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

5thEhsanur 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 BERT, SentiBERT, BERT-DenseNet, and RoBERTa 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, BERT, sentiBERT, BERT-denseNet,RoBERTa, tweet classification, Keyword extraction, emergency response, real-time analysis.**

I. Introduction

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 88.3% accuracy rate on a COVID-19 pandemic tweet dataset from four American towns, this method shows promise. It could serve as a valuable tool for public health organizations, helping them understand the impact of public health crises and develop mitigation strategies.

Balakrishnan et al. extensively studied transformer-based algorithms for disaster tweet classification. They compared different models, identified influencing factors, and offered recommendations. [6] Keyword-based methods are simple but prone to misinformation, while accurate machine learning methods can be complex and biased.

M. S. B. V. Pranay Kumar's study, "Hate Speech and Reality Check Analysis of Disaster Tweets Using BERT Deep Learning Model," effectively classifies disaster tweets. Keyword-based methods fall short, as disaster tweets often use unique terms and can be deceptive [7]. Machine learning solutions, like BERT, overcome these limitations by accurately categorizing disaster tweets. This automated approach has the potential to assist disaster response organizations in analyzing and coordinating their efforts more effectively.

In their study, "Bert-caps," Tulika Saha and Srivatsa Ramesh Jayashree propose a new deep learning model for categorizing disaster tweets. Keyword-based methods are inaccurate due to unique keywords and misleading messages in disaster tweets. The model combines BERT and Capsule Network designs, improving accuracy. It has the potential to assist disaster response organizations by automating the recognition of disaster-related tweets, aiding in analysis and coordination efforts [8].

Mohammed Ali Al Garadi et al. use natural language processing (NLP) to detect IPV-related tweets. [9] They review previous NLP methods and introduce a BERT-based model that achieves an 88% accuracy rate in identifying IPV tweets, including those related to physical, emotional, and sexual IPV. This research demonstrates the promise of NLP for IPV detection.

Le et al. developed a machine learning approach using BERT to classify disaster tweets with high precision. Their model, utilizing text and code data, achieved 92% accuracy on the Kaggle Natural Language Processing with Disaster Twitter competition dataset [10]. The authors suggest exploring BERT's potential for other disaster management tasks.

Shan et al. presented a real-time disaster damage estimation method using mobile phone-collected social media data. Their approach achieved 95.6% accuracy in classifying disaster-related text and 89% accuracy in assessing disaster damage in real-time [11]. 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 [12]. 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 [13]. 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 [14].

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 [15]. 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. Key terms like 'floods' prominently feature, highlighting prevalent themes.

This expansion enhances the dataset's depth, particularly in its geographic and thematic scope. Each tweet has been categorized through a precision-driven, crowd-sourced classification process, ensuring accuracy and relevance. Importantly, the balanced nature of this collection makes it especially suitable for advanced NLP models such as BERT, SentiBERT, BERT-DenseNet, and RoBERTa. As such, it provides a solid foundation for developing and testing these models, emphasizing the value of social media data in disaster relief and research. With its concise yet rich content, the dataset is an invaluable tool for researchers and technologists, aiding in the enhancement of disaster response strategies and the validation of both existing and emerging methods in social media analysis.

IV. Methodology:

The initiative to analyze Twitter dynamics for disaster response involves a multifaceted approach, employing advanced NLP models like BERT, SentiBERT, BERT-DenseNet, and RoBERTa. The first phase is data collection, where Twitter's API plays a crucial role in acquiring tweets. Targeted are tweets with specific keywords or hashtags linked to disasters, such as flooding or earthquakes, as well as those from regions affected by these events. This collected data undergoes a rigorous cleansing process, removing URLs, HTML tags, punctuation, emojis, and non-English content, followed by tokenization to break down tweets into individual words or tokens.

A critical step in this process is the manual labeling of tweets to determine their relevance to actual natural disasters. Despite its time-consuming nature, this step is vital for the effective training of supervised learning models. The dataset is divided into an 80% training set and a 20% test set, with the former used to train the models and the latter for performance evaluation.

The core of the analysis utilizes the BERT model, renowned for its proficiency in contextual word understanding, thanks to its training on an extensive text corpus. The Hugging Face Transformers Library provides access to pre-trained BERT models, which are further refined using both the N-gram model and BERT to enhance result accuracy. This fine-tuning process involves integrating the categorized disaster tweets, enabling the model to better understand the nuances of disaster-related communication.

Post-optimization, the model undergoes testing with metrics such as precision, accuracy, recall, and F1 scores to gauge its effectiveness in analyzing disaster-specific tweets. Once trained, it is deployed to identify live tweets in real-time, providing critical information for disaster response. This continuous learning from new tweets ensures the model remains attuned to the evolving language patterns in disaster-related communications.

Key features of this project include tweet classification (disaster or not, and type of disaster), a disaster map superimposed on open street maps, aid/needs classification, and analysis of disaster intensity based on tweet frequency, timing, location, and language used. The project also explores patterns in tweet composition, early warning signals of impending disasters, and multimodal content analysis including text, symbols, pictures, and videos. Additionally, it focuses on disaster preparedness, planning, response strategies, and a comprehensive analysis of words, sentences, and symbols within tweets.

Result and Analysis:

Conclusion:

Future Work:

Acknowledgement:

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] M. A. Al-Garadi, S.-J. Kim, Y. Guo, E. Warren, Y. Yang, S. Lakamana, and A. Sarker, "Natural language model for automatic identification of intimate partner violence reports from twitter," *Array*, vol. 15, p. 100217, 2022.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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