Project report

on

An Integrated Tool For Real-Time Responses In Disaster Management

Submitted to D Y Patil International University, Akurdi, Pune
in partial fulfilment of full-time degree.

B.Tech. Computer Science and Engineering

(Track-Data Science)

Submitted By:

Rohit Saktel 20190802096

Shivani Ambhore 20190802020

Under the Guidance of

Mr. Rajkumar Chaudhari

School of Computer Science and Engineering

D Y Patil International University, Akurdi,Pune, INDIA, 411044

[Session 2019-23]

D Y PATIL
INTERNATIONAL
UNIVERSITY
AKURDI PUNE

CERTIFICATE

This is to certify that the project entitled **An Integrated Tool For Real-Time Responses In Disaster Management** submitted by:

Rohit Saktel	20190802096
Shivani Ambhore	20190802020

is the partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering is an authentic work carried out by them under my supervision and guidance.

Chaudhari

Mr. Rajkumar Chaudhari
(Mentor)

Dr. Tamal Mondal
(Supervisor)

D Y PATIL INTERNATIONAL UNIVERSITY
AKURDI, PUNE-44.

Director
Dr. Bahubali Shiragapur
School of Computer Science Engineering & Application
D Y Patil International University, Akurdi
Pune, 411044, Maharashtra, India

DECLARATION

We, hereby declare that the following report which is being presented in the Major Project entitled as **An Integrated Tool For Real-Time Responses In Disaster Management** is an authentic documentation of our own original work to the best of our knowledge. The following project and its report in part or whole, has not been presented or submitted by us for any purpose in any other institute or organization. Any contribution made to the research by others, with whom we have worked at D Y Patil International University, Akurdi, Pune or elsewhere, is explicitly acknowledged in the report.

Rohit Saktel	(20190802096)	
Shivani Ambhore	(20190802020)	

ACKNOWLEDGEMENT

With due respect, we express our deep sense of gratitude to our respected guide and coordinator **Dr.Tamal Mondal**, for his valuable help and guidance. We are thankful for the encouragement that he has given us in completing this project successfully.

It is imperative for us to mention the fact that the report of major project could not have been accomplished without the periodic suggestions and advice of our project supervisor **Dr. Tamal Mondal** and project mentor **Mr. Rajkumar Chaudhari**.

We are also grateful to our respected Director, **Dr. Bahubali Shiragapur** and Hon'ble Vice Chancellor, DYPIU, Akurdi, **Prof. Prabhat Ranjan** for permitting us to utilize all the necessary facilities of the college.

We are also thankful to all the other faculty, staff members and laboratory attendants of our department for their kind cooperation and help. Last but certainly not the least; we would like to express our deep appreciation towards our family members and batch mates for providing support and encouragement.

Rohit saktel (20190802096)
Shivani Ambhore (20190802020)

Abstract

Social media, has a big impact on how individuals go about their daily lives. Platforms like social media and microblogging generate many billions of bytes of data in the form of text, audio, location, and other categories. These particulars are never employed in any form. We can use this information to provide rescue services to those who have been trapped in dreadful disasters. These can aid in the successful communication of all parties engaged in an incident. The use of human intelligence and information communications is integrated into the decision-making process in crisis and emergency situations. For effective situational awareness and reaction time in these circumstances, data gathering, manipulation, and analysis tasks require a mix of cognitive processes and information and communications technology. In contrast to news sources, especially microblogging services, which have the potential to be utilized as an extra tool for emergency services, have the capacity to deliver spontaneous information during emergency/disaster circumstances. In order to support these capabilities, we describe our real-time emergency response system, which was developed with the goal of managing the potential data torrents that could become available during a crisis and that, in the absence of technological intervention, could easily overwhelm human cognitive capacity. Our system is specifically designed to address the research challenges related to the real-time collection of relevant data, in particular, to live text data, making this data quickly accessible to a team of humans, and giving them the tools to manipulate, tag, and filter the most important information of relevance to the situation.

Keywords: Twitter, Disaster, catastrophes, disaster management, Machine learning, tweets, Situational and Non - Situational tweets.

TABLE OF CONTENTS

List of Figures

1. INTRODUCTION

In the present time it has become a very difficult responsibility for the rescue teams and the relevant government departments to provide rescue services for those who have been affected by disasters. The people who have been stuck in the catastrophes don't know what to do or where to go for finding rescue services. Many people want many different types of rescue materials but cannot get those because of the lack of rescue service availability due to the lack of information about the people.

Natural disasters can be characterized as a synthesis of vulnerabilities and natural hazards that put at risk weaker communities unable to overcome the adversities they bring. Unpredictability, the availability of few resources in affected areas, and dynamic environmental changes are the fundamental features of natural disasters. Natural disasters can't be foreseen with sufficient accuracy, according to the idea of unpredictable effects on people and property. When it is challenging to allocate sufficient resources in advance due to uncertainty, the issue of limited resources arises.

The main information which is lagging behind is about the exact location where they are stuck and what the particular items the disaster-affected people need and what is the quantity of the item needed. The rescue team also could not find the people where they are exactly in the casthrophe-affected area. For all these problems we can use an effective tool where the total natural disaster can be managed. In this tool application, real-time monitoring can provide quick and accurate information to rescue services, allowing for rapid responses in the event of disasters.

The rescue team can access real-time information about the disaster-affected area, such as the location of persons in need of rescue and the availability of relief supplies, using another module of the program. Additionally, social media sites like Twitter can be utilized to gather up-to-the-minute details about the catastrophe and plan rescue operations. Others modules like tweepy can be used to generate Twitter keywords containing disaster-related phrases in order to learn more about the needs of the impacted people in a certain location.

Real-time extraction of situational tweets during disasters poses several unique challenges. The high velocity and immense volume of tweets make manual extraction an impractical and time-consuming approach. Instead, leveraging automated techniques becomes essential for efficiently processing and categorizing tweets in real time. Natural Language Processing (NLP) techniques, combined with advanced machine learning algorithms, offer powerful tools to analyze tweet content, extract relevant features, and extract tweets based on their situational relevance.

Recent advancements in NLP and machine learning have paved the way for sophisticated text classification/extraction techniques. These methods exploit linguistic, contextual, and semantic features within tweets to infer their situational or non-situational nature. Features such as keywords, named entities, sentiment analysis, temporal information, user credibility, and geolocation can be harnessed to develop robust classification models. Machine learning algorithms, including traditional approaches such as Naive Bayes, Support Vector Classifier (SVC), as well as state-of-the-art deep learning architectures like Convolutional Neural Networks (CNNs) and Transformer models, can be trained on large labeled datasets to learn the underlying patterns and make accurate predictions in real time.

The benefits of real-time extraction of situational tweets during disasters are manifold. Firstly, it allows emergency response teams to efficiently filter and prioritize incoming information, ensuring that critical situational tweets are promptly identified and acted upon. This enables quicker decision-making processes, facilitates the deployment of resources to areas most in need, and enhances overall response coordination. Secondly, the ability to differentiate non-situational tweets helps in curbing the spread of misinformation, rumors, and irrelevant discussions, thus reducing confusion and enabling the dissemination of accurate and reliable information to the public. Thirdly, the extraction system can assist in tailoring targeted updates and alerts for specific user groups based on their situational relevance, improving the relevance and effectiveness of communication during disasters.

An integrated platform for real-time reactions in disaster management can promote better communication and collaboration amongst various stakeholders or parties and aid in managing disaster relief activities more successfully. It can lessen the effects of natural disasters and save lives by utilizing the power of technology. Such systems have the potential to support emergency responders, government agencies, and the public in making informed decisions, mitigating risks, and ensuring effective resource allocation during critical situations. As research continues in this field, the extraction of situational tweets will undoubtedly play a pivotal role in revolutionizing disaster management practices, fostering timely and effective responses, and ultimately, saving lives and minimizing the impact of disasters on affected communities.

1.1. Background

Real-time information sharing during emergencies and disasters has been greatly aided by social media platforms like Twitter. Disaster-related tweets can be analyzed to gain important information about the impacted areas, general sentiment, and developing situations. Extraction of situational tweets from the original tweets is an important step in the analysis of disaster-related tweets.

Extracting situational tweets offers in-the-moment information into how the situation is changing during a disaster. The intensity, impact, and urgent requirements of the affected population can be better understood by examining the content, location, and emotion of situational tweets. Authorities can better address issues, communicate with the public, and support impacted areas by adjusting their communication techniques.

1.2. Motivations

The motivation to define our project can be properly explain by considering a case study of Nepal.

" On April 25th, 2015, simply before noon, Nepal skilled an earthquake of value 7.Eight on the moment magnitude scale. The earthquake ripped via Kathmandu Valley, and a series of aftershocks leveled complete villages.

Straight away after the earthquake, volunteers from around the sector were instrumental in guiding emergency operations, the usage of satellite imagery to identify infrastructure destruction throughout the area.

But, human beings at the ground in Nepal have been additionally producing extraordinary amounts of information which might be of use to rescue operations, albeit less directly: on twitter. Among April twenty fifth and might twenty eighth, 33,610 tweets have been tweeted with the aid of human beings in Nepal. These tweets had been full of beneficial records, however 33,610 tweets is genuinely too many for a rescue operation to brush through."

Thus, this is the motivation of our project:

Can we take a large set of these tweets and from them extract a piece of useful information and a short summary that could have been useful to rescue teams for the further operations?

2. LITERATURE REVIEW

2.1. Literature review

Platforms like Twitter that allow for real-time information exchange and communication are essential during disasters. Crisis management, response coordination, and resource allocation can all be made more effective by categorising situational and non-situational tweets about disasters. The topic that we pick to explore hasn't been covered in any papers as far as we know. In addition, the topic we selected hasn't been the subject of a thorough literature review to our knowledge.

Following are few papers, which we have used for literature review for our project:

"Towards Automated Situational Awareness Reporting for Disaster Management—A Case Study" by Klaus Schwarz, Daniel Arias Aranda, and Michael Hartmann (2023).

