International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No. 1 (Mar-2022)

http://dx.doi.org/10.12785/ijcds/1101100

A Novel Framework for Assessing the Criticality of Retrieved Information

Ashwani Varshney[1], Yatin Kapoor[1], Vaishali Chawla[1] and Vibha Gaur[2]

[1]*Department of Computer Science, University of Delhi, New Delhi, India*
[2]*Department of Computer Science, Acharya Narendra Dev College, University of Delhi, New Delhi, India*

Received 12 Jun. 2021, Revised 21 Jan. 2022, Accepted 25 Feb. 2022, Published 31 Mar. 2022

Abstract: Data created by microblogging platforms provide an exceptional opportunity to mine valuable insights; however, their application in critical information retrieval is still at its inflection point. Taking advantage of Deep Learning (DL) and Natural Language Processing (NLP) techniques, this paper proposes a novel framework for retrieving critical information from Twitter to manage emergencies effectively. The proposed framework classifies the tweets into relevant and irrelevant classes using Bidirectional Encoder Representations from Transformers (BERT). Subsequently, relevant tweets are clustered using a k-means algorithm based on textual semantic similarity obtained using Universal Sentence Encoder (USE). Finally, the critical value of tweets is computed to segregate the relevant information that may assist the management teams to plan and organize their operations efficiently. The proposed work was tested on a real-world dataset of Uttarakhand Floods that occurred in February 2021. The critical information retrieved may be deployed to quickly manage disastrous situations and take the appropriate measures in time.

Keywords: Text Classification, BERT, k-means, Semantic Similarity, Clustering, Information Retrieval, Critical information.

1. Introduction

The advent of social media has marked a shift in information collection and dissemination during emergencies [1]. Researchers have leveraged social media to conduct numerous studies, including but not limited to outbreak detection [2], [3], information retrieval [4], evacuation behavior [5], [6], hazard assessment [6], and damage assessment [7], [8]. Twitter is one of the microblogging platforms that facilitate researchers in carrying out their studies on real-world data. It has 316 million users [9], empowering it with the capability of real-time feedback. Therefore, the potential of Twitter as a dependable and relevant data source is evident [10], [11]. It provides a platform for sharing crucial information and opinions on news updates, government and non-government initiatives and policies, and even requesting assistance during critical times [3], [12], [13]. Many studies [14], [15], [16], [17], [18] have proved that social media platforms like Twitter have also been beneficial to spread situational awareness during an emergency. Situational awareness may be defined as the available knowledge to assess and cope with a situation [18]. Thus, it plays an essential role in helping people during an emergency.

During critical times, government and non-government organizations look for related information to mitigate the adverse impact of the tragedies. Public authorities rely on timely and vital information to launch their operations timely for rescue management. The concerned authorities may send out alerts and learn the urgent needs of affected people to allocate the resources and take appropriate actions [8], [19]. Although social media platforms are promising data resources for emergencies, identifying helpful information for decision-making and action-taking is still challenging [20].

The vast volume of data on social media may contain an enormous amount of unwanted information, which can be overwhelming and bewildering to anyone trying to retrieve crucial details [21], [22]. Thus, it is essential to filter out the irrelevant information and segregate the relevant information to ensure that the precise information reaches the concerned authorities in time. Further, it may reduce the response time for relief and recovery measures during the emergency and increase situational awareness. As natural disasters are occurring more frequently than ever before, a system that could collect the data quickly during the crisis and facilitate assessing the severity of the situation

E-mail address: ashwani.mcs19.du@gmail.com, yatin.mcs19.du@gmail.com,vaishali.mcs18.du@gmail.com,vibhagaur@andc.du.ac.in

is the need of the hour. In such situations, Data Analytics (DA) and Artificial Intelligence (AI) can help to analyze the information rapidly from a large dataset to manage the disaster more efficiently. An effective information retrieval system that could quickly identify, extract, classify, and provide critical and crucial information from diverse and scattered sources to cope with the crisis is an urgent priority. Addressing the stated need, this paper proposes a novel framework to retrieve and segregate data from Twitter, considering the importance of extracting and prioritizing the critical information that may help the concerned authorities to respond timely. The tweets are classified to extract the relevant information and clustered based on textual semantic similarity. Finally, the critical values of the clustered tweets are computed to segregate the relevant information. The proposed framework facilitates the retrieval of critical information to support the timely management of rescue operations.

The proposed framework was tested on Uttarakhand Floods happened in February 2021 [23], and promising results were achieved. The relevant classified tweets were clustered, and k-optimal clusters were obtained. Additionally, the retrieved information was segregated and arranged in the descending order of the criticality so that the rescue management teams can plan their operations accordingly.

The main contribution of this study can be summarized as follows:

1) Classification of the extracted tweets into relevant if the information is valuable, irrelevant otherwise.
2) Semantically cluster the relevant classified tweets to target various rescue management operations.
3) Segregating the attained clusters to prioritize the rescue management based on the criticality of the information.

The organization of the paper is as follows: Section 2 includes the related work from the literature. Section 3 presents the proposed framework to assess and retrieve the critical information for planning and carrying out the rescue operations during the crisis. Section 4 shows the application of the proposed framework using the Uttarakhand Floods 2021 dataset. Section 5 concludes the paper and gives a direction to carry forward the future research.

2. Literature Review

In recent years, various research studies have been undertaken for information retrieval, including classifying information, analyzing its semantic features, and filtering out the relevant data for decision-making and action-taking. These studies present effective tools for obtaining holistic data and recognizing the concerns of affected stakeholders.

Classifying the information available on social media and extracting valuable insights from it have increasingly gained research interest. Studies indicate that supervised machine learning (ML) and deep learning (DL) approaches have been utilized for identifying crisis-related tweets during an emergency [14], [18], [24], [25]. ML techniques such as Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree [26] have been used for binary and multi-classification. However, the inherent disadvantage of ML models is the underlying training required from scratch for a domain-specific problem, making them unsuitable for application in diverse domains. Therefore, DL models quickly overtook ML counterparts owing to their ability to understand complex non-linear relationships and learn billions of trainable factors [27]. The advanced DL models need not be trained from scratch as they support transfer learning [28], making them more robust and valuable across varied domains. Some researchers used a monolingual BERT-based model for classification [29] and a universal model for fine-grained classification [30]. These models perform remarkably well on unseen data without further training. Researchers have also pursued a few multimodal approaches to address information retrieval during emergencies like disasters [31], [32], [33], [34] to identify the damage and assess its severity. Even though all these techniques emphasize information classification, retrieval of critical information has not been explored yet.

Further, various research studies have been carried out to retrieve and categorize relevant information required during the crisis to initiate the relief and recovery operations. A multi-classification model [35] for monitoring social media posts is presented that categorizes publicly available annotated datasets based on feature extraction and lexicon. However, the accuracy achieved by the above-mentioned classification models is inadequate and therefore not suitable for time-critical applications. A classifier for spreading situational awareness, which extracts syntactic features from the text of the tweets, is presented [36]. Their experiments showed that the classifier identifies situational awareness tweets with significantly higher accuracy than classifiers based on standard Bag-of-Words (BOW) models. However, the achieved accuracy of their classifier is not up to the mark since the model sometimes misclassifies the relevant tweets with a high error rate. The misclassification of tweets may be attributed to the lack of rigorous preprocessing on the dataset, as a large majority of the misclassified relevant tweets contained multiple fragments. Such a mix of relevant and personal sentiments in the same tweet brings down the overall objectivity of that tweet, resulting in the misclassification of critical information.

Various multimodal and DL approaches such as Long ShortTerm Memory (LSTM), Bidirectional Long Short Term Memory (Bi-LSTM), Convolutional Neural Networks (CNN), and BERT have been proven to give promising results for the classification task. However, classification alone is insufficient to aid the rescue management operations; therefore, it is essential to augment the classification techniques to obtain critical information.

Some researchers have utilized semantic similarity to retrieve the information and customize it as per the application [37]. A semantic analysis model [38] to identify the emergent knowledge related to natural disasters is presented using the Probabilistic Latent Semantic Analysis (PLSA) [39]. Their model clusters the information based on user preferences. However, it is a linear-distributional model and does not efficiently represent non-linear dependencies. A disaster event detection system [40] that utilizes a combination of semantic and sequential information of words using CNN is presented. Word embedding is performed to encode the words using word2vec [41], and the system performed well on a balanced dataset. However, their system misclassifies an unbalanced dataset; also, the word2vec algorithm does not detect Out-Of-Bag (OOB) words. Moreover, it was proved that Universal Sentence Encoder (USE) provides better sentence embedding than word2vec, GloVe, and ELMo [42].

