Social Media Data Analytics Framework for Disaster Management

Sabih Ahmad Ansari[1*], Ahmad Talha Siddiqui[2]

[1,2]School of Computer science & Information Technology Maulana Azad National Urdu University, Hyderabad, Telangana, India

Corresponding: sabihahmad444@gmail.com

Abstract -- Social media plays a significant role within the propagation of information throughout disasters. This paper essentially contains an investigation identifying with anyway people of Chennai utilized Social media especially Twitter, in light of the nation's most exceedingly awful flood that had happened recently. The tweets are collected & analysed by various machine learning algorithms like Random Forests, Naive Bayes and call Tree. By comparison the performances of all the three, it had been found that Random Forests is that the best algorithmic rule that may be relied on, throughout a disaster. This paper conjointly targeted the sources of the Twitter messages to explore the foremost influential users of Chennai flood.

Keywords: Random Forests, Naive Bayes, call Tree

I. INTRODUCTION

Communication suggests that sharing or exchange of thoughts, messages, visuals or signals. It needs a sender and a receiver. Ancient communication techniques are uncomplicated and are elementary in nature. However, these methodologies were restricted to hurry and distance. This was inundated by trendy communication technologies. Social media or social networking sites play a serious role in trendy communication techniques [2]. These techniques appear to be abundant quicker and are used additional oft than ancient technologies. Trendy technology succeeded in making an additional international society by enabling the folks to speak or move directly with each other from everywhere the planet. In style social media includes Twitter, Facebook, LinkedIn, blogs, YouTube, wikis etc. they supply a platform for the users to share or post their views, and data etc. during this era Facebook is one the biggest social medium network on the web in terms of recognition. It's one amongst the most effective tool for connecting folks everywhere the planet. Twitter could be a real-time microblogging network. LinkedIn could be a network that's dedicated to professionals. YouTube provides a platform for uploading or sharing videos and music among friends and family. For the past few years, Twitter is one amongst the quality medium for news and communication. Twitter, a micro-blogging service during which registered users will post messages known as tweets[3]. Registered users will broadcast tweets, follow alternative user's tweets etc. Twitter messages are solely one hundred forty characters long and that they are known as tweets. Tweets are often printed from multiple platforms and devices. The advantage of Twitter is that anyone will follow anyone on public Twitter. Tweets are delivered to users in real time. So as to attach to a general topic, users will add a hashtag(#) as keywords to their posts. Hashtag could be a Meta character that is expressed as #keyword or #hash tag. The hashtag helps folks to pursue their interested topics terribly simply and quickly. The hashtag can offer tweets associated with a typical or explicit topic. For eg. The keyword #flood can retrieve all the tweets that contain the keyword flood. During the time of natural calamities or natural disasters like earthquake, flood or cyclone, wherever all the wired connections are universally down, these social media tools seem to be additional beneficial [5]. Twitter plays a serious role in the speedy propagation of knowledge throughout disasters [11]. It permits accessing or dispersing crucial data or breaking news directly from the affected areas. Social media analysis is the method of collecting data or information from social media sites like Twitter, Facebook, LinkedIn, and YouTube etc. Associate in nursing an analysis is finished to urge helpful or purposeful results from it. The Twitter analysis includes scrutinizing the tweets of a specific event or matter [3]. Many machine learning algorithms like call Tree, Support Vector Machine, Random Forests, Naive Bayes, supplying Regression etc. are often applied to the data for analysis. These algorithms facilitate in getting helpful outcomes from the data and conjointly facilitate in visualizing the data in an exceedingly additional precise manner.

