Natural Hazards
https://doi.org/10.1007/s11069-021-04527-w

ORIGINAL PAPER

Content analyses of the international federation of red cross and red crescent societies (ifrc) based on machine learning techniques through twitter

Türkay Dereli[1] · Nazmiye Eligüzel[2] · Cihan Çetinkaya[3]

Received: 8 April 2019 / Accepted: 9 January 2021

Abstract

Intensity of natural disasters has substantially increased; disaster management has gained importance along with this reason. In addition, social media has become an integral part of disaster management. Before, during and after disasters; people use social media and large number of output is obtained through social media activities. In this regard, Twitter is the most popular social media tool as micro blogging. Twitter has also become significant in complex disaster environment for coordinating events. It provides a swift way to collect crowd-sourced information. So, how do humanitarian organizations use Twitter platform? Humanitarian organizations utilize resources and related information while managing disasters. The effective use of social media by humanitarian agencies causes increased peoples' awareness. The international federation of red cross and Red Crescent Societies (IFRC) is the most significant humanitarian organization that aims providing assistance to people. Thus, the aim of this paper is to analyze IFRC's activities on Twitter and propose a perspective in the light of theoretical framework. Approximately, 5201 tweets are passed the pre-processing level, some important topics are extracted utilizing word labeling, latent dirichlet allocation (LDA model) and bag of Ngram model and sentiment analysis is applied based on machine learning classification algorithms including Naïve Bayes, support vector machine SVM), decision tree, random forest, neural network and k-nearest neighbor (kNN) classifications. According to the classification accuracies, results demonstrate the superiority of support vector machine among other classification algorithms. This study shows us how IFRC uses Twitter and which topics IFRC emphasizes more.

Keywords Content analyses · IFRC · Topic labeling · LDA · Machine learning · Twitter

✉ Nazmiye Eligüzel
nazmiye@gantep.edu.tr

[1] Office of the President, Hasan Kalyoncu University, Gaziantep, Turkey

[2] Industrial Engineering, Gaziantep University, 27310 Gaziantep, Turkey

[3] Department of Management Information Systems, Adana Alparslan Turkes Science and Technology University, 01250 Adana, Turkey

Published online: 05 February 2021

1 Introduction

Big data describes the large volume of data both structured and unstructured. It can be described as a holistic approach to control, process and analyze the "5 Vs" to constitute actionable insights for continues value, performance measurement and competitive superiority (Papadopoulos et al. 2017). "5 Vs" is defined as Volume, Velocity, Variety, Veracity and Value. Several research studies have investigated the use of big data in crisis situations as follows: Monaghan et al. (2013) discussed big data phenomenon with its characteristics of volume, velocity, variety and veracity in humanitarian supply networks and emphasized the distinction between humanitarian aid and development aid. Humanitarian aid efforts and impact of delivering humanitarian relief was assessed in the light of big data techniques and technologies. Papadopoulos et al. (2017) proposed a theoretical framework to indicate the role of big data in disaster resilience for supply chains. They used unstructured big data included tweets, Google+, YouTube, news, Facebook, WordPress and also structured data involved in disaster relief activities after Nepal Earthquake in 2015. They examined the role of swift trust and quality information in supply chain networks unlike other studies. 36,422 tweets were collected for analyze and also 205 responses collected via disaster relief workers. They paid attention to the importance of public–private partnerships, swift trust, information quality, critical infrastructure resilience, community resilience and resource resilience using context of Nepal. An event detection system proposed by Cherichi et al. (2017) based on data obtained from Twitter. The aim of this system is determine new events, recognize temporal indicators and classify significant events using big data values in the case of social events. Semantic analysis was made into two clusters: positive class and negative class. Mulder et al. (2016) proposed the processes of "big data making" in their papers through crowdsourcing open data platforms, within the framework of two crises: 2010 earthquake in Haiti and the 2015 earthquake in Nepal. The study has included humanitarian response. Aim of the study was to discover what obstacles stand in the way of "big data making" approach. Gupta et al. (2017) proposed a systematic literature review study using Scopus database. They examined big data in humanitarian supply chain management using 28 journal papers after eliminating unnecessary studies. Classification was done according to the various theories such as the enablers for big data in humanitarian supply chain management (volume, variety, velocity, veracity organizational mindfulness), the concerns identified in humanitarian supply chain management (humanitarian logistics, remote sensing, information security, social media).

Recently, the first thing that comes to mind when it comes to big data is social media. Social media has gained significant attention because people want to get involved in the public issues within agenda. Social media platforms such as Twitter, Flickr, YouTube and LinkedIn generate social network data as big data to gain insight into an event. Information can be collected and disseminated via social media platforms. Social media analytics has also become prominent in natural disaster management. As the underlined by Xiao et al. (2018), an effective way for disasters and public sentiment is provided via rising development of big data and data mining technologies in social media. During crisis situations, situational awareness can be provided by social media easily. Hence, it has been widely utilized for crisis communication (Wang et al. 2017). In addition to user requests, calls from humanitarian organizations and officials are also very important. Accordingly, its importance is expressed as: Social media is a communication tool that the governments utilize for information dissemination and coordinate it with officials during emergencies (Malawani et al. 2020). Due to the increasing number of social media users, organizations make their share through social media platform. What these organizations do, what they will do and the information they share is very significant. In our opinions, the platform on which these are best shared is Twitter. Twitter is one of the most popular micro blogging websites of social media. The user can write his thoughts in limited 140 characters and other people can follow the blogs. There are 326 million total number of monthly active Twitter users worldwide at the third quarter of 2018.[1] Twitter has a critical role for disaster management activities. Sentiment analysis, opinion mining, topic extraction etc. are popular subjects performed via Twitter at recent times. A sentiment analysis of tweets during the disastrous Hurricane Sandy was performed by (Neppallia et al. 2017) to demonstrate how users' sentiment vary according to their location and based on distance from event. They were claimed that study would provide truly actionable information for official responders. Credibility is important against rumor and false information. Ikegami et al. (2013) evaluated the credibility of information by computing the ratio of same opinions to all opinions within a topic. They performed topic classification by Latent Dirichlet Allocation and sentiment analysis through a semantic orientation dictionary. Gitari Zuping et al. (2016) presented an approach that integrated topic modeling and Machine Learning algorithms to forms topics. They compared polarized detection classification accuracy using other topic models. Gul et al. (2018) analyzed Jammu and Kashmir floods through Twitter sentiment analysis. They provided an understanding how people use social media during natural calamities. Kim et al. (2014) presented an approach to understand public opinions on nuclear power. They made Twitter sentiment analysis creating positive and negative dictionary. Cody et al. (2015) analyzed tweets including the word "climate" between specific time intervals. They defined how collective sentiment changes in response to climate change, events, news and natural calamitous.