Furthermore, various techniques have been deployed to automatically prioritize useful information for response and relief during an emergency. These include categorizing the relevant tweets [13], [28], [43], [44], and extracting messages of eyewitnesses [45], and classifying sentiment polarity of the tweets [46], [47], [48]. Some researchers worked on informativeness classification models using various ML and DL techniques [49]. An in-depth analysis of multiple disaster-related datasets and experimentation with the models under cross-domain conditions is performed considering linguistic, emotional, and sentimental features. However, the inclusion of semantic features could help improve the accuracy of the model.

The research so far indicates that the work in the direction of rescue management needs remarkable improvement. Moreover, the studies mentioned above focus on a single goal, be it classification or information retrieval; however, there is a need for a framework that could blend these techniques to extract critical information quickly and efficiently to deploy it for rescue and relief as well as improve situational awareness. To address this objective, this paper proposes a framework that employs information classification followed by clustering based on semantic analysis to segregate the critical information, as described in the following section.

3. Proposed Framework To Retrieve Critical Information

During critical times, rescue management teams look for vital information to mitigate the adverse impact of the disasters. Although Twitter is a good data resource in such situations, recognizing and extracting valuable and critical information automatically is still challenging. The massive amount of data contains undesirable information, which can be a significant barrier for retrieving crucial details. Therefore, it is essential to extract relevant data and segregate it as per the criticality to ensure that the correct and organized information reaches the concerned authorities in time. It can aid the management teams to take rapid mitigating measures and increase situational awareness. The proposed framework employs following the AI techniques to refine and assess the critical information retrieved from social media.

1) Classification: The extracted tweets are classified using the BERT model into Relevant (1) and Irrelevant (0) classes. The useful tweets containing valuable information during critical times are labeled as Relevant and Irrelevant otherwise. The pre-trained BERT model is utilized due to its transfer learning capability and potential to understand language context while training.
2) Clustering: The relevant classified tweets are clustered based on textual semantic similarity using the k-means algorithm. The Semantic Similarity is calculated using the USE because it considers sentence-level features for encoding the sentences and gives faster results. K-means algorithm provides state-of-the-art results in document clustering. Additionally, it gives high-quality clusters as compared to hierarchical clustering algorithms.
3) Segregation: Finally, the critical value of each tweet from every cluster is calculated using the word frequency. The relevant information in each cluster is prioritized in order of critical value to aid the rescue management teams in organizing their operations better.

The framework to assess critical information from social media is depicted in Fig. 1, and the phases are described below.

A. Data Extraction and Formation

Data Extraction and Formation involves gathering raw data from the social media platforms to carry out the study. The data is classified according to the predicted label. Several datasets are available for the classification task [50]. However, for critical information retrieval, one can create a domain-specific dataset and annotate it as per the usage. The quality of data is highly crucial for good training. Therefore, preprocessing of the data is employed.

B. Data Preprocessing

The extracted tweets need to be preprocessed to extract the valuable features, thus making them ready for classification. Tweets have unique features called entities such as emojis (☺), emoticons (:o), user mentions (@), hashtags (#), cashtags ($), coupled with typical web constructs, such as URLs and email addresses, and other noisy data, such as translate noisy data into sentences that include converting text to lowercase, removing the special characters such as parentheses, brackets, URLs, punctuations, hashtags, emoticons, pictographs, and stop-words utilizing the RegEx module [51].

C. Data Annotation

Data annotation involves assigning labels to a dataset for classification. The annotation of the tweets is necessary

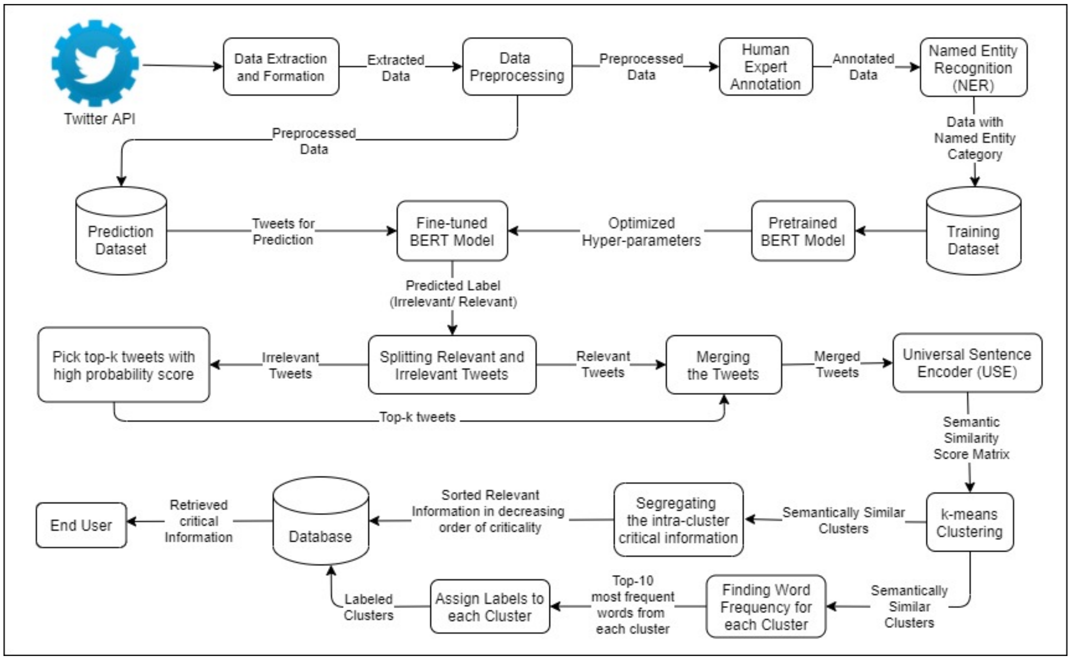


Figure 1. Proposed framework for retrieval of critical information

as the classification is essentially supervised. Annotation involves more than one annotator to obtain reliable and unbiased data. The tweets are assigned labels as Relevant (1) if the information present is valuable and Irrelevant (0) otherwise. Human annotation is required only once because once the data is annotated and the model learns on it, the same model can be used to generalize a different domain in the future. After annotating the tweets, it is essential to recognize the named entities to improve the quality of the training data.

D. Named Entity Recognition (NER)

NER involves determining the parts of a text that can be identified and categorized into predefined entities such as the names of individuals, locations, organizations, quantities, monetary values, expressions of times, and percentages [52]. NER is implemented using NLTK [53] and SpaCy [54] libraries of Python trained on the OntoNotes 5 corpus [55]. As the primary objective of this research is to gather critical information for disasters, it is necessary to treat each entity equally. NER is used to avoid any biasing during the training of the model and generalize the problem. NER enhances the overall training process of the model, thereby improving the accuracy of the model.

E. Model Training

The pre-trained BERT-base-uncased model [56] is utilized that can be hyper-tuned for the classification. The BERT model is chosen for classification due to its transfer learning capability and state-of-the-art performance [56]. BERT is a semi-supervised bidirectional language representation model based on transformers. It is built by transferring the encoder part of an encoder-decoder architecture for developing task-specific pre-trained models, which can be customized as per the application by only augmenting the output layer. Unlike other conventional models, the transformer encoder reads the entire sequence of tokenized words at once and understands the context better. Therefore, the model learns the representations in two directions simultaneously with a self-attention mechanism [57]. The pre-trained BERT model can be hyper-tuned as per the availability of the resources to get the optimal results. Once the classification is done, the labeled data can be clustered.

F. Merging the informative tweets

Since this study focuses on critical information retrieval; therefore, after the model training, the classified tweets are split as per their corresponding class labels. The relevant classified tweets contain crucial information; therefore, these tweets are considered for further processing. However, there is a possibility that the tweets that are classified as irrelevant may contain some valuable information. Therefore, even from the irrelevant classified tweets, it is vital to pick up the top-q tweets having a high probability score, where q is greater than or equal to the threshold value determined by the domain experts. Algorithm 1 for merging the informative tweets is described below.