In this study, the author suggested a method for automated situational awareness reporting that utilizes microblogging data (Twitter) through event recognition and summary. The first step is to extract relevant events from the microblogging data using an event detection algorithm (tfidf, top words, clustering). Second, various summarization libraries (Spacy, NLTK, and GenSim) are used to summarize the discovered events. Finally, the results of the event detection and summarization steps are combined to produce a report of the significant events.

"Social media data extraction for disaster management aid using deep learning techniques" by Trisha Vishwanath, Rudresh Deepak Shirwaikar (2023).

This paper presents the results of the study that suggests a method for gathering information from Twitter, the main social media source, in order to discover and disseminate the vital aid needed in emergency and catastrophe circumstances. Earthquakes, floods, cyclones, and wildfires are the four different forms of natural catastrophes that are classified in the study using a Convolution Neural Network (CNN) design and Transfer Learning (TL) architectures.

The study discovered that the suggested method was able to obtain an accuracy of 98.14 %, with precision, recall, and F1-score values of 0.82, 0.86, and 0.84 for cyclones; 0.96, 0.89, and 0.92 for earthquakes; 0.74, 0.95, and 0.83 for floods; and 0.97, 0.96, and 0.96 for wildfires.

"A multi-level analytic framework for disaster situational awareness using Twitter data" by Wei Zhai (2022).

In this paper, author describes a methodology for using Twitter data to improve disaster situational awareness. The methodology involves the following steps:

1. Data collection: Relevant Twitter information on the disaster incident is collected. This can be done by using a Twitter API to gather tweets that are geotagged to the crisis location, or by searching for tweets that mention keywords related to the disaster.
2. Data preprocessing: The collected data is preprocessed. This includes removing stop words, stemming words, and tokenizing the data.
3. Feature extraction: Features are extracted from the data. This can be done using a variety of methods, such as using bag-of-words, n-grams, or topic modeling.
4. Model training: A model is trained to classify the tweets into the three levels of situational awareness: perception-level, humanitarian-level, and action-level. This can be done using a variety of machine learning algorithms, such as Naive Bayes, decision trees, or support vector machines.
5. Model evaluation: The trained model is evaluated using a holdout set of data that was not used to train the model. The model's performance is evaluated using metrics such as accuracy, precision, and recall.

The methodology described in the research paper has been shown to be effective in improving disaster situational awareness. The framework can be used to track the public's perception of a disaster event, identify the specific needs of local communities, and quickly respond to the situation.

"Machine Learning in Disaster Management: Recent Developments in Methods and Applications" by Vasileios Linardos, Maria Drakaki, Panagiotis Tzionas, Yannis L. Karnavas(2022)

The paper discusses how machine learning (ML) and deep learning (DL) can be used to control disasters. The studies are grouped into categories by the authors based on the subphases of disaster management: disaster and hazard prediction, risk and vulnerability assessment, disaster detection, early warning systems, disaster monitoring, damage assessment, and post-disaster response. The authors concentrate on studies that have been published since 2017, and they do so by focusing on those that fall under each category.

"Situational awareness extraction: a comprehensive review of social media data classification during natural hazards" by Jirapa Vongkusolkit and Qunying Huang(2021)

In this paper, the use of social media data to support situational awareness (SA) during natural disasters.The authors examined research that analysed social media using various techniques and divided them into five groups: content, sentiment, user, temporal, and spatiotemporal.

The authors conclude that social media data can be a valuable source of information for SA during natural disasters. They recommend that future studies focus on developing more robust methods for extracting information from social media, and on using this information to improve disaster response and recovery.

"A Novel Method for Identifying the Damage Assessment Tweets during Disaster" by Sreenivasulu Madichetty · Sridevi M (2021).

In this paper, an author suggested a novel technique for locating damage assessment tweets during a crisis. The suggested approach effectively makes use of the syntactic, low-level lexical, and word properties with the highest frequency that are unique to damage assessment. Support Vector Regression (SVR) methods and simple linear regression are also used to weigh these features.

"Detection of situational information from Twitter during disaster using deep learning models" by SRIDEVI MUTHUKUMARASAMY (2020).

In this paper, the study uses a variety of deep learning architectures to detect situational tweets, including convolutional neural networks (CNN), long short-term memories (LSTM), bidirectional long short-term memories (BLSTM), and bidirectional long short-term memories with attention (BLSTM attention). For the purpose of recognising the situational information during a crisis, deep learning models are also applied to tweets in Hindi in addition to tweets in English.

"Summarizing Situational Tweets in Crisis Scenarios: An Extractive-Abstractive Approach" by Koustav Rudra, Pawan Goyal, Niloy Ganguly and Muhammad Imran (2019).

In this paper, a framework for automatically producing summaries of situational tweets in emergency situations is proposed. The framework initially divides tweets into various situational groups, such as information regarding missing, hurt, or deceased persons, infrastructure damage, and housing needs. A summary is subsequently produced for each class using an extractive-abstract methodology. While the abstractive summarising component creates a natural language summary of the extracted tweets, the extractive summarization component collects significant tweets from the class. It is demonstrated that the framework outperforms baseline techniques in terms of quality, coverage of events, locations, and utility using a dataset of tweets from the Hurricane Sandy disaster.

"Text mining on real time Twitter data for disaster responses" by Myneni Madhu Bala, K. Navya, P. Shruthilaya (2017).

In this paper, the authors attempted to predict coming harm in the days to come during the flood scenario by mining real-time data from Twitter TM. Twitter users share crucial information such as alerts, the location of an occurrence, and first-hand accounts. Such data is gathered, geo-located, preprocessed, and filtered. Based on the data gathered, geo-coded data is given precedence over text data. The data is then examined using regression analysis to see how the event unfolded. To verify the findings, they used information from the 2015 Chennai floods. Additionally, they projected the disaster curve to forecast the sites that would be vulnerable to damage in the next days.

2.2. Drawbacks of existing system

The existing systems for the extraction of situational information from large set of data in real-time have some drawbacks that need to be addressed for more accurate and efficient extraction. Some of the drawbacks include:

1. The availability of labeled training data specifically for real-time extraction of situational information is often limited. This can result in suboptimal performance of the classification models, as they may not be able to capture the full range of contextual nuances and emerging trends in real-time data.
2. Twitter is known for its fast-paced and evolving language, which includes slang, abbreviations, and trending hashtags. Existing systems may struggle to keep up with these language variations, leading to lower accuracy in extracting situational tweets, especially when new terms or expressions emerge during real-time events.
3. Tweets often lack explicit context, requiring deep understanding of the background or reference points to accurately classify them. Existing systems may struggle to capture the context accurately, leading to the misclassification of situational and non-situational tweets.
4. The distribution of situational and non-situational tweets is often imbalanced, with a significantly larger number of non-situational tweets. This class imbalance can bias the classification models towards the majority class, leading to lower performance in accurately identifying situational tweets.
5. Scalability and Processing Speed: Real-time tweet classification requires fast and scalable systems to handle the large volume of incoming tweets. Existing systems may struggle to keep up with the processing demands, resulting in delays and increased computational resources.

Addressing these drawbacks requires ongoing research and development to improve the accuracy, efficiency, and adaptability of existing systems for the real-time classification of situational and non-situational tweets.

2.3. Gaps Identified

Addressing the following gaps in the report will provide a comprehensive understanding of the limitations, challenges, and potential areas of improvement in the extraction of situational tweets in real time. It will also guide future research and development in this field.

There is a lack of extensive research specifically focused on the real-time extraction of situational tweets. Most existing studies and approaches have primarily focused on offline analysis or batch processing of tweet data. More research is needed to develop and evaluate real-time classification techniques that can handle the dynamic nature of Twitter data.

The availability of benchmark datasets specifically designed for real-time tweet classification is limited. These datasets should encompass a wide range of real-time events, varying levels of noise and ambiguity, and cover diverse domains and languages. The absence of such datasets hinders the evaluation and comparison of different real-time classification methods.

Existing evaluation metrics used for tweet classification may not fully capture the real-time nature of the task. Traditional metrics, such as accuracy or precision, may not adequately reflect the timeliness and responsiveness required for real-time analysis. Developing new evaluation metrics that consider the real-time aspect, such as response time or freshness of classification, can provide a more comprehensive assessment of the system's performance

Real-time tweet extraction requires fast and scalable machine learning models like Decision Tree, Random Forest, LSTM (Long-Short Term Memory), and Support Vector Classifier to handle the high volume and velocity of incoming tweets. Existing systems may face challenges in terms of computational resources, processing speed, and scalability. Developing efficient algorithms and system architectures that can process tweets in real-time, leverage parallel processing, and optimize resource utilization is essential to close this gap.

Overcoming these gaps through dedicated research and development efforts will contribute to more accurate, context-aware, and reliable models for the real-time extraction of situational tweets related to disasters. It will enhance the effectiveness of tweet analysis in disaster management, early warning systems, and crisis response efforts.

3. PROBLEM STATEMENT

When it comes to ensuring the welfare and security of those affected, natural disasters pose significant challenges. These natural disasters have wide-ranging negative impacts that have a significant impact on both persons and communities. The effects include the loss of livelihoods, deteriorating health conditions brought on by waterborne infections, challenges acquiring food and other necessities, and a critical need for blood donations. Catastrophe's destructive nature frequently causes the destruction of essential facilities, adding to the suffering of people affected.

A significant challenge in delivering efficient rescue services during natural disasters is the insufficient availability of accurate information regarding the whereabouts of individuals requiring assistance. Rescue teams and government agencies encounter obstacles in locating and accessing individuals who are stranded or trapped in areas affected by catastrophes. Consequently, the provision of rescue services is inadequate, leading to a tragic loss of lives due to the unavailability of timely aid.

Adding to the problem, individuals affected by catastrophes often face confusion and uncertainty regarding whom to reach out to or where to seek help. The lack of a dependable method to communicate their locations and requirements further obstructs the delivery of assistance. Consequently, those impacted are left feeling helpless and unsure of how to access the crucial support and resources they desperately need.

To get over these challenges, it is essential to create a real-time monitoring platform that can provide accurate and up-to-date information, aiding successful rescue operations. With the use of this tool, rescue teams would be able to compile crucial information on the precise locations and numbers of people in need of help. Rescue teams are better able to deploy their resources wisely and provide immediate assistance to people in need by using this up-to-the-minute information.

