MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Project Report

on

IDENTIFYING HASHTAGS IN DISASTER RELATED TWITTER DATA

Submitted By:
KARTIK SHARMA
0901CS191049

DHRUTI BHADORIA
0901CS191033

Faculty Mentor:
Mr. Mir Shahnawaz Ahmad
Asst. Prof. CSE, MITS

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957
MAY-JUNE 2022

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Project Report

on

IDENTIFYING HASHTAGS IN DISASTER RELATED TWITTER DATA

A project report submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

KARTIK SHARMA

0901CS191049

DHRUTI BHADORIA

0901CS191033

Faculty Mentor:
Mr. Mir Shahnawaz Ahmad
Asst. Prof. CSE, MITS

Submitted to:

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

CERTIFICATE

This is certified that **KARTIK SHARMA**(0901CS191049) has submitted the project report titled **IDENTIFYING HASHTAGS IN DISASTER RELATED TWITTER DATA**under the mentorship of**Prof Mir S. Ahmed**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior.

09/05/22

Mr. Mir Shahnawaz Ahmad
Faculty Mentor
Computer Science and Engineering

09/05/22

Dr. Manish Dixit
Professor and Head,
Computer Science and Engineering

Dr. Manish Dixit
Professor & HOD
Department of CSE
M.I.T.S. Gwalior

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

CERTIFICATE

This is certified that **DHRUTI BHADORIA**(0901CS191033) has submitted the project report titled **IDENTIFYING HASHTAGS IN DISASTER RELATED TWITTER DATA** under the mentorship of**Prof. Mir Shahnawaz Ahmad**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior.

09/05/22

Mr. Mir Shahnawaz Ahmad
Faculty Mentor
Computer Science and Engineering

09/05/22

Dr. Manish Dixit
Professor and Head,
Computer Science and Engineering

Dr. Manish Dixit
Professor & HOD
Department of CSE
M.I.T.S. Gwalior

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of**Prof. Mir ShahnawazAhmad ,CSE, MITS**

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

KARTIK SHARMA
0901CS191049
III Year,
Computer Science and Engineering

DHRUTI BHADORIA
0901CS191033
III Year,
Computer Science and Engineering

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Department of Computer Science and Engineering, for allowing** me to explore this project. I humbly thank **Dr. Manish Dixit**, Professor and Head, Department of Computer Science and Engineering, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of**Prof. Mir S Ahmed ,CSE, MITS**, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

K. Sh

KARTIK SHARMA
0901CS191049
III Year,
Computer Science and Engineering

Dhruti

DHRUTI BHADORIA
0901CS191033
III Year,
Computer Science and Engineering

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

ABSTRACT

During disasters, affected individuals often turn to social media platforms, such as Twitter and Facebook, to find the latest updates from government and response organizations, to request help or to post information that can be used to enhance situational. Nonetheless, the value of the information posted on social media platforms during disasters is highly unexploited, in part due to the lack of tools that can help filter relevant, informative, and actionable messages. Twitter is a major tool for communication during emergencies and disasters. This study aimed to investigate Twitter use during natural hazards and pandemics. The included studies reported the role of Twitter in disasters triggered by natural hazards. Tweet hashtags have the potential to improve the search for information during disaster events. However, there is a large number of disaster related tweets that do not have any user-provided hashtags. Moreover, only a small number of tweets that contain actionable hashtags are useful for disaster response. To facilitate progress on automatic identification (or extraction) of disaster hashtags for Twitter data, we use a dataset of disaster-related tweets annotated with hashtags useful for filtering actionable information. Using this dataset, we further investigate Machine Learning Random-Forest Classifier model with TF-IDF. With accuracy of about 76%.

KEYWORDS: Twitter, hashtag extraction, random forest model, disaster.