In this study, the tweets by International Federation of Red Cross and Red Crescent Societies (IFRC) are analyzed according to the content based on topic extraction models. This study demonstrates how IFRC, which is one of the leading humanitarian organizations, utilizes social media and which activities are conducted by IFRC. So, unspoken topics and other situations can be identified utilizing IFRC Twitter account. Therefore, reasonable precautions can be taken to ensure proper disaster management. The center of the Twitter account is Geneva, Switzerland. The remainder of this paper is organized as follows: Sect. 2 introduces related works in the literature, Sect. 3 introduces methodology including the process of getting tweets and steps of pre-processing, content analysis based on word frequencies and word labeling, latent dirichlet allocation (LDA) and bag of words (bag of Ngram) approaches and sentiment analyses of tweets. Section 4 demonstrates machine learning algorithms, Sect. 5 demonstrates comparison of classification algorithms including Naïve Bayes, SVM, decision tree, random forest, neural network and kNN. Section 6 introduces results and performance evaluation of classifiers. Finally Sect. 7 summarizes the conclusions.

[1] Statista, http://www.statistacom/statistics/282087/number-of-monthly-active-twitter-users/.

2 Related works

Adaptation of governmental humanitarian organizations and non-governmental humanitarian organizations to social media is important for interactive dialogue with stakeholders and creating novel opportunities. Unlike traditional media, mutual communication with the public is remarkably fast in social media. Khemka et al. (2017) examined usage of Twitter by Red Cross and Red Crescent organizations. They evaluated study in terms of three ways: adoption rates and influential factors, message frequencies and types, capability to access large audiences. They demonstrated geographic distributions of Red Cross/Red Crescent accounts. According to the study, half of the Red Cross and Red Crescent societies in the world have adopted to Twitter. Social media platforms have the opportunity to participate actively in the disaster management process. In the study proposed by Yu et al. (2019), capacity of a convolutional neural network model (CNN) was examined for real time Twitter text classification. The CNN model was compared to traditional machine learning methods. The results demonstrated that better accuracy is provided by CNN model. Pandey et al. (2018) proposed an interactive user-feedback system named “CitizenHelper-Adaptive”. It is based on streaming analytics. The aim of the “CitizenHelper-Adaptive” system is mining social media, news, other public Web data for humanitarian organizations and emergency services. Both humanitarian and disaster event data can be analyzed by the system. Reynard et al. (2019) utilized the American Community Survey 2017 geography to enable social-economic context. They relied on multinomial logit specification to analyze sentiments of tweets. Machine learning techniques were utilized to categorize geo-located tweets related disaster. Wukich et al. (2017) analyzed the two networks with their structure and antecedents. These networks operate in the policy field of emergency management. The first network comprises national-level government agencies that are responsible for recovery and response procedures. The other network comprises non-governmental organizations such as Red Cross and Red Crescent national societies. The contribution of the study is the understanding of how knowledge networks can created globally and how social media can provide that process.

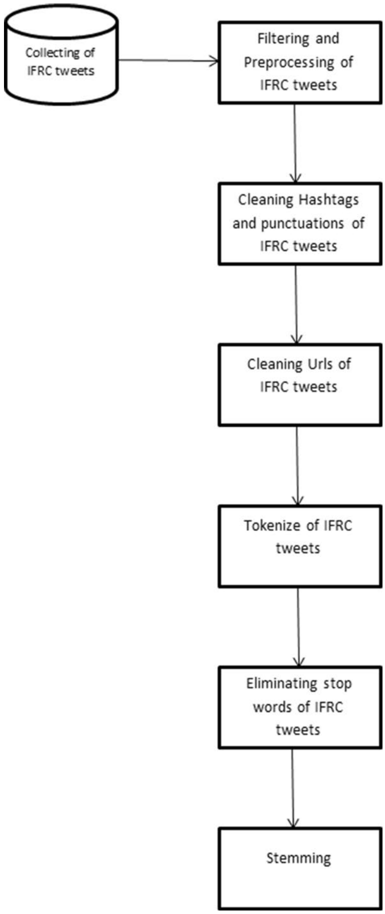
3 Methodology

This section explains the methodology for the problem. There are number of tweets that retrieved from IFRC Twitter account. Tweets are cleaned and important topics are extracted using some methods. Sentiment polarities of tweets are classified. In this study, machine learning techniques are utilized for the comparison. Parts of proposed context analyzes are explained in the following subsections:

3.1 Retrieving and cleaning tweets

IFRC is the significant humanitarian organization providing relief with no discrimination as to nationality, race, beliefs, religious, class or political aspects (International Federation of Red Cross and Red Crescent Societies(IFRC) 2018). IFRC performs relief operations to help victims. Organization claims that the works of IFRC focalize on four main fields: encouraging humanitarian values, disaster response, disaster preparedness and health and community care. So, Twitter is important in terms of being aware of how the IFRC works

Fig. 1 Steps of Pre-processing



and what kind of discourses are addressed by IFRC. In this study, tweets are obtained from IFRC Twitter account and analyzed according to its content. IFRC Twitter account was opened at 28th August 2008. Firstly, tweets have been collected from opening date of this account to the date of 27th July 2018. Tweets have been obtained using Python programming language via selenium library using '@federation' name. Approximately 5201 tweets are handled and evaluated. For obtaining accurate information, the data must be proper and need to be pre-processed. In Fig. 1, the steps of pre-processing methods are shown.