1) *Let I be the Irrelevant tweets, R be the Relevant tweets, and Prob() returns the probability score corresponding to the Relevant Class.*
2) $\forall$ *'a' in I*
3) *IF Prob(a)* $\geq q$*, where* $q < 0.5$
4) $R := R \cup a$
5) *Return R*

The classification model predicts the probability scores for each class label. Since the binary classification is used, given a tweet, two probability scores are returned by the model corresponding to each class label. The default threshold value (taken as h) used by the classification model is 0.5. If the probability score of any of the class labels is greater than 0.5, the class label is assigned to the given instance. To avoid information loss, the threshold value for classifying a tweet into a relevant class may be relaxed depending on the criticality of the application and the amount of misclassified data. Subsequently, the union of the model-classified relevant tweets and the Irrelevant tweets with the relaxed threshold value (h) is considered the final tweet set for further processing.

G. Cluster Formation using Semantic Similarity

The resultant merged tweets are clustered using the k-means algorithm based on semantic similarity as the tweets having the same meaning would convey related information. Google's USE is used to compute the pairwise semantic similarity scores of the tweets.

1) Universal Sentence Encoder (USE)

The USE encodes text into high-dimensional vectors for utilization in various NLP tasks such as semantic similarity, text classification, and clustering. The pre-trained USE is publicly available [58]. Once data is converted into a vector, the similarity between two data points is computed. While embedding a sentence and its words, the context of the whole sentence needs to be apprehended in that vector. The USE performs this task by inputting a variable-length English text, and the output is a 512-dimensional vector. The model is trained and optimized for sentences, phrases, or short paragraphs. The pairwise textual semantic similarity is calculated for every pair of tweets. Hence, the input to the pre-trained USE is the data frame of the merged tweets obtained in the previous section, and the output is a matrix of dimension: number of tweets $\times$ 512. Since the values are normalized, the inner product of encodings can be treated as a similarity matrix. This matrix is then passed onto the k-means module to cluster the tweets based on the obtained semantic similarity matrix. K-means clustering algorithm is employed for document clustering.

2) K-means clustering

K-means is one of the popularly used unsupervised learning algorithms which identifies k pre-specified, distinct, and non-overlapping subgroups (clusters) in the dataset [59]. It employs centroid as a representation of data points. The accuracy achieved by this algorithm is often better than many of its counterparts [60]. Therefore, it is chosen in this study to cluster the final relevant classified tweets. Algorithm 2 described below presents the k-means clustering algorithm in brief:

1) *Initially, k-points from the similarity matrix obtained using USE are designated as centroids.*
2) *k-clusters are formed by assigning each data point to its nearest centroid. It is done in the first pass over the similarity matrix.*
3) *Centroids are re-calculated for each cluster.*
4) *Repeat Step-2 and Step-3 until there is no change in centroids in Step-3.*

Algorithm 2 determines the number of clusters (taken as k) into which the given data should be clustered. The k-means algorithm works particularly well if the choice of k is optimal [59], which is the crucial step to the process. Several metrics are available for the same; however, the Elbow method [61] is the simplest and effective one; therefore, it is chosen for this study. It plots the number of clusters (k) versus inertia which is defined as the sum of squared distances of data points from the center of their closest cluster. An optimal number of clusters (k_{opt}) is the point after which inertia starts dropping linearly, also called the 'elbow' of the plot. As per the elbow of the plot, the optimal number of clusters (k_{opt}) is attained.

H. Assigning labels to the clusters

Once the optimal clusters are obtained, the next step is to assign suitable labels to them. Since a label reflects the underlying central theme of the cluster, it is assigned to each cluster so that the respective stakeholders can identify the desired cluster easily. A label may or may not be assigned from the words present in the cluster. For simplicity, this study makes use of essential words obtained in the cluster. The words which are occurring more frequently than others are inherently significant to the clusters. The most frequently occurring words in the clusters are finally used to assign an appropriate label to them.

I. Cluster segregation and computation of critical value

This study focuses on retrieving critical information for time-sensitive tasks. Therefore, it seems feasible to present vital information presented in the tweet based on the criticality value. Therefore, the clusters of relevant tweets obtained in the previous phase are further segregated. Algorithm 3 utilized to compute the criticality of a tweet is explained below.

1) *Let S(w, o) be a set of vocabulary with a word as w and occurrence of each w as o, T be a tweet, w' be a word in T, and C be the set of criticality values* $c_1, c_2, c_3, \ldots c_n$ *of each T.*
2) *∀ w' in T*
3) *IF w' in S(w,o)*
4) $c_i := c_i + o$
5) $C = C \cup c_i$
6) *Return* $C/max(c_i) \in [0,1]$

The frequency of each word is determined to calculate the criticality of each tweet, and the value of frequency is aggregated for every matched word in the tweet. The resulting aggregated value for each tweet is normalized in the range [0,1]. The relevant tweets can be arranged in the

Tweets before preprocessing	Tweets after preprocessing
Rajya Sabha observes a moment of silence in memory of people who have lost their lives in the #UttarakhandDisaster @NewIndianXpress @TheMornStandard	Rajya Sabha observes a moment of silence in memory of people who have lost their lives in the UttarakhandDisaster NewIndianXpress TheMornStandard
@SriNithyananda Prayer to Paramashiva for all those lives out there in #Uttarakhand #UttarakhandDisaster #nithyananda	SriNithyananda Prayer to Paramashiva for all those lives out there in nithyananda
#UttarakhandDisaster Made Many Livelihood into Unconditional and Drastic Situation Everything Should Get into Normal Condition if Any Help Need Too In Cause Plz Plz Dial Too 1070 Or 9557444486 Mainly Salute For the jawans And People Who joined in and Helped https://t.co/VaVHAfbFud	UttarakhandDisaster Made Many Livelihood into Unconditional and Drastic Situation Everything Should Get into Normal Condition if Any Help Need Too In Cause Plz Plz Dial Too 1070 Or 9557444486 Mainly Salute For the jawans And People Who joined in and Helped
20 bodies recovered in #Uttarakhand Only 02 have been identified so far 39 persons including workers and senior officials of a construction company trapped in #Tapovan tunnel 172 #NTPC officials and 12 villagers are missing	20 bodies recovered in Only 02 have been identified so far 39 persons including workers and senior officials of a construction company trapped in tunnel 172 NTPC officials and 12 villagers are missing
At least 15 people lost their lives and 203 are still missing as of now after a glacier broke in Uttarakhand's Chamoli triggered massive flooding of the Alaknanda and Dhauliganga rivers #newsmakhani #news #Uttarakhand #GlacierBurst #UttarakhandGlacierBurst #UttarakhandTragedy https://t.co/ibMdJUtoPS	At least 15 people lost their lives and 203 are still missing as of now after a glacier broke in Uttarakhands Chamoli triggered massive flooding of the Alaknanda and Dhauliganga rivers UttarakhandTragedy

Figure 2. A sample of tweets before and after preprocessing

decreasing order of their criticality scores. The final results can be saved onto a database to ensure ease of access to the respective stakeholders. The tweets containing valuable and critical information are segregated and can be utilized by various rescue management teams to launch their operations efficiently.

4. Results and Analysis

This section presents the experimental setup for the undertaken case study to illustrate the application of the proposed framework for critical information retrieval. The framework was trained and tested on the real-world data obtained from the Uttarakhand Floods 2021. The purpose of taking a case study of the disaster was that it is incredibly time-sensitive. Therefore, the rescue management and disaster response teams must respond and start their operation at the earliest. The experimental setting and their respective results are discussed below.

The data used in this study was crawled from Twitter using the Twitter search API [62]. The extraction of tweets was based on the hashtags #UttarakhandFloods, #UttarakhandDisaster, #Tapovan, #GlacierBurst, #Joshimath, and #RishiGangaPowerProject, which were trending on Twitter in India during the first week of February as soon as the disaster hit the state. A total of 7500 tweets during February 7-10, 2021, were obtained. All the tweets were scraped from the public profile of Twitter users. For this research, tweets only in the English language were considered, while tweets in a language other than English were removed.

Additionally, after removing the duplicate tweets and the tweets with unknown words and slang terms, the count of the final tweets was reduced to 3472. It was also observed that the tweets contained a lot of noisy non-disaster related data. Therefore, a filtering step was incorporated to keep the tweets containing the following disaster-specific keywords: rescue, help, shelter, disaster, underground, lives, death, emergency, contact, trap, response, relief, government, food, and cloth. Finally, 2323 tweets were considered and preprocessed to eliminate non-essential entities, such as stop words, punctuation, special symbols, hashtags, and URLs. A sample of the tweets before and after preprocessing is shown in Fig. 2.