सार

आपदाओं के दौरान, प्रभावित व्यक्ति अक्सर सरकार और प्रतिक्रिया संगठनों से नवीनतम अपडेट प्राप्त करने के लिए ट्विटर और फेसबुक जैसे सोशल मीडिया प्लेटफॉर्म की ओर रुख करते हैं, मदद का अनुरोध करने के लिए या ऐसी जानकारी पोस्ट करने के लिए जिसका उपयोग स्थिति को बढ़ाने के लिए किया जा सकता है। फिर भी, आपदाओं के दौरान सोशल मीडिया प्लेटफॉर्म पर पोस्ट की गई जानकारी का मूल्य अत्यधिक अप्रयुक्त है, आंशिक रूप से ऐसे उपकरणों की कमी के कारण जो प्रासंगिक, सूचनात्मक और कार्रवाई योग्य संदेशों को फ़िल्टर करने में मदद कर सकते हैं। आपात स्थिति और आपदाओं के दौरान संचार के लिए ट्विटर एक प्रमुख उपकरण है। इस अध्ययन का उद्देश्य प्राकृतिक खतरों और महामारियों के दौरान ट्विटर के उपयोग की जांच करना था। शामिल अध्ययनों ने प्राकृतिक खतरों से उत्पन्न आपदाओं में ट्विटर की भूमिका की सूचना दी। ट्वीट हैशटैग में आपदा की घटनाओं के दौरान सूचना की खोज को बेहतर बनाने की क्षमता होती है। हालाँकि, बड़ी संख्या में आपदा से संबंधित ट्वीट्स हैं जिनमें कोई उपयोगकर्ता द्वारा प्रदत्त हैशटैग नहीं है। इसके अलावा, केवल कुछ ट्वीट्स जिनमें कार्रवाई योग्य हैशटैग होते हैं, आपदा प्रतिक्रिया के लिए उपयोगी होते हैं। ट्विटर डेटा के लिए आपदा हैशटैग की स्वचालित पहचान (या निष्कर्षण) पर प्रगति की सुविधा के लिए, हम कार्रवाई योग्य जानकारी को फ़िल्टर करने के लिए उपयोगी हैशटैग के साथ एनोटेट किए गए आपदा-संबंधित ट्वीट्स के डेटासेट का उपयोग करते हैं। इस डेटासेट का उपयोग करते हुए, हम TF-IDF के साथ मशीन लर्निंग रैंडम-फ़ॉरेस्ट क्लासिफायर मॉडल की और जाँच करते हैं। लगभग 76% की सटीकता के साथ।

TABLE OF CONTENTS

Chapter 1: INTRODUCTION

During disasters, affected individuals often turn to social media platforms, such as Twitter and Facebook, to find the latest updates from government and response organizations, to request help or to post information that can be used to enhance situational awareness.

Even in this COVID-19 pandemic, people used social media to get help. Nonetheless, the value of the information posted on social media platforms during disasters is highly unexploited, in part due to the lack of tools that can help filter relevant, informative, and actionable messages.

1.1 Real World Example

Examples of tweets that illustrate the diverse ways in which people use hashtags to highlight information during disasters are shown in Fig 1. Specifically, the user provided hashtags, when available, are shown in blue color in the table. As can be seen, the first two tweets do not have any user-provided hashtags. The third tweet has a general disaster-name hashtag, #HurricaneIrma. While this hashtag is useful in recognizing that the tweet was posted during Hurricane Irma, it is not useful in identifying situational awareness (e.g., damage, power loss, blocked street) or the type of disaster response requests. The fourth tweet, which explicitly reports damage, has disaster-name, location, and weather as hashtags, but no specific hashtag about damage. Finally, the fifth tweet is at the other extreme, in that it has a large number of hashtags (specifically, 11), some of them representing lexical variations of the same base word.