There are some steps for pre-processing. Cleaning hashtags, removing punctuations andurls, removing stop words ('the', 'on', etc.), stemming, converting all text to lower case are the first steps of the pre-processing to convert the text a uniform format. After conducting these steps, all tweets have been tokenized. Tokenization can be referred as lexicon analysis. It is the act of splitting the strings into pieces such as words, keywords, phrases, symbols and other elements called tokens.[2] Finally stemming process has been conducted.

[2] Techopedia,https://www.techopediacom/definition/13698/tokenization.

Table 1 Tokenized and Pre-processed of Tweets

(5057,1)	11 tokens	albino hunt kill tanzania burundi red cross red crescent help protect
(5058,1)	7 tokens	interest onlinedeb host economist theeconomist intern aid
(5059,1)	13 tokens	tell non profit Twitter well ask mrtweet pleas support feder local red crosscresc
(5060,1)	15 tokens	red cross clean mokupa beach plant 2000 mangrov tree indonesian red cross world ocean confer
(5061,1)	9 tokens	defend albino right life superstit led kill albino burundi
(5062,1)	9 tokens	reliefweb mln displac climaterel natur disast 2008 check studi
(5063,1)	13 tokens	just move close anoth fraudul web site come suspici site email pleas contact
(5064,1)	13 tokens	red cross volunt provid lifelin shanti town work haitian red cross volunt danger
(5065,1)	9 tokens	portrait volunt evelyn koroma uyubi volunt motto servic mankin
(5066,1)	13 tokens	pakistan ifrc call 239 million swiss franc help 140,000 displac peopl intern federat

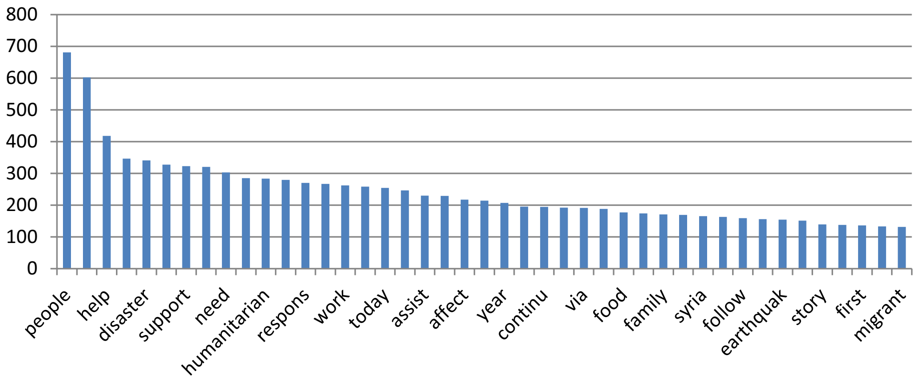


Fig. 2 Word Counts of Text

Stemming is the process of separating a word to its word stem such as affixes, suffixes or roots. It is important natural language processing.[3] Stemming process decreases the data redundancy. In Table 1, tokenized and pre-processed tweets between 5057 and 5066 can be seen as a sample.

Pre-processing is the first stage of Twitter content analysis in this study to extract word frequencies and relationship between words.

3.2 Word labeling

Following of the pre-processing operation, all the words in the text have been counted and word frequencies have been obtained. The highest mentioned words; 'red', 'red cross', 'red crescent', 'crescent' and 'cross' are eliminated because of being the organization's name. As seen in Fig. 2, 'people', 'voluntary', 'help', 'community', 'disaster', 'new', 'support', 'day', 'need','haiti' is the highest mentioned words within top 10 words. In Fig. 3, word cloud of these words including organization's name can be seen. Furthermore, in Fig. 4,

[3] TechTarget, https://searchenterpriseaitechtarget.com/definition/stemming.

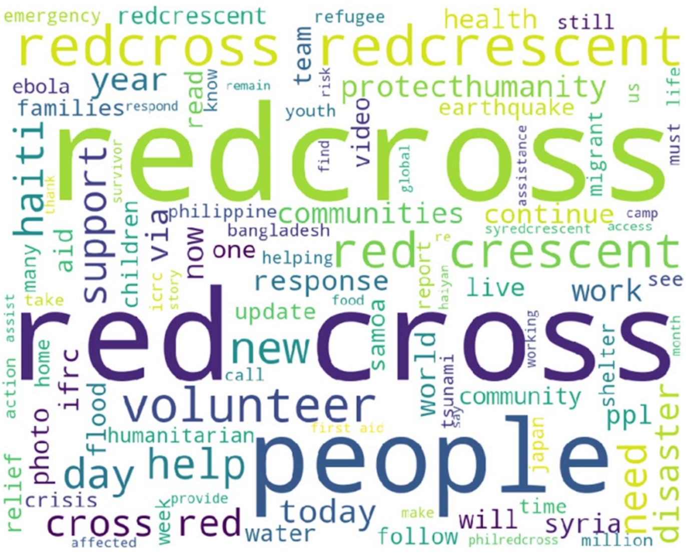


Fig. 3 Word Cloud of Text

sub-words of topics and their word frequencies can be seen. In Fig. 5, ratio of mentioned categories as percent can be seen.

Humanitarian logistic plays critical role in disaster management activities according to the IFRC Twitter account analysis.

3.3 Latent dirichlet allocation (LDA)

Latent Dirichlet Allocation (LDA) is a generative probabilistic model of a corpus Blei et al. (2003). The main idea is that documents are demonstrated as random mixtures over latent topics, where every single topic is described by a distribution over words. It finds underlying topics in a collection of documents and deduces probabilities of words in topics[4] LDA is unsupervised learning algorithm. If data is labelled, it turns supervised learning.

According Fig. 6:

- α is the Dirichlet-prior concentration parameter of the per-document topic distribution
- β is the same parameter of the per-topic word distribution
- θ (d) is the topic distribution for document d
- z (d,n) is the topic assignment for w (d,n)
- w (d,n) is the n^{th} word in the d^{th} document
- K is the number of topics
- N is the number of words in the document
- M is the number of documents to analyze
- D is the corpus of collection M documents

[4] MathWorks, https://www.mathworks.com/help/textanalytics/ref/fitlda.html.