For the past few years, many analyses of tweets are being disbursed on the disaster-affected areas across different countries, from this, it's evident that tweets generated

throughout the time of natural disasters are thought of, preponderantly, to urge important data relating to the accepted areas [8]. This work targeted tweets relating to the 2015 Chennai Flood. The tweets are analyzed so as to urge purposeful data from it. It had been found that the tweets obtained throughout the time of flood principally fall into 5 classes. It includes would like to facilitate, Relief Measures, categorical feeling, Complaints, and alternative. Tweets collected got as input to wood hen. It could be a free data processing analytical tool [4]. The tweets were classified victimization 3 machine learning algorithms like Random Forests, call trees, and Naive Bayes. The performance was disbursed in terms of exactness, Recall, and F-measure. It conjointly targeted the sources of Twitter messages to explore the foremost influential users. High 10 influential users of the dataset were known on the premise of the count of tweets, re-tweets and therefore the followers that every user has. the remainder of the paper is structured as follows. Section a pair of focuses on Literature Survey and connected works. A background study of Chennai flood is conferred in Section three. The methodology of machine learning algorithms like Random Forests, Naive Bayes and call tree is mentioned in Section four. It conjointly includes the analysis of the tweets generated from Twitter. Section five contains the results of research. This section principally contains the performance comparison of all the 3 algorithms supported exactness, Recall, and F-measure. It conjointly identifies the foremost influential users of Chennai flood on the premise of the count of tweets, re-tweets and therefore the followers that every user has.

II. LITERATURE SURVEY AND PAST WORKS

Twitter plays a serious role within the dissemination of knowledge throughout natural disasters. The assorted techniques that are applied to extract information from Twitter are delineated below.

The work of Star bird et.al reveals the utilization of Twitter throughout the Red Flood that occurred within the Red depression in central North America [3]. They collected all the tweets relating to the Red Flood victimization the keyword #redriver. Every tweet contains attributes like mastermind of the tweet, retweet, geographical location, affiliation and approximate distance from the event.

The work of Aisha et.al principally targeted on the 2014 Asian country Flood, that had ravaged Asian country totally. It had been the worst flood they'd witnessed within the recent decades. The quantity of Twitter messages throughout the flood amount was higher compared to the conventional amount.This principally focuses on the satisfaction gained by the Twitter user from sharing of knowledge throughout the flood.

Case Study of 2011 Thai Flood [5] studied the utilization of Twitter throughout the Thai Flood. Twitter vie a serious role in providing first-hand data throughout this disaster. #thaiflood tweets were collected and analysed victimization the keyword analysis and Rule primarily based approach. Tweets constitute one of the five classes that embrace Situational Awareness, requests for help, announcements for support, requests for data and alternative. They conjointly knew the influential users of the Thai Flood matter in Twitter, by scrutinizing the sources of the tweets. The bulk of the highest users were from government or non-government organizations associated with disaster.

The work of Vieweget.al [7] principally concerned 2 natural hazards Grassfires in Oklahoma and Red Flood. Information for analysis was obtained through Twitter Search API. The keywords #redriver and #red watercourse were want to get the Red Flood tweets whereas #oklahma, #okfire, #grassfire were want to retrieve Oklahoma fire tweets. Tweets obtained have some options related to it for the understanding. It includes geo—location, situational updates, location primarily based referencing, markedness and retweets. Geo-location contains data like address of street, city, highways, country and landmarks. The work principally targeted on however Twitter has contributed effciently for Situational Awareness.

The work of Beak et.al [7] concerned the reaction of public to a specific disaster. This principally targeted on the anxiety of public that arises throughout a disaster. The tweets relating to the nice East Japan Earthquake were collected. A strategy was developed for measurement and evaluating the general public anxiety throughout the time of disaster. Their work conjointly targeted on the kind of communication that occurred between the voters and government throughout the time of disaster.

The work of Takahashi et.al [7] principally targeted on the storm Haiyan within the Philippines. Here the tweets relating to the storm were collected victimization #Haiyan or #PrayforthePhilippines. A chi sq. analysis is finished on the tweets together with the attributes like geographic location, sort of user (Individuals, Celebrities, Journalists, and News organisations, Government or NGOs), time of use, request use and used report use. The association between every attribute and sort of social media use, is disbursed.