<table>
<tr><th>No.</th><th>Tweet text</th></tr>
<tr><td rowspan="2">1.</td><td>we need help in Houston. our apartments are surrounded with water like an island we need rescue 10373 N Sam Houston Pkwy E</td></tr>
<tr><td>need help | Houston | need rescue</td></tr>
<tr><td rowspan="2">2.</td><td>@houstonpolice please help I'm stranded with my kids I need help fast my address is 8618 Banting st. houston tx 77078.</td></tr>
<tr><td>stranded | need help | houston</td></tr>
<tr><td rowspan="2">3.</td><td>Big tree fell on power lines and blocking Brown Ave near Washington St in Orlando's Thornton Park neighborhood. #HurricaneIrma</td></tr>
<tr><td>power lines | blocking | Orlando | #HurricaneIrma</td></tr>
<tr><td rowspan="2">4.</td><td>Very extensive damage sustained throughout #Wilmington, #ncwx... from #hurricane #Florence. Lots of trees split or uprooted, siding ripped from homes, powerlines down, flooding of downtown streets, etc.</td></tr>
<tr><td>extensive damage | #Wilmington | #hurricane | #Florence | powerlines down</td></tr>
<tr><td rowspan="2">5.</td><td>I am evacuated from my house but I'm safe. #fire #CampFire #WoolseyFire #wildfire #safe #Evacuation #evacuations #EVACUATED #scary #ThousandOaks #Camarillo</td></tr>
<tr><td>evacuated | #WoolseyFire | #ThousandOaks | #Camarillo</td></tr>
</table>

Fig 1: Examples of tweets posted during disasters. The original user-provided hashtags, when available, are shown in blue color for each tweet. Relevant and topically more informative hashtags manually identified to have the potential to retrieve actionable disaster tweets are highlighted in a light red box.

1.2 Applications

Tweets related to disaster can be used in many aspects. They can be used to inform the concerned authorities for help such as fire, electricity cut/loose electric wires, for evacuation purposes. These tweets can also be used in damage estimations. For example - according to Villegas, Martinez, and Krause (2018), more than 5, 200 rescue requests made on social media were missed by the first responders, while about 46% of the critical damage information posted on social media during Harvey Hurricane was missed by FEMA in their original damage estimates (that is almost half of the total costs of $125 billion estimated for this hurricane).

As an official explained: “It’s very labor intensive to watch [social media] and because of the thousand different ways people can hashtag something or keyword something, trying to sort out what’s relevant and what’s not and what’s actionable is very, very difficult”

People use a variety of ways to hashtag tweets, that they may not understand or know how to hashtag tweets. User-provided hashtags tend to be either too general or too specific. An analysis of a large corpus of tweets that was used in this work revealed that most of the hashtags in a disaster-related tweet corpus simply represent disaster names and locations, and that approximately half of the tweets do not have any hashtags at all.

To address these limitations of tweet retrieval based on user-provided hashtags or keywords, we envision a system that learns to identify relevant tweets and extract relevant hashtags directly from the content of the disaster tweets

Chapter 2: LITERATURE REVIEW AND SURVEY

An analysis of a large corpus of tweets that was used in this work revealed that most of the hashtags in a disaster-related tweet corpus simply represent disaster names and locations, and that approximately half of the tweets do not have any hashtags at all.

Thus, filtering based on user-provided hashtags is not helpful for disaster response or people on the ground to quickly find relevant information.

Similarly, filtering tweets during disasters based on keyword match is not expected to work well since keywords can be ambiguous and can lead to noisy results, e.g., a search for the keyword "Harvey" will retrieve tweets about the hurricane, but also about people whose name is "Harvey."

To address these limitations of tweet retrieval based on user-provided hashtags or keywords, we envision a system that learns to identify relevant and topically informative hashtags and extract them directly from the content of the disaster tweets, capturing three main elements in a tweet:

1) disaster name;

2) location; and

3) situational awareness information.

After analyzing a lot of tweets and conducting online surveys we reached to the following conclusion that we have to-

1. Filter twitter data using hashtags and key-phrases to provide more specific situational information in times of any disaster or natural calamity.
2. This will in turn help the people in need to reach to more people and organizations who could provide help and assistance. Also, situational information from time to time will give the status of the people and their livelihood affected by the disaster.