4. OBJECTIVES

The broad objective of an integrated tool for real-time response in disaster management is to improve the effectiveness of disaster response by providing decision-makers with timely and accurate information.

1. **The extraction of Situational Information from large volume of data :**

   The extraction of situational information in real time provides emergency responders with critical information on the severity and impact of a disaster event. By distinguishing between situational tweets that provide real-time information on the disaster and non-situational tweets that are unrelated to the event, responders can quickly identify areas that require urgent attention and allocate resources accordingly. This extraction can greatly enhance the effectiveness and efficiency of disaster response efforts.

2. **Training and Comparing the Models and Improving Accuracy :**

   It is essential to continuously train and improve these models for improving accuracy and precision. This is because the language used in tweets can be diverse and rapidly evolving, making it challenging to accurately and precisely classify them. By regularly training, comparing and improving models using new and relevant data, it is possible to enhance their accuracy and precision in real-time tweet extraction. This can lead to more effective and efficient disaster response efforts, ultimately improving outcomes for affected communities.

3. **Visualization of Plots:**

   Using plots we can visualize real-time responses so that it can provide emergency responders with a comprehensive and easy-to-understand view of disaster events. Plots can visualize real-time data and enable responders to quickly identify patterns, trends, and critical areas requiring attention. This can facilitate better decision-making, enhance communication between response teams, and ultimately lead to more effective disaster response efforts. Therefore, developing and implementing plots that can display real-time responses visually can greatly enhance the effectiveness and efficiency of disaster management.

5. METHODOLOGY

5.1. Proposed Methodology

Social media platforms have recently become important sources of information during times of disasters and emergencies. Twitter, in particular, has become a significant platform due to its real-time communication and information sharing capabilities. The substantial volume of data generated through tweets during disaster events presents a unique opportunity to leverage this information for the enhancement of disaster management efforts.

The proposed methodology provides a comprehensive outline of a detailed process for obtaining real-time Twitter tweets using the Twitter API and authentication. Subsequently, a raw data set will be constructed, which will serve as the fundamental basis for developing a predictive model capable of identifying pertinent situational tweets pertaining to disaster management.

1. **Twitter API and Authentication:** The first step in this methodology involves accessing the Twitter API for real-time data collection. The Twitter API is a powerful tool that can be used to gather a wealth of information about what is happening on the platform in real-time. By authenticating with the API and acquiring the necessary API keys and tokens, developers can establish a secure connection and begin collecting data. This data can then be used to track trends, identify emerging issues, and respond to disasters in a timely manner.
2. **Real-time Data Collection:** We can begin gathering real-time tweets via the Twitter API after the authentication process is finished. We can focus our data collection to only include pertinent tweets by specifying pertinent keywords, hashtags, or geographic regions connected to particular disasters. We were able to compile a thorough dataset including multiple crisis occurrences because of the API's access to historical and real-time streams of tweets. And the dataset accumulated via Twitter API includes textual content(text), date, and location as attributes(functions).
3. **Base Dataset Collection:** We gathered and combined the baseline dataset from multiple sources such as crisesNLP and Kaggle. The dataset collected is an unlabelled dataset that requires unsupervised learning to categorize the data. The data in the dataset is related to a different category of natural disasters such as earthquakes, Floods, hurricanes, landslides, etc. Date, Location, and Tweet text are taken into consideration as features for model building and extraction.
4. **Data Preprocessing:** After collecting a significant amount of raw data, the next step is to preprocess it to prepare it for model training. Data preprocessing includes a number

of crucial tasks, such as cleaning the text data by eliminating pointless symbols, special characters, emotional punctuations and URLs. The text data can also be normalized and regulated using methods like tokenization, stop-word elimination, and stemming. We may make sure that the data is in an appropriate format for subsequent analysis by carrying out these pretreatment processes.

5. **Feature Extraction:** Finding relevant characteristics in the preprocessed tweets is essential for creating an efficient predictive model. For feature extraction, a variety of methods can be used, including word embeddings like CountVectorizor, TF-IDF (Term Frequency-Inverse Document Frequency), and bag-of-words representation. These methods convert text data into numerical representations while preserving the tweets' semantic content and social context. The next modeling stage can better capture patterns and relationships within the data by removing significant elements from the textual material.

6. **Model Building:** The next stage after feature extraction is to create a predictive model that can categorize situational tweets about disaster management using n-gram features as a training sets. Logistic Regression, Random Forest, Support Vector Classifiers, and deep learning models like Recurrent Neural Networks (RNNs) are just a few examples of machine learning algorithms that can be used. The amount of the dataset and the difficulty of the challenge influence the model selection. Using a labeled dataset in which each tweet is classified as being relevant or irrelevant to the disaster situation, the model is trained on the preprocessed data. We used the TextBlob algorithm to identify and analyze the sentiments towards the dataset used, as a general cause.

7. **Model Evaluation:**

   After training the model, to verify its accuracy in correctly predicting situational tweets, it is essential to assess its performance. The model's effectiveness can be assessed using evaluation measures like accuracy, precision, recall, and F1-score. In order to evaluate the model's resilience and generalizability, additional methods like cross-validation may be used. To get the model to perform as well as possible, it is essential to refine it iteratively by experimenting with various validation methods and hyperparameters.

8. **Further Analysis and Validation:** Implementing further analysis techniques likecross-validationn and Hyperparameter tuning is responsible for the validation of the results and it ensures that the robustness and generalizability of the results seem good or sufficient.

9. **Real-time tweet Prediction:** Once the model is trained and evaluated, It can be used to predict how relevant real-time tweets will be during emergencies. The trained model can be used to identify new tweets as either relevant or irrelevant to the current catastrophe situation. New tweets are collected using the Twitter API and can be fed through the model. With the use of these real-time prediction capabilities, situational tweets can be

identified in a timely manner, offering important information to support disaster management efforts.

5.2. Block diagram

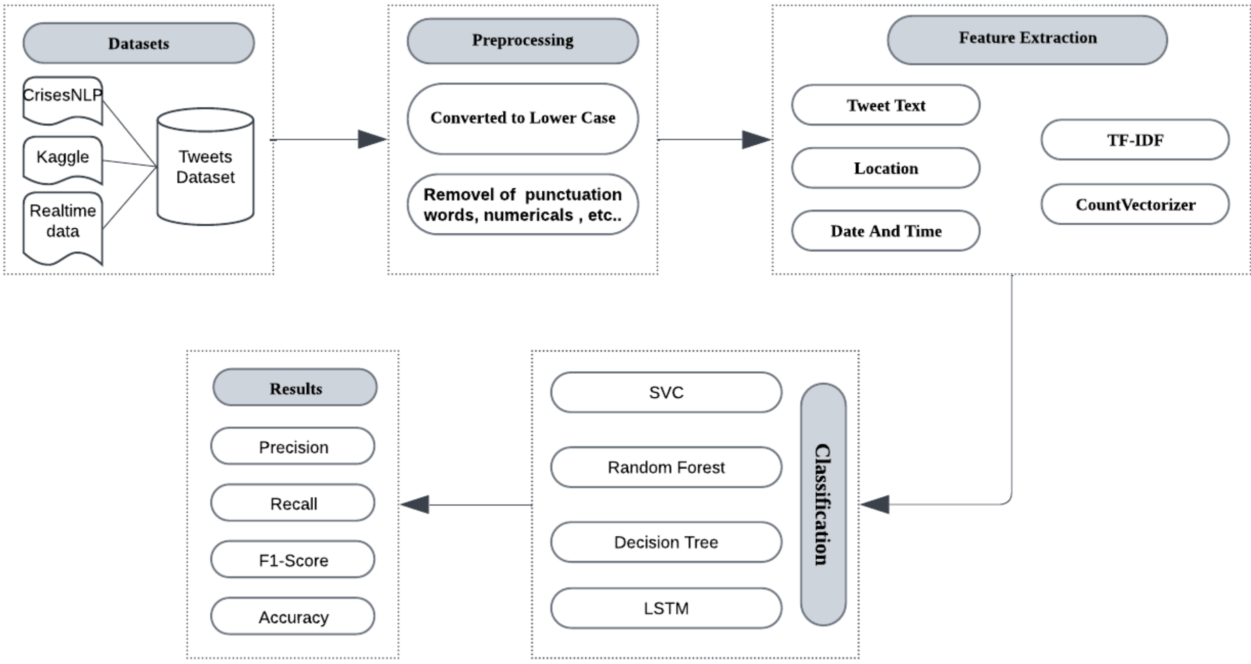

Figure 5.1: Model Block diagram

Here, the illustration of Model Block Diagram is as follows:

- we begin by gathering the raw dataset from the sources.
- The preparation processes that follow entail data cleaning, stopword removal, and removal of unnecessary words (punctuations, emotions, special characters, etc).
- After that, using feature algorithms, the cleaned data is transformed into formats that can be read by models.
- We take actions to develop models when they are transformed into features. When the model is finished being built, it is evaluated for better prediction.
- Once everything functions properly, real-time data is now loaded into the training model, which subsequently predicts the situational data.

5.3. Tools Used

Python: Due to its ease of use and adaptability, Python is a well-liked programming language for machine learning. Python is a fantastic choice for creating models because it has many machine-learning and deep-learning libraries built in.

Tweepy: Tweepy is a popular Python library that simplifies the process of accessing the Twitter API and retrieving data from the platform. It provides an easy-to-use interface for authenticating with the Twitter API, making API requests, and handling the received data.

NLTK: NLTK is a toolkit built for working with NLP in Python. It provides various text processing libraries with a lot of test datasets. A variety of tasks can be performed using NLTK such as tokenizing, parse tree visualization, etc.

Scikit-learn: is a well-known Python machine learning library that offers a number of tools for data preprocessing, feature selection, and model evaluation. It is frequently combined with deep learning libraries to create end-to-end pipelines for machine learning tasks.

Tensor Flow: TensorFlow is an open-source machine learning framework. For developing and deploying machine learning models, it offers a complete collection of tools, frameworks, and APIs. Deep learning, neural networks, computer vision, natural language processing, and other activities all make extensive use of TensorFlow.

