource tweets during a disaster",2020 Multimedia Tools and Applications

[11] Hien To, Sumeet Agrawal, Seon Ho Kim, Cyrus Shahabi,"On Identifying Disaster-Related Tweets:Matching-based or Learning-based",2017 IEEE third international conference on multimedia big data (BigMM)

[12] Mehdi Jamali, Ali Nejat, Souparno Ghosh, Fang Jin, Guofeng Cao,"Social Media Data and Post-Disaster Recovery ",2019 International Journal of Information Management,pg 25-37

[13] Devaraj, Ashwin, Murthy, Dhiraj, Dontula, Aman, "Machine-learning methods for identifying social media-based requests for urgent help during hurricanes",International Journal of Disaster Risk Reduction; Volume 51, December 2020, 101757

[14] lokabhiram dwarakanath ,amirrudin kamsin ,rasheed abubakar rasheed ,anitha anandhan ,and liyana shuib,"Automated Machine Learning Approaches forEmergency Response and Coordination via Social Media in the Aftermath of a Disaster: A Review",2021 Journals & Ma. gazines ;EEE Access ;Volume: 9,Page(s): 68917 - 68931

[15] Sumona Deb a, Ashis Kumar Chanda ,"Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data",Machine Learning with Applications,Volume 7, 15 March 2022, 100253

[16] Madichetty, Sreenivasulu; Sridevi, M",A novel method for identifying the damage assessment tweets during disaster",Future Generation Computer Systems,Volume 116, March 2021, Pages 440-454

[17] Kruspe, Anna, "Detecting novelty in social media messages during emerging crisis events",2020 International Conference on Information Systems for Crisis Response and Management (ISCRAM)

[18] Bai, Hua; Yu, Hualong; Yu, Guang; Rocha, Alvaro; Huang, Xing; "Analysis on an Auto Increment Detection System of Chinese Disaster Weibo Text.",February 2021 Journal of Universal computer science 27(2):230-252

[19] Cody Buntain,Jennifer Golbeck ,"Automatically Identifying Fake News in Popular Twitter Threads",2017 IEEE International Conference on Smart Cloud

[20] Viktor Pekar,,Jane Binner, Hossein Najafi, Chris Hale, Vincent Schmidt†, "Early Detection of Heterogeneous Disaster Events Using Social Media",March 2019 Journal of the Association for Information Science and Technology 71(1)

[21] Muhammad Arslan, Ana-Maria Roxin, Christophe Cruz, Dominique Ginhac, "A Review on Applications of Big Data for Disaster Management",2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)

[22] Muniz-Rodriguez, Kamalich; Ofori, Sylvia K; Bayliss, Lauren C; Schwind, Jessica S; Diallo, Kadiatou; Liu, Manyun; Yin, Jingjing; Chowell, Gerardo; Fung, Isaac Chun-Hai,"Social media use in emergency response to natural disasters: a systematic review with a public health perspective",Disaster Medicine and Public Health Preparedness , Volume 14 , Issue 1 , February 2020 , pp. 139 - 149

[23] Kankanamge, Nayomi; Yigitcanlar, Tan; Goonetilleke, Ashantha,"How engaging are disaster management related social media channels? The case of Australian state emergency organisations",International Journal of Disaster Risk Reduction,Volume 48, September 2020, 101571

[24] Vinay Chamola, Senior Member, Vikas Hassija, Sakshi Gupta, Adit Goyal, Mohsen Guizani, Fellow, and Biplab Sikdar, Senior Member,"Disaster and Pandemic Management Using Machine Learning: A Survey",IEEE Internet Things J. 2021 Nov 1; 8(21): 16047–16071.

[25] Imran, Muhammad; Ofli, Ferda; Caragea, Doina; Torralba, Antonio,"Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions",Information Processing & Management,Volume 57, Issue 5, September 2020, 102261

[26] Mondal, Tamal; Pramanik, Prithviraj; Bhattacharya, Indrajit; Boral, Naiwrita; Ghosh, Saptarshi,"Analysis and early detection of rumors in a post disaster scenario",Information Systems Frontiers 20, 961–979 (2018)

[27] Ullah, Irfan; Khan, Sharifullah; Imran, Muhammad; Lee, Young-Koo,"RweetMiner: Automatic identification and categorization of help requests on twitter during disasters",Expert Systems with Applications,Volume 176, 15 August 2021, 114787

[28] Fan, Chao; Wu, Fangsheng; Mostafavi, Ali,"A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters",January 2020,IEEE Access PP(99):1-1

[29] Kamari, Mirsalar; Ham, Youngjib,"AI-based risk assessment for construction site disaster preparedness through deep learning-based digital twinning",February 2022;Automation in Construction 134(4):104091 pages

[30] Zijiang Zhu, Yu Zhang, "Flood disaster risk assessment based on random forest algorithm",Neural Computing and Applications 34, 3443–3455 (2022)

[31] S. Choi and B. Bae, "The real-time monitoring system of social big data for disaster management", Computer Science and its Applications, pp. 809-815. Springer Berlin Heidelberg, 2015.

