Grenze International Journal of Engineering and Technology, June Issue

AI-Driven Automated Crisis Distress Classification System using Natural Language Processing

R. Siva[1] and Aishvar Amudhan[2]
[1-2]SRM Institute of Science and Technology Kattankulathur, Tamil Nadu 603203
Email: sivar@srmist.edu.in, aa4674@srmist.edu.in

***Abstract*— The social media platforms have been of great essence as sources of real time information in case of disaster and emergencies. People tend to provide immediate updates on the necessity of rescue, medical emergency, infrastructure damage, and resource shortages using social networks like Twitter. Despite the usefulness of the situational information in these messages, the emergency response teams find it challenging to monitor the large amount of unorganized social media data manually. The proposed research is a crisis distress classification system based on artificial intelligence, which automatically analyzes the messages in social media to determine disaster related information and prioritize distress signals. The suggested framework employs the BERTweet transformer model that is specifically created to interpret informal language that is widespread in social media posts. The pipeline being used in the system processes tweets in the Kaggle Disaster Tweets dataset by using a modular pipeline that includes text preprocessing, generation of contextual embedding, crisis classification, urgency prediction, and location extraction via Named Entity Recognition. The crisis messages are divided into the classes of disaster-related and another classifier forecasts the urgency levels to prioritize the response activities. The organized outcomes are combined with the UiPath Robotic Process Automation workflow, which automatically works with the output and sends email alerts on emergency monitoring. The offered solution emphasizes the possibility of integrating the transformer-based Natural Language Processing with the automation technologies to help to provide intelligent disaster management systems and real-time monitoring of the crisis.**

***Index Terms*— Crisis Detection, Disaster Tweet Classification, BERTweet, Natural Language Processing, Named Entity Recognition, Robotic Process Automation.**

I. INTRODUCTION

Social media has emerged as a very important means of communication in times of disasters and emergencies. Whenever there is a flood, earthquake, fire or a major accident, people tend to post real-time information about the rescue operation, damage to infrastructure, medical crises, and lack of necessary resources using social networks like Twitter, Facebook, and Instagram. Such user generated messages offer a good situational information that can help an emergency response team and a disaster management authority to comprehend the current events and organize rescue missions. Nevertheless, the sheer mass of communications that are caused by crisis situations poses significant problems in terms of extracting the information effectively. It is possible that thousands of messages can be posted in a few days, and it is almost impossible to track them manually and find the messages of critical distress among them.

Grenze ID: 01.GIJET.12.2.199_24

Also, the messages shared in the social media are usually informal and unstructured and include use of abbreviations, slang, emojis, misspellings, and irregular grammar. Such features render it hard to precisely decipher the significance of such messages using the conventional form of filtering that relies on keywords. The recent progress in Natural Language Processing (NLP) and deep learning has brought a considerable enhancement in the capacity of a machine to process the textual data and obtain valuable insights. Transformer-based language models include BERT, RoBERTa, and BERTweet, which have been shown to perform better in sentiment analysis, text classification, and information extraction. In these models, self-attention systems are used to extract contextual relationships between words in a sentence so that they can comprehend more about the semantic meaning of text. BERTweet is one of these models that are specifically trained and pretrained on massive twitter data. This renders it especially useful in the analysis of social media messages which include informal language patterns. Using such models as the means of crisis detection can greatly enhance the effectiveness of the classification of messages related to the disaster and allow classifying the messages based on the urgent signs of distress automatically.

A. Background and Motivation

Over the past few years, social media has become a more significant part of disaster response and crisis management. Social media platforms are regularly used by the affected individuals in case of an emergency scenario to demand help, express their damage, and discuss real-time information about their environment. The analysis of this information can be useful to emergency response organizations and humanitarian agencies to detect the affected areas and efficiently allocate resources. Although such data is available, it is difficult to extract actionable information out of the social media streams. The main challenge is to find pertinent messages about crisis related information amidst the high number of irrelevant posts that are usually posted during big events. Moreover, despite the detection of crisis messages, it is necessary to determine the urgency of each message to prioritize the actions of response. The classical methods of machine learning are based on the feature engineering methods like bag-of-words and TF-IDF. Though these techniques may be able to offer rudimentary classification, they do not tend to represent the contextual meaning of language, particularly where the message is short and noisy, as in social media messages. Transformer models address these weaknesses by learning contextual word representations. The models examine the connection among words in a sentence with the help of attention mechanisms and, therefore, they understand more complex linguistic patterns. The combination of such models with crisis detection systems will go a long way in improving the automatic identification and classification of messages about disasters. Besides proper classification, real-world crisis monitoring systems must have the capability of being automated so that they can respond fast. The addition of Natural Language Processing models to Robotic Process Automation (RPA) platforms would allow raising an automated alert and sharing information with the appropriate authorities. This drives the creation of a smart system that integrates text analyzing with transformers alongside automation technology.

