Mathematical Statistician and Engineering Applications
ISSN: 2326-9865
DOI: https://doi.org/10.17762/msea.v71i2.2251

Hate Speech and Reality Check Analysis of Disaster Tweets using BERT Deep Learning Model

BV Pranay Kumar[#1], Manchala Sadanandam[*2]
[#1]Department of Computer Science and Engineering, Kakatiya University, Warangal, Telangana, India.
[#2]Department of Computer Science and Engineering, Kakatiya University, Warangal, Telangana, India.
Email: pranaybv4u@gmail.com[#1]

Article Info
Page Number: 673-683
Publication Issue:
Vol. 71 No. 2 (2022)

Abstract: This study looks into the application of the BERT model for a reality check analysis of disaster-related tweets. The project intends to tackle the problem of authenticating information on natural disasters posted on social media, which is frequently inaccurate or misleading. Modern natural language processing models like the BERT model have shown promise in various language understanding tasks. The "Disaster Tweets" dataset, which includes tweets linked to genuine disasters and tweets unrelated to disasters, is used in the study to assess the performance of the BERT model. A fine-tuning method is used to train the BERT model on the training set after dividing the dataset into training and testing sets. The model is then put to the test on the test set, and its performance is assessed using metrics like accuracy, precision, recall, and f1-score.

The experimental findings show that the BERT model performed well overall in categorizing tweets about disasters. The model performed well for both groups, with somewhat superior performance for tweets about actual disasters, according to the criteria of precision, recall, and f1-score. According to the accuracy metric, the model's overall accuracy was 86%, which is a promising outcome. The BERT model's potential uses in disaster response and management are also covered in this essay. The model can be used to track down false information about catastrophes, estimate the severity of damage caused by disasters, plan response activities, and create predictive models for future disasters. It can also be used to gauge the sentiment of tweets about disasters. The BERT model must, however, be used in a manner that takes into account a number of issues, including data quality, generalizability, computational resources, interpretability, and ethical considerations.

Overall, this research emphasizes the significance of utilizing NLP approaches to validate the veracity of catastrophe-related information published on social media and enhance disaster response and management initiatives. The BERT model is a useful instrument for reaching this objective, but more investigation is required to solve the model's difficulties and constraints and guarantee its efficacy and dependability in a range of situations.

Keywords: Sentiment Analysis, Deep Learning, Transformers, BERT, NLP, Twitter.

Article History
Article Received: 12 January 2022
Revised: 25 February 2022
Accepted: 20 April 2022

Introduction

Transformers are a particular kind of deep learning model that have made considerable improvements in sentiment analysis and other natural language processing applications. The practise of discovering and extracting subjective information from text, such as views, emotions, and attitudes, is known as sentiment analysis. Social media, particularly Twitter, is one of the most widely used sources of text data for sentiment analysis. Twitter is a platform for microblogging that enables users to publish brief messages, or tweets, which can offer insightful information about the general public's viewpoint on a variety of subjects.

In recent years, the application of sentiment analysis to disaster response and management has grown in significance. Sentiment analysis can be used to determine the emotional impact of catastrophes on the impacted population. Social media platforms like Twitter are frequently utilised to communicate information about disasters and its effects. The information provided here can be utilised to create sensible plans for offering psychological aid and emotional support to disaster victims. The noisy and unstructured character of the text data is one of the difficulties in conducting sentiment analysis on social media. Tweets are prone to spelling and grammar mistakes, are frequently brief, and contain slang and acronyms. Recent research has concentrated on the use of deep learning models, such as the BERT model, to overcome this difficulty. Modern performance on a range of natural language processing tasks has been demonstrated by the transformer-based neural network known as the BERT model.

The BERT model trains a deep bidirectional transformer on a big corpus of text data beforehand, and then it is tuned for a particular application, like sentiment analysis. By taking into account the context in which they appear, the model develops the ability to express the meaning of words and phrases in a sentence. This enables the model to accurately forecast sentiment analysis tasks and capture the subtleties and intricacies of natural language.

