



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: IV Month of publication: April 2022

DOI: <https://doi.org/10.22214/ijraset.2022.42006>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Informing Disasters Management by Using Twitter tweets

Ashish Vijay Naikwade¹, Sachin N. Deshmukh²

^{1, 2}Department of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University Aurangabad-431001, India

Abstract: Recently social media plays a major role and providing information during disasters. This paper mainly focuses on how people used social media, especially Twitter, in response to the country's worst flood, Earthquake that had occurred recently.

And these tweets collecting analyzed using machine learning algorithms such as Naïve Bayes, Random Forests, Decision Tree, sentiment Analysis .during the disaster social media provides a surplus of information which includes information about the natural disaster, affected people's emotions, and relief efforts.

And collect the tweets relating to disasters and build the sentimental classifier to categorize the user's emotions during disaster based on various distress levels.

Various analysis techniques are applied in collecting tweets.

Keywords: Twitter Analysis, Sentiment Analysis, Random Forests, Naive Bayes, Decision Tree.

I. INTRODUCTION

Microblogging can form a short chat that allows the user to share chat messages by using the internet. There are many microblogging services available such as Twitter, Facebook, what's app, etc.

But my focus on this research paper is only on Twitter tweets, which allows only short message means tweets these tweets are generally 140 characters or less.

Microblogging is mainly focused on sharing information and tracking general people's opinions during any social, political, Natural Disaster, etc.

During any event of a natural disaster, social media has gained a lot of attention and additional crisis communication. And Twitter is most popular in microblogging sites by users sharing the messages, photos, videos for that disaster crises site globally. Twitter is real-time data generated by the user community. The behavior and emotions that users express from natural disaster crises site to be distributed to users across the globe.

II. RELATED WORK

Twitter plays a major role transmit information during natural disasters. The various methods and techniques are used to extract data from Twitter are described below

The work of Starbird, chatter used Twitter during the red river flood in central North America [1] They have collected all the tweets regarding the Red River Flood using the keyword #redriver.

And each tweet contains such as a tweet, retweet, geographical location approximation distance from the event, and tweets analysis under one of the following categories hopeful, fear, support, and humor.

The work of T.S.Aisha[2] mainly focuses on the 2014 Malaysia Flood that had fully destroyed Malaysia. the number of tweets messages during the flood period was very high as compared to the normal period. This mainly focuses on the satisfaction gained by the Twitter user from sharing information during the flood.

Most useful data for this work in [8],[9], and[10] analysis of user emotion and distributing the geographical location by using Twitter tweets.

Twitter is the best channel for a communication system for natural disasters [13]. Another people's sentiment analysis for disaster victim emotion and relief measurement [14].

Following Related Work as below

No.	Disaster	Finding	Country	Ref.no
1	Flood	Starbird et.al collect the Twitter tweets during red river valley in north America .and tweets were analyzed and categorized including fear, support, hopeful	North America	(K. Starbird, et al. Chatter 2010). [1]
2	Flood	Aisha et.al analyzed Malaysia Flood in 2014 .analyzed before or after the flood response of the users as compared to the normal period.	Malaysia	(Aisha et al 2014).[2]
3	Earthquake	The work of baek et.al collected tweets of great east japan earthquake-related tweets and analyzed the anxiety of the public during the time of the earthquake.	Japan	(Baek et.al 2013).[15]
4	Typhoon	Takahashi et.al mainly focus typhoon haiyan in Philippines. Collect the tweets and observe the geographical location and different types of users.	Philippines	(Takahashi 2015).[16]
5	Flood and Grassfire	S.Vieweg mainly focuses on the red river flood and Grassfire .mainly focus on how to twitter effective contribution to situation awareness about the natural disaster.	Oklahoma	(S. Vieweg 2010)[7]
6	Hurricane	J.B.Lee worked for sentiment analysis and anxiety in the people before and after the hurricane.	Philippines	(J.B.Lee 2014).[10]
7	Hurricane	B.mandel, A.culotto , A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J.rodrique work for sentiment analysis and anxiety in the people before and after during hurricane.	Philippines	(B.mandel , A.culotto , A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J.rodrique 2012).[14]
8	Hurricane	S. Kumar, G. Barbier, M. A. Abbasi, and H. Liu, work for tweet tracker and analysis tool for hurricane and disaster relief during the hurricane.	Philippines	(S. Kumar, G. Barbier, M. A. Abbasi, and H. Liu 2011).[18]
9	Flood	V. Lorini, C. Castillo, F. Dottori, M. Kalas, D. Nappo, P. Salamon works show that EFAS flood forecasts can be successfully used to inform satellite imagery acquisition and social media monitoring, improving the timeliness and relevance of the information collection and analysis and prediction of natural disaster-affected areas.	-	(V. Lorini, C. Castillo, F. Dottori, M. Kalas, D. Nappo, P. Salamon,2019).[19]
10	Flood, earthquake	Alexander (2014) identified seven ways in which social media can be used in disaster risk reduction and crisis response.	-	(Alexander DE 2014).[20]