B. Problem Statement

Despite the fact that social media has useful real-time information in case of disaster situations, there are a number of challenges that restrict its efficient use in crisis response: Manual monitoring is not practical because of the large number of posts on their social media during the disaster. Text in social media is usually unstructured, informal, and noisy and it is not easy to interpret any messages using traditional ways of using key words. Current machine learning methods cannot provide contextual correlations in short text messages. There are numerous systems that are currently in place, which are only dedicated to the detection of a disaster and lack urgency prioritization and automated response systems. Thus, the intelligent system that can analyze the messages of social media automatically, detect crisis-related messages, prioritize the pressing distress messages, and support automated response procedures are in need.

C. Proposed Approach

To solve the challenges raised above, this study suggests the use of an AI-based crisis distress categorization system, in which transformer-based Natural Language Processing is used to process disaster-related messages on social media. The proposed framework takes the Kaggle Disaster Tweets data as input and processes the tweets in a multi-stage pipeline, which includes data preprocessing, contextual embedding generation, crisis classification, urgency detection, and location extraction. The system is based on the BERTweet transformer model that is optimized to comprehend the informal language of social media. Messages are classified into disaster-related categories with the contextual embeddings produced by the transformer model. Moreover, an urgency prediction module determines the level of urgency of each message; emergency responders are able to concentrate on the distress signals that are the most important. In order to prove the practical applicability, the structured output of the NLP model is merged with UiPath Robotic Process Automation that automatically processes the results and creates an email alert to monitor the results. This integration points out to the possibilities of implementing such systems in the actual disaster monitoring conditions.

II. Literature Review

A number of research studies have investigated the application of machine learning and natural language processing methods to the analysis of disaster-related social media data. The transformer-based semantic role labeling framework that was proposed by Ariyanto et al. (2025) can be utilized to analyze crisis events on Twitter datasets [1]. Their method enhanced semantic understanding through establishing connections among entities and textual actions. Nevertheless, it was a system that was oriented to event extraction, as opposed to crisis type classification.
Kar et al. (2025) conducted a comparative study of the classical machine learning models such as Naive Bayes, Support vector Machines, Logistic regression, random Forest and Decision Trees to classify disaster tweets [2]. The research stated that logistic regression was the most effective models compared to the other models that were tested. Nevertheless, the models were very dependent on manual feature extraction techniques like TF-IDF and did not have much contextual information.
The paper by Puhazhendhi et al. (2025) is a real-time media monitoring system created on the basis of machine learning tools and rule-based alert systems [3]. The system was more efficient than the other system in monitoring efficiency in detection of crises but gave higher false-positive rates and failed to include transformer-based contextual modeling.
Nath et al. (2025) suggested a hybrid approach to represent the TF-IDF features and Long Short-Term Memory (LSTM) networks to classify the type of crisis and the level of urgency of the tweet [4]. Despite the moderate classification performance of the model, LSTM architectures do not scale to long-range contextual relations as well as transformer-based models. Multimodal frameworks of disaster response have also been reviewed recently.
Hussain et al. (2024) introduced an idea of a deep learning framework that combines textual analysis performed with BERT and visual analysis performed with convolutional neural networks [6]. Although the method enhanced the accuracy of damage detection, it needed multimodal data and was more complex to compute. These studies, though showing improvement in crisis informatics, most of the existing systems are only involved with classification and fail to include urgency prioritization or structured information extraction. Thus, a holistic system that incorporates transformer-based classification, urgency detection and location extraction is required to enhance disaster response support.

III. Methodology

The system suggested has a modular structure that is meant to analyze messages on the social media and produce structured crisis information. The system comprises a number of stages that are data ingestion, preprocessing, classification using transformers, urgency identification, location acquisition, and structured output.