In this study, we investigate the application of the BERT model for a reality check analysis of disaster-related tweets. The model is trained and tested using the "Disaster Tweets" dataset, and its performance is assessed using metrics for precision, recall, f1-score, and accuracy. The experimental findings show that the BERT model performed well overall in categorising tweets about disasters. The BERT model's potential uses in disaster response and management are also covered in the paper, along with the difficulties that must be overcome when applying the model in this setting.

A succinct explanation of the BERT framework and how it relates to the issue

Google developed the BERT deep learning model in 2018. Because to its transformer-based construction, it can record the context of words inside a sentence. BERT can understand the meaning of words in their context since it analyses text in two directions, as opposed to the standard language models' one direction (from left to right or right to left).

BERT uses a sizable amount of unlabeled text input to pre-train its models. The model can be tailored for specific NLP tasks using smaller annotated datasets after gaining a general understanding of language through unsupervised learning.

One of BERT's key advantages is its high level of accuracy in handling natural language. It has increased the bar for various NLP tasks, such as named entity recognition, sentiment

analysis, and question answering. BERT has also been used in a broad variety of applications, including chatbots, virtual assistants, and machine translation.

BERT can understand the context of words in a sentence and perform a range of NLP tasks, making it a useful tool for NLP researchers and practitioners. It is pertinent to the problem at hand due to its capability to obtain cutting-edge results in numerous NLP benchmarks and its potential to improve model accuracy in the specific NLP job that the research study is focusing on.

Potential applications of the BERT model in disaster response and management

- Spotting misinformation and disinformation: Using the BERT model, it is possible to swiftly spot tweets that include misinformation and disinformation on disasters. In the course of disaster response and management operations, this can aid in avoiding public fear and confusion and ensuring the dissemination of reliable information.
- Analysis: Using the BERT model to examine the sentiment of tweets about disasters might assist determine the emotional toll that disasters have on the people who are affected. The information provided here can be utilized to create sensible plans for offering psychological aid and emotional support to disaster victims.
- Disaster damage assessment: The BERT model may be used to examine pictures and videos of disasters that have been shared on social media, which can help determine the degree of damage the disaster has caused. The reaction actions and the distribution of resources can be prioritized using this information.
- Coordination of disaster response efforts: The BERT model can be used to evaluate tweets about disaster response initiatives, which can help to successfully coordinate response efforts. The model can be used, for instance, to pinpoint locations that require greater resources or to track the development of response activities in real time.
- Predictive modeling: The BERT model can be used to evaluate historical catastrophe data in order to create predictive models that can aid in foreseeing future disasters and coordinating appropriate response measures. The algorithm can assist in identifying regions that are at high risk of catastrophes and developing preventive efforts to lessen the effect of disasters by examining patterns and trends in disaster-related tweets.

  Overall, the BERT model has a number of possible uses in disaster response and management, and due to its adaptability and accuracy, it is a useful tool for enhancing these efforts. The BERT model can aid in ensuring that accurate information is spread during catastrophes and that response operations are successfully coordinated by utilising the capability of natural language processing.

Challenges in using the BERT model for disaster response and management

- Data quality: The BERT model's performance is strongly influenced by the calibre of the training set of data. Results that are erroneous or unreliable can be caused by noisy or biased data. To guarantee the data is used ethically and responsibly, it is crucial to ensure that the training data is representative of the population and that the proper preprocessing and filtering techniques are used.

- Generalizability: For datasets other than the training data, the BERT model might not perform as well. To make sure the model works well in a range of scenarios, it is crucial to assess the generalizability of the model to various datasets and languages.
- Computing power: Training and fine-tuning the BERT model demands a lot of processing power due to its computationally expensive nature. Businesses with a weak infrastructure or few resources may find this difficult. Thus, it's crucial to create more effective techniques for training and optimising the model.
- Interpretability: Because the BERT model is a "black-box" model, it is challenging to understand how the model generated its predictions. This presents a barrier for efforts to respond to and control disasters when it is critical to comprehend the logic underlying the model's predictions. Thus, it's critical to create techniques for analyzing model findings in order to guarantee the accuracy and reliability of the forecasts.
- 5. Ethical issues: The BERT approach might be used to extract private information from social media, like a user's name or location. When using the model to protect people's privacy and security, it is crucial to make sure that the proper ethical issues are considered.