III. METHODS

A. Naive Bayes

Naive Bayes is a classification technique that gives prime importance to conditional independency between predictors. It is based on the assumption that all the attributes are conditionally independent. It is a statistical classifier that performs probabilistic prediction that predicts the class membership probabilities. Classification is based on Bayes Theorem. It is very useful for high-dimensional datasets and they are easy to build. The posterior probability $P(c|x)$ is calculated from the likelihood $P(x|c)$, class prior probability $P(c)$, and predictor prior probability $P(x)$. $P(c|x) = (P(x|c) * P(c))/P(x)$. Features of Naive Bayes are

- 1) Prediction of a class for a test set is simple and easy.
- 2) Multiclass prediction is a key feature of Naive Bayes.
- 3) Class conditional independency is a limitation of Naive Bayes. This is because in real life it is difficult to get completely independent class predictors.
- 4) It works well for categorical data than numerical data

B. Decision Tree

The decision tree is a graph .and a decision tree is a branching method to exhibit every possible output for the decision. The tree is constructed in a top-down recursive divide and conquer manner. Initially, all the training samples are placed as the root element. They are partitioned recursively based on the selected attribute. The decision tree mainly consists of 2 nodes namely leaf nodes and decision nodes. A decision node or internal node contains 2 or more subspaces or branches, where each denotes the test on a particular attribute. Leaf node contains the resulting class label, classification, or decision. The decision tree is a classification technique that can handle both numerical and categorical data.

C. Random Forests

Random Forests is a machine learning algorithm used for both classification and regression. It is an ensemble learning method that involves the construction of multiple decision trees. Collections of decision tree classifiers are called Forests. Individual decision trees are generated during the training time by the random selection of attributes in each node, which will determine the split. Each tree will give a classification result that is, all the decision trees will give their vote individually. The forests choose the most popular class. Benefits of Random Forests classification algorithm are:

- 1) Random Forests can handle data that contain outliers.
- 2) It is not mandatory that data need to be pre-processed. They can handle missing values automatically.
- 3) They have a high tolerance to overtraining.
- 4) Random Forests can be built in a faster manner.
- 5) Takes less time to predict the output.

IV. COMPARISON OF PREVIOUS WORK ANALYSIS

Sr.no	Disaster type	Ref.no	Finding	Comparison
1)	Flood	(K. Starbird, et al. Chatter 2010). [1]	Starbird et.al collected the Twitter tweets during red river valley in north America and tweets were analyzed and categorized including fear, support, hopeful	As compared to my work analysis and identify most people's fear, anger, helpful emotion and find the positive-negative, and neutral tweets
2)	Earthquake	(Baek et.al 2013).[15]	The work of baek et.al collected tweets of great east japan earthquake-related tweets and analyzed the anxiety of the public during the time of the earthquake.	During earthquake lots of positive, negative impacts of classified and categorized in different aspect views and analysis

3)	Flood	(V. Lorini, C. Castillo, F. Dottori, M. Kalas, D. Nappo, P. Salamon,2019).[19]	V. Lorini, C. Castillo, F. Dottori, M. Kalas, D. Nappo, P. Salamon works show that EFAS flood forecasts can be successfully used to inform satellite imagery acquisition and social media monitoring, improving the timeliness and relevance of the information collection and analysis and prediction of natural disaster-affected areas. By using satellite images for the particular affected area to find and analysis	As compared to easy identify the disaster affected areas using classification methods identify twitter tweets and affected areas and provide help services to that areas priority on basis of that area.
4)	Typhoon	(Takahashi 2015).[16]	Takahashi et.al mainly focus typhoon haiyan in Philippines. Collect the tweets and observe the geographical location and different types of users.	As compared to similar to the main focus on floodwater density high areas or low areas to identify and observe geographical location.
5)	Flood and Grassfire	(S. Vieweg 2010)[7]	S.Vieweg mainly focuses on the red river flood and Grassfire .mainly focus on how to twitter effective contribution to situation awareness about the natural disaster.	My work is to help disaster victims people to categorize by using support help groups, NGOs, and other people's help and suggestion to analyze.

V. ANALYSIS SYSTEM:

- 1) *Tweeter Data Collection*: Data collection with the help of data API. This API collect live or flood, Earthquake disaster-related keywords collect disaster related tweet, retweet, geographical location, likes, followers, etc.
- 2) *Sorting*: Sorting data in different disaster categories like a flood, Earthquake, Forest fire than to sort data in different attributes needed for help, Relief measure NGO, complaints, suggestion, and most Affected area and worst area and low affected areas to analysis and identify and sort, etc.
- 3) *Preprocessing*: After that collect the pure data remove a bag of words or unused symbols, numbers, the text then collects the pure data and is ready to preprocess data using
- 4) *Classification Data*: After that collect data the first step is to segregate data .like floods, earthquakes, Forest fires. After that classify different aspect-wise tweets and sentiment analysis of users.
- 5) *Analysis Data*: By using Methods analysis for result showing output. And sentiment analysis of people. And frequency determines and analysis of geographical area accuracy and result.