Each tweet, after preprocessing was given a label depending on the valuable information it contains concerning disaster management. Being considerably mindful of the disaster, the class label, either Relevant (1) or Irrelevant (0), was identified for each tweet. Five anonymous domain experts were included in this study for annotating the scraped tweets. Guidelines were shared among the annotators to limit and resolve the disagreement. The annotation process did not involve any of the authors of this research. The guidelines for annotation are listed below:

1) The information present in the tweets should be considered without any personal bias.
2) The tweet contains information regarding the whereabouts and count of the trapped people, supply of food, medicine, and essential services, count of deaths, any helpline number, ongoing rescue operation, and donation or helps in any form should be labeled as Relevant (1).
3) Tweets containing prayer and condolence messages were of no use to the rescue management organizations; hence, they should be labeled Irrelevant (0).

As five annotators were involved, Fleiss's Kappa was utilized to calculate the Inter Annotator Agreement (IAA) [63], [64]. The manually annotated corpus achieved an IAA of 95.09%, and the majority label was finalized as the gold standard. The count of the irrelevant and relevant tweets was 1582 and 741, respectively. Since the data distribution was almost 2:1 for Irrelevant and Relevant class labels, which was highly unbalanced, the irrelevant labeled tweets were inspected. It was found out that most of these tweets were delivering redundant information. Therefore, the irrelevant labeled data was under-sampled to 741. Finally, the count of the resultant tweets was 1482 with a balanced class distribution. The authors of the study did not manipulate the information in any way. This study's sole purpose was to provide high-priority information to rescue management organizations preserving the originality of the information without loss.

As the names of the locations and people were mentioned in the tweets, it was essential to consider and replace them with their corresponding category types to avoid information bias during model training. The snapshot of one such tweet is presented in Fig. 3.

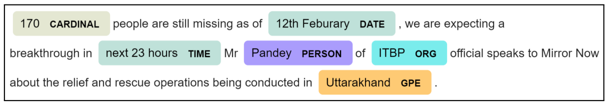


Figure 3. Snapshot of the Named Entity recognition performed

In Fig. 3, CARDINAL, DATE, TIME, PERSON, ORG, and GPE were recognized as NER representing numeral, absolute or relative date, time shorter than a day, people including fictional names, organization, and state, respectively [52]. The snapshot of the final dataset for training the model used for the classification task is shown in Fig. 4. The column *"Tweets after NER"* in Fig. 4 depicts the recognized entities replaced with their corresponding categories. The columns *"Tweets after preprocessing"* and *"Tweets after NER"*, as depicted in Fig. 4, were utilized separately for the classification task to compare the performance in both scenarios.

The classification task was carried out using the BERT-base-uncased model. The BERT model was implemented using TensorFlow 2.0 and the Transformers library in Python. The snapshot of the code is shown in Fig. 5. The pre-trained BERT model was fine-tuned for the undertaken binary classification task. A dropout layer with a dropout rate of 50% was added to the output of the BERT layer to avoid overfitting. Subsequently, a Dense layer of 768 neurons with Tanh activation function was added, followed by a second dropout layer with the rate of 50%, and a Dense layer with two neurons was utilized with Softmax activation function. The output of the BERT model resulted in probability scores, and the tweets were classified into two categories: irrelevant and relevant based on these scores. The Sparse Categorical Cross Entropy was used as the loss function to optimize the performance of the BERT model. The training phase utilized the different values of hyper-parameters such as Optimizer function, Epochs, number of layers, Learning Rate for performing experiments several times. Eventually, the model was fine-tuned with the optimal values, as shown in Table I.

After the model was trained and fine-tuned, it was used to make predictions. This study compared two variations of training data during the model training: data with NER and without NER. The performance observed in both cases is shown in Fig. 6. The metrics utilized to compare both the model configurations are described below [63].

- Accuracy: The ratio of correctly predicted observations to the total observations.

TABLE I. HYPER-PARAMETERS OF THE BERT MODEL USED IN THE STUDY

Hyper-parameters	**Optimal Value**
Max. Sequence Length	128
Optimizer	Adam
Loss Function	Sparse categorical cross entropy
No.of Epochs	2
No. of Layers	12
Batch Size	16
Learning Rate	3e-5
Activation Function	Tanh (hidden layer), Sigmoid (output layer)
Train, test, and validation split	70%, 18%, and 12%, resp.
Dropout	0.5

- Precision: The ratio of correctly predicted positive observations to the total predicted positive observations.
- Recall: The ratio of correctly predicted positive observations to all observations in the actual class.
- F1-measure: The weighted average or Harmonic mean of Precision and Recall.

The higher the value of metrics, the better is the performance of the model. Fig. 6 clearly shows that the performance recorded in the case of model training using NER was better, strengthening the fact that NER helps extract crucial information from unstructured data, thus improving the overall training of the model. For instance, the accuracy achieved by the model with NER was almost 95%, suggesting that, on average, the model can correctly classify almost 95 out of every 100 tweets containing relevant information. Similarly, Precision, Recall, and F1-measure values were higher with NER indicating improved performance before training the model. Even the smallest piece of information can prove vital during crucial and sensitive times like disasters. Hence, the high value of accuracy ensured that most of the relevant information was correctly classified and passed on to the next stage to obtain semantically similar clusters.

In times of calamity, even a minuscule amount of information can prove to be a life savior. Therefore, once the tweets were classified into their respective Relevant (1) and Irrelevant (0) classes, the Relevant and top-q tweets from the Irrelevant class were merged. Even though some irrelevant tweets may get classified into relevant, it does not affect the disaster response operations; however, the converse is not

Tweet after preprocessing	Tweet after NER	Label
such a horrifying incidence took place nature is beautiful but at the same time is more disastrous itbp official teams in rescue mission near tapovanusing mountaineering equipment hope all things gets back on the track soon uttarakhanddisaster	horrifying incidence took place nature beautiful time disastrous <GPE> official teams rescue mission near tapovanusing mountaineering equipment hope things gets back track soon	0
marcos team on standby 16 in delhi 40 at mumbai for uttarakhand rescue and relief operation jaihind	<PERSON> team standby <CARDINAL> delhi <CARDINAL> mumbai rescue relief operation jaihind	1
while china took 30 hours to send a rescue team to rescue its personnel trapped in the mine and 10 died indias rescue team evacuated everyone from the tunnel in 4 hours who is more efficient and serious uttarakhanddisaster	<GPE> took <TIME> send rescue team rescue personnel trapped mine <CARDINAL> died indias rescue team evacuated everyone tunnel <TIME> efficient serious	1
...	...	...
aerial rescue relief missions resume as helicopters fly to with disaster relief teams on board	aerial rescue relief missions resume helicopters fly disaster relief teams board	0
army is helping people of uttrakhand rescueing and clearing road still number of people are missing	<ORG> helping people uttrakhand rescueing clearing road still number people missing	1
30 people dead rescue operations still on at the tapovan tunnel	<CARDINAL> people dead rescue operations still tapovan tunnel	1

Figure 4. Final Training data

```
train_InputExamples, validation_InputExamples = convert_data_to_examples(
                                    train, test, DATA_COLUMN, LABEL_COLUMN)

train_data = convert_examples_to_tf_dataset(list(train_InputExamples),
                                                         tokenizer)
train_data = train_data.shuffle(100).batch(16).repeat(2)

validation_data = convert_examples_to_tf_dataset(list(validation_InputExamples)
                                                         , tokenizer)
validation_data = validation_data.batch(16)

model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=3e-05,
                                                 epsilon=1e-08,
                                                 clipnorm=1.0),
          loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
          metrics=[tf.keras.metrics.SparseCategoricalAccuracy('accuracy')])

model.fit(train_data, epochs=2, validation_data=validation_data)
```

Figure 5. Snapshot of the configuration of the BERT model for classification

true. Missing out any relevant information may prove to be fatal and may affect several lives. Therefore, the Irrelevant class tweets with a probability score of more than q were also picked up. The value of q was decided by domain experts and was taken as 0.35. Finally, the merged data contained the tweets from the Relevant class (complete) and Irrelevant class (top-q). The tweets after merging are shown in Fig. 7 with the highlighted top-q irrelevant tweets.

Further, the resultant dataset was clustered to be utilized by the different rescue teams. Several teams operate simultaneously during an emergency, involving response operation teams, SOS helpline coordinators, relief camps supervisors, and donation heads. Therefore, it is essential to retrieve information that could be utilized simultaneously and circulated among different teams. So, after classification, the relevant tweets were clustered so that each cluster provides the critical information in concern to a specific team.