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 11. Issue. 1, PP. 45-73, November 2023
https://doi.org/10.22555/pjets.v11i1.1004

[32] I. L. Janis and L. Mann, "Emergency decision making: a theoretical analysis of responses to disaster warnings", Journal of human stress, vol. 3, no. 2, pp. 35-48, 1977.
[33] Vasileios Linardos, Maria Drakaki , Panagiotis Tzionas and Yannis L. Karnavas "Machine Learning in Disaster Management: Recent Developments in Methods and Applications",2022 Machine Learning and Knowledge Extraction 4(2):446-473
[34] Rania Rizki Arinta , Emanuel Andi W.R ,"Natural Disaster Application on Big Data and MachineLearning: A Review",2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)
[35] Oliver Mauroner; Anna Heudorfer,"Social media in disaster management: How social media impact the work of volunteer groups and aid organizations in disaster preparation and response" ,International Journal of Emergency Management,2016 Vol.12 No.2
[36] Karla Saldana Ochoa, Tina Comes. "A Machine learning approach for rapid disaster response based on multimodal data",2021 arXiv:2108.00887 [cs.LG]
[37] Saptarsi Goswami, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti, Basabi Chakraborty" A review on application of data mining techniques to combat natural disaster",Ain Shams Engineering Journal Volume 9, Issue 3, September 2018, Pages 365-378
[38] Saloni Jain,"Real-Time Social Network Data Mining For Predicting The Path For A Disaster", 2015.
[39] Himanshu Shekhar , Shankar Gangisetty ,"Disaster Analysis Through Tweets", IEEE International Conference on Advances in Computing, Communications and InformaticsAt: Kerala, IndiaVolume: 2015
[40] reenivasulu Madichetty, Sridevi Muthukumarasamy, "Detection of situational information from Twitter during disaster using deep learning models",December 2020 Sadhana 45(1)
[41] Venkata Kishore Neppalli, Cornelia Caragea, Doina Caragea,"Deep Neural Networks versus Naïve Bayes Classifiers for Identifying Informative Tweets during Disasters". International Conference on Information Systems for Crisis Response and Management,2018
[42] Jason Christopher Chan (RPO) ,"Social media in disaster management: How social media impact the work of volunteer groups and aid organizations in disaster preparation and response", International Journal of Emergency Management 2016 Vol.12 No.2
[43] Md. Yasin Kabir, Sanjay Madria "A Deep Learning Approach for Tweet Classification and Rescue Scheduling for Effective Disaster Management", Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information SystemsNovember 2019 ;Pages 269–278
[44] Aldo Hernandez-Suarez, Gabriel Sanchez-Perez,Karina Toscano-Medina, Hector Perez-Meana , Jose Portillo-Portillo, Victor Sanchez and Luis Javier García Villalba,"Using Twitter Data to Monitor Natural Disaster Social Dynamics: A Recurrent Neural Network Approach with Word Embeddings and Kernel Density Estimation",Sensors 2019, 19(7), 1746;
[45] Erol, M.H.; Bulut, F, "Real-time application of traveling salesman problem using Google Maps API",. Proceedings of the IEEE Electric Electronics, Computer Science, Biomedical Engineering's' Meeting (EBBT), Istanbul, Turkey, 20–21 April 2017; pp. 1–5.
[46] Mohammad Reza Faisal, Radityo Adi Nugroho, Rahmat Ramadhani,Friska Abadi, Rudy Herteno and Triando Hamonangan Saragih,"Natural Disaster on Twitter: Role of Feature Extraction Method of Word2Vec and Lexicon Based for Determining Direct Eyewitness",TRENDS IN SCIENCES 2021; 18(23): 680
[47] https://www.databricks.com/glossary/neural-network
[48] https://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier/
[49] https://medium.com/analytics-vidhya/ensemble-models-baggings-vs-boosting-8affa6d18098
[50] Mohammed Alsaqer,Salem Alelyani,Mohamed Mohana,Khalid Alreemy and Ali Alqahtani ,"Predicting Location of Tweets Using Machine Learning Approaches",Appl. Sci. 2023, 13(5), 3025
[51] Premkumar Duraisamy; M Duraisamy; M Periyanayaki; Yuvaraj Natarajan,"Predicting Disaster Tweets using Enhanced BERT Model",2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)
[52] Rani Koshy & Sivasankar Elango ,"Multimodal tweet classification in disaster response systems using transformer-based bidirectional attention model",Neural Computing and Applications 35, pages 1607–1627 (2023)pages
[53] Courtney J. Powers, Ashwin Devaraj, Kaab Ashqeen, Aman Dontula, Amit Joshi, Jayanth Shenoy, Dhiraj Murthy; "Using artificial intelligence to identify emergency messages on social media during a natural disaster: A deep learning approach". Volume 3, Issue 1, April 2023, 100164
[54] "Study on typhoon disaster assessment by mining data from social media based on artificial neural network", Natural Hazards volume 116, pages 2069–2089 (2023)