3. So eventually, data could be filtered easily and people could reach out to help more quickly. For example, if we take this COVID crisis, we are getting time to time updates of the number cases, recoveries, medication and the vaccine under test for the virus.

The dataset that we used is https://www.kaggle.com/c/nlp-getting-started/data?select=train.csv. The ML model uses 7613 tweets from this dataset in all. We use the train test split function to split the data into training set and test set. The test size was set to 18% of the entire dataset and the rest was used as training set to train the model. The dataset, code, and other resources from this work are made available on GitHub.

Application programming interfaces (APIs), which support these connections. MuleSoft, in May 2018, was acquired by Salesforce, a software as a service (SaaS) provider. Salesforce now uses MuleSoft technology as part of its Salesforce Integration Cloud.

Chapter 3: PRELIMINARY DESIGN

3.1 Gaps Identified

An analysis of a large corpus of tweets that was used in this work revealed that most of the hashtags in a disaster-related tweet corpus simply represent disaster names and locations, and that approximately half of the tweets do not have any hashtags at all.

Thus, filtering based on user-provided hashtags is not helpful for disaster response or people on the ground to quickly find relevant information.

Similarly, filtering tweets during disasters based on keyword match is not expected to work well since keywords can be ambiguous and can lead to noisy results, e.g., a search for the keyword “Harvey” will retrieve tweets about the hurricane, but also about people whose name is “Harvey.”

Therefore, after carrying out some research, we were able to find out that most of the proposed work suggested to use hashtags as keywords to obtain actionable information. But we saw the limitations of user provided hashtags. Retrieval of tweets using only hashtags will lead to a very low accuracy of only about 50-55%.

Hence in this project we focus on keywords extraction from both the hashtags provided by user but also from the tweets as in whole. And we will run the both the models separately and then by testing on multiple values we will estimate minimum values of the probability for each model separately and combined may give rise to more accurate results.

3.2 Proposed Work

We have seen how people use a variety of ways to hashtag a tweet and how this will pose a problem while retrieval of relevant tweets. Also, keyword matching cannot be used since keywords are ambiguous and can lead to noisy results.

To address these limitations of tweet retrieval based on user provided hashtags or keywords, we envision a system that learns to identify relevant tweets and extract relevant hashtags directly from the content of the disaster tweets

3.3 Methodology

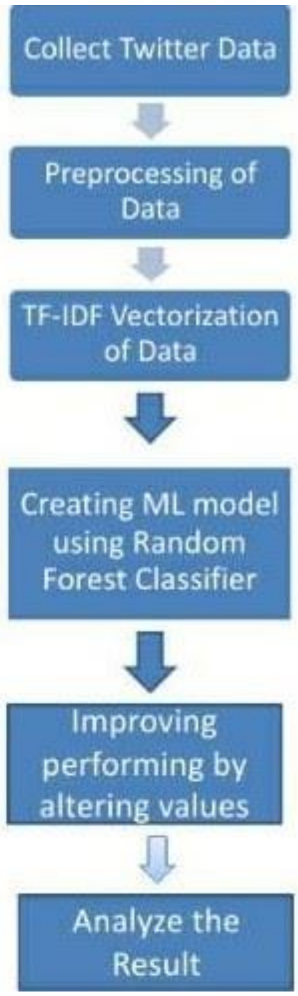
We first collected tweets related to multiple disasters and disaster types (e.g., hurricane, flooding) and then craft a lexicon, which were used together with the hashtags from the tweets, whenever available, to annotate a large dataset. Using this dataset, we will further investigate a powerful machine learning model, for key-phrase extraction from general tweets, and evaluate its performance capability for hashtag extraction from disaster-related tweets. This model, a Random-Forest Classifier model will be trained using assembled dataset will do the required work.

We will implement our model two times: -

1) For extraction of relevant keywords from the hashtags provided.

2) For extraction of keywords from the entire content of the tweet itself.