Fig. 4 Sub-categories and Sub-words Frequencies for Text

The LDA model involves a three-level model. In Fig. 6, the outer box shows documents, while the inner box shows the repeated selection of topics and words inside a document. Probability of corpus can be seen in Eq. 1.

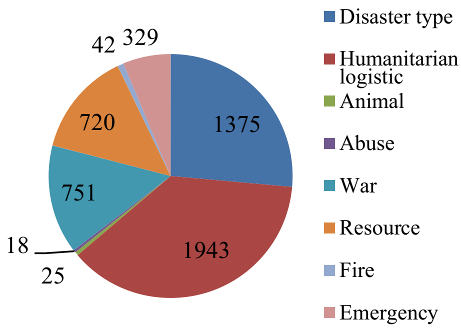


Fig. 5 Ratio of Text Sub-categories

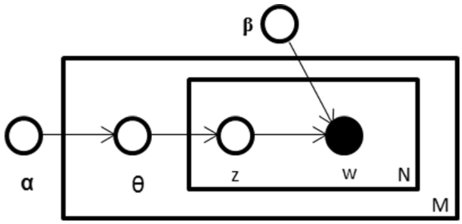


Fig. 6 Structure of LDA Model

$$p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta d|\alpha)\left(\prod_{n=1}^{Nd} \Sigma p(zdn|\theta d)p(wdn|zdn, \beta)\right)d\theta d \tag{1}$$

After pre-processing operations for IFRC Twitter data, LDA process is implemented using MATLAB software. K = 50 topics in this model and Dirichlet hyper-parameters β = 0.1 and α = 50 / *K* (Tong et al. 2016). Topic distributions are generated for each document. These distributions indicate relevance of each document with each topic. Some selected topics composed by IFRC topic distribution can be seen in Table 2.

3.4 Bag of ngram model

Our text is evaluated via bag ofNgram model lastly. N-grams are generalization of set of words (BOW). It is an adjacent order of n items in a given text sample. A bag-of-n-grams model records the number of times that each n-gram seems in every single document of a corpus.[5] In Table 3, five topics obtained with Bag of Ngram model can be seen.

3.5 Sentiment analyses

Sentiment analysis determines sentiments and opinions represented in the text. Sometimes it can be referred as opinion mining. It is the positive or negative orientation that a writer states versus some object (Jurafsky et al. 2018). In this study, positive (Hu et al. 2004) and negative (Liu et al. 2005) word lists are used for determining sentiment of tweets. The words in the lists are compared to our text corpora using Python Programming Language

[5] MathWorks, https://www.mathworks.com/help/textanalytics/ref/bagofngrams.html.

and overlapped words are obtained both negative and positive. In Fig. 8 and 9 positive and negative words in the text can be seen.

In the text, while number of 3281 positive words is counted number of 3010 negative words is counted. The word 'cross' in negative word list is eliminated because IFRC includes this word within its name. So, number of 2406 negative words is counted.

4 Machine learning classification algorithms

Classification is appointing a class label to set of unclassified cases (Sharma et al. 2017). For the supervised classification, set of possible classes is known in advance. For the unsupervised learning, set of possible classes is not known in advance. After classification, names are given to classes. In this study, Naïve Bayes algorithm, support vector machine (SVM), decision tree, random forest, neural network and k-Nearest Neighbor (KNN) classifier approaches are performed to classify text after sentiment analyses.

4.1 Naïve bayes

Naive Bayes is a probabilistic classifier, that means for a document d, out of all classes $c \in C$ the classifier returns the class c^* which has the maximum back probability determined the document (Jurafsky et al. 2018). Prediction of correct class can be seen in Eq. 2.

$$c^* = \text{argmax } P(c|d) \tag{2}$$

Bayes' rule is demonstrated in Eq. 3.

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{3}$$

In natural language processing to get text classification, we can substitute Eq. 1 to Eq. 2. The new equation can be seen in Eq. 4.

$$c^* = \text{argmax } P(c|d) = \text{argmax} \frac{P(d|c)P(c)}{P(d)} \tag{4}$$

Based on above equations, Multinomial Naïve Bayes algorithm is implemented. It is used in text classification where the data are typically demonstrated as word vector counts[6] It is based on an assumption. Given the predicted value, the attributes utilized for producing a prediction are independent of each other (Frank et al. 2000).

4.2 Support vector machine (SVM)

It is binary classification algorithm. Given a set of points of 2 types in N dimensional place, SVM creates a (N—1) dimensional hyperplane to divide those points into 2 groups (Le 2019). Objective of this algorithm is to find a plane that has the maximum margin. Hyperplanes help classify the data points as decision boundaries. Support vectors are data points

[6] scikit learn, org/stable/modules/naive_bayes html.

and they influence the orientation and position of the hyperplane. SVM is developed from sound theory to the application and experiments. It is the non-parametric model that means there are no parameters at all for SVM. (Kecman 2014). “Learning” is the significant issue for SVM such as training and tuning, identification, selection and estimation. The number of parameters is determined according to the utilized training data. Namely, parameters are not pre-defined.

4.3 Decision tree

The decision tree includes the number of nodes that create a rooted tree. Namely, it is a directed tree with a root node that has no incoming edges. The rest of nodes have one incoming edge. Internal or test node is a node that has outgoing edges. The other nodes are referred as leaves (Rokach et al. 2005). Decision tree builds regression or classification model based on a tree structure.[7] Decision tree algorithm uses if–then rule set that is learned sequentially utilizing the training data on at a time. Decision tree can be used to represent decision and decision making visually. The aim of the algorithm is to discover the optimal decision tree to minimize the generalization error.

4.4 Random forest

It refers a trademark term for an ensemble of decision trees (Ray 2019). In this classification algorithm, there is collection of decision trees known as ‘Forest’. Each tree gives a classification to classify a novel object based attributes. Actually, decision trees are ideal representative for ensemble methods because they often have low bias and high variance and this makes them very probably to benefit from the averaging process. Random forests methods introduce random complexities into the induction procedure. They mostly differ from each other in this way (Louppe 2014). Random forests are attractive from a computational perspective, because they are relatively quick to predict and train, they naturally obtain both regression and classification, they can be utilized clearly for high-dimensional problems, they can simply be applied in parallel, they depend solely one or two tuning parameters (Cutler et al. 2012).