The textual semantic similarity was calculated for all the tweets. A sample of 15 tweets was selected, as shown in Fig. 8, and similarity between a pair of tweets using a Heatmap is shown in Fig. 9. A heatmap is a valuable tool to view the similarity between the tweets; the high intensity of the color indicates higher similarity among tweets.

As shown in Fig. 8, tweet 12 (*"rescue operations underway in tapovan tunnel"*) is highly semantically similar to tweet 7 (*"disasterarmy opens blocked tunnel at tapovan rescue operations on"*). At the same time, it is highly dissimilar to tweet 9 (*"pls tag relief organizationsngos assisting with the thanks"*), and the heatmap representation in Fig. 9 well supports the results.

Once the semantic similarity scores were obtained in the form of a matrix, it was fed to the k-means clustering algorithm as input. Before clustering, finding the optimal number of clusters is necessary to form high-quality clusters [61]; hence, the Elbow method was applied. As shown in Fig. 10, the optimal value of k was obtained as 3. Therefore, the tweets were grouped into three optimal clusters, and the tweets in the dataset were assigned a cluster number to identify the cluster to which it belongs.

The respective word clouds shown in Fig. 11 were formed to better visualize the clusters, where the size of the word indicates its frequency in the cluster [65].

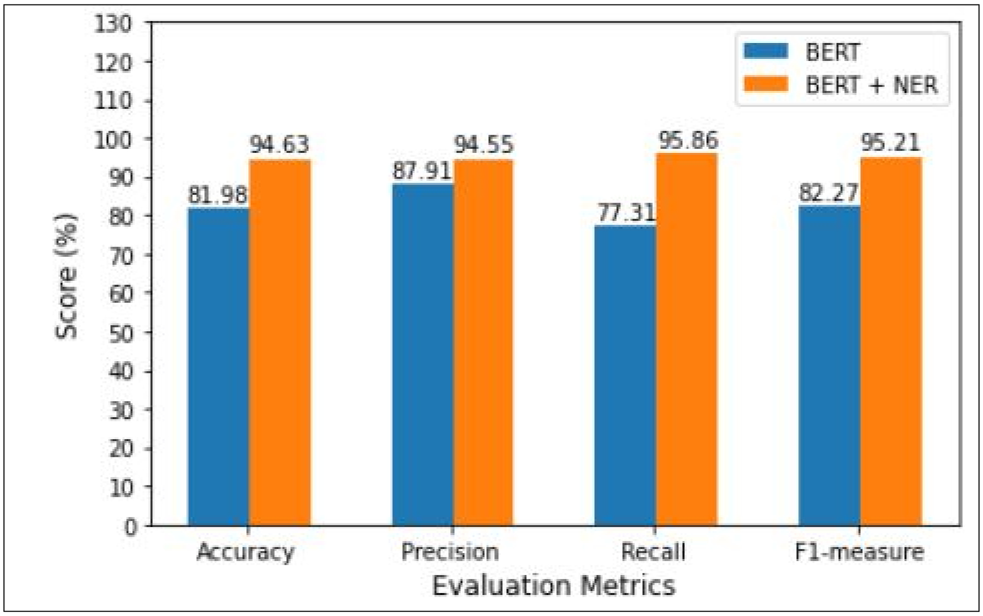


Figure 6. Comparison of model performance with and without NER

Tweets after NER	Probability Score
starting <CARDINAL> <DATE> <CARDINAL> mi <CARD...	0.968765
stranded affected areas need help please conta...	0.968254
<ORG> state overseeing disaster management flo...	0.971516
<CARDINAL> people still missing expecting brea...	0.851694
<CARDINAL> bodies recovered169 still missingre...	0.975972
...	...
glacialburst aerial view devastation site cham...	0.960852
<CARDINAL> dead <CARDINAL>0 injured rescue ope...	0.374731
<CARDINAL> dead <CARDINAL> people still trappe...	0.431067
dgp <PERSON> explains <ORG> leading mass devas...	0.382648
extraordinary footage <GPE> official rescue op...	0.417408

Figure 7. Dataset after merging the tweets

Further, to efficiently organize the retrieved information for its rapid utilization, labels were assigned to each cluster so that various rescue management teams could utilize them as per their need. Labels were decided based on the top-10 most frequent words from each cluster, as depicted in Fig. 12.

Each cluster was assigned a meaningful label shown in Table II for easy and quick access to the relevant information. Once each cluster was assigned a suitable label, the tweets within each cluster were segregated in decreasing order of criticality values using Algorithm 3. The range of criticality was defined and finalized after consulting with the domain experts to achieve the segregation shown in Table III, where C is the criticality value. The optimized criticality values were decided after rigorous experimentation, which came out to be sound as per the domain experts, strengthening the framework's capability to segregate the critical information. Finally, the tweets were segregated into High, followed by Medium and Low severity categories.

Fig. 13 shows the distribution of the tweets according to their criticality values in which the dotted lines depict the criticality cut-off as described in Table III. The results for

Tweet Number	Tweet
1	28 bodies recovered169 still missingrescue operation underway joshimath
2	rescue teams clear debris at tunnel entry more videos
3	rescue work underway at itbp official tsrawatbjp
4	ntpcs under construction hydropower project was damaged trapped chamoli
5	garuda aerospace drones to be deployed in chamoli rescue op
6	update ndrfhq incident site utilising pmoindia hmoindia bhallaajay26 pibhomeaffairs pibdehradun
7	disasterarmy opens blocked tunnel at tapovan rescue operations on
8	civil police at participating in the rescue operation
9	pls tag relief organizationsngos assisting with the thanks
10	153 missing from tapovan project sites rescue operation underway chamoliflood
11	photos glacierrescue operations underway in chamoli
12	rescue operations underway in tapovan tunnel
13	death toll rises to 32 rescue operation underway
14	over 20 dead 200 injured rescue operations continue
15	14 dead and 30 people still trapped in a tunnel

Figure 8. A sample of tweets

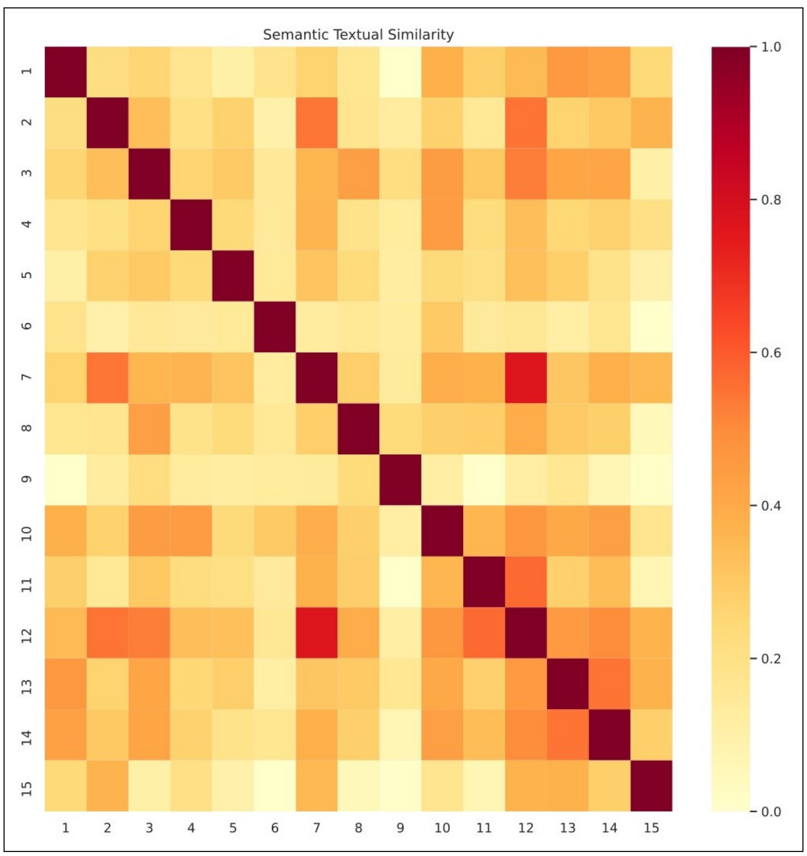


Figure 9. Heatmap of the semantic similarity scores for a sample of 15 tweets

one of the clusters containing tweets arranged in the non-increasing order of their criticality are shown in Fig. 14. As shown in the Fig. 14, the information was segregated into H, M, and L categories. Further, the segregated tweets were stored in a database for ease of access and retrieval. This information can be shared with the concerned rescue management teams working on disaster-related operations. It would help these teams prioritize their actions based on the severity of the situation and the information conveyed through the tweet.