We get the results for these two operations separately and we calculate the weighted average of both the results. This will increase the accuracy of the model since it not relies on the user provided hashtags, but also checks the entire tweet for topically informative keywords.



3.4 Algorithm

Step 1. Import twitter data using the twitter API.

Step 2. Preprocessing of the text, example removing stepwords using NLP.

Step 3. Count Vectorization using TF-IDF.

Step 4. Splitting the dataset into Training set and Test set using the train_test_split function in sklearn.model_selection.

Step 5. Creating two models, one for hashtags and another one for the entire tweet.

Step 6. Testing of both the models

Step 7. Obtaining results using classification_report form sklearn.metrics.

Step 8. Calculating the weighted average of the two results obtained**.**

3.5 Code Snippets

Tokenize the text, clean it, lemmatization etc. Converting the textual words to their numeric representation.

```
for sentence in range(0, len(features)):
    # Remove all the special characters
    hashtag=re.findall(r"#(\w+)", str(features[sentence]))
    processed_feature = re.sub(r"#(\w+)", ' ', str(features[sentence]))
    processed_feature = re.sub(r'\W', ' ', processed_feature)
    # remove all single characters
    processed_feature= re.sub(r'\s+[a-zA-Z]\s+', ' ', processed_feature)
    # Remove single characters from the start
    processed_feature = re.sub(r'\^[a-zA-Z]\s+', ' ', processed_feature)
    # Substituting multiple spaces with single space
    processed_feature = re.sub(r'\s+', ' ', processed_feature, flags=re.I)
    # Removing prefixed 'b'
    processed_feature = re.sub(r'^b\s+', '', processed_feature)
    # Converting to Lowercase
    processed_feature = processed_feature.lower()
    # print(processed_feature)
    hash=''
    for word in range(0,len(hashtag)):
        hash=hash+hashtag[word]+' '
    processed_features.append(processed_feature)
    hashtags.append(hash)
```

Doing TF-IDF vectorization.

```
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer (max_features=2500,min_df=5,stop_words=stopwords.words('english'))
vectorizer2 = TfidfVectorizer (max_features=1500,min_df=2,stop_words=stopwords.words('english'))
processed_features = vectorizer.fit_transform(processed_features).toarray()
hashtags = vectorizer2.fit_transform(hashtags).toarray()
```

Training and Testing of Random Forest Classifier Model.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(processed_features, labels, test_size=0.18, random_state=4)
X2_train, X2_test, y2_train, y2_test = train_test_split(hashtags, labels, test_size=0.18, random_state=4)
from sklearn.ensemble import RandomForestClassifier
model =RandomForestClassifier(random_state=0)
model2 =RandomForestClassifier(random_state=0)
model.fit(X_train, y_train)
model2.fit(X2_train,y2_train)
prediction = model.predict(X_test)
prediction2 = model2.predict(X2_test)
```

Extracting Tweets from twitter based on location and date.

```
lat=input(" Latitiude ")
lon=input(" Longitude ")
rad=input(" Radius ")
query=lat+','+lon+','+rad+"mi"
cou=int(input("Number of tweets"))
ter=input(" Related term ")
dat=input(" Date should be formatted as YYYY-MM-DD. ")
tweets=api.GetSearch(geocode=query,lang='en',count=cou,term=ter,since=dat)
```

Tweet classification based on model's prediction.