6. MODEL AND DESIGN

6.1. Model Implemented

Support Vector Classifier (SVC):

SVC, or Support Vector Classifier, is a supervised machine learning algorithm typically used for classification tasks [5]. SVC works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes.

The Support vector machine classifier works by finding the hyperplane that maximizes the margin between the two classes. The Support vector machine algorithm is also known as a max-margin classifier. Support vector machine is a powerful tool for machine learning and has been widely used in many tasks such as hand-written digit recognition, facial expression recognition, and text classification. Support vector machine has many advantages over other machine learning algorithms, such as robustness to noise and the ability to handle large datasets.

SVM can be used to solve non-linear problems by using kernel functions [10]. For example, the popular RBF (radial basis function) kernel can be used to map data points into a higher dimensional space so that they become linearly separable. Once the data points are mapped, SVM will find the optimal hyperplane in this new space that can separate the data points into two classes.

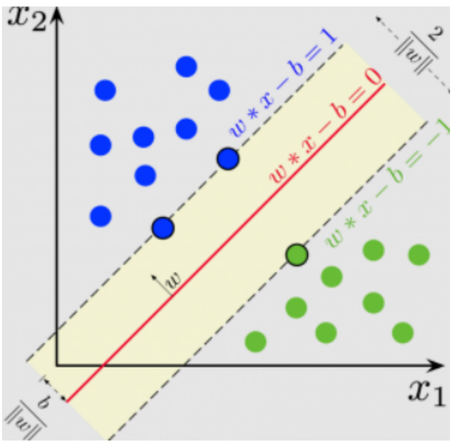

Figure 6.1: Support Vector Classifier (SVC)[5]

Random Forest Classifier:

It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

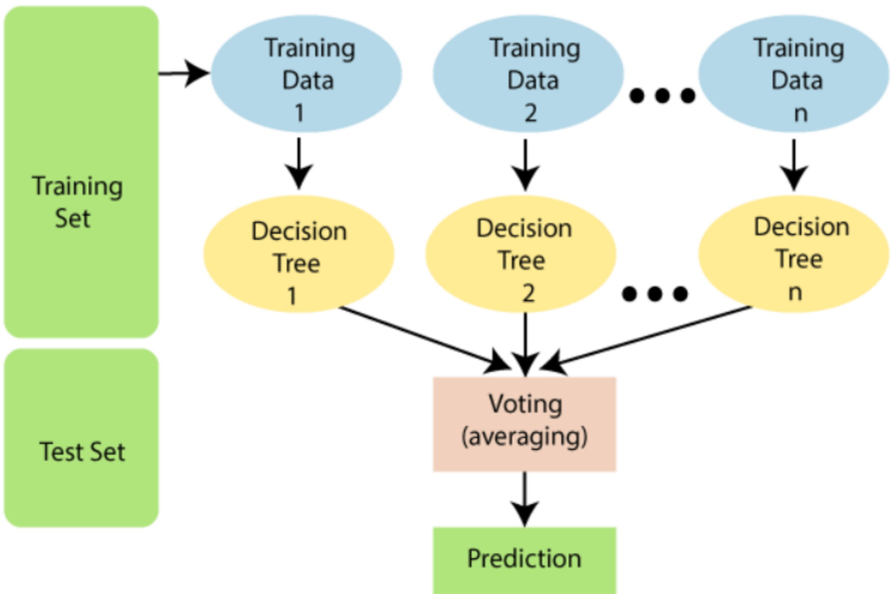

Figure 6.2: Random Forest Classifier

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

Logistic regression:

Logistic regression is a widely used machine learning technique for classifying data into two categories. It involves establishing a relationship between a group of independent variables (features) and a binary outcome (target variable) by calculating the probabilities of the outcome belonging to a specific class.

The logistic regression algorithm utilizes a sigmoid or logistic function to transform the linear combination of the independent variables into a value ranging from 0 to 1. This transformed value represents the probability of the outcome belonging to a particular class. The logistic function's distinctive "S" shaped curve enables it to map the predicted values to probabilities accurately.

The formula for the logistic function is:

p = 1 / (1 + e*(-z))

Where:

the predicted probability (p) of the outcome belonging to the positive class, e is the base of the natural logarithm (approximately 2.718), and z is the linear combination of the independent variables and their corresponding coefficients.

During the learning process of logistic regression, the algorithm determines the optimal values for the coefficients by minimizing a cost function, such as maximum likelihood estimation. The goal is to find the best-fit "S" shaped curve that maximizes the likelihood of the observed data.

When making predictions with logistic regression, a threshold is typically chosen. If the predicted probability is above the threshold, the outcome is classified as the positive class; otherwise, it is classified as the negative class.

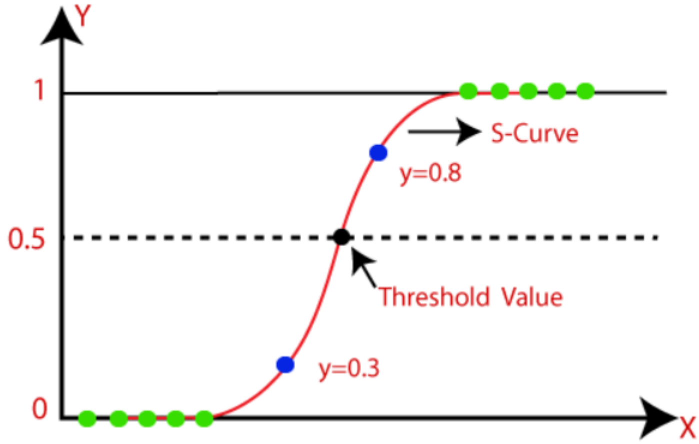

Figure 6.3: Logistic Regression [17]

Long - Short Term Memory (LSTM):

RNNs include long short-term memories (LSTM) depicts the LSTM's structural layout. LSTM has a number of activation functions and operations inside four interactive layers, as opposed to normal RNN, This just has one activation function and keeps the initial input data. The "Gate" part of an LSTM structure is where information is selectively passed, and the "Cell State" path is where information flows constantly along the length of the LSTM cell. Each gate is constructed from computations and functions such as sigmoid and tanh [10]. After receiving the input value for the current time-step and the output value for the previous time-step, the input gate, which is situated in the second layer, performs the sigmoid and tanh processes to decide what data should be stored. Depending on the cell state of the current time-step, the fourth layer determines what information should be output, while the third layer updates the cell state output value of the previous time-step to the cell state of the present timestep.

To address the long-term reliance issue with RNNs, LSTM was developed. As previously said , it seeks to provide output values that are more accurate by appropriately mixing relevant historical and present data across numerous layers [6]. Given that LSTM is a subtype of RNN, it is advantageous for time series data models and, when the input data length is very large, it often achieves superior prediction performance than RNN.

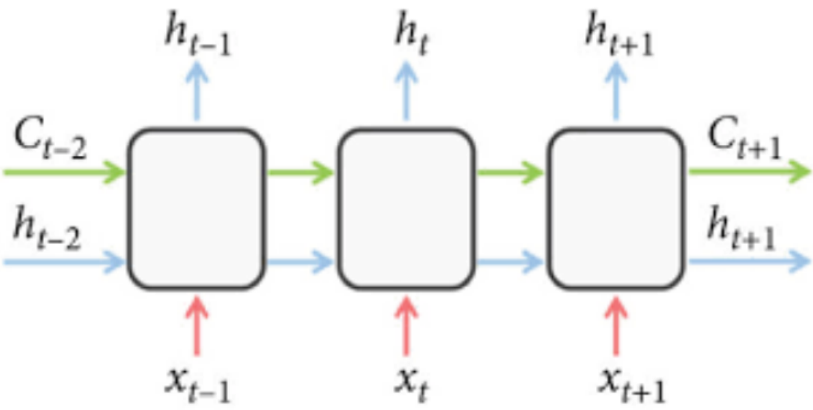

Figure 6.4: Long - Short Term Memory (LSTM) [6]

N-Gram

N-grams are sequences of consecutive items, which can consist of various elements like words, numbers, symbols, or punctuation, extracted from a textual document. These n-gram models have practical applications in text analysis tasks where the order of words holds significance, such as sentiment analysis, text classification, and text generation. Utilizing n-gram modeling is one approach employed to convert unstructured text into a structured format. Alternatively, word embedding techniques like word2vec can also be utilized for this purpose. The classification of n-grams is determined by the value of 'n.' For example, a unigram is used when 'n' equals 1, while a bigram is used when 'n' is set to 2, and so on.

N-grams are commonly employed in feature extraction to represent text data, functioning as features that capture patterns, dependencies, and semantic details within the text. Varying the value of 'n' allows for the creation of distinct context levels, enabling the capture of diverse information types. These n-grams are valuable features applicable to various machine learning models, including logistic regression, as they effectively capture local context and word dependencies. Their utilization proves beneficial in tasks such as sentiment analysis, text classification, language modeling, and more.