VI. DATASET DESCRIPTION

Twitter social media is the source of the dataset. The dataset of the text mining data set collects tweets, retweets, likes, followers, and comparisons on a count of tweets. disaster-related data for flood, forest fires, Earthquake, and sentiment analysis of people.

VII. CONCLUSION

This paper presents a systematic Literature review on disaster management. And used a total of 20 research papers on this field and reviewed their proposed system and manage the disaster victims and sentiment analysis of disaster victims save people's lives in the future situation how to manage and precautions that type of situation and Twitter helps to identify the most disaster-affected area and low affected area are easily find and provide help and Twitter Social media plays a major role in Natural Disaster Management help for society and government for quick action in future work.

REFERENCES

- [1] K. Starbird, et al. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work 2010; 241–250
- [2] T.S. Aisha, et al. Exploring the use of social media during the 2014 flood in Malaysia. *Procedia-Social and Behavioral Sciences* 2015; 211: 931–937.
- [3] Y. Huang, et al. A scalable system for community discovery in Twitter during hurricane sandy. In: *Cluster, Cloud, and Grid Computing (CCGrid)*, 2014 14th IEEE/ACM International Symposium on 2014; 893–899.
- [4] K. Kandasamy, P. Korothe. An integrated approach to spam classification on Twitter using URL analysis, natural language processing, and machine learning techniques. In: *Electrical, Electronics and Computer Science (SCEECS)*, 2014 IEEE Students Conference on 2014; 1–5.
- [5] H. Dong, et al. Social media data analytics applied to hurricane sandy. In: *Social computing (SocialCom)*, 2013 international conference on IEEE. 2013; 963–966.
- [6] S.R. Kalmegh. Comparative analysis of weka data mining algorithm random forest, random tree, and bad tree for classification of indigenous news data. *International Journal of Emerging Technology and Advanced Engineering* 2015;5(1): 507–517
- [7] S. Vieweg, et al. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 2010; 1079–1088.
- [8] R. E. Cohn and W. A. Wallace, *The Role of Emotion in Organizational Response to a Disaster: An Ethnographic Analysis of Videotapes of the Exxon Valdez accident*. Natural Hazards Research and Applications Information Center, Institute of Behavioral Science, University of Colorado, 1992
- [9] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue, “A demographic analysis of online sentiment during hurricane irene,” in *Proceedings of the Second Workshop on Language in Social Media*. Association for Computational Linguistics, 2012, pp. 27–36.
- [10] J. B. Lee, M. Ybanez, M. M. De Leon, and M. R. E. Estuar, “Under- ~ standing the behavior of Filipino Twitter users during the disaster,” *Journal on Computing (JoC)*, vol. 3, no. 2, 2014.
- [11] S. Srivastava. Weka: a tool for data preprocessing, classification, ensemble, clustering, and association rule mining. *International Journal of Computer Applications* 2014; 88(10).
- [12] J. B. Lee, M. Ybanez, M. M. De Leon, and M. R. E. Estuar, “Under- ~ standing the behavior of Filipino Twitter users during the disaster,” *Journal on Computing (JoC)*, vol. 3, no. 2, 2014
- [13] J. Sutton, L. Palen, and I. Shklovski, “Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires,” in *Proceedings of the 5th International ISCRAM Conference*. Washington, DC, 2008, pp. 624–632.
- [14] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue, “A demographic analysis of online sentiment during hurricane irene,” in *Proceedings of the Second Workshop on Language in Social Media*. Association for Computational Linguistics, 2012, pp. 27–36.
- [15] S.J. Baek, et al. Disaster anxiety measurement and corpus-based content analysis of crisis communication. In: *Systems, Man, and Cybernetics (SMC)*, 2013 IEEE International Conference on 2013; 1789–1794.
- [16] B. Takahashi, et al. Communicating on Twitter during a disaster: An analysis of tweets during typhoon Haiyan in the Philippines. *Computers in Human Behavior* 2015; 50 392–398.
- [17] A. Bruns, Y.E. Liang. Tools and methods for capturing Twitter data during natural disasters. *First Monday* 2012; 17(4).
- [18] S. Kumar, G. Barbier, M. A. Abbasi, and H. Liu, “Tweettracker: An analysis tool for humanitarian and disaster relief.” in *ICWSM*, 2011
- [19] V. Lorini, C. Castillo, F. Dottori, M. Kalas, D. Nappo, P. Salamon, Integrating Social Media into a Pan-European Flood Awareness System: A Multilingual Approach, *Proceedings of the 16th ISCRAM Conference – València, Spain*, (2019)
- [20] Alexander DE 2014, Social Media in Disaster Risk Reduction and Crisis Management, *Science and Engineering Ethics*, vol. 20, no. 3, pp. 717–733



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)