Machine Learning Framework for Analyzing Disaster-Tweets

R. Manimegalai S. Kavisri M. Vasundhra R. Kingsy Grace
Department of Computer Science and Engineering
PSG Institute of Technology and Applied Research
Coimbatore, India.
corresponding author : drrm@psgitech.ac.in

***Abstract* - During natural disasters and catastrophes, Twitter is becoming a more popular source of information exchange. It is primarily used to share the status of disaster recovery efforts initiated by humanitarian and disaster relief organizations, to report and request or provide volunteer services, and to update on the scope of geographic phenomena. This paper supports the creation of future automated crisis management systems as well as the planning and preparation of effective disaster responses by teams working on disaster mitigation. This work focuses on developing a comprehensive framework for text processing and analysis on tweets posted in Twitter during natural catastrophes using natural language processing techniques. Disaster-related tweets are categorized into precautionary tweets, educational tweets, and recovery tweets. The algorithms which are used to develop the framework are Naïve Bayes based on Bayes theorem, Logistic Regression based on Sigmoid function, Random Forest based on decision trees, Extreme Gradient Boosting is based on bagging and boosting, Support Vector Machine is based on hyperplane. Five performance metrics, namely, accuracy, precision, recall, F1-score, and time, are calculated to assess how well the algorithms perform. The data set is split into training set and testing set as 75:25, 63:37, and 50:50. This comparison is to provide insights about the performance of algorithms in terms of efficiency with time bound actions and reactions.**

Keywords— social media, tweet processing, text classification, disaster response, machine learning

I. Introduction

AI Technologies based on machine learning are built around the notion that a machine can analyse data, spot trends, and draw conclusions with little to no human involvement. The approach makes the system to think and react like humans i.e. the ability to learn. Gathering data, processing that data, selecting a model, learning, assessment, hyper-parameter tuning, then output are the seven steps of any machine learning application [1, 2]. During natural disasters and emergencies, Twitter is progressively being used to update and communicate the degree of regional occurrences, report the affected communities, request or provide practical support, and share the condition of the mitigation phase initiated by emergency relief and catastrophe organisations. During natural disasters and emergencies, Twitter is progressively being used to update and communicate the degree of regional occurrences, report the affected communities, request or provide practical support, and share the condition of the mitigation phase initiated by emergency relief and catastrophe organizations.

Tweet-analysis offer usable insights to crisis prevention and mitigation teams in the meticulous planning of successful catastrophe responses, as well as to improve the effectiveness of upcoming automation systems for contingency planning. Tens of hundreds of tweets shared on Twitter throughout catastrophic events are analyzed to provide useful insights for disaster recovery. This paper classifies disaster-based tweets into three categories, namely, precautionary tweets during disaster, informative tweets during disaster, recovery related tweets during disaster. The main objectives of the proposed work include: *i*) classify tweets into disaster and non-disaster tweets, *ii*) further classify disaster related tweets into three categories based on mitigation, preparedness, response, and recovery, *iii*) compare the performance of various algorithms based on accuracy, precision, recall, F1-score and time. Based on the above, the proposed system provides support to assist legislators, public affairs experts, emergency preparedness organizations in using Twitter to communicate with the public. During catastrophes, risk communication tries to avoid and limit disaster harm, alerts the people well before the catastrophe, communicates information during disasters, and facilitates recovery.

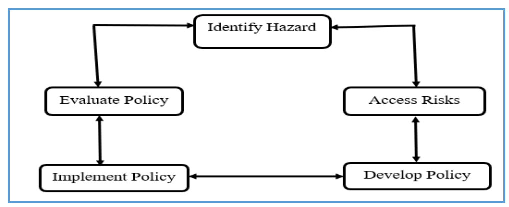


Fig. 1. Risk Communication in Disaster Management

Figure 1 shows the workflow of risk communication in disaster management. Effective risk communication assists various partners with overseeing hazard all the more viably, help individuals in danger, and to assume a more dynamic part in various phases of disaster management. At the point when the danger is known, it works on the client's productivity to utilize Twitter. It is important to frame arrangements at different levels to the use of hazard correspondence and to utilization of the vital devices, including Twitter, for planning and assessment. The hazards that are involved in disaster management needs to be identified. Then the risks are assessed to identify which is of higher impact and which is of lower impact. Then the relevant policies are developed for managing the risks and approval is obtained for those that are feasible. Then the policies are implemented and then evaluated again. This work classifies the disaster-based tweets into different categories which can be further applied to its intended purpose as quick as possible since timely action is considered as the most important factor when it comes to disaster relief and recovery. The following algorithms, are Naïve Bayes, SVM, Logistic Regression, Xgboost, Random Forest and compared performance of each algorithm based on accuracy, precision, recall, f1-score and time.

Section 2 presents details of the various AI techniques reported in the literature for tweet analysis. Section 3 presents the proposed methodology of how tweets can be used for disaster management by classifying them according

2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS) | 979-8-3503-3583-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICISCOIS56541.2023.10100450

to the purpose using ML algorithms. Section 4 provides the implementation details along with the results and comparison between various algorithms in terms of accuracy, performance, recall score, F1-score. Section 5 concludes the paper and gives directions for future enhancements.