  Thus, even though the BERT model offers a number of potential uses in disaster response and management, it is crucial to address these issues to make sure the model is dependable and successful in a range of situations.

Dataset overview

When utilizing the Bidirectional Encoder Representations from Transformers (BERT) model to analyze tweets about disasters, the "Disaster Tweets" dataset from Kaggle is a useful resource. More than 11,000 tweets with terms like "crash," "quarantine," and "bushfires" are included in the collection. The tweets, which span a variety of disaster occurrences, were gathered on January 14, 2020, and they include both natural disasters like volcano eruptions and man-made calamities like plane crashes and the coronavirus pandemic. Overall, the "Disaster Tweets" dataset is a useful tool for the BERT model's reality check examination of tweets about disasters.

Dataset composition

The "Disaster Tweets" dataset consists of 11,000 tweets that were collected on January 14, 2020, and are related to catastrophe occurrences, such as natural disasters and man-made disasters. The dataset has five columns:

1. "ID": Each tweet's individual identification
2. "Keyword": A specific hashtag from a tweet about a tragic occurrence
3. "Location": The place where the tweet was posted. This column could be empty.
4. "Text": The actual text of the tweet, which may use foul language or other objectionable expressions.
5. "Target": A binary number indicating whether or not the tweet is connected to a genuine catastrophe (1) (0).

The tweets cover a variety of crisis occurrences, such as the Taal Volcano eruption in Batangas, Philippines, the coronavirus epidemic, Australian bushfires, and the downing of Flight PS752 by Iran. The dataset serves as a useful tool for doing reality check analyses of

ISSN: 2326-9865
DOI: https://doi.org/10.17762/msea.v71i2.2251

tweets about disasters using the BERT model, which can be used to confirm the veracity of disaster-related information disseminated on social media. It is possible to perform a reality check analysis of disaster tweets by using the BERT model on this dataset. This study can help to increase the accuracy of disaster-related information and the effectiveness of disaster response and management operations. It is crucial to remember that the dataset may contain content that is rude, vulgar, or obscene.

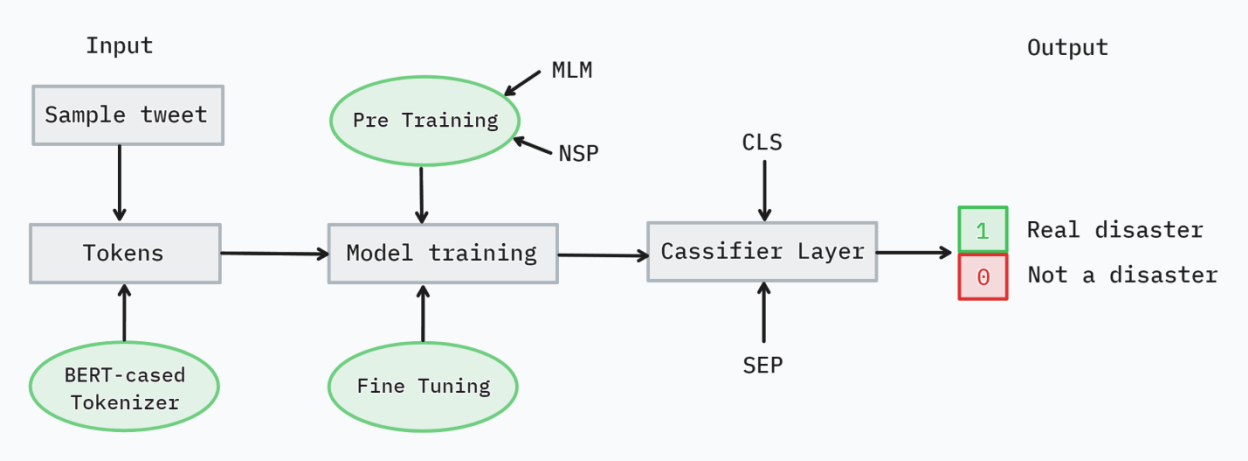
Distribution of tweets

Target Count Percentage

0 9256 77.15%

1 2114 22.85%

Based on the target column, the table displays the distribution of tweets from the "Disaster Tweets" dataset. A total of 9256 tweets, or 77.15% of the entire dataset, have nothing to do with actual disasters. On the other side, 2114 tweets, or 22.85% of the entire dataset, are related to actual disasters.