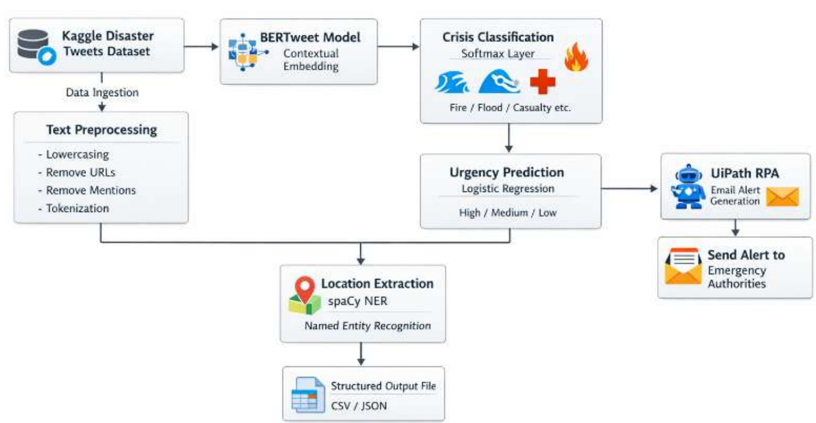

Figure 1. System Architecture Diagram

A. Dataset

In this paper, the dataset was taken under the Natural Language Processing with Disaster Tweets dataset in Kaggle. The dataset has thousands of tweets that are classified on whether the message is about a real disaster event or not. These tweets contain some types of informal language that are prevalent on social media sites.

TABLE I: DATASET DESCRIPTION

Feature	Description
Dataset Source	Kaggle Disaster Tweets
Total Tweets	~10,000
Text Field	Tweet message
Label	Disaster / Non-Disaster
Additional Fields	Keyword, Location

B. Text Preprocessing

Social media messages could be very noisy and may have irrelevant contents that may have an adverse impact on model performance. Hence, preprocessing is conducted to process and clean the text and normalize it before inputting it into the NLP model. The preprocessing processes involve:

- All text can be changed to lowercase.
- Eliminating URLs and hyperlinks.
- Cutting the user mentions and hashtags.
- Eliminating special characters and emojis.
- Tokenizing the text

These measures assist in standardization of the input data and enhance the capacity of the model to acquire significant patterns.

C. Transformer-Based Contextual Encoding

The BERTweet transformer model is the main element of the suggested system. Transformer models are premised on an attention mechanism that enables the model to attend to various sections of a sentence during representation generation. Transformers do not process one word at a time in order as is the case in traditional neural networks, but process an entire sentence at once. BERTweet model is pretrained using large scale twitter data and is specifically trained to deal with informal languages used in social media posts. The BERTweet tokenizer is used to tokenize each message and transform it into token embeddings. These embeddings are further fed into several transformer layers that calculate the contextual relationships among words. The transformer gives out a contextual embedding of each token. The embedding of the [CLS] token is the whole sentence and it is taken as the feature representation to perform classification.

D. Crisis Classification

The contextual embedding that is acquired via the transformer model is inputted in a fully connected neural network classification layer. The outputs of the model are transformed into the probabilities of each category of crisis by using a softmax activation function. The softmax function is as follows:

$$(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{1}$$

where z i is the score of the output of the class i. The model has the following categories of messages: Fire, Flood, Natural Disaster, Casualty, Non-Disaster. AdamW optimizer is used to train the model and it enhances the convergence by using weight decay regularization.

E. Urgency Detection

The system forecasts the urgency of the messages after crisis classification. Urgency detection is done based on embeddings produced by the transformer model. The input features of a logistic regression classifier consist of these embeddings. The classifier is divided into three levels of urgency: High urgency; Medium urgency; Low urgency. Urgency prediction assists emergency responders in ranking messages that are urgent to attend to.

F. Location Extraction

The system recognizes the affected regions by doing Named Entity Recognition (NER) via spaCy NLP library. NER recognizes the entities in the text and classifies them into a set of predefined entity types. The entities in this system are labeled as GPE (Geopolitical Entity) and LOC (Location) are taken out of the message. This data is used to guide emergency officials on the areas that need to be assisted.