II. Liteature Survey

Micro-blogging platforms provide significant information resources during emergency conditions, especially catastrophic and man-made disasters [3]. It's critical to develop system driven approaches for identifying tweets that help with situational awareness. Natural Language process based techniques are used to get the tweets. The challenge for natural language processing is that the approach within which tweets are accessible; due to size-limit, tweets contain abbreviations, informal words, and lots of other unnecessary information. Performance of the classifier is reduced in inter classification, i.e., when a classification model is educated on tweets about previously occurred incidents and then used to categorise tweets about presently occurring events. 169, 186, l32, and 198 tweets were identified as Situational Awareness tweets for the HBlast, UFlood, TBopha, and SHshoot datasets respectively, out of the 500 tweets for each event. From the twitter posts that were recognised as non-Situational Awareness, an equal number is chosen. As a result, the collection contains a total of 1048 tweets, with majority of them labelled as Situational Awareness tweets and the rest as non-Situational Awareness tweets.

Hien To et al [4] have developed the classification system by claiming that social networks such as Facebook and Twitter have been extensively used for social interaction during emergency events such as catastrophes. During calamities, Twitter serves as a venue for raising alertness. The details shared on Twitter by those requests for assistance and warnings, can assist first rescuers, judgement call, and the general public in learning about the circumstance. Although there are ample tweets available that can serve as a source for study and analysis, it is very difficult to identify the messages automatically that are relevant because tweets are small and don't have any specific format which results in inappropriate classification performance. The proposed effective algorithm uses two approaches matching based algorithm and learning based algorithm. A five-step process for analysing and interpreting tweets during disasters, is developed.

A categorization strategy is suggested by Beverly Estephany Parilla-Ferrer et al. [5] by creating ML models that can automatically identify disaster-related information tweets. A random subset of the collected data set is manually marked as informative or non-informative to provide background truth and automatically classifying tweets. These models are rated according to a number of factors, including measurements such as F1, accuracy indicators, accuracy, recall, and the area under the curve. Shriya Goswamia et al. have [6] projected that the social media information about disasters can save thousands of lives by informing people so that evasive action can be taken. The iniquitousness of smartphones, laptops, and tablets has enabled people to speak the occurrence of disasters experienced in real-time.

The importance of social media has been stated by Muhammad Imran et al. [1], who claim that micro-blogging platforms have become a vital tool to exchange information on the web, specifically during time-critical situations such as catastrophic and human-made disasters. Babak Abedin et al. [7], have advanced the meaning of online media by expressing that web-based media destinations are assuming a critical part in the rapid propagation of information when disasters occur. Data trade is significant during the disaster, the executives' cycles like wave, tremor, fountains of liquid magma and particularly the reaction stage. Disaster researchers and emergency management specialists typically rely on the four-stage classification of disaster reduction, emergency planning, action, and mitigation to analyse and handle crises [8]. The goal is to look into the nature of the information of twitter posts posted during a disaster and prepare a set of categories based on the knowledge gained throughout the crisis phase, which includes disaster planning, disaster relief, and recovery. This document proposes an alternative coding scheme for categorizing tweets into different topics in order to establish an understanding of geographic location, and a structure that will be used to classify tweets into these group. Archana Gopnarayan et al. [9] have proposed the importance of classifying disaster-related tweets and noted that disaster researchers and emergency managers use data from social media as a reference for their analysis to discover various changes and management of disasters over several stages. Social networks also came to the conclusion that a timely, efficient solution is required. SVM, K-nearest neighbourhood, logistic regression, and data mining algorithms are applied to tweet classification to finally determine the most accurate algorithm.

Jyoti Prakash Singh et al. [10] have proposed the incident identification and geographical estimation of tweets during tragic events. Disaster-related tweets also often warn and inform people of preventive measures. In order to assist victims, their actual location need to be included in the twitter posts, which is another critical consideration in an emergency. There are many research work [11]-[21] that discuss the tweets and tweets analysis.

III. The Proposed ML based Framework for Analysing Disaster Tweets

In this digital era, massive amount of data available on social media is changing people's lives every day to a larger level. Twitter is one of the globally used platform where people connect and share about events happening in and around them. During disaster times, twitter can be used as the best mode to find when and where disasters are happening. The government officials, NGO, aid workers, health system will get notified whenever there is meteorological information regarding disasters are posted to reach out to the people in particular region to create awareness and also there are agencies that can provide necessary support to people who are affected by disaster. The workflow of the proposed disaster tweet management system is illustrated in Figure 2. The initial step in the proposed methodology is to process the request for getting Twitter API by submitting the required project details. The next step is to acquire the data, i.e. tweets, from twitter by giving the credentials. Then, NLP techniques are applied to categorize the and group the related tweets. After collecting disaster related tweets, they are formatted and grouped into a dataset. Pre-processing is done after removing duplicate tweets. Pre-processing includes removing URL's, removing hash tags, removing username, spell check and corrections, replacing colloquial words with proper English words, replacing abbreviations, removing stop words, lemmatization, removing spaces. Then, the tweet is checked whether it belongs to disaster related tweet category. The final step is to categorize them according to the intended purposed such as response, recovery, informative or not relevant. The flowchart of the proposed system is shown in Figure 2.

ML models cannot accept the textual data for processing. It accepts only numerical data. Therefore, the textual data needs to be featured. The two features applied are count vectorizer and TF-IDF. Several ML algorithms such as Naïve Bayes, Support Vector Machine, Logistic Regression, Extreme Gradient Boosting, Random Forest are implemented. The Naive Bayes classifier is a set of algorithms that use Bayes' theorem to classify data.