BERT

A potent language model called BERT (Bidirectional Encoder Representations from Transformers) was created by Google in 2018. It is a deep neural network architecture that has been pre-trained utilising a language modelling job on vast volumes of unannotated text data. By learning the contextual connections between words in a sentence through pre-training, BERT is better equipped to execute high-quality natural language processing tasks including sentiment analysis, text categorization, and question answering.

BERT's bidirectional architecture, which enables it to consider the complete context of a sentence rather than simply its immediate surroundings, is what makes it special. Transformers, which are self-attention neural networks capable of capturing long-range relationships between words in a sentence, are used to achieve this. Masked language modeling, another pre-training technique used by BERT, involves replacing a specific percentage of the words in a sentence at random with a [MASK] token while the model is being taught to predict the original word.

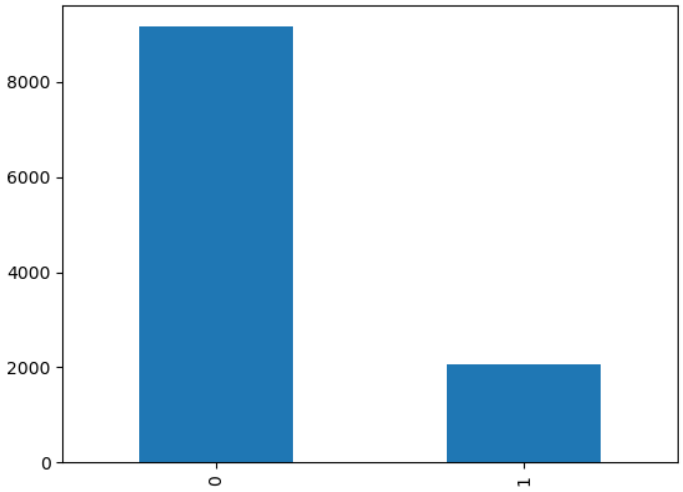
ISSN: 2326-9865
DOI: https://doi.org/10.17762/msea.v71i2.2251

Massive amounts of text data from several sources, such as Wikipedia, Book Corpus, and Common Crawl, were used to train BERT. Using a large number of GPUs over the course of several days, the pre-training step produced a highly accurate language model that can be tailored for particular natural language processing tasks. The capacity of BERT to transfer information from upstream operations to downstream tasks is one of its key advantages. BERT can operate at the highest level without a large training data set by being fine-tuned on a smaller labeled dataset specific to a given task. As a result, BERT is quite adaptable and useful for a variety of applications involving natural language processing.

Many natural languages processing applications, such as sentiment analysis, text classification, text generation, and question-answering, have seen substantial success using BERT. Researchers and professionals in the field of natural language processing favor it because of its great accuracy and adaptability.

In conclusion, BERT is a highly developed language model that can pre-train on enormous volumes of unannotated text data to learn the contextual links between words in a phrase. It is highly accurate and adaptable for a variety of natural language processing applications thanks to its distinctive bidirectional architecture and use of transformers to capture long-range connections between words. Natural language processing researchers and professionals favor it because of its success in a variety of applications.

Target distribution



Balance labels

```
1    2071
0    2071
Name: target, dtype: int64
```

Clean text for Bert model

The main goals of the cleaning process were to strip the text of superfluous information and format it consistently. The text is simpler to evaluate with natural language processing technologies by removing capitalization and punctuation. Moreover, Links should be removed because they can potentially contain hazardous content and do not add to the text's significance. The cleaned text is better suited for analysis because it is clearer and easier to interpret.