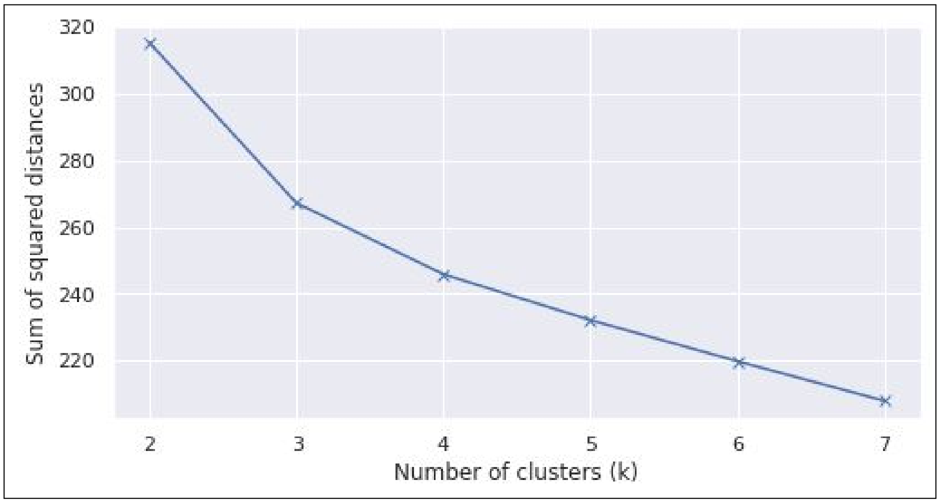
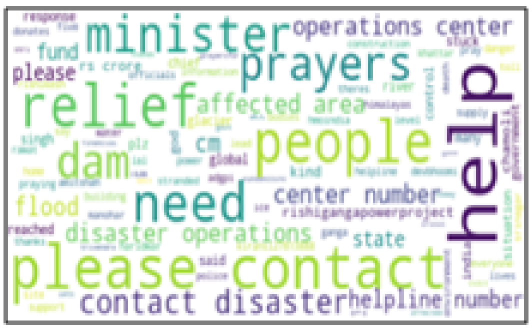
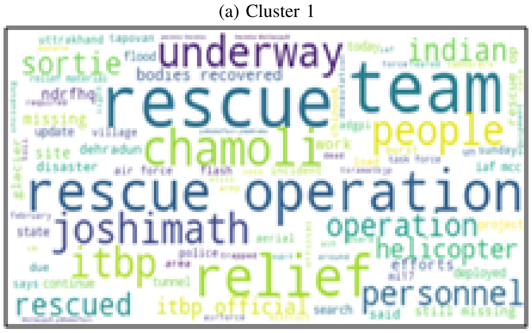


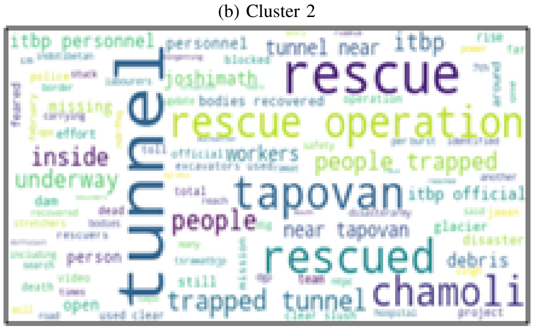
Figure 10. Sum of squared distances vs. No of clusters forming an elbow at k = 3



(a) Cluster 1



(b) Cluster 2



(c) Cluster 3

Figure 11. Word Clouds

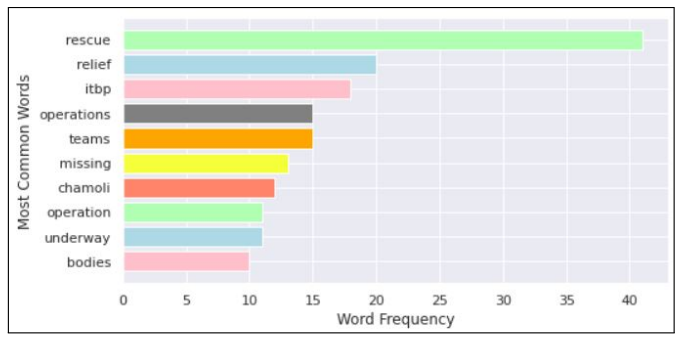


Figure 12. The count of the top-10 most frequent words

TABLE II. LABELS ASSIGNED TO THE CLUSTERS ACCORDING TO THE TOP-10 MOST FREQUENT WORDS

Cluster	Label
0	Tweets related to help line contact numbers
1	Tweets related to relief operations, rescue teams
2	Tweets related to trapped people, rescued people

A. Comparative Analysis

This research study proposes a novel framework for retrieving critical information from diverse and scattered tweets, especially in time-critical situations. Although studies in the past have utilized ML techniques [66] and Reinforcement Learning methods [67] for efficient information retrieval, this study takes information retrieval a step further for emergencies like disasters, segregating the critical part of the retrieved information to aid rescue management.

A multi-classification model [35] for monitoring social

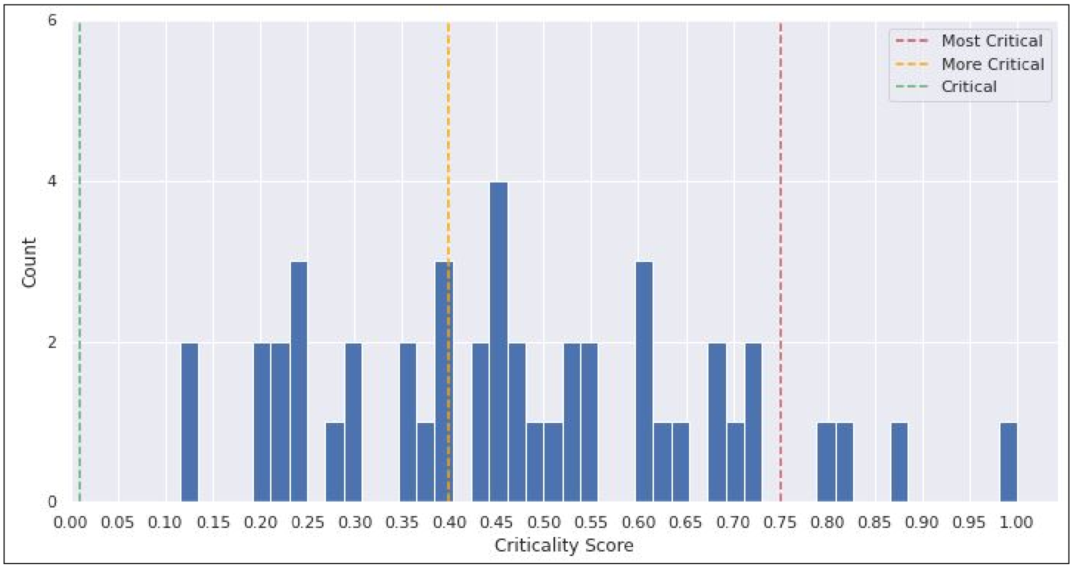


Figure 13. Segregation thresholds

TABLE III. RANGE OF SEGREGATION OF THE CLUSTERS AND THEIR ASSIGNED CATEGORIES

S. No.	Criticality Value	Category	Meaning
1	≥ 0.75	H	High severity
2	0.40 ≥ C > 0.75	M	Medium severity
3	< 0.40	L	Low severity

media posts is presented that categorizes publicly available annotated datasets based on feature extraction and lexicon. Similarly, a reputation analysis of the product is performed from the extracted information to identify the positive and negative parts from the ratings, achieving an accuracy of around 86% [68]. However, the accuracy achieved by the above-mentioned classification models is inadequate and therefore not suitable for time-critical applications. This study uses NER before model training, which helps the classification model achieve better accuracy than direct model training.

5. Conclusion and Future Scope

Social media is a powerful tool for extracting and analyzing data for various time-sensitive tasks, including enhancing situational awareness, planning and controlling operations during a crisis. Twitter is a social media platform that facilitates the rapid dissemination of information in real-time during a disaster. However, extracting valuable information from the tweets is a challenging task. Therefore, this paper proposed a novel framework for retrieving information from Twitter and segregating it for rescue management operations, considering the importance of extracting critical information during a disaster. The proposed approach consists of various phases, including data formation, data preprocessing, NER, classification, cluster-formation, and prioritizing information. The data was extracted from Twitter using Twitter search API followed by preprocessing to eliminate non-essential entities, such as stop words, punctuation, special symbols, hashtags, and URLs. Additionally, NER was performed to replace the recognized entities with their corresponding categories to avoid information bias during model training for classification.