III. BACKGROUND: CHENNAI FLOODING

A large space of South Republic of India as well as Tamil Nadu and state receive most of its rain from northeast monsoon throughout the months of October to Gregorian calendar month. However in 2015 the quantity of rain they received exceeded the conventional monsoon by ninety per cent. This excessive rain was because of the El Nino development. At intervals a hundred years this was the worst

rain that the capital state of Tamil Nadu, Chennai had witnessed. This disaster has caused severe destruction and this was intense because of black-market urban development and poorly designed system. it had been calculable that over five hundred folks were dead and gone.8 million folks were displaced. The damages caused were within thevary of 15000 large integer to a hundred thousand large integer.

IV. METHODLGY

Flooding in Chennai occurred throughout the month of November and Gregorian calendar month. The quantity of tweets relating to the Chennai flood has magnified considerably throughout this era. In Twitter nearly 70%-75% of tweets were relating to Chennai Flood. to grasp the kind of knowledge that has been disseminated in Twitter throughout and when the disaster, tweets relating to flood were retrieved from 10/11/2015 to 25/03/2016 with the assistance of the keyword #chennaiflood. The hash tag is that the # image that's wont to seek for tweets that have similar content [9]. Dataset contains solely flood connected tweets that have been documented victimization the #chennaifloodhashtag.

Tweets of Chennai flood were collected together with seventeen attributes. The attributes were Twitter ID of the user, text- it's the content or data that the user can tweet relating to a specific topic, Language during which text is tweeted, date at that it's created, Favourite Count is that the count of likes that the tweet has obtained, Is Re-tweet that's whether or not the text is re-tweeted or not, this can be a Boolean price that is either true if the text is re-tweeted or false if it's not, Re-tweet Count- it's the quantity of times the text is re-tweeted, User Name, User Screen Name, date at that Twitter user created his account, User Tweets is that the variety of tweets that the user has tweeted, User Favourites Count is that the count of tweets that the user likable the foremost, User Followers Count- it shows the quantity of individuals WHO are following the user, User Following Count is that the count of the quantity of individuals that the user follows, User Listed Count shows the count of variety of individuals WHO have extra that specific user to his or her favourites list, User Lang is that the language in that the text is tweeted and class- it's the category to which a specific tweet belongs. The tweets with seventeen attributes got as input to the wood hen tool. The tweets obtained principally belong to 5 categories particularly categorical feeling, Relief, Complaint, would like for facilitate and alternative. Victimization keyword analysis the tweets obtained were classified into the higher than 5 different classes.

The 5 categories or classes thought of during this work are as follows:
Need for facilitate -- This class contains tweets relating to any sort of assistance; like requests for additional volunteers,

medical support, food, water and alternative basic requirements.

Relief measures -- Several NGOs have come out throughout the disaster to produce facilitate to the accepted folks. Indian Army conjointly vie a serious role relating to this concern. They undertook many relief measures throughout and when the flood to build the Madras. This class includes Twitter messages that contain relief measures.

Express feeling -- This class includes messages that categorical a deep feeling towards those that have whole heartedly supported and helped the disaster victims.

Complaints -- Chennai Flood was caused because of the El Nino development; however this was intense or reached to its peak because of the black-market urban development and poorly designed system. Therefore this class contains the complaints relating to the untrustworthiness of the govt. relating to this matter.

Other --This class includes all alternative messages like lost and located, general comments and expression of opinions.