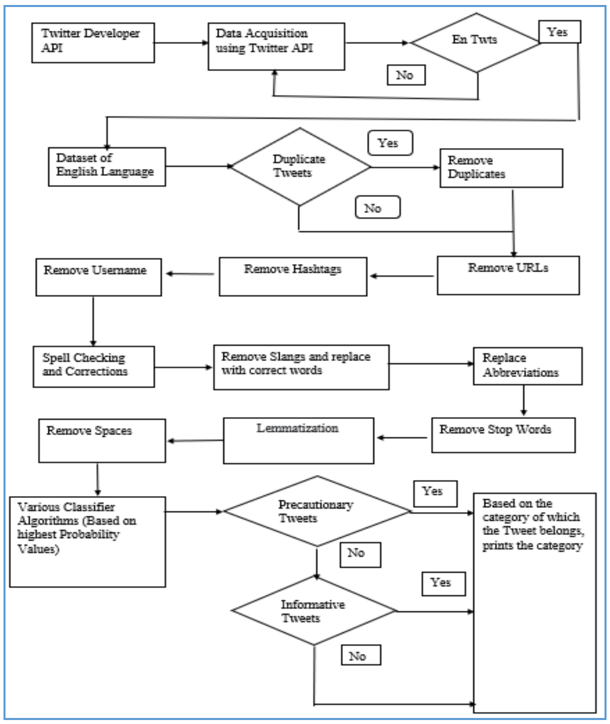


Fig. 2 Flowchart of the Proposed Disaster Tweet Analyzer

Random Forest Algorithm is a classifier that evolves from decision trees. Random forest model is a bagging-type ensemble of decision trees that trains several trees in parallel and uses the majority decision of the trees as the final decision of the random forest model. Random Forest Regression has become a commonly used tool in multiple prediction scenarios due to their high accuracy and ability to handle large features with small samples.

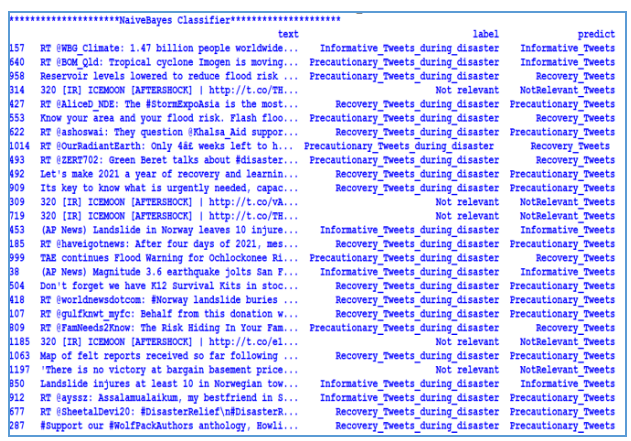
Logistic regression is an important machine learning algorithm because it has the ability to predict probabilities and classify new data with both continuous and discrete data sets. Logistic regression requires a large set of samples. Sigmoid Function is implemented as cost function and is used for predicting the values of probabilities The category is run through maximum likelihood.

SVM is a machine learning method that can classify and predict data. In the case of a linear support vector machine, a straight line is drawn between the two classes. It can handle a variety of static and dynamic analysis. The two approaches applied for multiclass classification are one-to-one approach and one-to-many approaches. One-to-one approach divides the multiclass problem into a number of binary classification problems. For each pair of classes, a binary classifier is used. One-to-many approach assigns a binary classifier for each class. Extreme Gradient Boosting (XGBoost) is a fully accessible toolkit that makes gradient boosting methods more efficient. Gradient augmentation refers to a set of built-in machine learning methods that may be utilised for classification and regression prediction modelling. Weight is assigned to all independent variables, which are entered into the decision tree and predict the outcome. If the weight of the variable is incorrectly predicted, the second decision tree is used.

IV. Experimental Results

The proposed system for disaster- tweets analysis is implemented using Python. The features identified for tweet analysis are countvector and TF-IDF. The implementation for applying count vector feature with parameter of analyser should be in word format and the pattern for the word character is specified using regular expression. The implementation for applying TF and IDF for both n-gram level and word level with its parameter. The dataset is split into test set and training set of various sizes. The Navie-Bayes algorithm, logistic regression, random forest algorithm and SVM are implemented and the obtained results are compared with the parameters such as accuracy value, precision, recall value and the F1 – score.

The output of Naïve Bayes Algorithm, logistic regression, random forest algorithm and SVM are shown in the Figures 3 – 7. In Figure 3, the dataframe is displayed which has the first column as Tweet ID, the next column as original tweet in the Dataset. Second column is manually defined label and the last column is the predicted Label. Table 1, 2, 3, 4, 5 depicts the inferences of varying the data set size gradually and calculation of all performance metrics along with the time taken to execute the algorithms such as Naïve Bayes Algorithm, logistic regression, random forest algorithm and SVM.