6.2. Flow Chart

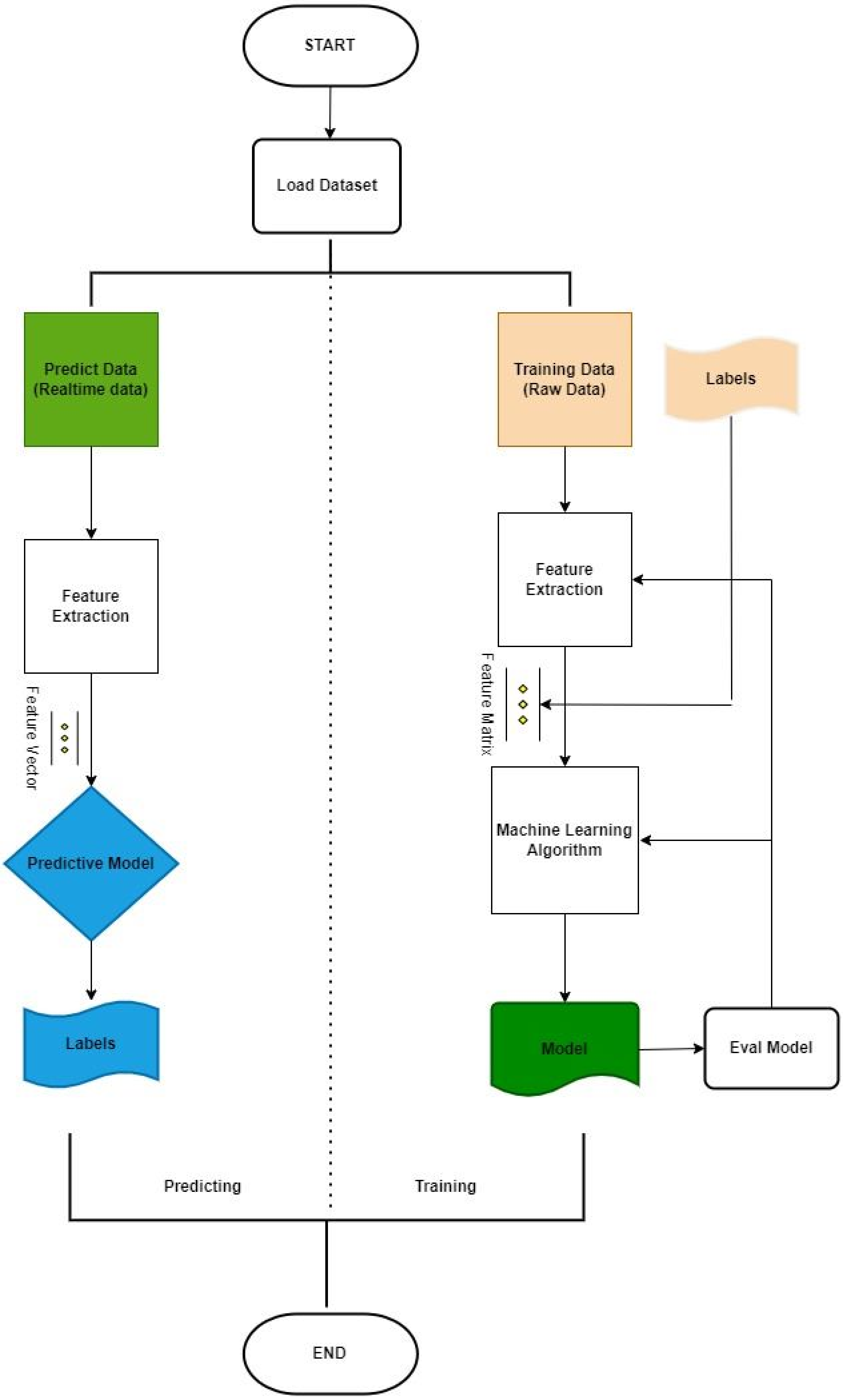

Figure 6.5: Flow Chart

In this diagram, we have the main data flow components:

- First, we fetched the real-time tweets using Twitter API and also collected a raw dataset from crisis NLP and crisis MMD.
- Then we preprocessed our data, in text preprocessing we tokenized the tweet based texts and also removed stop words.
- Then in the data cleaning step, we removed punctuations, and special characters and also lowered case all the words in the tweet text.
- In feature extraction we did an n-gram analysis and also did the count vectorization with respect to the n-gram analysis.
- At last we applied the models(Decision Tree, Random forest, LSTM, SVC) to get the accuracy for the classification of situational and non-situational tweets.

7. RESULTS ANALYSIS AND DISCUSSIONS

With the help of tweets fetched in real-time, the proposed project sought to predict a machine learning-based model for the extraction of situational tweets along with other features. In particular, we developed a machine learning (Support Vector Classifier, Random Forest, Logistic Regression and LSTM) model for the extraction of situational tweets in real time.

7.1. Results Analysis

- **Single Feature (n-gram = 3)**

```
from sklearn.feature_extraction.text import TfidfVectorizer

# Create an instance of TfidfVectorizer with trigram (ngram_range=(3, 3))
tfidf_vectorizer = TfidfVectorizer(n_gram_range=(3,3))

# Perform feature extraction on the training data
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)

# Transform the testing data using the learned vocabulary from the training data
X_test_tfidf = tfidf_vectorizer.transform(X_test)
```

Figure 7.1: Single Feature (n-gram = 3)

In the code snippet above, we create an instance of TfidfVectorizer with ngram_range = (3, 3), which specifies that we want to extract trigrams. This means that the vectorizer will consider sequences of three consecutive words as features. You can adjust the ngram_range to capture different lengths of n-grams, such as bigrams (ngram_range=(2, 2)) or a combination of unigrams, bigrams, and trigrams (ngram_range=(1, 3)).

The fit_transform method is used to perform feature extraction on the training data (X_train), and the transform method is used to transform the testing data (X_test) using the vocabulary learned from the training data. The resulting X_train_tfidf and X_test_tfidf will contain the TF-IDF weighted representations of the trigram features.

Classification Report

The classification report offers a comprehensive evaluation of model performance in a binary classification scenario, providing insights into each class. It encompasses crucial metrics like precision, recall, F1-score, and support. Analyze the classification report findings for each model, emphasizing any discrepancies in performance among different classes. Assess the precision and recall values to gain an understanding of the model's accuracy in classifying positive and negative instances.

1. **SVC Model**

```
Accuracy of the model: 99.4%
--------------------------------------------------
Classifictaion Report
              precision    recall  f1-score   support

  Earthquake       1.00      0.99      1.00      1828
       Flood       0.99      0.94      0.96        97
       Other       0.99      1.00      0.99      1060

    accuracy                           0.99      2985
   macro avg       0.99      0.98      0.98      2985
weighted avg       0.99      0.99      0.99      2985
```

Figure 7.2: Classification Report of SVC

The SVC model achieved an accuracy of 99.4% in the classification task, demonstrating strong performance. It exhibited high precision and recall for all classes, with perfect precision for the "Earthquake" class and high precision for the "Flood" and "Other" classes. The model accurately predicted instances across the board, indicating its effectiveness in accurate classification.

2. **Random Forest**

```
Accuracy of the model: 99.4%
--------------------------------------------------
Classifictaion Report
              precision    recall  f1-score   support

  Earthquake       1.00      0.99      1.00      1828
       Flood       0.99      0.94      0.96        97
       Other       0.99      1.00      0.99      1060

    accuracy                           0.99      2985
   macro avg       0.99      0.98      0.98      2985
weighted avg       0.99      0.99      0.99      2985
```

Figure 7.3: Classification Report of SVC

The Random Forest model achieved an accuracy of 95.0% in the classification task, demonstrating strong performance. It exhibited high precision, recall, and F1-scores for all classes, indicating accurate classification across the board.

3. **Logistic Regression**

Accuracy of the model: 98.7%

Classifictaion Report

	precision	recall	f1-score	support
Earthquake	1.00	0.99	0.99	1828
Flood	1.00	0.86	0.92	97
Other	0.97	1.00	0.98	1060
accuracy			0.99	2985
macro avg	0.99	0.95	0.97	2985
weighted avg	0.99	0.99	0.99	2985

Figure 7.4: Classification Report of Logistic Regression

The logistic regression model achieved an accuracy of 98.7% with high precision, recall, and F1-scores for all classes. The ”Earthquake” class had a precision, recall, and F1-score of 1.00, 0.99, and 0.99 respectively. The ”Flood” class achieved perfect precision, a recall of 0.86, and an F1-score of 0.92. The ”Other” class had a precision of 0.97, recall of 1.00, and an F1-score of 0.98. Overall, the logistic regression model exhibited exceptional performance in accurately classifying instances.

Confusion Matrix

A confusion matrix represents the prediction summary in matrix form. It shows how many predictions are correct and incorrect per class. It helps in understanding the classes that are being confused by the model as other classes.

$$Accuracy = TP + TN/TP + FP + TN + FN \tag{7.1}$$

- True Positive (TP) data point
- True Negative (TN) data point
- False Positive (FP) data point
- False Negative (FN) data point