During the time of disaster that's from tenth November to Gregorian calendar month last there was magnified variety of Twitter messages on the primary a pair of classes that's would like for facilitate and Relief measures. Nearly 620 tweets obtained throughout these months were relating to Relief measures and want for facilitate. when the disaster (during the month of Gregorian calendar month and February) that's, once the water level started residing there was magnified variety of Twitter messages for following three classes which incorporates categorical feeling, Complaints and alternative. Throughout this point tweets obtained were additional focused to remaining three categories. Nearly 505 tweets were relating to these 3 categories and solely a really few tweets belongs to initial 2 categories. Messages announce throughout the amount of flood enclosed current water level in numerous areas and roads ; requests for additional variety of volunteers to pack survival kits that embrace basic necessities; requests for medical help and requests for food, water provides and boats. Wood hen could be a assortment of visualization tools and algorithms for data processing tasks. It's a free analytical tool used for the analysis of information and prophetic model. Wood hen contains tools for classification, clustering, information pre-processing and visualisation [17]. It's authorised below antelope General Public License. Tweets collected are classified victimization machine learning algorithms like call trees, Random Forests and Naive Bayes. Performances of all the 3 algorithms are compared in terms of exactness, Recall and F-measure. The most notions behind preferring these 3 algorithms are:
- All these algorithms run efficiently for giant datasets.

- They are one in every of the correct machine learning algorithms.
- They can handle information sets that contain each categorical and numerical data.
- They are easy, straightforward to handle and interpret.

IV.a. Random Forests

Random Forests could be a machine learning algorithmic rule used for each classification and regression [13]. It's Associate in nursing ensemble learning technique that involves construction of multiple call trees. Collections of call tree classifiers are known as Forests. Individual call trees are generated throughout the coaching time by the random choice of attributes in every node, which is able to confirm the split. Throughout classification, for a brand new object to be classified from the input vector, the new object is passed among all the trees within the forests. Every tree can provides a classification result that's, all the choice trees can offer their vote separately. The forests select most well liked category.

Benefits of Random Forests classification algorithmic rule are:

- Random Forests will handle information that contains outliers.
- It is not necessary that information have to be compelled to be pre-processed. they will handle missing values mechanically.
- They have high tolerance to over coaching.
- Random Forests are often in-built a quicker manner.
- Takes less time to predict the output.

IV.b. Decision Tree

Decision tree could be a graph. It follows branching technique to exhibit each attainable outcome for a decision. Tree is made in an exceedingly top-down algorithmic divide and conquer manner. All the traning samples are placed because the root component. They're divided recursively supported the chosen attribute. Decision tree principally consists of two nodes particularly leaf nodes and call nodes. a call node or internal node contains a pair of or additional subspaces or branches, wherever every denotes the take a look at on a specific attribute. Leaf node contains the ensuing category label, classification or call. Wood hen has enforced the C4.5 call tree algorithmic rule victimization J48. C4.5 has been evolved from ID3 algorithmic rule. The most difference between each is C4.5 will handle numeric attributes whereas ID3 cannot. So C4.5 doesn't need discretization of numeric attributes. The tweets got as input to wood hen and call tree was applied.

The main benefits of Decision tree are:

- Since it's depicted in an exceedingly tree format it's easy and straightforward to grasp.

- Classification techniques of call tree are easy and quick.
- They do not need domain data

IV.c. Naive Bayes

Naive Bayes could be a classification technique that provides prime importance to conditional freedom between predictors. It's supported the idea that each one the attributes are not absolutely freelance. It'san applied mathematics classifier that performs probabilistic prediction that predicts the category membership possibilities. Classification is predicated on Bayes Theorem. It's terribly helpful for top dimensional datasets and that they are straightforward to create. The posterior likelihood $P(c|x)$ is calculated from the chance $P(x|c)$, category previous likelihood $P(c)$ and predictor previous likelihood $P(x)$.

$$P(c|x) = (P(x|c) * P(c))/P(x).$$

Features of Naive Bayes are:

- Prediction of a category for a take a look at set is straightforward and straightforward.
- Multi category prediction could be a key feature of Naive Bayes.
- Class conditional freedom could be a limitation of Naive Bayes. This can be as a result of in world it's difficult to urge category predictors that are utterly freelance.
- It works well for categorical information than numerical information.