********************NaiveBayes Classifier********************

	text	label	predict
157	RT @WBG_Climate: 1.47 billion people worldwide...	Informative_Tweets_during_disaster	Informative_Tweets
640	RT @BOM_Qld: Tropical cyclone Imogen is moving...	Precautionary_Tweets_during_disaster	Informative_Tweets
958	Reservoir levels lowered to reduce flood risk ...	Precautionary_Tweets_during_disaster	Recovery_Tweets
314	320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/TH...	Not relevant	NotRelevant_Tweets
427	RT @AliceD_NDE: The #StormExpoAsia is the most...	Recovery_Tweets_during_disaster	Precautionary_Tweets
553	Know your area and your flood risk. Flash floo...	Precautionary_Tweets_during_disaster	Recovery_Tweets
622	RT @ashoswai: They question @Khalsa_Aid suppor...	Recovery_Tweets_during_disaster	Precautionary_Tweets
1014	RT @OurRadiantEarth: Only 4â£ weeks left to h...	Precautionary_Tweets_during_disaster	Recovery_Tweets
493	RT @ZERT702: Green Beret talks about #disaster...	Precautionary_Tweets_during_disaster	Recovery_Tweets
492	Let's make 2021 a year of recovery and learnin...	Recovery_Tweets_during_disaster	Precautionary_Tweets
909	Its key to know what is urgently needed, capac...	Recovery_Tweets_during_disaster	Precautionary_Tweets
309	320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/vA...	Not relevant	NotRelevant_Tweets
719	320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/TH...	Not relevant	NotRelevant_Tweets
453	(AP News) Landslide in Norway leaves 10 injure...	Informative_Tweets_during_disaster	Informative_Tweets
185	RT @haveigotnews: After four days of 2021, mes...	Recovery_Tweets_during_disaster	Precautionary_Tweets
999	TAE continues Flood Warning for Ochlockonee Ri...	Precautionary_Tweets_during_disaster	Recovery_Tweets
38	(AP News) Magnitude 3.6 earthquake jolts San F...	Informative_Tweets_during_disaster	Informative_Tweets
504	Don't forget we have K12 Survival Kits in stoc...	Recovery_Tweets_during_disaster	Precautionary_Tweets
418	RT @worldnewsdotcom: #Norway landslide buries ...	Recovery_Tweets_during_disaster	Precautionary_Tweets
107	RT @gulfknwt_myfc: Behalf from this donation w...	Recovery_Tweets_during_disaster	Precautionary_Tweets
809	RT @FamNeeds2Know: The Risk Hiding In Your Fam...	Precautionary_Tweets_during_disaster	Recovery_Tweets
1185	320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/el...	Not relevant	NotRelevant_Tweets
1063	Map of felt reports received so far following ...	Recovery_Tweets_during_disaster	Precautionary_Tweets
1197	'There is no victory at bargain basement price...	Not relevant	NotRelevant_Tweets
850	Landslide injures at least 10 in Norwegian tow...	Informative_Tweets_during_disaster	Informative_Tweets
912	RT @ayssz: Assalamualaikum, my bestfriend in S...	Informative_Tweets_during_disaster	Precautionary_Tweets
677	RT @SheetalDevi20: #DisasterRelief\n#DisasterR...	Recovery_Tweets_during_disaster	Precautionary_Tweets
287	#Support our #WolfPackAuthors anthology, Howli...	Recovery_Tweets_during_disaster	Precautionary_Tweets

Fig. 3 Result of the Naïve Bayes Algorithm

Table 1. Inferences of Naïve Bayes Algorithm

Varying Test Set Size		25 %	37%	50%
METRICS	Accuracy	0.97689	0.95545	0.93564
	Precision	0.98392	0.96005	0.93920
	Recall	0.97689	0.95545	0.93564
	F1- Score	0.78306	0.76539	0.74849
Time for Execution		0.557 secs	0.467 secs	0.993 secs

```
********************Logistic Regression*************************
                                                    text                              label               predict
953   RT @littlelouv: last new yearâs eve, we were...   Informative_Tweets_during_disaster   Informative_Tweets
663   RT @HFI1995: Relief items delivered by @Humani...      Recovery_Tweets_during_disaster      Recovery_Tweets
957   Know your area and your flood risk. Flash floo...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
15    RT @MMC_Global: We believe that climate change...      Recovery_Tweets_during_disaster      Recovery_Tweets
842   What if we started practising now for differen...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
215   North Queensland on alert for Cyclone Imogen |...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
1117  320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/vA...                          Not relevant   NotRelevant_Tweets
1204  Brass and Copper in Cataclysm &amp; AfterShock...                          Not relevant   NotRelevant_Tweets
433   RT @FamNeeds2Know: The Risk Hiding In Your Fam...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
534   SHV extends Flood Warning for the Ouachita Riv...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
111   RT @MikeKrueger7: There may be an isolated thu...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
646   Cyclone Imogen makes landfall near Karumba and...    Informative_Tweets_during_disaster   Informative_Tweets
275   Reliefline provides a wide range of shelter to...      Recovery_Tweets_during_disaster      Recovery_Tweets
452   #NaturalDisasters are estimated to have killed...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
1121  @KJForDays I'm seeing them and Issues at after...                          Not relevant   NotRelevant_Tweets
281   RT @DeMarcoWriter: #Support our #WolfPackAutho...      Recovery_Tweets_during_disaster      Recovery_Tweets
802                  Aftershock https://t.co/jV8ppKhJY7                          Not relevant   NotRelevant_Tweets
1057  RT @severeweatherEU: 2020 has ended with a new...    Informative_Tweets_during_disaster   Informative_Tweets
168   Body believed to be third flood victim found 1...      Recovery_Tweets_during_disaster      Recovery_Tweets
1179  Aftershock was the most terrifying best roller...                          Not relevant   NotRelevant_Tweets
337   'The first man gets the oyster the second man ...                          Not relevant   NotRelevant_Tweets
483   ð· My Goal is to Set Up The MRT Emergency Ce...     Recovery_Tweets_during_disaster      Recovery_Tweets
725   &gt;&gt; $15 Aftershock : Protect Yourself and...                          Not relevant   NotRelevant_Tweets
1099  RT @khalsaaid_india: `ASSAM FLOOD RELIEF, 2019...      Recovery_Tweets_during_disaster      Recovery_Tweets
49    (AP News) Landslide in Norway leaves 10 injure...    Informative_Tweets_during_disaster   Informative_Tweets
```

