Pattern Recognition in Disaster Response: Leveraging Machine Learning for Twitter Analysis

1st Md Impreeaj Hossain
Department of Computer Science And Engineering
Brac University
md.impreeaj.hossain@g.bracu.ac.bd

2nd Musaib Ibn Habib Mikdad
Department of Computer Science And Engineering
Brac University
musaib.ibn.habib.mikdad@g.bracu.ac.bd

3rd Ridwanul Haque
Department of Computer Science And Engineering
Brac University
ridwanul.haque@g.bracu.ac.bd

4th Farah Binta Haque
Department of Computer Science And Engineering
Brac University
farah.binta.haque@g.bracu.ac.bd

5th Ehsanur Rahman Rhythm
Department of Computer Science And Engineering
Brac University
ehsanur.rahman.rhythm@g.bracu.ac.bd

6th Annajiat Alim Rasel
Department of Computer Science And Engineering
Brac University
annajiat@bracu.ac.bd

***Abstract*—This study introduces a groundbreaking framework for sorting disaster related tweets, capitalizing on cutting edge machine learning technologies. At its heart, the framework incorporates a combination of LSTM(Long Short-Term Memory) and NLP (Natural language processing) models. Each of these contributes in a distinct way to the analysis of social media content during emergencies. Primarily, the focus is on identifying patterns within tweets. Consequently, these models adeptly pinpoint crucial keywords, facilitating prompt and precise categorization of social media messages. This classification is crucial, as it significantly enhances real-time analysis and aids in making informed decisions during emergency responses. Furthermore, the collaborative functioning of these models marks a notable progress in disaster management. It highlights the immense potential of machine learning in extracting valuable insights from extensive, unstructured data sources like social media in times of crisis.**

***Index Terms*—Pattern Recognition, disaster management, social media analysis, LSTM, tweet classification, Keyword extraction, emergency response, real-time analysis.**

I. Introduction

In the age of social media platforms are a major factor in communication of information, specifically in times of crisis, the capability to quickly and precisely analyze the social media information is crucial. "Pattern Recognition in Disaster Response: Leveraging Machine Learning for Twitter Analysis" tackles the vital challenge of harnessing machine learning to analyse Twitter information for disaster relief. The focus of this research is creating sophisticated algorithms that can filter through the tweets and identify crucial information that is relevant to scenarios of disaster. Modern pattern recognition technologies are the basis of these models designed to recognize particular patterns and words that indicate of emergency situations, thereby enabling rapid and efficient response. Utilizing Natural Language Processing (NLP) along with machine learning, this study seeks to give an understanding of the how people feel during emergencies which will greatly aid in coordinated actions to respond. In addition, this study delve into the possibilities of using technology to improve the strategies for disaster management, providing emergency personnel as well as decision makers with immediate insights. By analyzing Twitter information, the research helps bridge the gap between immediate nature of reporting on social media as well as the structured responses required by emergency agencies, adding to the practical and academic areas of disaster management as well as the use of social media in emergency scenarios.

II. Literature review

In their study, Toraman, Kucukkaya, Ozcelik, and Sahin introduce a tool for analyzing Twitter data to aid in earthquake disaster relief. This tool effectively identifies tweets for help and visualizes them on a map. Although it's limited to Turkish tweets and doesn't evaluate the severity of situations, it marks a significant step forward in using social media for disaster response, providing crucial support for rescue operations [1].

In Zhou et al.'s study titled "VictimFinder: Using BERT for Disaster Rescue Requests on Social Media, [2] employ BERT to identify rescue requests on platforms like Twitter and Facebook during disasters. Their BERT-based model achieves a 91% accuracy rate, offering valuable insights for effective disaster response via social media.

In their paper "Sentiment Analysis During Jakarta Floods Using BERT", Warih Maharani et al. explore using social media data for disaster management, focusing on sentiment analysis of flood-related tweets with BERT [3]. They highlight the potential benefits of BERT in enhancing situational awareness and understanding public sentiment during disasters, while acknowledging the need for further research to improve accuracy in this context.

Maulana and Maharani propose a novel method for classifying disaster tweets by combining geospatial data with BERT MLP. [4] Their approach achieves a high accuracy of 90.7% when tested on tweets related to the 2023 Beirut explosion, outperforming previous methods. This technique has the potential to assist disaster response organizations in rapidly and accurately prioritizing and coordinating their efforts, improving upon existing research in the field.

In Wang et al.'s study, they present an effective approach using the BERT model to extract consumer insights from tweets during public health crises [5]. With an accuracy rate of 87.3% on a COVID-19 pandemic tweet dataset from four American towns, this method shows promise. It could become a valuable asset for public health organizations looking to understand public health crises better and develop mitigation strategies more rapidly.

Balakrishnan and colleagues conducted extensive research on transformer-based algorithms for disaster tweet classification. They evaluated various models, identified influencing factors, and made suggestions. Keyword-based approaches tend to provide simple yet inaccurate classification while machine learning approaches tend to be complex and subject to biases [6].