G. Output Structuring

The last phase of the system systematizes the predictions and presents them in form of structured outputs that consist of:

- Crisis category
- Urgency level
- Entity of location extracted.
- Confidence scores

H. Automation with UiPath RPA

Besides the machine learning model, the proposed system will incorporate a lightweight Robotic Process Automation (RPA) workflow, which is based on UiPath. This integration is aimed at showing how the forecasts made by the NLP model can be used automatically in a working process. Once the crisis classification and urgency prediction phases have been finalized, the output is exported to a formatted output file with the text of the tweet, predicted crisis type, urgency score, and extracted location objects. This structured output is read by the UiPath automation workflow and automated operations including creating alerts or a monitoring dashboard are undertaken. The RPA integration shows how the given system can be integrated with the already existing disaster management tools and automated response pipelines. Though the RPA element is small in the existing implementation, it demonstrates the usefulness of the system in the real-life emergency response setting.

IV. METHODOLOGY

In this section, the experimental assessment of the suggested crisis distress classification system will be provided, and the efficiency of the approach based on transformers to study the disaster-related social media messages will be discussed. The experiments were performed by the use of the Kaggle Disaster Tweets dataset that consists of thousands of tweets that are either considered to be a disaster-related message or not. The data was prepared and stained to train the suggested transformer-based classification model. The training process was done to fine-tune the BERTweet model that is designed to analyse social media text. The AdamW optimizer was used to train the model, and standard classification metrics such as precision, recall, F1-score, accuracy, and Hamming loss were used to evaluate the results.

A. Training Performance

The model was found to converge stably in the training stage with a steady decrease in training loss over the epochs. The values of loss are decreasing which means that the model learned significant patterns in the dataset.
The decrease in training loss shows better model learning and parameter optimization in the course of the fine-tuning process. The application of the AdamW optimizer helped with stable convergence since it is able to manage the weight updates and regularization. The performance of the classification is evaluated using the performance of the training set.

TABLE II: EPOCH TRAINING LOSS

Epoch	Loss
1	0.2203
2	0.0753
3	0.0527

B. Classification Performance

The performance of the classification is based on the performance of the training set. The F1-score that is one of the most common metrics of classification tasks with imbalanced datasets was used to determine the performance of the crisis classification model.
The following equation is used to compute the F1-score:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

Where:
The ratio of the correctly predicted positive observations is referred to as precision.

Recall is used to determine the percentage of true positives that were correctly recognized.

C. Multi-label Evaluation

The Hamming Loss measure was employed to assess the general performance of classification. Hamming Loss is a figure of the number of labels that are falsely predicted.

$$HammingLoss = \frac{1}{N \times L} \sum_{i=1}^{N} \sum_{j=1}^{L} I(y_{ij} \neq \hat{y}_{ij}) \tag{3}$$

Where:

- N= number of samples
- L= number of labels
- yij= true label
- $\hat{y}ij$= predicted label

The Hamming Loss of the model was 0.003, which means that the rate of misclassification between labels is very low. This finding illustrates that the model is useful in identifying the various categories of disasters.

D. Urgency Performance Prediction

The system forecasts the level of urgency of each message besides classifying the messages as crises. Urgency prediction assists in prioritizing messages that are urgent. The urgency classifier divided the messages into three categories:

- High urgency
- Medium urgency
- Low urgency

The urgency prediction module was found to have a general accuracy of 84.06. Whereas the accuracy of urgency prediction is a bit less than the accuracy of classification, the outcome is anticipated since the identification of urgency is frequently necessitated by additional contextual meaning of language.

E. System Integration and Automation

In order to prove the practical relevance of the suggested system, the structured output produced by the NLP model was combined with UiPath Robotic Process Automation (RPA). The tasks that the workflow undertakes include: Reads the structured output file with classification results. Finds disaster related high urgency tweets. Creates an email notification with the relevant information automatically. Forwards the alert to the responsible surveillance officers. This integration shows the possibility of implementing the proposed system in a real-time disaster monitoring setting, where automated alert systems can support an emergency response team.

F. Comparison to Existing Approaches

The model performance was also compared to the traditional machine learning methods applied in the past to assess the effectiveness of the proposed approach. The findings indicate that the suggested BERTweet-based method is much more effective than the conventional machine learning models. This advantage can be credited to the fact that transformer models are able to learn contextual relationships of text with the help of attention mechanisms.