Fig. 4 Output for Logistic Regression Algorithm

Table 2. Inferences of Logistic Regression Algorithm

Varying Test Set Size		25 %	37%	50%
METRICS	Accuracy	0.98019	0.97104	0.92807
	Precision	0.98132	0.97247	0.92928
	Recall	0.98019	0.97104	0.92807
	F1- Score	0.97847	0.97023	0.92687
Time for Execution		1.475 secs	1.202 secs	1.674 secs

```
*******************************RandomForest Classifier*******************************
                                                    text                              label               predict
718   320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/TH...                          Not relevant   NotRelevant_Tweets
1115                 Aftershock https://t.co/xMWODFMtUI                          Not relevant   NotRelevant_Tweets
1102  The purpose of the #RedCross is to meet the di...      Recovery_Tweets_during_disaster      Recovery_Tweets
39    RT @EIB: Together w/ @UNOPS, EIB works towards...    Informative_Tweets_during_disaster   Informative_Tweets
319   Aftershock â¢ (2010) Fullâ¢ Streaming - YouT...                         Not relevant   NotRelevant_Tweets
1092  Glad to have played a small part in the relief...      Recovery_Tweets_during_disaster      Recovery_Tweets
976   Body believed to be third flood victim found 1...      Recovery_Tweets_during_disaster      Recovery_Tweets
272   I wonder if @fema or @RedCross could replicate...      Recovery_Tweets_during_disaster      Recovery_Tweets
800   Brass and Copper in Cataclysm &amp; AfterShock...                          Not relevant   NotRelevant_Tweets
1021  RT @insan_honey: Deeply saddened to know about...    Informative_Tweets_during_disaster   Informative_Tweets
947   Meadow Street Coventry flood protection and tr...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
278   RT @RuchiKhurana10: Till date, many #DisasterR...      Recovery_Tweets_during_disaster      Recovery_Tweets
134   NWS Little Rock AR issued a Flood Warning for ...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
746   Bedroom clean  bathroom clean  laundry done .....                          Not relevant   NotRelevant_Tweets
409   RT @FamNeeds2Know: Learn how to spot the signs...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
175   I'm stucked on a heavy traffic jam due to the ...      Recovery_Tweets_during_disaster      Recovery_Tweets
9     RT @scienceat60: Widespread devastations have ...    Informative_Tweets_during_disaster   Informative_Tweets
415   RT @Mohamme27859094: Any big natural disasters...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
173   RT @haveigotnews: After four days of 2021, mes...      Recovery_Tweets_during_disaster      Recovery_Tweets
348   GEARS OF WAR 1!(preview member) Come chat! XB1...                          Not relevant   NotRelevant_Tweets
584   RT @Stansberry: PwC becomes latest to highligh...    Informative_Tweets_during_disaster   Informative_Tweets
665   RT @HFI1995: A trailer load of relief items de...      Recovery_Tweets_during_disaster      Recovery_Tweets
1099  RT @khalsaaid_india: `ASSAM FLOOD RELIEF, 2019...      Recovery_Tweets_during_disaster      Recovery_Tweets
1203                         @OnFireAnders I love you bb                          Not relevant   NotRelevant_Tweets
1065  RT @HFI1995: Relief items delivered by @Humani...      Recovery_Tweets_during_disaster      Recovery_Tweets
166   RT @melisa_idris: The annual monsoon in recent...    Informative_Tweets_during_disaster   Informative_Tweets
1175  @afterShock_DeLo im speaking from someone that...                          Not relevant   NotRelevant_Tweets
542   RT @codeofvets: Army Vietnam Era Vet Michael 1...    Informative_Tweets_during_disaster   Informative_Tweets
```

Fig. 5 Output for the Random Forest Algorithm

Table 3. Inferences of Random Forest Algorithm

Varying Test Set Size		25 %	37%	50%
METRICS	Accuracy	0.9702	0.9665	0.8993
	Precision	0.9714	0.9671	0.9023
	Recall	0.9702	0.9665	0.8993
	F1- Score	0.9762	0.9727	0.9191
Time for Execution		2.052 secs	1.582 secs	1.161 secs

```
                                                    text                              label               predict
397   @JadeForMKX You should be happy I don't use Af...                          Not relevant   NotRelevant_Tweets
434   RT @StayYoungMedia: Preparing for medical emer...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
909   Its key to know what is urgently needed, capac...      Recovery_Tweets_during_disaster      Recovery_Tweets
167   Neptune Commercial Flood offers you greater li...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
876   Geo-Hazard Report (December 28, 2020)\nWeekly ...    Informative_Tweets_during_disaster   Informative_Tweets
782   @KJForDays I'm seeing them and Issues at after...                          Not relevant   NotRelevant_Tweets
1134  Praise God that we have ministry that tells it...                          Not relevant   NotRelevant_Tweets
992   RT @littlelouv: last new yearâs eve, we were...   Informative_Tweets_during_disaster   Informative_Tweets
934   Monitoring the local river levels in Braithwai...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
433   RT @FamNeeds2Know: The Risk Hiding In Your Fam...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
670   I support Croatia Earthquake Fund - 2020 Disas...      Recovery_Tweets_during_disaster      Recovery_Tweets
917   @saradibazmi Prayers for your relatives and al...      Recovery_Tweets_during_disaster      Recovery_Tweets
714   Stop saying 'I Wish' and start saying 'I Will'...                          Not relevant   NotRelevant_Tweets
576   RT @NST_Online: #NSTnation Perak flood evacuee...      Recovery_Tweets_during_disaster      Recovery_Tweets
888   RT @Bob_Mayer: These are key documents you nee...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
1193  320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/TH...                          Not relevant   NotRelevant_Tweets
375   Stop saying 'I Wish' and start saying 'I Will'...                          Not relevant   NotRelevant_Tweets
56    RT @BiIndia: @pabsgill 8. Windstorms in Europe...    Informative_Tweets_during_disaster   Informative_Tweets
1185  320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/e1...                          Not relevant   NotRelevant_Tweets
660   In this week's City News... Support Local Busi...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
302   @afterShock_DeLo im speaking from someone that...                          Not relevant   NotRelevant_Tweets
368   'The harder the conflict the more glorious the...                          Not relevant   NotRelevant_Tweets
270   RT @HFI1995: A trailer load of relief items de...      Recovery_Tweets_during_disaster      Recovery_Tweets
406   RT @BRINKNewsNow: "We have an opportunity to r...  Precautionary_Tweets_during_disaster  Precautionary_Tweets
```