V.a. Analysis of Tweets

From the tweets collected, the foremost influential users of Chennai Flood were known. For that, the Twitter users were graded on the premise of their count of tweets, count of followers that every user have and on the premise of count of re-tweets. Table 1 shows the leading ten users on the premise of count of their tweets.

Table 1: Leading ten Users in terms of variety of Tweets.taken from[10]

Username	Number of tweets	No of retweets	No of followers
Informant	1296859	35	2622
News On Fire	747739	192	2735
Ajinkya Ambre	658716	28	3389
Khalid Rafiq	608473	0	1032
Indian Tv Shows	388769	1	770
The Solid Trader	366970	159	3504
India Today	311998	3911	2479530
DT Next	310560	3890	2451306
Options Tip	300723	67	387
Judy Sale	282608	355	7070

From Table 1 it's clear that the list contain users that contain people and non-government organisations. Table1 shows high ten users from the dataset on the premise of count of tweets they need tweeted relating to Chennai flood. So it's necessary to extract or establish the relevant information out of clanging or faux information. A way to find the relevant information is thru its mastermind. The credibleness of a Twitter user depends upon the quantity of followers he/she has. The additional the quantity of followers, the additional credibleness or the additional responsibility of the user is proved. Table 1 a pair of shows the leading ten Twitter users supported their variety of followers.

Table 2: Leading ten Users supported the quantity of Followers that every user has.taken from[10]

Username	Number of tweets	No of retweets	No of followers
India Today	2479530	311998	3911
DT Next	2451306	310560	3890
The Indian Express	1625049	187604	3312
Adam Braun	307461	8143	1881
Agnishwar Jayaprakash	170288	326	1
KV KunalVimal	124595	37430	281
Asianet	109640	19773	29
Future Cape Town	81104	60299	705
South Scope	70354	16174	461
Andy Burton	64843	88610	784

From Table 2 a pair of it's clear that Republic of India nowadays has the utmost variety of followers. Republic of India nowadays is that the official twitter page of the news channel. But considering solely the quantity of followers isn't a whole answer for the matter. Even supposing the quantity of followers can advocate the users responsibility, this cannot account for the users actual repercussion. The credibleness of a Twitter message lies once it's re-tweeted many times, particularly once the message is from a well-liked author or user. A message is re-tweeted suggests that it's reached resolute many readers. Table 3 shows the leading ten Twitter users on the premise of the count of re-tweets.

Table 3: Leading ten Users supported the quantity of Re-tweets. Taken from[10]

Username	Number of tweets	No of retweets	No of followers
India Today	3911	311998	2479530
DT Next	3890	310560	2451306
The Indian Express	3312	187604	1625049
Safe World for Women	2501	195435	56851
Parvez Ahmed	2282	99404	2348
Insurance Sharer	1889	174251	2461
Adam Braun	1881	8143	307461
Madanmohan Rao	1425	40657	7533
Khalsa Aid	1046	103305	8618
Irene Koehler	870	35661	21216

With this list of reliable users, we tend to obtain a thought relating to that set of writers posts need to be followed throughout natural disasters, so as to get the credible, relevant and up-to-date data. as an example, Republic of India nowadays could be a verified Twitter page of the news channel and therefore the credibleness is proved by the quantity of followers it's. From all the tables it's clear that Republic of India nowadays and DT Next are the foremost influential users of Chennai flood. this can be as a result of whereas considering all the three factors, they need a good variety of tweets relating to Chennai flood, most variety of followers and tweets are retweeted a good quantity of your time. From the collected tweets of Chennai flood, supported the mix of the count of tweets, followers and re-tweets, the foremost relevant tweets are chosen for analysis.