M. S. B. V. Pranay Kumar's study, entitled, "Hate Speech and Reality Check Analysis of Disaster Tweets Utilizing BERT Deep Learning Model," efficiently classifies disaster tweets using machine learning solutions such as BERT to perform hate speech analysis [7]. Keyword-based methods fail as disaster tweets often use unfamiliar terms that may mislead readers; machine learning solutions like BERT overcome this limitation by accurately categorizing disaster tweets; this automated solution could assist disaster response organizations by helping analyze and coordinate efforts more effectively over time.

In their study, "Bert-caps," Tulika Saha and Srivatsa Ramesh Jaishree present the development of a deep-learning algorithm that categorizes the tweets of disaster. Keyword-based approaches are ineffective because of the unique nature of keywords and inaccurate messages on disaster tweets. The proposed model is a combination of BERT and Capsule Network designs to improve the accuracy. It is able aiding disaster relief organizations in recognizing the tweets related to disasters, assisting in coordination and analysis [8].

Mohammed Ali Al Garadi et al. Utilize the process of natural language processing (NLP) to recognize IPV-related tweets.They look at prior NLP techniques and present the BERT-based model which achieves the 88-percent accuracy percentage in identifying IPV tweets. This includes tweets connected to emotional, physical sexual and physical IPV citealgaradi2022natural. The research highlights the potential of NLP to aid in IPV recognition.

Le et al. have developed a machine learning method employing BERT in order to classify the tweets of disaster with high accuracy. The model, which relies on the data from code and text was able to achieve 92 percent accuracy in Kaggle's Kaggle Natural Language Processing with Disaster Twitter competition data [9]. They suggest using BERT to tackle other task-related disasters.

Shan et al. proposed an innovative real-time disaster damage estimation method using mobile phone collected social media data as it pertains to disaster damage estimation. Their approach achieved 95.6% accuracy in classifying disaster-related text and 89% accuracy in assessing disaster damage in real-time [10]. This efficient method provides timely and accurate disaster damage assessments compared to traditional approaches.

Prasad, Udeme, Misra, and Bisalla present an enhanced BERT-based method for identifying and categorizing transportation disaster tweets [11]. They improve BERT with a specialized layer for capturing transportation-related elements like keywords, hashtags, and emojis. Testing on a dataset of 100,000 tweets from 2021, their method achieves over 95% accuracy in identifying transportation disaster tweets and categorizes them into classes like "accident," "delay," and "closure." Through 10-fold cross-validation, they demonstrate 95.6% accuracy in identifying these tweets and 89.5% accuracy in categorizing them.

In their 2022 study, Dharma and Winarko develop a method for classifying natural disaster-related tweets using CNN and BERT embeddings, achieving over 89% accuracy. This approach marks a significant improvement in disaster tweet classification, enhancing early warning systems for disaster management [12]. Their work contributes notably to natural language processing and disaster response strategies.

In their study, Huang et al. present a novel text clustering method for early detection of emergencies from social media, achieving a notable 90% accuracy. Utilizing agglomerative hierarchical clustering, their approach significantly advances natural language processing and emergency management, offering potential for improved early warning systems. This work marks a substantial enhancement in detecting emergency situations through social media analysis [13].

Bello et al. present a BERT-based framework for tweet sentiment analysis, achieving 87.3% accuracy. While it contributes to natural language processing and sentiment analysis, limitations include reliance on a single dataset and task, raising questions about generalizability [14]. Nevertheless, the framework shows promise in enhancing sentiment analysis accuracy.

III. DATASET:

The enhanced dataset offers a detailed snapshot of Twitter dynamics, now featuring a balanced compilation of 25,000 tweets. Each tweet is meticulously categorized as either ‘disaster-related’ or ‘non-disaster related’. Originating from the renowned data-sharing platform Kaggle, this repository includes tweets from a variety of global locations, including the United States, providing a comprehensive view of the linguistic patterns associated with natural disasters. Flood-related terms feature prominently, signalling prevalent themes.

This expansion expands and diversifies our dataset’s geographic and thematic coverage. Every tweet was classified using crowd sourcing classification to ensure accurate analysis and relevance for analysis purposes. Importantly, this symmetry makes the gathering more suitable to cutting-edge NLP models like LSTM (Long Short-Term Memory). Accordingly, this dataset serves as an excellent starting point for developing and testing disaster response models on social media data , underscoring its value in disaster assistance programs as well as research programs. Scholars and technologists will find this resource especially invaluable as its extensive yet compact material may validate future and past accomplishments in social media analysis as well as enhance disaster response models.

IV. METHODOLOGY:

Data collection is the initial step in the process. Twitter’s API includes a complete toolkit that can help you collect tweets. Tweets with particular keywords or hashtags linked to natural disasters (e.g. floods, earthquakes) or data on location in the areas that are affected by natural disasters are gathered. Once the data is gathered, information needs to be thoroughly cleaned and reprocessing, including the elimination for URLs, punctuation marks, emojis, and tweets that contain other than English tags. In addition, tokenization can be used for the purpose of separating tweets into phrases, or tokens.