Fig. 6 Output the Support Vector Machine Algorithm

Table 4. Inferences of the Support Vector Machine

Varying Test Set Size		25 %	37%	50%
METRICS	Accuracy	0.96039	0.94877	0.93894
	Precision	0.96102	0.95276	0.94432
	Recall	0.96039	0.94877	0.93894
	F1- Score	0.95962	0.95794	0.751736
Time for Execution		1.232 secs	3.852 secs	1.291 secs

```
                                                    text                              label               predict
481   Billions lost in damages due to #ExtremeWeathe...    Informative_Tweets_during_disaster   Informative_Tweets
1057  RT @severeweatherEU: 2020 has ended with a new...    Informative_Tweets_during_disaster   Informative_Tweets
280   We collect donations for people who are made h...      Recovery_Tweets_during_disaster  Precautionary_Tweets
122   RT @opcmiaintl: @BuildTogether_ : Investing in...      Recovery_Tweets_during_disaster  Precautionary_Tweets
388   Sometimes you face difficulties not because yo...                          Not relevant   NotRelevant_Tweets
463   RT @BiIndia: @pabsgill 5. Floods in India â ...    Informative_Tweets_during_disaster   Informative_Tweets
910   VOX: Fires, floods, hurricanes, and locusts: 2...    Informative_Tweets_during_disaster   Informative_Tweets
323   Sometimes you face difficulties not because yo...                          Not relevant   NotRelevant_Tweets
1014  RT @OurRadiantEarth: Only 4â£ weeks left to h...  Precautionary_Tweets_during_disaster      Recovery_Tweets
163   @hollyhopkins_ Particularly like the 'flood de...  Precautionary_Tweets_during_disaster      Recovery_Tweets
183   RT @ayssz: Assalamualaikum, my bestfriend in S...      Recovery_Tweets_during_disaster  Precautionary_Tweets
557   RT @malaymail: Flood victims in Pahang, Johor ...      Recovery_Tweets_during_disaster  Precautionary_Tweets
290   RT @Opoyis: The Dzoukou range, a prominent tou...      Recovery_Tweets_during_disaster  Precautionary_Tweets
568   RT @jongjinsbeanie: Guys pls help pray for mal...    Informative_Tweets_during_disaster   Informative_Tweets
651   Cyclone warning for Gulf country and far north...  Precautionary_Tweets_during_disaster      Recovery_Tweets
18    RT @nordicreporter: So far nine injured, none ...    Informative_Tweets_during_disaster   Informative_Tweets
719   320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/TH...                          Not relevant   NotRelevant_Tweets
1035  .@FijiAG Fiji ð«ð¯ today commences  its re...      Recovery_Tweets_during_disaster   Informative_Tweets
1124  #WisdomWed BONUS - 5 Minute Daily Habits that ...                          Not relevant   NotRelevant_Tweets
784   320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/TH...                          Not relevant   NotRelevant_Tweets
718   320 [IR] ICEMOON [AFTERSHOCK] | http://t.co/TH...                          Not relevant   NotRelevant_Tweets
14    RT @worldnewsdotcom: #Norway landslide buries ...      Recovery_Tweets_during_disaster  Precautionary_Tweets
357   @thrillhho jsyk I haven't stopped thinking abt...                          Not relevant   NotRelevant_Tweets
906   #DisasterPreparedness #earthquake #tsunami: Ge...  Precautionary_Tweets_during_disaster      Recovery_Tweets
50    (AP News) Quake aftershocks keep people out of...      Recovery_Tweets_during_disaster  Precautionary_Tweets
255   Map of felt reports received so far following ...      Recovery_Tweets_during_disaster  Precautionary_Tweets
352                          you wrecked my whole world                          Not relevant   NotRelevant_Tweets
468   RT @BiIndia: @pabsgill From the bushfires in A...    Informative_Tweets_during_disaster   Informative_Tweets
41    RT @taymarch: Landslide injures at least 10 in...    Informative_Tweets_during_disaster   Informative_Tweets
```

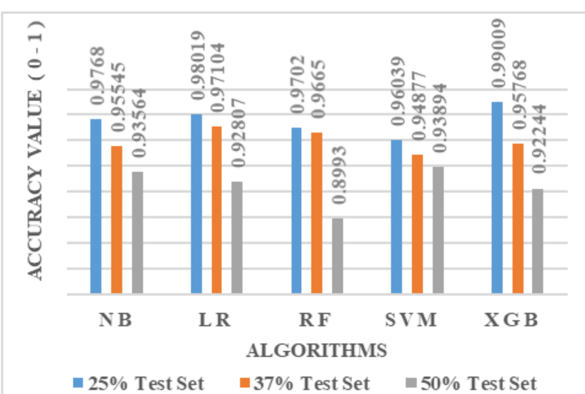
Fig. 7 Output for the Extreme Gradient Boosting Algorithm

Table 5. Inferences of the Extreme Gradient Boosting Algorithm

Varying Test Set Size		25 %	37%	50%
METRICS	Accuracy	0.99009	0.95768	0.92244
	Precision	0.99023	0.95845	0.92321
	Recall	0.99009	0.95768	0.92244
	F1- Score	0.98977	0.96569	0.93618
Time for Execution		3.760 secs	3.344 secs	2.372 secs

The above mentioned ML algorithms are implemented using the trained_model which accepts various parameters and gives the result of efficiency measures such as accuracy value, precision, recall value and the F1–score. Comparison between theose algorithms with 25%, 37%, 50% test data set are presented in Figures 8-12.