4.5 Neural network

It is supervised learning algorithm. Neural networks need the desired outputs for a given set of inputs which is what authorize it to learn from the data. Neural network resembles the brain in two ways: through a learning process, knowledge is obtained by the network from its environment and to store the obtained knowledge, synaptic weights are utilized. Good performance is provided by utilizing a massive interconnection of basic computing cells called “neurons” or “processing units” (Simon Haykin 2014). It is attractive method because of its capability to learn and massively parallel distributed architecture. MLPClassifier is imported as a class from sklearn.neural_network library.

[7] Towards Data Science, https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623.

4.6 kNN (k-Nearest Neighbors)

It can be utilized both regression and classification problems but it is widely preferred in classification problems. All available cases are stored via this algorithm that classifies new cases by a majority vote of its k neighbors (Ray 2019). kNN is a non-parametric algorithm because of avoiding priority assumptions related with the shape of class boundary. Therefore, it can adapt to non-linear boundaries as the training data size enhances (Bzdok et al. 2018). It has higher variance compared with support vector machine. Generated classes adapt to any boundary, so it causes having the advantage for kNN.

5 Performance evaluations

In this study, classification accuracy (ACC) and F-measure have been implemented as the evaluation metrics. In Eq. 5, formulation of classification accuracy can be seen.

$$\mathrm{ACC} = \frac{TN + TP}{TP + FP + FN + TN} \tag{5}$$

where TP,FP,FN,TN represent the number of true positives, false positives, false negatives and true negatives, respectively.

The other widely used measure is F-measure for performance evaluation of classification algorithms. It is the harmonic mean of the precision and recall of a classification algorithm. Higher values of F-measure mean better predictive performance. In this study macro-averaged F-measure is used, which defines the average F-measure across all one-versus-all classes, is computed as in the Eq. 6 (Onan 2018).

$$\mathrm{Macro-averaged\ F-measure} = \frac{1}{n}\sum_{i=1}^{n} \frac{2 * Precisioni * Recalli}{Preciisoni + Recalli} \tag{6}$$

where precision is the number of true positive divided by the number of true positive plus the number of false positive, recall is the number of true positive divided by the number of true positive plus the number of false negative.

Formulation of precision and recall of a classification algorithm can be seen in Eq. 7 and 8, respectively.

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{8}$$

The experimental analysis is performed with the Python 3.6 version.

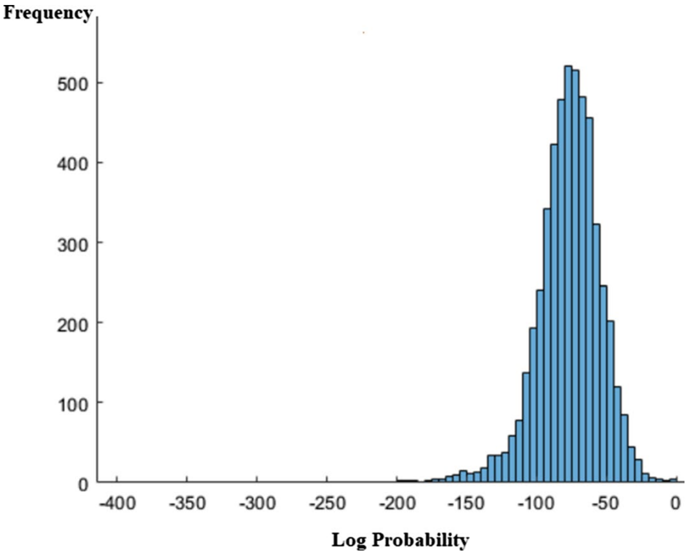


Fig. 7 Log-probabilities for IFRC Twitter Data

Fig. 8 Positive Words for IFRC Tweets

6 Results and discussions

In the pre-processing, all operations are performed through Python programming language. According to the word frequencies, content are especially related voluntarily help to people and supplying water, food etc. in disaster response activities. Haiti and Syria are the highest mentioned countries.

Fig. 9 Negative Words for IFRC Tweets

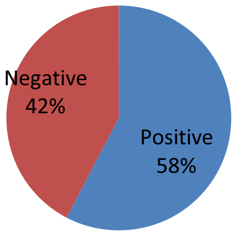


Fig. 10 Percent of Positive and Negative Words

IFRC focuses on people and their needs especially during and after disasters rather than before disasters. Although social media is utilized at every stage of disaster management (before, during and after), proposed analysis shows that IFRC considers post-disaster activities. Rather than words such as precaution, prevention, education that can be utilized before a disaster, other words such as aid, response, help etc. used during and after the disaster are more common according to the word cloud and topic analysis. Eight main categories are determined intuitively considering disaster management activities. All main categories include specific words. 'Disaster type', 'humanitarian logistic', 'animal', 'abuse', 'war','resource', 'fire' and 'emergency' are determined main category names.

As in the seen in Fig. 4, 'flood' is the highest mentioned disaster type and 'earthquake' follows it. For the humanitarian logistic topic, consecutively 'people', 'help' and 'aid' words are highest mentioned words. This category is related to response and relief activities of IFRC. 'Animal' and 'abuse' categories are mentioned merely. It demonstrates that there can be small care for animals and violence. 'Fire' category is also mentioned very few among other categories. 'War' is the category that mentioned significantly but the situation of Syria plays an important role in this. The category 'emergency' is related with accidental events and the highest mentioned word is 'death' as seen. The category 'resources' include necessary items after disaster events. 'Water', 'food' and 'shelter' are the most mentioned words respectively. But, the important point is that no discuss about baby supplies such as diaper and baby food.