V.b Performance analysis and Comparison
As mentioned earlier, the dataset comprised of 5 classes or category labels particularly would like for facilitate, Relief measures, categorical feeling, Complaints and alternative. Every algorithmic rule can predict the category labels of given instances. Properly categorised instances are those instances whose class labels are foretold properly by the algorithmic rule. Whereas incorrectly categorised instances are those whose class labels are foretold incorrectly by the algorithmic rule. From the Table 4 it's terribly clear that Random Forests is that the best algorithmic rule in analysing the tweets compared to the opposite 2. Random Forests has the smallest amount variety of misclassification instances.

Table 4: The number of correctly and incorrectly classified instances of all 3 algorithms.Taken from[10].

	Random forests	Decision Tree	Naive Bayes
Correctly classified instances	1154	1133	796
Incorrectly classified instances	3	24	361
Correctly classified instances in percentage	99.74%	97.92%	68.79%

Incorrectly classified instances in percentage	0.25%	2.07%	31.20%

Table 5 shows the confusion matrix of all the three algorithms. This offers U.S. a thought relating to that are the misclassified instances gift all told the 3 algorithms. The performances of all the 3 classification algorithms were evaluated supported exactness; Recall and F- live as shown in Table. 6.

A → categorical feeling; B → Relief; C → criticism; D → would like for facilitate; E → alternative;

Precision suggests that accuracy. That is, what proportion of instances that the categorised tagged as positive is truly positive? Weighted average for the exactness of the entire class label instances of Random Forests, call Tree and Naive Bayes are zero.997, 0.98, and 0.771 severally. Recall suggests that completeness. That's what proportion of positive.

Table 5: The performance matrix for all three algorithms.Taken from [10]:

Random forests					
Classified	A	B	C	D	E
A	175	0	0	0	0
B	1	231	0	0	0
C	0	0	283	0	0
D	0	0	0	132	0
E	0	0	2	0	333
Decision Tree					
Classified	A	B	C	D	E
A	171	4	0	0	0
B	2	229	0	0	1
C	0	0	282	0	1
D	0	0	0	132	0
E	3	5	8	0	319
Naïve Bayes					
Classified	A	B	C	D	E
A	164	2	4	3	2
B	145	34	1	23	29
C	72	2	176	18	15
D	4	4	0	124	0
E	15	3	8	11	298

Table 6: The performance evaluation results of all the 3 algorithms with respect to precision, recall and F-measure. Taken from [10]

Random forests			
Class	Precision	Recall	Fmeasure
Express Gratitude	0.994	1	0.997
Relief	1	0.996	0.998
Complaint	0.993	1	0.996
Need for help	1	1	1

Other	1	0.994	0.997
Weighted Average	0.997	0.997	0.997
Decision Tree			
Class	Precision	Recall	Fmeasure
Express Gratitude	0.972	0.977	0.974
Relief	0.962	0.987	0.974
Complaint	0.972	0.996	0.984
Need for help	1	1	1
Other	0.994	0.952	0.973
Weighted Average	0.98	0.979	0.979
Naive Bayes			
Class	Precision	Recall	Fmeasure
Express Gratitude	0.41	0.937	0.57
Relief	0.756	0.147	0.245
Complaint	0.931	0.622	0.746
Need for help	0.693	0.939	0.797
Other	0.866	0.89	0.878
Weighted Average	0.771	0.688	0.663

Instances the classifier tagged as positive. Weighted average for the recall of all the category label instances of Random Forests, call Tree and Naive Bayes are zero.997, 0.979, and 0.688 severally. It suggests that Random Forests identifies ninety nine.7% of positive instances, call Tree ninety seven.9% and Naive Bayes sixty eight.8%. F-measure is that the mean value of exactness and recall. From all this it's clear that Random Forests is that the best among the three for analysing the tweets.

VI. CONCLUSION

Social media plays a serious role in the propagation of knowledge throughout natural disasters. Throughout the Chennai flood, social media like Twitter has contributed tons in providing up so far data from the accepted areas. When examining and discovering the tweets identifying with the flood, the results can be used for managing and coming up with disaster relief measures within the future. During this paper, Twitter messages are classified victimization keyword analysis and a comparative study of 3 machine learning algorithms like Random Forests, call tree and Naive Bayes is disbursed. The performances of all the 3 classification algorithms are evaluated in terms of exactness, Recall, and F-measure. It had been found that Random Forests was additionally suited to Twitter analysis and classification.