The pre-trained BERT-base-uncased model was used to classify the tweets into relevant and irrelevant classes. The model was tested on a real-world dataset of the Uttarakhand Floods 2021, and promising results were achieved. The tweet classification task attained an accuracy of 94.9%. Further, the top-q tweets from the irrelevant class were merged with the relevant class tweets to minimize the chances of losing critical information. Furthermore, the merged tweets were semantically encoded using USE and clustered by the k-means algorithm, and three optimal clusters were obtained. Each cluster was assigned a label, based on the top-10 most frequent words that appeared in each cluster, to ease the utilization of information for the rescue management teams. Finally, the critical value of the tweets from each cluster was calculated, and the information was arranged in decreasing order of the criticality to assist the rescue management teams in planning and performing their operations faster. The segregated data contained 5.92% extremely critical tweets, followed by 27.63% moderately critical, and remaining 66.45% tweets were mildly critical.

The proposed technique can further be extended for future improvements. The size of the training data can be increased and generalized for different domains. Moreover, a granular level of preprocessing can be performed to

Cluster Number	Tweet	Criticality	Code
1	nair hena 1 kewal nilesh pat priyakhi2 am dili...	1.000000	H
1	iaf mcc update starting from 0645 hrs today 6 ...	0.996892	H
1	starting from 645 am today 6 mi 17 sorties 1 a...	0.963481	H
1	indianairforce resumed aerial rescue relief ...	0.592852	M
1	amit shah on personnel of 5 teams of ndrfhq 8 ...	0.515929	M
...	...	...	...
1	airnews relief aerial view of the devastation...	0.082362	L
1	glacier burst aftermathrescue operations halt...	0.065268	L
1	update ndrfhq teams airlifted for incident si...	0.055167	L
1	photos glacierrescue operations underway in ch...	0.045066	L
1	ntpcs under construction hydropower project w...	0.043512	L

Figure 14. Tweets with Segregation Category: H, M, and L

enhance the classification model for training. Tweets in local languages may be extracted to address location-specific disasters and emergencies, followed by multilingual models for training.

Acknowledgment

We thank the anonymous annotators for annotating the dataset utilized in the study. Moreover, we thank the anonymous reviewers for their valuable suggestions.

References

[1] C. Reuter and M. A. Kaufhold, "Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics," *Journal of Contingencies and Crisis Management*, vol. 26, no. 1, pp. 41–57, 2018.

[2] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Extracting information nuggets from disaster- Related messages in social media," *ISCRAM 2013 Conference Proceedings - 10th International Conference on Information Systems for Crisis Response and Management*, no. May, pp. 791–801, 2013.

[3] M. Imran, P. Mitra, and C. Castillo, "Twitter as a lifeline: Human-annotated Twitter corpora for NLP of crisis-related messages," *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, pp. 1638–1643, 2016.

[4] H. Li, D. Caragea, C. Caragea, and N. Herndon, "Disaster response aided by tweet classification with a domain adaptation approach," *Journal of Contingencies and Crisis Management*, vol. 26, no. 1, pp. 16–27, 2018.

[5] D. Metaxa-Kakavouli, P. Maas, and D. P. Aldrich, "How social ties influence hurricane evacuation behavior," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, 2018.

[6] L. Burks, M. Miller, and R. Zadeh, "Rapid estimate of ground shaking intensity by combining simple earthquake characteristics with tweets," *NCEE 2014 - 10th U.S. National Conference on Earthquake Engineering: Frontiers of Earthquake Engineering*, 2014.

[7] K. Stowe, J. Anderson, M. Palmer, L. Palen, and K. Anderson, "Improving Classification of Twitter Behavior During Hurricane Events," pp. 67–75, 2019.

[8] B. Resch, F. Usländer, and C. Havas, "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment," *Cartography and Geographic Information Science*, vol. 45, no. 4, pp. 362–376, 2018.

[9] D. Olanoff, "Twitter Monthly Active Users Crawl To 316M, Dorsey: "We Are Not Satisfied"," 2015.

[10] D. Gayo-Avello, "Political opinion," *Twitter: A Digital Socioscope*, pp. 52–74, 2015.

[11] M. J. Paul and M. Dredze, "You Are What You Tweet: Analyzing Twitter for Public Health," *Icwsm*, no. 20, pp. 265–272, 2011.

[12] T. Situations, "978-1-107-13576-5 - Big Crisis Data : Social Media in Disasters and Time-Critical Situations Carlos Castillo Frontmatter More information © in this web service Cambridge University Press Cambridge University Press 978-1-107-13576-5 - Big Crisis Data : Soc."

[13] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, *Processing Social Media Messages in Mass Emergency: Survey Summary*, 2018, vol. 47, no. 4.

[14] M. Basu, A. Shandilya, P. Khosla, K. Ghosh, and S. Ghosh, "Extracting Resource Needs and Availabilities from Microblogs for Aiding Post-Disaster Relief Operations," *IEEE Transactions on*

Computational Social Systems, vol. 6, no. 3, pp. 604–618, 2019.

[15] K. Rudra, P. Goyal, N. Ganguly, P. Mitra, and M. Imran, “Identifying sub-events and summarizing disaster-related information from microblogs,” *41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018*, pp. 265–274, 2018.

[16] K. Rudra, N. Ganguly, and P. Goyal, “Extracting and Summarizing Situational Information,” *ACM Transactions on WEB*, vol. 12, no. 3, p. 35, 2018.

[17] S. Madichetty and S. Muthukumarasamy, “Detection of situational information from Twitter during disaster using deep learning models,” *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 45, no. 1, 2020.

[18] M. Sreenivasulu and M. Sridevi, “Comparative study of statistical features to detect the target event during disaster,” *Big Data Mining and Analytics*, vol. 3, no. 2, pp. 121–130, 2020.

[19] N. B. Sarter and D. D. Woods, “Situation Awareness: A Critical But Ill-Defined Phenomenon,” *The International Journal of Aviation Psychology*, vol. 1, no. 1, pp. 45–57, 1991.

[20] C. Granell and F. O. Ostermann, “Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management,” *Computers, Environment and Urban Systems*, vol. 59, pp. 231–243, 2016.

[21] A. H. Tapia, K. Bajpai, B. J. Jansen, and J. Yen, “Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations,” *8th International Conference on Information Systems for Crisis Response and Management: From Early-Warning Systems to Preparedness and Training, ISCRAM 2011*, no. January, 2011.

[22] C. Reuter, A. L. Hughes, and M. A. Kaufhold, “Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research,” *International Journal of Human-Computer Interaction*, vol. 34, no. 4, pp. 280–294, 2018.

[23] D. S. Rathore. (2021) Uttarakhand glacier burst. Accessed Feb. 12, 2021. [Online]. Available: https://www.indiatoday.in/india/story/uttarakhand-glacier-burst-flood-joshimath-alert-haridwar-rishikesh-live-updates-1766778-2021-02-07

[24] S. Madichetty and S. M, “Identification of medical resource tweets using Majority Voting-based Ensemble during disaster,” *Social Network Analysis and Mining*, vol. 10, no. 1, 2020.

[25] S. Cresci, M. Tesconi, A. Cimino, and F. Dell’Orletta, “A linguistically-driven approach to cross-event damage assessment of natural disasters from social media messages,” *WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web*, pp. 1195–1200, 2015.

[26] M. R. Nair, G. R. Ramya, and P. B. Sivakumar, “Usage and analysis of Twitter during 2015 Chennai flood towards disaster management,” *Procedia Computer Science*, vol. 115, pp. 350–358, 2017.

[27] L. Arnold, S. Rebecchi, S. Chevallier, and H. Paugam-Moisy, “An introduction to deep learning,” *ESANN 2011 - 19th European Symposium on Artificial Neural Networks*, pp. 477–488, 2011.

[28] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, *Artificial Neural Networks and Machine Learning – ICANN 2018*. Springer International Publishing, 2018, vol. 11141, no. November.

[29] G. Ma, “Tweets Classification with BERT in the Field of Disaster Management,” pp. 1–15.