From the Figure 8, illustrated that XGBoost has the highest accuracy 0.99 with 25% test set and Random Forest with 50% test set has the least accuracy 0.89. From Figure 9, it can be observed that XGBoost has the highest precision 0.99 with 25% test set and Random Forest with 50% test set has the least precision 0.90. From Figure 10, it is concluded that XGBoost has the highest recall score 0.99 with 25% test set and Random Forest with 50% test set has the least recall score 0.89. From Figure 11, we can conclude that XGBoost has the highest F1-score 0.98 with 25% test set and Naïve Bayes with 50% test set has the least F1-score 0.74. From the Figure 12, we can observe that SVM takes the longest time of 3.852 mins with 37% test set and Naïve Bayes with 37% test set takes the least time of 0.467mins. From figures 8, 9, 10, 11, and 12, it is concluded that XGBoost is efficient in terms of accuracy, precision, recall score and F1-score whereas Naïve Bayes is efficient in terms of execution time.



t.

Fig. 8 Comparison based on Accuracy

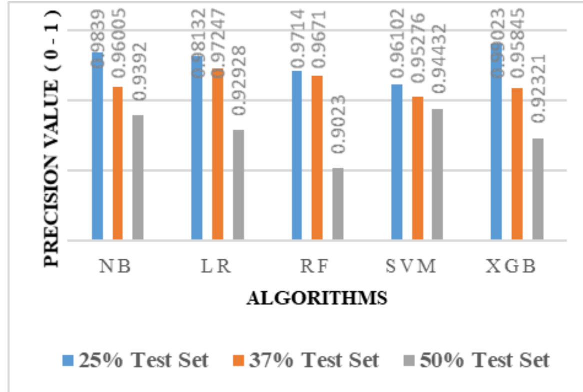


Fig. 9 Comparison based on Precision

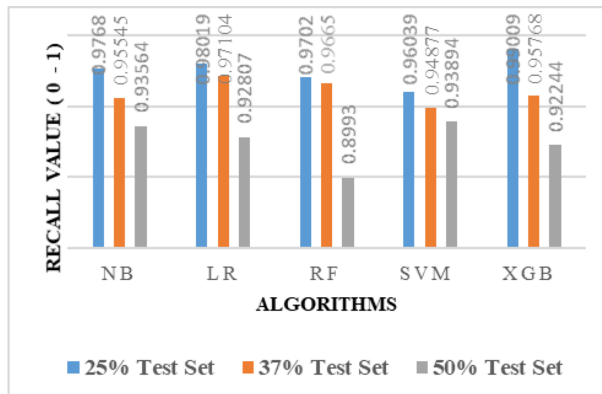


Fig. 10 Comparison based on Recall

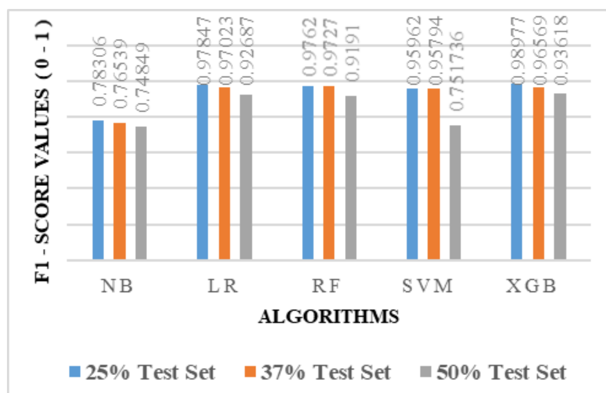


Fig. 11 Comparison based on F1-Score

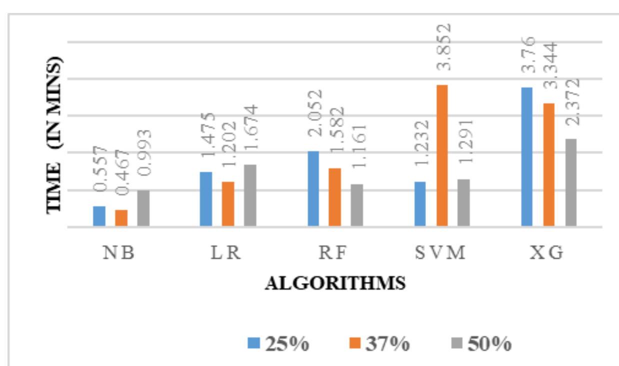


Fig. 12 Comparison based on Time for Execution

V. CONCLUTIONS

The paper enabled the visualisation of the emotional reaction of the population in case of a catastrophic situations, by analyzing disaster-tweets. It is recommended to identify the influential people in each country and community before the disaster occurs, in order to disseminate the information necessary to face the disaster. Help can be provided by providing rescuing community initiatives with clear, and reliable information. Interpretation of Twitter huge data might be a useful tool for determining the mental health of a community, estimating the amount of people harmed, estimating the items and effects of natural disasters, and visualising epidemics. As a result, many entities such as governments, non-governmental enterprises, humanitarian workers, and healthcare organizations can use this data to make decisions and carry out appropriate measures. Tweets are ranked through keyword analysis and comparative research on five machine learning algorithms including Naive Bayes, Logistic Regression, Random Forest, SVM, and XGBoost. Analysis tools are used to compare the above mentioned five algorithms. The performance of five classification algorithms is evaluated in terms of accuracy rate, accuracy rate, recall rate and F1 measurement. On the basis of the results, it may be concluded that social media data can be a valuable addition during disaster management. Analysis can be made by using the other social media networks such as YouTube and Facebook to increase the database's information. Multi-label classification of English and multilingual tweets is also essential for extracting relevant information that ultimately helps improve situational awareness. The real-time system can detect and filter disaster-related tweets, which can be further developed for effective and efficient disaster response management organizations. A prediction model that can predict disaster trends in diverse places is an interesting open problem.