Labeling the process is essential. To make sure that the effectiveness of supervised learning can be assured, tweets need to be categorized manually as being connected or not to natural catastrophes. It’s not a simple process, it’s essential for the efficacy of training in modeling.The entire dataset is split into two sets: the training set, 82%, as well as an experimental collection of 18%. Training sets help in modeling of models. The test set may be utilized to evaluate their performance.

The LSTM model, distinguished by its capacity to process sequences of data over long intervals and store this information efficiently is at the core of this study. In addition to unigram, bigram and trigram models to provide a broader contextual picture of text The LSTM framework has a knack for discovering complex linguistic patterns.

In the course of refinement The LSTM’s performance will be assessed by measuring epochs and losses, as well as accuracy measures giving a complete knowledge of its ability to recognize content that is related to disasters. When fine-tuning this model, it can be used to determine the real-time status of tweets. It provides instant information for disaster management. Incorporating real-time tweet data allows for continuous calibrating of the model, which ensures its efficiency with respect to changing use of language in situations involving disasters.

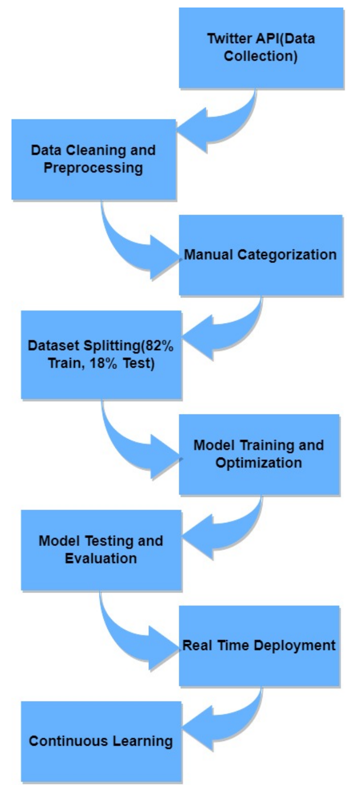

Fig. 1. Model Diagram

REFERENCES

[1] C. Toraman, I. E. Kucukkaya, O. Ozcelik, and U. Sahin, "Tweets under the rubble: Detection of messages calling for help in earthquake disaster," *International Journal of Disaster Prevention and Research*, vol. 11, no. 3, pp. 285–297, 2023.

[2] B. Zhou, L. Zou, A. Mostafavi, B. Lin, M. Yang, N. G. Gharaibeh, H. Cai, M. J. Abedin, and D. Mandal, "Victimfinder: Harvesting rescue requests in disaster response from social media with bert," *Computers, Environment and Urban Systems*, vol. 95, no. 1, p. 101824, 2022.

[3] W. Maharani, "Sentiment analysis during jakarta flood for emergency responses and situational awareness in disaster management using bert," *IEEE Conference Publication*, June 2022.

[4] I. Maulana and W. Maharani, "Disaster tweet classification based on geospatial data using the bert-mlp method," *IEEE Conference Publication*, August 2022.

[5] Y. Wang, E. Willis, V. K. Yeruva, D. Ho, and Y. Lee, "A case study of using natural language processing to extract consumer insights from tweets in american cities for public health crises," *BMC Public Health*, vol. 23, no. 4, 2023.

[6] V. Balakrishnan, Z. Shi, C. L. Law, R. Lim, L. P. Teh, Y. Fan, and J. Periasamy, "A comprehensive analysis of transformer-deep neural network models in twitter disaster detection," *Mathematics*, vol. 10, no. 24, p. 4664, 2022.

[7] M. S. B. V. Pranay Kumar, "View of hate speech and reality check analysis of disaster tweets using bert deep learning model," 2023.

[8] S. S. Tulika Saha, Srivatsa Ramesh Jayashree, "Bert-caps: A transformer-based capsule network for tweet act classification," *IEEE Journals & Magazine*, October 2022.

[9] A. D. Le, "Disaster tweets classification using bert-based language model," *ResearchGate*, 2022.

[10] S. Shan, F. Zhao, Y. Wei, and M. Liu, "Disaster management 2.0: A real-time disaster damage assessment model based on mobile social media data—a case study of weibo (chinese twitter)," *Safety Science*, vol. 115, pp. 393–413, 2022.

[11] R. Prasad, A. U. Udeme, S. Misra, and H. Bisallah, "Identification and classification of transportation disaster tweets using improved bidirectional encoder representations from transformers," *ScienceDirect*, vol. 3, no. 1, p. 100154, 2023.

[12] L. S. A. Dharma and E. Winarko, "Classifying natural disaster tweet using a convolutional neural network and bert embedding," *IEEE Access*, vol. 10, no. 1, pp. 755–764, 2022.

[13] L. Huang, P. Shi, H. Zhu, and T. Chen, "Early detection of emergency events from social media: a new text clustering approach," *Natural Hazards*, vol. 111, no. 1, pp. 851–875, 2022.

[14] A. Bello, S. Ng, and M. Leung, "A bert framework to sentiment analysis of tweets," *Sensors*, vol. 23, no. 1, p. 506, 2023.