Fig. 11 Graphic of Sentiment Analysis

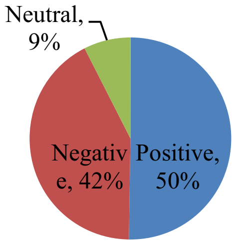


Table 2 Some Selected Topics Composed by IFRC Topic Distribution

Topic 7	Topic 14	Topic 15	Topic 16	Topic 20	Topic 35	Topic 47
"first"	"support"	"water"	"know"	"food"	"commun"	"peopl"
"team"	"ppl"	"access"	"humanitarian"	"need"	"help"	"help"
"aid"	"provid"	"sanit"	"onlin"	"crisi"	"build"	"assist"
"meet"	"children"	"safe"	"cours"	"shelter"	"resili"	"evacu"
"volunt"	"famili"	"need"	"law"	"water"	"women"	"mediterranean"
"provid"	"psychosoci"	"clean"	"basic"	"drought"	"disast"	"didyouknow"
"wai"	"commit"	"suppli"	"listen"	"emerg"	"futur"	"move"
"hope"	"girl"	"right"	"croixroug"	"kenya"	"like"	"survivor"
"hear"	"long"	"popul"	"head"	"kenyaredcross"	"better"	"100"
"visit"	"redcrescenttr"	"address"	"redtalk"	"almost"	"develop"	"relief"
0.5932	0.7084	0.5962	0.4281	0.6333	0.5987	0.5843

Scores of total words for each topic can be seen in bottom row in Table 2. Seven topics are selected randomly. The words in the same topic tend to be parallel. These associated words can be labeled with topic names. For example; topic 7 is about aid by volunteers, topic 14 is about children. It is related with providing psychological support to them. Topic 15 is about needs. It is related with providing sanitary water. Topic 16 is about education. It is related with increasing awareness about humanitarian activities. Topic 20 is about aid materials for Kenya. Topic 35 is about resilience. It is related with returning back the normal life after disaster. Finally, topic 47 is about evacuation. Maybe it can be related with immigrants. It is hard to label this topic. Logarithmic probabilities can be seen in Fig.7 for IFRC Twitter data. A low log-probability may propose that the document may be an outlier.

The indices of the three documents with the lowest log-probability are 106, 10,298. These three tweets may be referred as outlier.

There are 36,328 Ngrams and Ngram length is selected as 2 seen in the Table. Number of document equals to number of tweets that means there are 5201 documents. In fact, the organization name in tweets affects accuracy of extraction topics. ‘Disaster risk’, ‘affect area’, ‘risk reduction’, ‘voluntary staff’, ‘relief effort’, ‘humanitarian need’, ‘humanitarian response’, ‘water sanitation’,’psychological support’, ‘climate change’, ‘continuous assistance’, ‘save live’ etc. are the most used successive words in text corpora. These successive words can provide accurate insights about disaster management. For example, water is important but clean water is more important.

Table 3 Five Topics provided by Bag of Ngram Model

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
Word	Score	Word	Score	Word	Score	Word	Score	Word	Score
"redcross redcresc"	0,00,475	"red cross"	0,03,153	"redcross redcresc"	0,01,434	"red cross"	0,03,153	"cox bazar"	0,00,318
"latest photo"	0,00,361	"red crescent"	0,02,818	"psychosoci support"	0,00,351	"red cross"	0,01,787	"world disast"	0,00,318
"disast risk"	0,00,289	"cross red"	0,01,945	"first aid"	0,00,330	"redcross redcresc"	0,00,695	"disast report"	0,00,308
"risk reduct"	0,00,258	"redcross redcresc"	0,01,188	"climat chang"	0,00,290	"cross volunt"	0,00,531	"save live"	0,00,278
"peopl live"	0,00,227	"aid worker"	0,00,469	"save live"	0,00,220	"million peopl"	0,00,376	"field hospit"	0,00,278
"redcross help"	0,00,217	"around world"	0,00,421	"search rescu"	0,00,220	"samoa red"	0,00,367	"social media"	0,00,258
"thousand peopl"	0,00,196	"volunt staff"	0,00,249	"relief effort"	0,00,220	"peopl affect"	0,00,357	"staff volunt"	0,00,219
"affect area"	0,00,175	"crescent volunt"	0,00,230	"humanitarian need"	0,00,210	"redcross volunt"	0,00,338	"continu assist"	0,00,209
"new photo"	0,00,175	"cross societi"	0,00,220	"humanitarian respons"	0,00,210	"water sanit"	0,00,338	"onlin cours"	0,00,199
"everi dai"	0,00,165	"sri lanka"	0,00,220	"red cross"	0,00,806	"photo dai"	0,00,328	"red cross"	0,00,806

Table 4 Confusion matrixes of Algorithms

Naive Bayes			SVM			Decision Tree			Random Forest			Neural Network			kNN		
36	3	14	35	1	17	26	0	27	30	1	22	29	5	19	12	9	32
1	1	4	2	2	2	1	1	4	2	1	3	4	0	2	0	1	5
9	7	75	12	2	77	10	10	71	10	1	80	12	3	76	9	6	76

Table 5 Classification Accuracies of Algorithms

Algorithms	ACC (Classification Accuracy) (%)
Naive Bayes	74.6
SVM	76
Decision Tree	65.3
Random Forest	74
Neural Network	70
kNN	59.3

Table 6 F-measure of Naïve Bayes

Naïve Bayes	precision	recall	F-measure
Positive	0.78	0.68	0.73
Negative	0.09	0.17	0.12
Neutral	0.81	0.82	0.82
Average/total	0.77	0.75	***0.76***

Table 7 F-measure of SVM

SVM	precision	recall	F-measure
Positive	0.71	0.66	0.69
Negative	0.40	0.33	0.36
Neutral	0.80	0.85	0.82
Average/total	0.75	0.76	***0.76***

Table 8 F-measure of Decision Tree

Decision Tree	precision	recall	F-measure
Positive	0.70	0.49	0.58
Negative	0.09	0.17	0.12
Neutral	0.70	0.78	0.74
Average/total	0.67	0.65	***0.66***

Table 9 F-measure of Random Forest

Random Forest	precision	recall	F-measure
Positive	0.71	057	0.63
Negative	0.33	0.17	0.22
Neutral	0.76	0.88	0.82
Average/total	0.73	0.74	***0.73***

Table 10 F-measure of Neural Network

Neural network	precision	recall	F-measure
Positive	0.64	0.55	0.59
Negative	0.00	0.00	0.00
Neutral	0.78	0.84	0.81
Average/total	0.70	0.70	***0.70***

Table 11 F-measure of kNN

kNN	precision	recall	F-measure
Positive	0.57	0.23	0.32
Negative	0.06	0.17	0.09
Neutral	0.67	0.84	0.75
Average/total	0.61	0.59	***0.57***

In the positive word cloud, some of promising words such as ‘peace’,’hearth’, ‘free’,’well’,’hope’ are drawn the attention. In the negative word cloud, some of disappointed words such as ‘damage’, ‘fight’, ‘lost’, ‘dead’, ‘fear’ are drawn attention. In Fig. 10, percent of positive and negative words in text can be seen.