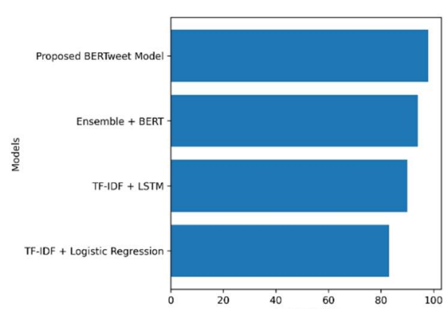

Figure 2. Comparison of model performance

G. Discussion

The obtained experimental evidence proves that Natural Language Processing models with transformers are effective in detecting and classifying a crisis. The high performance of the BERTweet model was because it was able to produce contextual embedding based on text in social media. The transformer-based approach has much better classification accuracy and lower misclassification rates than the traditional machine learning techniques. The system is also further supplemented by the integration of urgency detection which allows prioritization of distress messages that are critical. Nevertheless, there are still certain constraints. Urgency prediction task is complex in its nature since urgency usually relies on the minor linguistic clues and context. Also, the existing system concentrates more on textual information and tweets in the English language. The next steps to follow might involve using multilingual datasets, increasing the size of the dataset, and using multimodal data, e.g., images or videos uploaded at times of disasters.

V. Conclusion and Future Scope

In the presented research, an AI-based crisis distress classification system was proposed, which is aimed at processing social media messages related to disasters automatically and identifying the signs of severe distress. The transformer-based Natural Language Processing methods proposed use the unstructured and informal text that is prevalent on social media platforms. With the help of the BERTweet transformer model that is specifically adapted to the Twitter information, the system is able to capture the contextual relationships in brief social media messages. System has been built based on the Kaggle Disaster Tweets dataset. The transformer model produced contextual embeddings that were utilized to categorize tweets under disaster-specific profiles, and an urgency detection module based on logistic regression was applied to rank the messages that needed to be addressed urgently.

The experimental results showed high classification, and F1 scores were over 96 percent on several crisis categories and Hamming Loss was 0.003, which means that the misclassification is very low. The urgency prediction module recorded a total accuracy of about 84 and is therefore able to identify high-priority distress messages.

Additionally, the UiPath Robotic Process Automation allowed the automated processing of the results of model outputs, creating email alerts in response to identified crisis events. This integration brings out the possibility of implementing the system in automated monitoring processes in emergency response applications. Comprehensively, the suggested solution proves that a combination of the transformer-based NLP models and automation technologies can help to enhance the efficiency of the crisis monitoring system significantly. This capability to recognize disaster-relevant messages and prioritize urgent cases automatically can help the emergency response teams to make more timely and informed decisions in the case of a critical situation.

Further research and development will be directed to the expansion of the system to accommodate the multilingual disaster messages, addition of multimodal data i.e. images and videos, and incorporation of real time social media streamline to provide continuous monitoring of the crisis.

References

[1] A. D. P. Ariyanto et al., “Transformer-Based Semantic Role Labeling for Crisis Events,” IEEE, 2025.
[2] D. Kar et al., “Comparative Analysis of ML Models for Disaster Tweets,” IEEE, 2025.
[3] P. T. Puhazhendhi et al., “Real-Time Media Monitoring System,” IEEE, 2025.
[4] B. Nath et al., “Crisis and Urgency Classification Using TF-IDF and LSTM,” IEEE, 2025.
[5] N. Shanthi et al., “Enhanced Ensemble Framework for Disaster Tweets,” IEEE, 2024.
[6] I. Hussain et al., “Multi-Modal Deep Learning Framework for Disaster Response,” IEEE, 2024.
[7] D. M. Kamble et al., “Evaluation of Tweets for Disaster Management,” IEEE, 2025.
[8] P. Daga et al., “Sentiment Analysis of Hazardous Events,” IEEE, 2025.
[9] J. L. R. Julanta et al., “BERTopic-Based Disaster Tweet Clustering,” IEEE, 2024.
[10] S. Soni et al., “Rule-Based Information Extraction for Disaster Damage,” IEEE, 2024.
[11] K. S. Gill et al., “ML-Based Tweet Classification for Disaster Predic- tion,” IEEE, 2024.
[12] S. Kumawat et al., “Evaluation of ML Models for Disaster Tweets,” IEEE, 2024.
[13] M. S. Adhantoro et al., “BERT and Temporal Clustering for Event Detection,” IEEE, 2025.