1. **SVC Model**

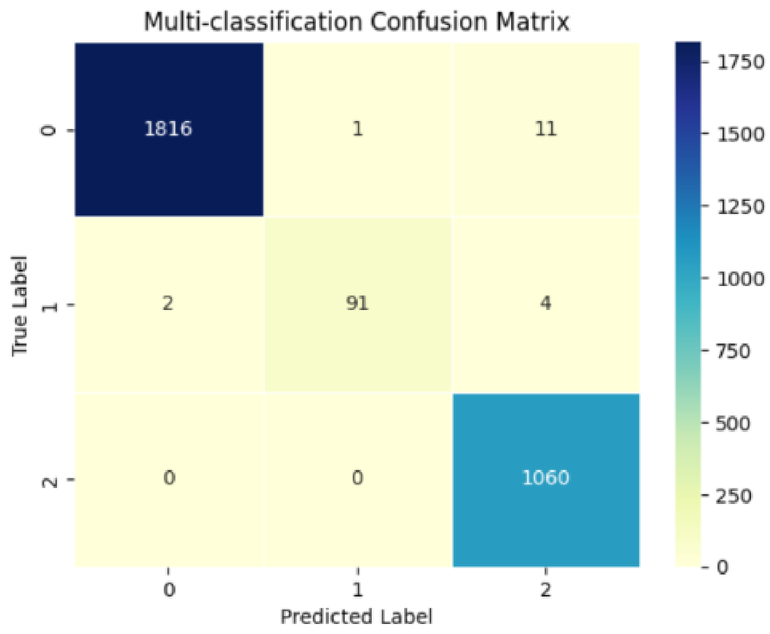

Figure 7.5: Confusion Matrix of SVC

2. **Random Forest**

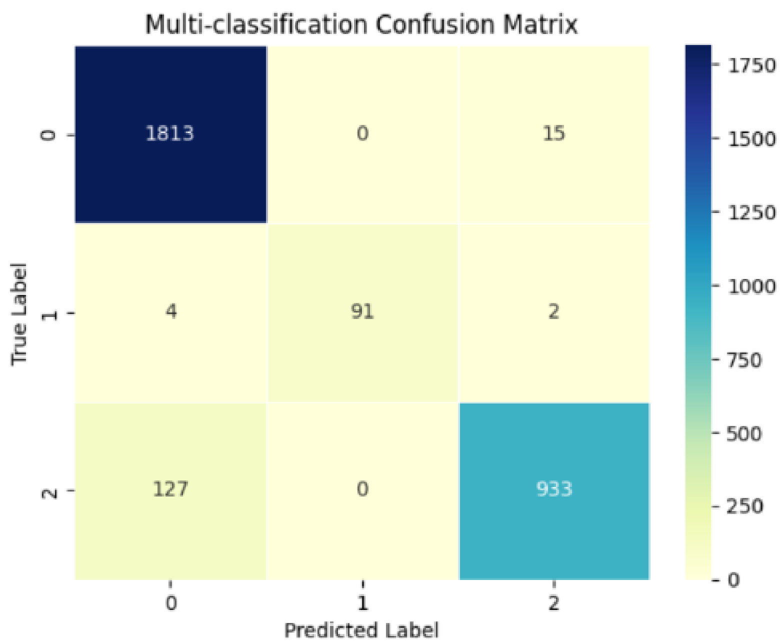

Figure 7.6: Confusion Matrix of Random Forest

3. **Logistic Regression**

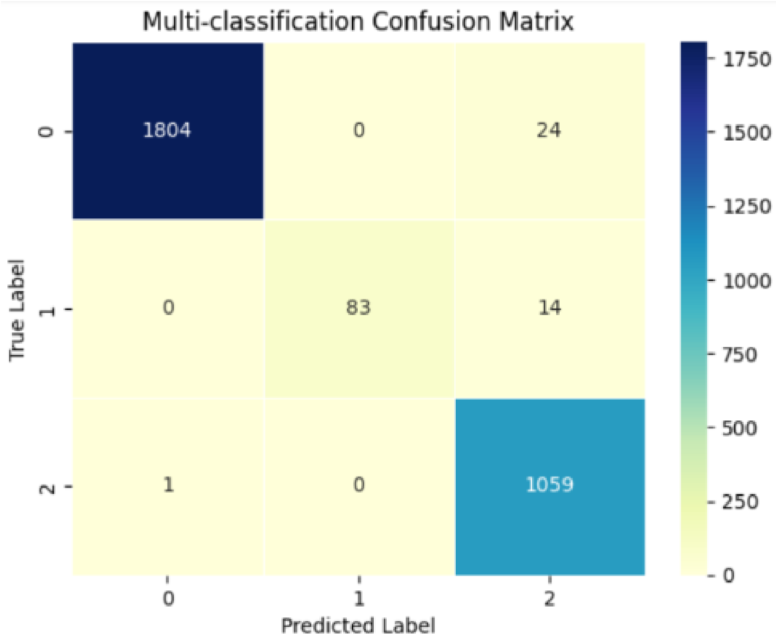

Figure 7.7: Confusion Matrix of Logistic Regression

Model Comparision Table: (Single Feature)

Below is a comparison table summarizing the performance of different models:

Model	Accuracy	Precision	Recall	F1-score
SVC	99.4%	99.14%	97.72%	98.40%
Random Forest	95.0%	97.16%	93.67%	95.26%
Logistic Regression	98.7%	98.83%	94.72%	96.58%

Figure 7.8: Model Comparision Table: (Single Feature)

This table provides an overview of the accuracy, precision, recall, and F1-scores achieved by each model. It allows for easy comparison and evaluation of their performance in the classification task. Overall, SVC model shows outstanding performance. Therefore we consider SVC model as a primary model for prediction.

- **Multiple Features (n-gram = 3)**

```
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(n_gram_range = (3,3))
feature_1 = vectorizer.fit_transform(data['date'])
feature_2 = vectorizer.fit_transform(data['user_location'])
feature_3 = vectorizer.fit_transform(data['clean_tweet'])

# Concatenate the features
features = pd.concat([pd.DataFrame(feature_1.toarray()), pd.DataFrame(feature_2.toarray()),pd.DataFrame(feature_3.toarray())], axis=1)
```

Figure 7.9: Multiple Features (n-gram = 3)

In the code snippet above, an instance of TfidfVectorizer is created with ngram_range=(3, 3), indicating that trigrams (sequences of three consecutive words) will be used as features.The fit_transform method is called on the vectorizer object for three different columns of the data DataFrame: 'date', 'user_location', and 'clean_tweet'. This method transforms the text data in each column into a feature matrix using the TF-IDF (Term Frequency-Inverse Document Frequency) representation.

The resulting feature matrices (feature_1, feature_2, and feature_3) are converted to dense arrays using the toarray() method, and then converted into DataFrame objects. Finally, these DataFrames are concatenated along the columns axis using pd.concat() to create a single DataFrame called features. This DataFrame contains the extracted features from the three columns of the data DataFrame.

Classification Report

The classification report offers a comprehensive evaluation of model performance in a binary classification scenario, providing insights into each class. It encompasses crucial metrics like precision, recall, F1-score, and support. Analyze the classification report findings for each model, emphasizing any discrepancies in performance among different classes. Assess the precision and recall values to gain an understanding of the model's accuracy in classifying positive and negative instances.

1. **SVC Model**

Classifictaion Report

	precision	recall	f1-score	support
Earthquake	1.00	1.00	1.00	3582
Flood	1.00	0.98	0.99	208
Hurricane	0.00	0.00	0.00	2
Landslide	0.00	0.00	0.00	1
Other	0.99	1.00	1.00	2176
accuracy			1.00	5969
macro avg	0.60	0.60	0.60	5969
weighted avg	1.00	1.00	1.00	5969

Figure 7.10: Classification Report of SVC Model

The SVC model achieved high performance on the given dataset with accuracy, precision score and recall score of 0.99. The classification report shows precision, recall, and F1-score for each class. Most classes had perfect or near-perfect scores, except for ”Hurricane” and ”Landslide,” where the scores were reported as 0,

2. **Random Forest Classifier**

Classifictaion Report

	precision	recall	f1-score	support
Earthquake	0.99	1.00	0.99	3582
Flood	1.00	0.98	0.99	208
Hurricane	0.00	0.00	0.00	2
Landslide	0.00	0.00	0.00	1
Other	0.99	0.98	0.99	2176
accuracy			0.99	5969
macro avg	0.60	0.59	0.59	5969
weighted avg	0.99	0.99	0.99	5969

Figure 7.11: Classification Report of Random Forest Classifier

The Random Forest model achieved high accuracy (0.99) on the given dataset. However, it encountered difficulties in correctly predicting instances for the ”Hurricane” and ”Landslide” classes, resulting in precision, recall, and F1-score scores of 0 for those classes. The ”Earthquake,” ”Flood,” and ”Other” classes had better performance, with precision, recall, and F1-scores ranging from 0.98 to 1.00. Overall, the Random Forest model demonstrated strong accuracy

3. **Logistic regression**

	precision	recall	f1-score	support
Earthquake	1.00	0.99	1.00	3582
Flood	1.00	0.98	0.99	208
Hurricane	0.00	0.00	0.00	2
Landslide	0.00	0.00	0.00	1
Other	0.98	1.00	0.99	2176
accuracy			0.99	5969
macro avg	0.60	0.59	0.59	5969
weighted avg	0.99	0.99	0.99	5969

Figure 7.12: Classification Report of Logistic regression

The Logistic Regression model achieved high accuracy (0.99) on the given dataset. However, it encountered difficulties in correctly predicting instances for the ”Hurricane” and ”Landslide” classes, resulting in precision, recall, and F1-score scores of 0 for those classes. The ”Earthquake,” ”Flood,” and ”Other” classes had better performance, with precision, recall, and F1-scores ranging from 0.98 to 1.00. Overall, the Logistic Regression model demonstrated strong accuracy

Confusion Matrix

A confusion matrix represents the prediction summary in matrix form. It shows how many predictions are correct and incorrect per class. It helps in understanding the classes that are being confused by the model as other classes.

$$Accuracy = TP + TN / TP + FP + TN + FN \tag{7.2}$$

- True Positive (TP) data point
- True Negative (TN) data point
- False Positive (FP) data point
- False Negative (FN) data point