REFERENCES

[1] Imran, Muhammad & Elbassuoni, Shady & Castillo, Carlos & Diaz, Fernando & Meier, Patrick, "Practical Extraction of Disaster-Relevant Information from Social Media", Proceedings of the 22nd International Conference on World Wide Web Social Web for Disaster Management (SWDM), 2013, doi: 10.1145/2487788.2488109.

[2] Kaur, A. "Analyzing Twitter Feeds to Facilitate Crises Informatics and Disaster Response During Mass Emergencies" Dissertation M.Sc. in Computing (Data Analytics), TU Dublin, 2019. URL: https://arrow.tudublin.ie/scschcomdis/166/

[3] A. Sen, K. Rudra and S. Ghosh, "Extracting Situational Awareness from Microblogs During Disaster Events," 7th International Conference on Communication Systems and Networks, 2015, pp. 1-6, doi: 10.1109/COMSNETS.2015.7098720.

[4] H. To, S. Agrawal, S. H. Kim and C. Shahabi, "On Identifying Disaster-Related Tweets: Matching-Based or Learning-Based?," 2017 IEEE Third International Conference on Multimedia Big Data (BigMM), 2017, pp. 330-337, doi: 10.1109/BigMM.2017.82.

[5] Parilla-Ferrer, Beverly & Fernandez, Proceso & IV, Jaime, "Automatic Classification of Disaster-Related Tweets", International conference on Innovative Engineering Technologies, 2014.

[6] Goswami, Shriya and Raychaudhuri, Debaditya, "Identification of Disaster-Related Tweets Using Natural Language Processing", International Conference on Recent Trends in Artificial Intelligence, IOT, Smart Cities & Applications, 2020.

[7] B. Abedin, A. Babar and A. Abbasi, "Characterization of the Use of Social Media in Natural Disasters: A Systematic Review," 2014 IEEE Fourth International Conference on Big Data and Cloud Computing, 2014, pp. 449-454, doi: 10.1109/BDCloud.2014.17.

[8] Huang Q, Xiao Y "Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery", ISPRS Journal of Photogrammetry and Remote Sensing, 2015, pp. 1549–1568.

[9] Gopnarayan, Archan, Deshpande, Sachin, "Tweets Analysis for Disaster Management: Preparedness, Emergency Response, Impact, and Recovery", Innovative Data Communication Technologies and Application Book, Jan 2020, pp -760 – 764.

[10] Singh, Jyoti & Dwivedi, Yogesh & Rana, Nripendra & Kumar, Abhinav & Kapoor, Kawal., "Event Classification and Location Prediction from Tweets during Disasters.", Annals of Operations Research, 2019, Vol.283, doi: 10.1007/s10479-017-2522-3.

[11] T. Funayama, Y. Yamamoto, M. Tomita, O. Uchida and Y. Kajita, "Disaster Mitigation Support System using Twitter and GIS," 2014 Twelfth International Conference on ICT and Knowledge Engineering, 2014, pp. 18-23, doi: 10.1109/ICTKE.2014.7001528.

[12] H. Shekhar and S. Setty, "Disaster Analysis through Tweets," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015, pp. 1719-1723, doi: 10.1109/ICACCI.2015. 7275861.

[13] Boaz, John & Ybañez, Michael & De Leon, Marlene & Estuar, Ma.Regina.. "Understanding the Behavior of Filipino Twitter Users during Disaster" GSTF Journal on Computing (JoC), 2013, Vol:3, doi: 10.7603/s40601-013-0007-z.

[14] Kumar, Shamanth & Barbier, Geoffrey & Abbasi, Mohammad Ali & Liu, Huan, "TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief.", Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011.

[15] Stowe, Kevin & Paul, Michael & Palmer, Martha & Palen, Leysia & Anderson, Kenneth, "Identifying and Categorizing Disaster-Related Tweets", Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media, 2016, pp.1-6, doi: 10.18653/v1/W16-6201.

[16] Acar, Adam & Muraki, Yuya, "Twitter for Crisis Communication: Lessons learned from Japan's Tsunami Disaster. IJWBC", International Journal of Web Based Communities, 2011, Vol.7, pp.392-402, doi: 10.1504/ IJWBC.2011.041206.

[17] Son Doan, Bao-Khanh Ho Vo, Nigel Collier, "An Analysis of Twitter Messages in the 2011 Tohoku Earthquake", International Conference on Electronic Healthcare, 2012, Volume: 91, pp 58-66, ISBN: 978-3-642-29261-3

[18] Kanhabua, Nattiya & Nejdl W, "Understanding the Diversity of Tweets in the Time of Outbreaks", Proceedings of the 22nd International Conference on World Wide Web, 2013, pp.1335-1342.

[19] Go,A., Richa, B. and Lei,H.(2009). "Twitter Sentiment Classification using Distant Supervision. In Processing", 2009, pp.1–6. URL: http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

[20] Bhat, F., Oussalah, M., Challis, K., & Schnier, T., "A Software System for Data Mining with Twitter", Proceedings of the 2011 IEEE 10th International Conference on Cybernetic Intelligent Systems (CIS) London, United Kingdom, 2011, pp.139–144, doi: IEEE.10.1109/CIS.2011.6169149.