[30] J. Ray Chowdhury, C. Caragea, and D. Caragea, “Cross-Lingual Disaster-related Multi-label Tweet Classification with Manifold Mixup,” pp. 292–298, 2020.

[31] A. K. Gautam, L. Misra, A. Kumar, K. Misra, S. Aggarwal, and R. R. Shah, “Multimodal analysis of disaster tweets,” *Proceedings - 2019 IEEE 5th International Conference on Multimedia Big Data, BigMM 2019*, pp. 94–103, 2019.

[32] F. Alam, F. Ofli, and M. Imran, “Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria,” *Behaviour and Information Technology*, vol. 39, no. 3, pp. 288–318, 2020.

[33] H. Mouzannar, Y. Rizk, and M. Awad, “Damage Identification in Social Media Posts Using Multimodal Deep Learning,” *Proceedings of the International ISCRAM Conference*, vol. 2018-May, no. May, pp. 529–543, 2018.

[34] M. Agarwal, M. Leekha, R. Sawhney, and R. R. Shah, “Crisis-DIAS: Towards Multimodal Damage Analysis - Deployment, Challenges and Assessment,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 346–353, 2020.

[35] S. S. M. Win and T. N. Aung, “Target oriented tweets monitoring system during natural disasters,” *International Journal of Networked and Distributed Computing*, vol. 5, no. 3, pp. 133–142, 2017.

[36] A. Sen, K. Rudra, and S. Ghosh, “Extracting situational awareness from microblogs during disaster events,” *2015 7th International Conference on Communication Systems and Networks, COMSNETS 2015 - Proceedings*, pp. 0–5, 2015.

[37] C. Corley and R. Mihalcea, “Measuring the semantic similarity of texts,” no. June, pp. 13–18, 2005.

[38] Z. Zhang, S. Zhao, X. Li, D. Conga, and D. Suna, “An information service model for remote sensing emergency services,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 42, no. 2W7, pp. 187–191, 2017.

[39] T. Hofmann, “Probabilistic Latente Semantic Analysis,” 2004.

[40] Z. Lin, H. Jin, B. Robinson, and X. Lin, “Towards an accurate social media disaster event detection system based on deep learning and semantic representation,” *14th Australasian Data Mining Conference, AusDM 2016*, no. December, pp. 15–23, 2016.

[41] L. Ma and Y. Zhang, “Using Word2Vec to process big text data,” *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, pp. 2895–2897, 2015.

[42] M. Hendre, P. Mukherjee, R. Preet, and M. Godse, “Efficacy of Deep Neural Embeddings based Semantic Similarity in Automatic Essay Evaluation,” 2020.

[43] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, “Practical extraction of disaster-relevant information from social media,” *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*, pp. 1021–1024, 2013.

[44] M. Imran, P. Mitra, and J. Srivastava, “Enabling Rapid Classification of Social Media Communications During Crises,” *International Journal of Information Systems for Crisis Response and Management*, vol. 8, no. 3, pp. 1–17, 2017.

[45] K. Zahra, M. Imran, and F. O. Ostermann, “Automatic identification of eyewitness messages on twitter during disasters,” *Information Processing and Management*, vol. 57, no. 1, p. 102107, 2020.

[46] C. Caragea, A. Squicciarini, S. Stehle, K. Neppalli, and A. Tapia, “Mapping moods: Geo-mapped sentiment analysis during hurricane sandy,” *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management*, no. May, pp. 642–651, 2014.

[47] Y. Wang and J. E. Taylor, “Coupling sentiment and human mobility in natural disasters: a Twitter-based study of the 2014 South Napa Earthquake,” *Natural Hazards*, vol. 92, no. 2, pp. 907–925, 2018.

[48] S. Sinnappan, C. Farrell, and E. Stewart, “Priceless tweets! A study on twitter messages posted during crisis: Black Saturday,” *ACIS 2010 Proceedings - 21st Australasian Conference on Information Systems*, 2010.

[49] D. Graf, W. Retschitzegger, W. Schwinger, B. Pröll, and E. Kapsammer, “Exploiting Twitter for Informativeness Classification in Disaster Situations,” in *Transactions on Large-Scale Data- and Knowledge-Centered Systems XLV: Special Issue on Data Management and Knowledge Extraction in Digital Ecosystems*, A. Hameurlain, A. M. Tjoa, R. Chbeir, Y. Manolopoulos, H. Ishikawa, S. Ilarri, and A. Papadopoulos, Eds. Berlin: Springer Berlin Heidelberg, 2020, pp. 27–55.

[50] C. Dua, D. and Graff. (2019) UCI Machine Learning Repository. Accessed Feb. 10, 2021. [Online]. Available: https://archive.ics.uci.edu/ml/datasets.php

[51] Python RegEx. Accessed Sep. 01, 2020. [Online]. Available: https://www.w3schools.com/python/python_regex.asp

[52] EntityRecognizer. Accessed Feb. 15, 2021. [Online]. Available: https://spacy.io/api/entityrecognizer

[53] E. Loper and S. Bird, “Nltk,” pp. 63–70, 2002.

[54] spaCy. Accessed Feb. 15, 2021. [Online]. Available: https://spacy.io/

[55] OntoNotes 5 corpus. Accessed Feb. 12, 2021. [Online]. Available: https://www.ldc.upenn.edu/language-resources/data

[56] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[57] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in Neural Information Processing Systems*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[58] D. Cer, Y. Yang, S. y. Kong, N. Hua, N. Limtiaco, R. St. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar, Y. H. Sung, B. Strope, and R. Kurzweil, “Universal sentence encoder for English,” *EMNLP 2018 - Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Proceedings*, pp. 169–174, 2018.

[59] J. Wu, “Cluster Analysis and K-means Clustering: An Introduction 1.1 The Emergence of Data Mining,” *Advances in K-means Clustering*, pp. 1–16, 2012.

[60] I. Mokris and L. Skovajsova, “Comparison of Document Clustering Techniques,” no. June, pp. 75–78, 2008.

[61] T. M. Kodinariya and P. R. Makwana, “Review on determining number of Cluster in K-Means Clustering,” *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 2321–7782, 2013.

[62] (2021) Twint Project. Accessed Feb. 11, 2021. [Online]. Available: https://github.com/twintproject/twint

[63] H. Dalianis and H. Dalianis, “Evaluation Metrics and Evaluation,” *Clinical Text Mining*, no. 1967, pp. 45–53, 2018.

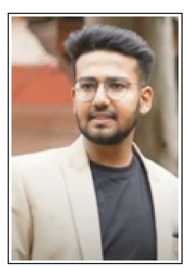
[64] T. R. Nichols, P. M. Wisner, G. Cripe, and L. Gulabchand, “Putting the kappa statistic to use,” *Quality Assurance Journal*, vol. 13, no. 3-4, pp. 57–61, 2010.

[65] Y. Jin, “Development of Word Cloud Generator Software Based on Python,” *Procedia Engineering*, vol. 174, pp. 788–792, 2017.

[66] A. C. Mazari and A. Djeffal, “A Combined Method of Naïve-Bayes and Pooling Strategy for Building Test Collection for Arabic/English Information Retrieval,” *International Journal of Computing and Digital Systems*, vol. 10, no. 1, pp. 629–638, 2021.

[67] W. Zhang, X. Zhao, L. Zhao, D. Yin, G. H. Yang, and A. Beutel, “Deep Reinforcement Learning for Information Retrieval: Fundamentals and Advances,” *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2468–2471, 2020.

[68] I. C. Doo, H. D. Shin, and M. H. Park, “Automated product review collection and opinion analysis methods for efficient business analysis,” *International Journal of Computing and Digital Systems*, vol. 10, no. 1, pp. 37–45, 2021.



Ashwani Varshney recently completed Masters of Science, Computer Science from Department of Computer Science, University of Delhi, India. His research interests include Machine Learning, Deep Learning, Natural Language Processing, Data and Business Analytics.

Yatin Kapoor recently completed Master of Science, Computer Science from Department of Computer Science, University of Delhi, India. His research interests include Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing, and Data Science.

Vaishali Chawla completed her postgraduate degree in Computer Science from Department of Computer Science, University of Delhi, India. Her research interests include Machine Learning, Deep Learning, and Digital Image Processing.

Dr. Vibha Gaur is a professor in the Department of Computer Science, Acharya Narendra Dev College, University of Delhi. She has published more than 40 papers in international journals and conferences. Her research interests include Requirement Engineering, Software Quality, Fuzzy Logic. She is a corresponding author of the article.