1. **SVC Model**

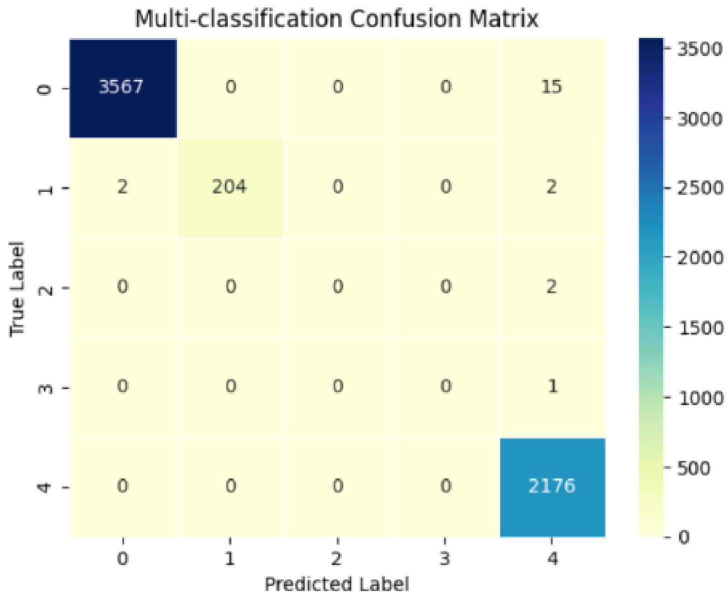

Figure 7.13: Confusion Matrix of SVC Model

2. **Random Forest**

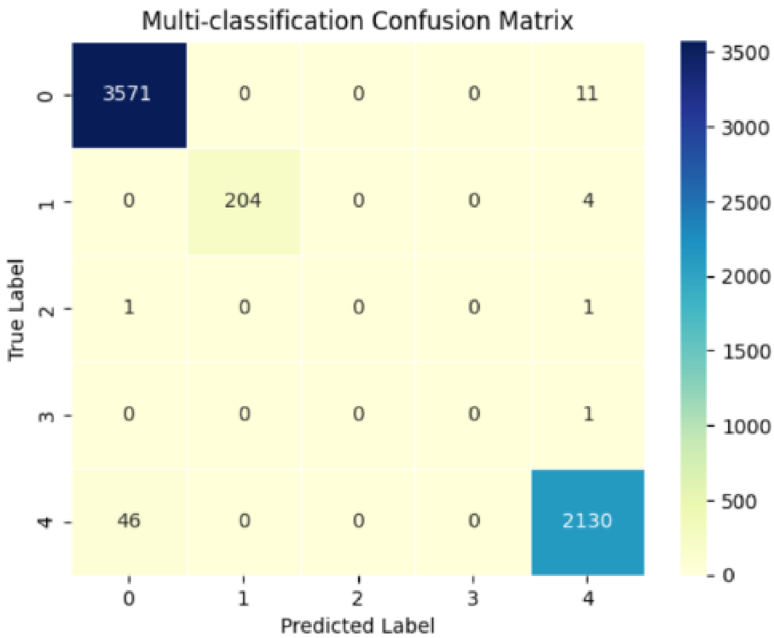

Figure 7.14: Confusion Matrix of Random Forest

3. **Logistic Regression**

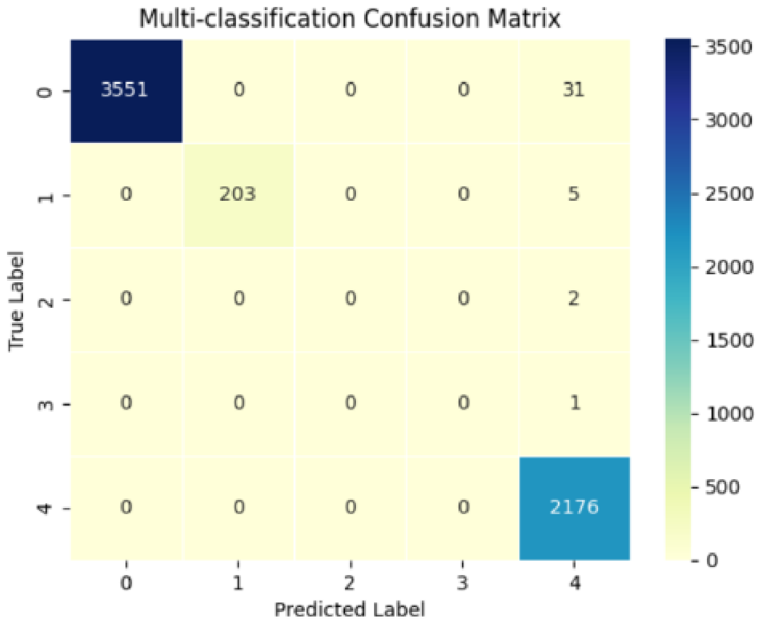

Figure 7.15: Confusion Matrix of Logistic Regression

Model Comparision Table: (Multiple Features)

Below is a comparison table summarizing the performance of different models:

Model	Accuracy	Precision	Recall	F1-score
SVC	0.996314	0.848066	0.845316	0.836668
Logistic Regression	0.993466	0.766479	0.753461	0.754921
Random Forest	0.989278	0.775818	0.791312	0.783532

Figure 7.16: Model Comparision Table: (Multiple Features)

The table summarizes the accuracy, precision, recall, and F1-score for each model. The SVC model achieved the highest accuracy of 0.996314, followed by the Logistic Regression model with an accuracy of 0.993466, and the Random Forest model with an accuracy of 0.989278. The SVC model also achieved the highest precision, recall, and F1-score among the three models. Therefore, for multiple features we select SVC as our primary predicting model.

LSTM Model

Model Evaluation

```
7/7 [==============================] - 0s 50ms/step - loss: 1.0914 - accuracy: 0.6306
Test set
  Loss: 1.091
  Accuracy: 0.631
```

Figure 7.17: Model Evaluation of LSTM Model

From the above figure, we can see that the Loss score of the LSTM model for real-time data is 1.091, and the accuracy score is 0.631 correspondingly.

Comparing plot: Model Loss and Model Accuracy (Single Features)

We additionally plotted the model Accuracy and model loss of the LSTM model. We have been given 0.62 training model accuracy, 0.64 of validation model accuracy with the aid of setting epoch 5 and we were given 0.69 of training model loss, 0.65 of validation modal loss by using placing epoch 5.

Model Loss

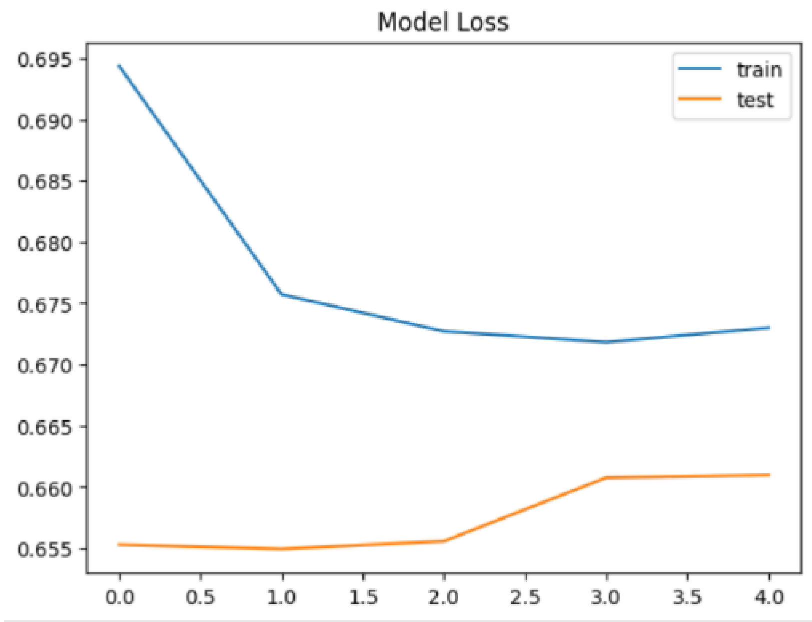

Figure 7.18: LSTM Model Loss

Model Accuracy

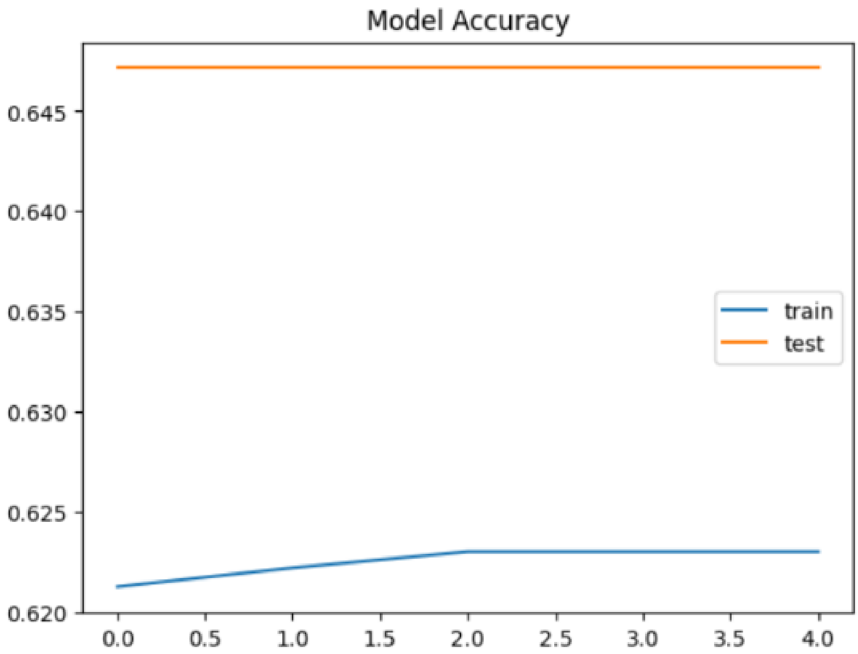

Figure 7.19: LSTM Model Accuracy

7.2. Discussion

Using the Twitter API and authentication to retrieve real-time Twitter tweets, then creating apredictive model to find situational tweets pertinent to crisis management, provides a number of important concerns and ramifications. We'll talk about the importance, difficulties, and potential uses of this methodology in this part.

First and foremost, leveraging real-time Twitter data has become increasingly important in disaster management. The ability to access and analyze tweets as they are posted provides valuable insights into the evolving situation on the ground. It enables emergency responders, government agencies, and organizations involved in disaster management to obtain timely and relevant information. This information can aid in assessing the severity of a disaster, identifying affected areas, and coordinating response efforts efficiently. The Twitter API's rate limits need to be carefully managed to ensure a continuous flow of data without exceeding the allowed quotas. Moreover, identifying and filtering relevant tweets from the vast stream of data requires sophisticated algorithms and techniques. Data preprocessing becomes crucial to remove noise, handle duplicate or irrelevant tweets, and extract meaningful features that contribute to predicting situational tweets accurately.

The constructed raw data set plays a pivotal role in training the predictive model. The data set needs to be diverse, representative, and comprehensive to capture different types of disaster-related tweets. It should include tweets that provide situational updates, reports of damage, calls for help, and other relevant information. Careful consideration must be given to ensure that biases, such as geographic or linguistic biases, are minimized during data collection and selection.

Building a predictive model based on the raw data set involves applying appropriate machine learning techniques. Supervised learning algorithms, such as classification models, can be trained on the labeled data set to predict whether a tweet is situational or not. This model can assist in automatically identifying tweets that contain critical information about the ongoing disaster, allowing for swift response and decision-making.

Advantages

Correct statistics Filtering:

Category enables clear out non-situational tweets, noise, and irrelevant facts from the large volume of social media data for the duration of a catastrophe. This improves the efficiency of records management and ensures that applicable and dependable records reaches the proper channels and selection-makers.

Real-Time scenario cognizance:

Through extracting situational tweets in actual time, selection-makers benefit precious insights into the unfolding occasions and evolving dynamics of a disaster. This enhances their situational cognizance, letting them make informed selections concerning deployment of emergency offerings, and ordinary disaster management.