Even if the tweets are related with disaster, the reason of more positive words is referred the good works the IFRC has conducted. Based on positive and negative words, 500 samples are selected within 5201 tweets. Positive tweets are referred as 1, negative tweets are referred as -1 and neutral tweets are referred as 0. This process is done intuitively because of obtaining accurate results. Graphic of sentiment analysis can be seen in Fig.11.

According to the test sample, ratio of positive tweets is 50%, ratio of negative tweets is 42% and ratio of neutral tweets is 9%. Negative tweets mostly include despair, positive tweets mostly include help and relief operations done by IFRC and neutral tweets mostly include some announcements.

Test size is taken as 0.30 (30%). Program is run via Python and count vectorizer that converts a collection of text documents to a matrix is taken as 1000.

Confusion matrixes of proposed classification algorithms can be seen in Table 4.

In the light of confusion matrixes, F-measures and accuracies of classifiers are evaluated. Classification accuracies of algorithms are given by Table 5.

The higher value of accuracy is obtained from SVM as 76%. Accuracies of Naïve Bayes, random forest and neural network classification algorithms have been obtained as 74.6%, 74% and 70%, respectively. These results indicate consistency of sentiment analysis. However, satisfactory accuracy results have not been obtained from decision tree and k-nearest neighbor algorithms. Evaluation of ACC is not sufficient by itself. F-measures are also examined to support the results. F-measures of Naïve Bayes algorithm, support vector machine (SVM), decision tree, random forest, neural network and k-nearest neighbor (KNN) classifier approaches can be seen in Tables 6, 7, 8, 9, 10 and 11, respectively.

As seen in the Tables, the higher values of F-measures are obtained from Naïve Bayes and SVM as 76%. F-measures of random forest and neural network classification algorithms have been obtained as 73% and 70%, respectively. Driven results from ACC and F-measure are almost same with aforementioned four algorithms. Results from ACC and F-measure indicate consistency of sentiment analysis. Decision tree and kNN algorithms are not affective to provide satisfactory results.

7 Conclusion

It is important to use social media effectively to increase people's awareness about humanitarian activities. Thus in this study, The International Federation of Red Cross and Red Crescent Societies' (IFRC) activities on Twitter are analyzed. With this Twitter analysis, effective use of Twitter by IFRC is identified. From opening date of IFRC Twitter account at 28th August 2008 to 27th July 2018, number of 5201 tweets is analyzed in total. After the pre-processing operations (cleaning hashtags, removing punctuations and urls, removing stop words, stemming, converting all text to lower case and tokenization) word frequencies are obtained and in the light of these words a word labeling study is done. Inspired by the bag of words model, some main topics and their sub-words are determined. According to this labeling, while 'humanitarian logistic' activities, 'emergency' actions, 'disaster', 'war' and 'resources' are on the agenda of IFRC organization, there are not many speeches related with 'violence' and 'animals' activities. Despite mentioning about the resources in general, there is no relevant speech about resources for babies such as 'baby foods' and 'diaper'. Later, sentiment analysis is performed. According to the test sample, ratio of positive tweets is 50%, ratio of negative tweets is 42% and ratio of neutral tweets is 9%. Negative tweets mostly include despair such as being inadequate, assault on volunteers, lack of blood etc., positive tweets mostly include help and relief operations done by IFRC and neutral tweets mostly include some announcements such as online courses. According to the latent dirichlet allocation (LDA) fifty topics is obtained. Some selected topics are analyzed and labelled such as 'aid by volunteers', 'psychological support to children', 'aid materials', 'resilience' and 'immigrants'. Also to provide accurate insights about disaster management bag of Ngram model is implemented and most used successive words are obtained.

IFRC Twitter content is especially related with voluntarily help to people and supply water, food etc. in the disaster response activities. Apparently, works in the preparedness phase of disaster are not as common as post-disaster works. Haiti and Syria are the highest mentioned countries in this content because of confusion in Syria and Haiti earthquake. Furthermore, earthquake and flood is the most mentioned disaster types.

Naïve Bayes, support vector machine, decision tree, random forest, neural network and k-nearest neighbor algorithms are implemented for classification after sentiment analyses. The higher value of accuracy is obtained from SVM as 76%. Results from ACC and F-measure are almost same with aforementioned six classification algorithms. Results from ACC and F-measure indicate consistency of sentiment analysis for SVM, Naïve Bayes, random forest and neural network algorithms. Decision tree and kNN algorithms are not affective to provide satisfactory results.

In general, it is seen in the study that IFRC takes disaster activities into consideration during and after phases of the disaster management. IFRC can talk about pre-disaster activities and trainings about disaster management and can operate pre-disaster activities more often. Besides activities for visible disasters, war, etc., it can also draw more attention to incidents such as abuse. In addition, more attention can be given to babies at war, sick and hunger-stricken babies.

In future, the study would be integrated more sophisticated models and sentiment classification can be improved. In addition, content analyses can be performed with more specific label words and software application can be developed in order to have comprehensive perspective for humanitarian activities via social media analyses.

Acknowledgements The authors did not receive support from any organization for the submitted work.