Future work, we tend to are progressing to implement all the 3 algorithms on on-line streaming information, that is, as and once the info comes the analysis is disbursed so we'll acquire a more robust classification results and a more robust data are often inferred.

REFERENCES

[1] B. Abedin, et al. Characterization of the use of social media in natural disasters: a systematic review. In: Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on 2014; 449–454.

[2] L. Palen, et al. Twitter-based information distribution during the 2009 red river valley flood threat. Bulletin of the American Society for Information Science and Technology 2010; 36(5): 13–17.

[3] N.A. lateef Saeed, et al. The use of social media in knowledge integration for improving disaster emergency management task performance: Review of flood disasters. Indian Journal of Science and Technology 2016; 9(34).

[4] 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.

[5] 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

[6] K. Kandasamy, P. Koroth. 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.

[7] H. Dong, et al. Social media data analytics applied to hurricane sandy. In: Social computing (SocialCom), 2013 international conference on IEEE. 2013; 963–966.

[8] S.R. Kalmegh. Comparative analysis of weka data mining algorithm randomforest, randomtree and ladtree for classification of indigenous news data. International Journal of Emerging Technology and Advanced Engineering 2015;5(1): 507–517.

[9] 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.

[10] Nair MR, Ramya GR, Sivakumar PB (2017) Usage and analysis of Twitter during 2015 Chennai flood towards disaster management. Procedia Comput Sci 115:350–358

[11] A. Kongthon, et al. The role of twitter during a natural disaster: Case study of 2011 thai flood. In: Technology Management for Emerging Technologies (PICMET), Proceedings of PICMET12: IEEE 2012; 2227–2232.

[12] 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.

[13] 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.

[14] 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.

[15] A. Bruns, Y.E. Liang. Tools and methods for capturing twitter data during natural disasters. First Monday 2012; 17(4).

[16] S. Srivastava. Weka: a tool for data preprocessing, classification, ensemble, clustering and association rule mining. International Journal of Computer Applications 2014; 88(10).

[17] G. Biau. Analysis of a random forests model. Journal of Machine Learning Research 2012; 1063–1095.

[18] N. Bhargava, et al. Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering 2013; 3(6).

[19] G. Ramya, P.B. Sivakumar. Advocacy monitoring of women and children health through social data. Indian Journal of Science and Technology 2016; 9(6).

[20] K.L. Devi, et al. Tweet sentiment classification using an ensemble of machine learning supervised classifiers employing statistical feature selection methods. In: Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO-2015). Springer 2015; 1–13.

[21] T.R. Patil, S. Sherekar. Performance analysis of naive bayes and j48 classification algorithm for data classification. International Journal of Computer Science and Applications 2013; 6(2): 256–261

[22] S. Sathyamoorthy, Data Mining and Information Security in Big Data, International Journal of Scientific Research in Computer Science and Engineering Vol.5, Issue.3, pp.86-91, June (2017)

Author's Profile

Sabih Ahmad Ansari completed his Bachelor of Technology in Computer Science from Maulana Azad National Urdu University (MANUU), Gachibowli, Hyderabad (Telangana) in 2017. He is currently pursuing Master of Technology in Computer Science from MANUU.

Mr. Ahmad Talha Siddiqui received MCA, M.Tech degree in Computer Science from Jamia Humdard in 2007 and 2012 respectively. Presently, he is working as an Assistant Professor in School of Computer Science and Information Technology of Maulana Azad National Urdu University (MANUU), Gachibowli, Hyderabad (Telangana). He has more than 4 year of teaching experience. His areas of interests include Cryptography and Mobile Computing.