Progressed aid Allocation:

Correct category of situational tweets assists in optimizing resource allocation during a disaster, authorities can allocate all necessary resources and refuge more successfully. This reduces reaction time and guarantees that resources are directed in which they're maximum needed.

Advanced Situational awareness models:

The category of situational tweets contributes to the improvement and enhancement of situational cognizance insights, analyzing actual-time social media information, those insights may be continually up to date to enhance their accuracy and responsiveness, in the long run leading to more effective disaster control strategies.

Studies and evaluation:

Actual-time class of situational tweets offers a treasured dataset for researchers and analysts to study the styles, developments, and impact of screw ups. This records may be used to perceive vulnerabilities, verify the effectiveness of response strategies, and develop predictive results for destiny catastrophe situations.

Disadvantages

1. Restricted contextual understanding of disaster-related tweets.
2. Challenges in generalizing throughout one of a kind catastrophe sorts or contexts.

8. CONCLUSION AND FUTURE WORKS

In order to provide an improved understanding of the crisis situation, the model's future development may incorporate adding real-time data inputs from Twitter. As a result, the model might be able to quickly adjust to changing conditions and give rescue services the most recent information on the emergency situation as it develops. To increase its ability to predict outcomes and make sure it can manage a variety of crisis scenarios, the model might be trained on extensive simulated disaster scenarios. This might involve generating actual crisis situations utilizing cutting-edge simulation techniques like Random Forest, Nave's Bayes Classifier, Decision Tree, and K-Nearest Neighbour, and training the model on these scenarios to improve accuracy and efficiency.

Additionally, improving the data visualization and reporting capabilities of the model may give emergency responders a deeper knowledge of crisis scenarios, enabling them to take targeted measures and make more educated judgments. This could entail creating interactive dashboards and visualization tools to enable responders to evaluate real-time data and spot patterns and trends that could aid in decision-making.

Developing real-time tweet classification has the potential to significantly improve disaster response operations, to sum up. Accuracy, real-time analysis, and data visualization improvements can give emergency responders the tools they need to quickly comprehend and control catastrophic situations. These technologies can be used to enhance disaster response efforts, which will benefit the impacted communities. Real-time tweet classification will need to keep developing and being used if emergency response and catastrophe management are to improve.

8.1. Future Works

Certain limitations related to our work can be overcome with the help of this future work' effort. With many different search terms or queries, we can retrieve tweets in real time. Additionally, we can able to create a real-time system for receiving tweets that doesn't stop until all of the tweets have been found. Building educational plots or dashboards is another idea we can come up with. A dashboard should be constructed in such a way that it can include details, including the precise date and time, concerning recent natural disasters that have occurred in a specific location. The usage of world maps is another option, as they make it simple for viewers to identify situations.

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30]

References

[1] K. Schwarz, D. A. Aranda, and M. Hartmann, "Towards automated situational awareness reporting for disaster management—a case study," *Sustainability*, vol. 15, no. 10, p. 7968, 2023.

[2] T. Vishwanath, R. D. Shirwaikar, W. M. Jaiswal, and M. Yashaswini, "Social media data extraction for disaster management aid using deep learning techniques," *Remote Sensing Applications: Society and Environment*, vol. 30, p. 100961, 2023.

[3] W. Zhai, "A multi-level analytic framework for disaster situational awareness using twitter data," *Computational Urban Science*, vol. 2, no. 1, p. 23, 2022.

[4] V. Linardos, M. Drakaki, P. Tzionas, and Y. L. Karnavas, "Machine learning in disaster management: Recent developments in methods and applications," *Machine Learning and Knowledge Extraction*, vol. 4, no. 2, pp. 446–473, 2022.

[5] R. Suwaileh, M. Imran, T. Elsayed, and H. Sajjad, "Are we ready for this disaster? towards location mention recognition from crisis tweets," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 6252–6263.

[6] S. Madichetty and M. Sridevi, "A novel method for identifying the damage assessment tweets during disaster," *Future Generation Computer Systems*, vol. 116, pp. 440–454, 2021.

[7] S. Madichetty and S. Muthukumarasamy, "Detection of situational information from twitter during disaster using deep learning models," *Sādhanā*, vol. 45, pp. 1–13, 2020.

[8] J. Vongkusolkit and Q. Huang, "Situational awareness extraction: a comprehensive review of social media data classification during natural hazards," *Annals of GIS*, vol. 27, no. 1, pp. 5–28, 2021.

[9] K. Rudra, P. Goyal, N. Ganguly, M. Imran, and P. Mitra, "Summarizing situational tweets in crisis scenarios: An extractive-abstractive approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 5, pp. 981–993, 2019.

[10] M. M. Bala, K. Navya, and P. Shruthilaya, "Text mining on real time twitter data for disaster response," *Int. J. Civ. Eng. Technol*, vol. 8, no. 8, pp. 20–29, 2017.

[11] I. Sirbu, T. Sosea, C. Caragea, D. Caragea, and T. Rebedea, “Multimodal semi-supervised learning for disaster tweet classification,” in *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, pp. 2711–2723.

[12] S. Goswami and D. Raychaudhuri, “Identification of disaster-related tweets using natural language processing: International conference on recent trends in artificial intelligence, iot, smart cities & applications (icaisc-2020),” *IOT, Smart Cities & Applications (ICAISC-2020)(May 26, 2020)*, 2020.

[13] N. Altay and W. G. Green III, “Or/ms research in disaster operations management,” *European journal of operational research*, vol. 175, no. 1, pp. 475–493, 2006.

[14] W. Sun, P. Bocchini, and B. D. Davison, “Applications of artificial intelligence for disaster management,” *Natural Hazards*, vol. 103, no. 3, pp. 2631–2689, 2020.

[15] B. W. Robertson, M. Johnson, D. Murthy, W. R. Smith, and K. K. Stephens, “Using a combination of human insights and ‘deep learning’for real-time disaster communication,” *Progress in Disaster Science*, vol. 2, p. 100030, 2019.

[16] S. Madichetty and M. Sridevi, “Re-ranking feature selection algorithm for detecting the availability and requirement of resources tweets during disaster,” *International Journal of Computational Intelligence & IoT*, vol. 1, no. 2, 2018.

[17] A. Kumar, J. P. Singh, and S. Saumya, “A comparative analysis of machine learning techniques for disaster-related tweet classification,” in *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129)*. IEEE, 2019, pp. 222–227.

[18] M. Avvenuti, S. Cresci, F. Del Vigna, T. Fagni, and M. Tesconi, “Crismap: a big data crisis mapping system based on damage detection and geoparsing,” *Information Systems Frontiers*, vol. 20, pp. 993–1011, 2018.

[19] S. Roy, S. Mishra, and R. Matam, “Classification and summarization for informative tweets,” in *2020 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE, 2020, pp. 1–4.

[20] M. Abavisani, L. Wu, S. Hu, J. Tetreault, and A. Jaimes, “Multimodal categorization of crisis events in social media,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 679–14 689.

[21] F. Alam, T. Alam, M. Imran, and F. Ofli, “Robust training of social media image classification models for rapid disaster response,” *arXiv preprint arXiv:2104.04184*, 2021.

[22] S. Bidari, “Categorization of disaster related tweets using multimodal approach,” Ph.D. dissertation, Pulchowk Campus, 2021.

[23] A. K. Gautam, L. Misra, A. Kumar, K. Misra, S. Aggarwal, and R. R. Shah, "Multimodal analysis of disaster tweets," in *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2019, pp. 94–103.

[24] P. Prasad, V. J. Loveson, B. Das, and M. Kotha, "Novel ensemble machine learning models in flood susceptibility mapping," *Geocarto International*, vol. 37, no. 16, pp. 4571–4593, 2022.

[25] A. Lin, H. Wu, G. Liang, A. Cardenas-Tristan, X. Wu, C. Zhao, and D. Li, "A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster," *International Journal of Disaster Risk Reduction*, vol. 49, p. 101682, 2020.

[26] S. S. M. Win and T. N. Aung, "Target oriented tweets monitoring system during natural disasters," in *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. IEEE, 2017, pp. 143–148.

[27] L. S. Gopal, R. Prabha, D. Pullarkatt, and M. V. Ramesh, "Machine learning based classification of online news data for disaster management," in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2020, pp. 1–8.

[28] J. Longhini, C. Rossi, C. Casetti, and F. Angaramo, "A language-agnostic approach to exact informative tweets during emergency situations," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 3739–3475.

[29] L. S. Snyder, M. Karimzadeh, R. Chen, and D. S. Ebert, "City-level geolocation of tweets for real-time visual analytics," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 2019, pp. 85–88.

[30] H. J. Scholl, S. Ballard, S. Carnes, A. Herman, and N. Parker, "Informational challenges in early disaster response: the massive oso/sr530 landslide 2014 as case in point," 2017.

D Y PATIL
INTERNATIONAL
UNIVERSITY
AKURDI PUNE

CERTIFICATE

This is to certify that the project entitled **An Integrated Tool For Real-Time Responses In Disaster Management** submitted by:

Rohit Saktel	20190802096
Shivani Ambhore	20190802020

is the partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering is an authentic work carried out by them under my supervision and guidance.

Mr. Rajkumar Chaudhari
(Mentor)

Dr. Tamal Mondal
(Supervisor)

D Y PATIL INTERNATIONAL UNIVERSITY ★ AKURDI, PUNE-44. ★

Director
Dr. Bahubali Shiragapur
School of Computer Science Engineering & Application
D Y Patil International University, Akurdi
Pune, 411044, Maharashtra, India

DECLARATION

We, hereby declare that the following report which is being presented in the Major Project entitled as **An Integrated Tool For Real-Time Responses In Disaster Management** is an authentic documentation of our own original work to the best of our knowledge. The following project and its report in part or whole, has not been presented or submitted by us for any purpose in any other institute or organization. Any contribution made to the research by others, with whom we have worked at D Y Patil International University, Akurdi, Pune or elsewhere, is explicitly acknowledged in the report.

Rohit Saktel	(20190802096)	
Shivani Ambhore	(20190802020)	