References

Blei D, Jordan M, Ng AY (2003) Latent Dirichlet allocation. Journal of Machine Learning Research. 3(4–5):993–1022

Bzdok D, Krzywinski M, Altman N (2018) Machine learning: supervised methods. Nat Methods 15:5–6

Cherichi S, Larodec RF (2017) Using big data values to enhance social event detection pattern. Proceedings of IEEE/ACS International Conference on Computer Systems Applications AICCSA, org/10 1109/AICCSA 2016 7945732

Cody EM, Reagan AJ, Mitchell L, Dodds PS, Danforth CM (2015) Climate change sentiment on Twitter: An unsolicited public opinion poll. PLoS ONE 10:1–18

Cutler A, Cutler DR, Stevens JR (2012) Random Forests. Ensemble Machine Learning, Springer, US, pp 157–175

Frank E, Trigg L, Holmes G, Witten IH (2000) Technical note: naive bayes for regression. Machine Learning 41:5–25

Gitari ND, Zuping Z, Herman W (2016) Detecting polarizing language in twitter using topic models ML algorithms. International Journal of Hybrid Information Technology 9:211–222

Gul S, Shah TA, Ahmad S, Mubashir M (2018) Twitter sentiments related to natural calamities: Analysing tweets related to the Jammu Kashmir floods of 2014. The Electronic Library 36:38–54

Gupta S, Altay N, Luo Z (2017) Big data in humanitarian supply chain management: a review further research directions. Ann Oper Res 5:1–21

Hu M, Liu (2004) Mining summarizing customer reviews. In Proceedings of the ACM SIGKDD International Conference on Knowledge. Seattle, Washington, USA. PP 38–46

Ikegami Y, Kawai K, Namihira Y, Tsuruta S (2013) Topic opinion classification based information credibility analysis on twitter. Proceedings - 2013 IEEE International Conference on Systems Man Cybernetics SMC. 2013; 4676–4681

International Federation of Red Cross Red Crescent Societies(IFRC) (2018) ifrc org/en/who-we-are/vision- -mission/. Accessed 28 August 2018

Jurafsky D, Martin JH (2018) Speech language processing. 3rd edition draft, org/10 1162/089120100750105975

Kecman V (2014) Basics of machine learning by support vector machines. Springer-Verlag, Berlin Heidelberg, Real World Applications of Computational Intelligence, pp 148–217

Khemka A, Wukich C (2017) Social media adoption message content reach: an examination of Red Cross Red Crescent national societies. Int J Emergency Manage 13:89–116

Kim DS, Kim JW (2014) Public opinion sensing trend analysis on social media: a study on nuclear power on Twitter. International Journal of Multimedia and Ubiquitous Engineering 9:373–384

Le J (2019) KDnuggets. The 10 Algorithms Machine Learning Engineers Need to Know. kdnuggets com/2016/08/10-algorithms-machine-learning-engineers html. Accessed 14 March 2019

Liu B, Hu M, Cheng J (2005) Opinion observer: analyzing and comparing Opinions on the web. International World Wide Web Conference Committee (IW3C2). Chiba, Japan

Louppe G (2014) Understanding Random Forests: From Theory to Practice. Dissertation, University of Liège

Malawani AD, Nurmandi A, Purnomo EP, Rahman T (2020) Social media in aid of post disaster management. Transforming Government: People, Process and Policy 14(2):237–260. https://doi.org/10.1108/TG-09-2019-0088

Wang B, Zhuang J (2017) Crisis information distribution on Twitter: a content analysis of tweets during Hurricane Sandy. Nat Hazards 89:161–181. https://doi.org/10.1007/s11069-017-2960-x

Xiao Y, Li B, Gong Z (2018) Real-time identification of urban rainstorm waterlogging disasters based on Weibo big data. Nat Hazards 94:833–842. https://doi.org/10.1007/s11069-018-3427-4

Monaghan A, Lycett M (2013) Big data humanitarian supply networks: Can Big Data give voice to the voiceless? Proceedings of the 3rd IEEE Global Humanitarian Technology Conference GHTC. 8; 432–437

Mulder F, Ferguson J, Groenewegen P, Boersma K, Wolbers J (2016) Questioning Big Data: Crowdsourcing crisis data towards an inclusive humanitarian response. Big Data & Society. org/10 1177/2053951716662054

Neppallia VK, Caragea C, Squicciarini A, Tapia A, Stehle S (2017) Sentiment analysis during Hurricane Sandy in emergency response. International Journal of Disaster Risk Reduction. 8:213–222

Onan A (2018) Biomedical Text Categorization Based on Ensemble Pruning Optimized Topic Modelling. Computational Mathematical Methods in Medicine. org/10 1155/2018/2497471

Pandey R, Purohit H (2018) Citizenhelper-adaptive: Expert-augmented streaming analytics system for emergency services humanitarian organizations. Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis Mining ASONAM 2018. 630–633. org/10 1109/ASONAM 2018 8508374

Papadopoulos T, Gunasekaran A, Dubey R, Altay N, Childe SJ, Fosso-Wamba S (2017) The role of Big Data in explaining disaster resilience in supply chains for sustainability. Journal of Cleaner Production. 142:1108–1118

Ray S (2019) Commonly used Machine Learning Algorithms. Analytics Vidhya. analyticsvidhya com/blog/2017/09/common-machine-learning-algorithms/. Accessed 17 March 2019

Reynard D, Shirgaokar M (2019) Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? Transportation Research Part D: Transport Environment 1–15. org/10 1016/j trd 2019 03 002

Rokach L, Maimon O (2005) Decısıon Trees. Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA, pp 165–192

Sharma P, Agrawal A, Lalit A, Garg A (2017) Challenges Techniques in Preprocessing for Twitter Data. International Journal of Engineering Science and Computing 7:6611–6613

Simon, H (2014) Neural Networks. Learning Machines Third Edition. org/978–0131471399

Tong Z, Zhang H (2016) A Text Mining Research Based on LDA Topic Modelling The Sixth International Conference on Computer Science, Engineering and Information Technology 201–210

Wukich C, Siciliano MD, Enia J, Boylan B (2017) The Formation of Transnational Knowledge Networks on Social Media. International Public Management Journal 20:381–408

Yu M, Huang Q, Qin H, Scheele C, Yang C (2019) Deep learning for real-time social media text classification for situation awareness– using Hurricanes Sandy, Harvey and Irma as case studies. International Journal of Digital Earth. org/10 1080/17538947 2019 1574316

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.