```
X_predict = vectorize.transform(processed_tweets).toarray()
Y_predict = model.predict_proba(X_predict)
X2_predict = vectorize2.transform(hashtags).toarray()
Y2_predict = model2.predict_proba(X2_predict)
predictions =[]
# print(Y_predict)
# print(Y2_predict)
for pr in range(0,len(Y2_predict)):
    # print(Y2_predict[pr][1])
    if Y2_predict[pr][1]==0.4161057735809483:
        if Y_predict[pr][1]>=0.5:
            predictions.append(1)
        else:
            predictions.append(0)
    else:
        if (Y2_predict[pr][1]+Y_predict[pr][1]>=1 or Y2_predict[pr][1]>=0.75 or Y_predict[pr][1]>=0.65):
            predictions.append(1)
        else :
            predictions.append(0)
```

Chapter 4: RESULTS AND CONCLUSIONS

To address these limitations of tweet retrieval based on user-provided hashtags or keywords, we envisioned a system that learns to identify relevant tweets and extract relevant hashtags directly from the content of the disaster tweets.

We first collected tweets related to multiple disasters and disaster types (e.g., hurricane, flooding) and then craft a lexicon, which were used together with the hashtags from the tweets, whenever available, to annotate a large dataset.

Using this dataset, we further investigated a powerful machine learning model, for key-phrase extraction from general tweets, and evaluate its performance capability for hashtag extraction from disaster-related tweets. This model, a Random-Forest Classifier model was trained using assembled dataset did the required work

4.1 Random Forest Classifier

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes become our model’s prediction.

Our model gathers a corpus of tweets from the Twitter API. The dataset that we used is https://www.kaggle.com/c/nlp-getting-started/data?select=train.csv.

The ML model uses 7613 tweets from thus dataset in all. We use the train test split function to split the data into training set and test set. The test size was set to 18% of the entire dataset and the rest was used as training set to train the model.

4.2 Results

We set the test size to be 18% of the entire dataset, and the remaining was used as training set to train the model.

And after training the model with the training set and passing the test data through it, we were able to get the classification report as follows:

```
PS C:\Users\HP\Minor Project> & C:/Users/HP/AppData/Local/Programs/Python/Python37-32/python.exe "c:/
              precision    recall  f1-score   support

           0       0.76      0.86      0.80       793
           1       0.76      0.62      0.68       578

    accuracy                           0.76      1371
   macro avg       0.76      0.74      0.74      1371
weighted avg       0.76      0.76      0.75      1371

              precision    recall  f1-score   support

           0       0.61      0.97      0.75       793
           1       0.79      0.14      0.24       578

    accuracy                           0.62      1371
   macro avg       0.70      0.56      0.49      1371
weighted avg       0.68      0.62      0.53      1371

PS C:\Users\HP\Minor Project> 
```

Our model was successfully able to predict the disaster related tweets with an accuracy of 76%.

Chapter 4: CONCLUSION

4.1 Conclusion & Future Scope

In this project, we introduce a model that identifies whether the tweet is disaster related or not based on ML model. It also contains disaster related twitter hashtags dataset.

We believe that our dataset will foster research in this domain, will enable the design of deep learning models, and will help response organizations to make better use of social media data contributed by individuals affected by disasters, and will contribute to better decision-making during disasters when resources are limited.

Also, incorporating other ML models like CNN and neural network into the models can improve performance further.

4.2 Future Scope

1. Sending info to the concerned authority for example in case of fire there can be info sent down to fire station, in case of terrorist attacks police authorities can be informed, in case of evac emergencies, the evacuation teams can be sent the info along with the current location from which the tweet was posted.
2. To encourage visitors to vote on which hashtags are most useful in hopes of bringing multiple conversations together. (example: #pediatric, #pediatrics, #pediatrician, #pediatricians)
3. Show hashtags that are related to one another.
4. Determine which healthcare hashtags overall are most popular at any given time. And to give suggestions.
5. We aim to update our database dynamically.

Reference-

Jishnu Ray Chowdhury, Cornelia Caragea, and Doina Caragea - Research paper On Identifying Hashtags in Disaster Twitter Data. https://arxiv.org/pdf/2001.01323v1.pdf

Changxiu Cheng, Nikita Murzintcev - Disaster Hashtags in Social Media https://www.mdpi.com/2220-9964/6/7/204/htm

Hien To, Sumeet Agrawal, Seon Ho Kim, Cyrus Shahabi - On Identifying Disaster-Related Tweets: Matching-based or Learning-based? - https://infolab.usc.edu/DocsDemos/to_bigmm2017.pdf