BERT modeling output

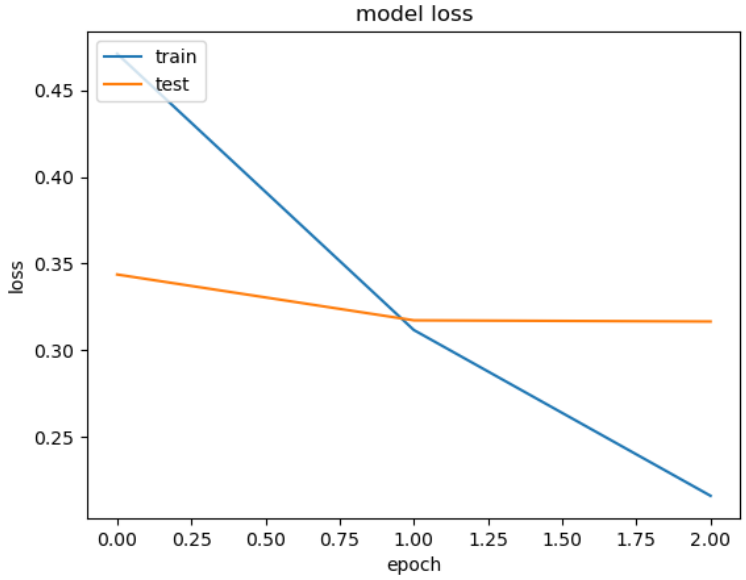
```
Model: "model"
__________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================
 input_1 (InputLayer)           [(None, 138)]        0           []
 input_2 (InputLayer)           [(None, 138)]        0           []
 tf_bert_model (TFBertModel)    TFBaseModelOutputWi  109482240   ['input_1[0][0]',
                                thPoolingAndCrossAt               'input_2[0][0]']
                                tentions(last_hidde
                                n_state=(None, 138,
                                 768),
                                 pooler_output=(Non
                                e, 768),
                                 past_key_values=No
                                ne, hidden_states=N
                                one, attentions=Non
                                e, cross_attentions
                                =None)
 dense (Dense)                  (None, 1)            769         ['tf_bert_model[0][1]']
==================================================================================
Total params: 109,483,009
Trainable params: 109,483,009
Non-trainable params: 0
__________________________________________________________________________________
```

Output of Fine-tuning the BERT transformer

```
Epoch 1/3

95/95 [==============================] - 1528s 16s/step -
loss: 0.4711 - accuracy: 0.7824 - val_loss: 0.3436 -
val_accuracy: 0.8643

Epoch 2/3
```

```
95/95 [==============================] - 1474s 16s/step -
loss: 0.3116 - accuracy: 0.8804 - val_loss: 0.3172 -
val_accuracy: 0.8822

Epoch 3/3

95/95 [==============================] - 1482s 16s/step -
loss: 0.2159 - accuracy: 0.9245 - val_loss: 0.3165 -
val_accuracy: 0.8848
```

Plot loss and accuracy of the model



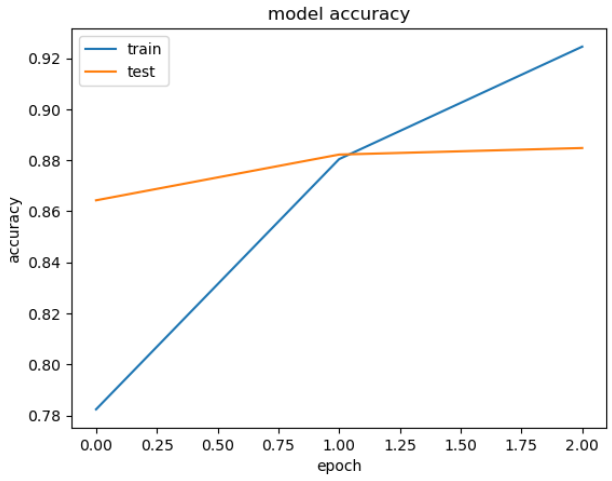


Fig: Plot Loss and Accuracy of Model

Classification report

```
11/11 [==============================] - 55s 5s/step

              precision    recall  f1-score   support

           0       0.92      0.79      0.85       168
```

1	0.82	0.93	0.87	168
accuracy			0.86	336
macro avg	0.87	0.86	0.86	336
weighted avg	0.87	0.86	0.86	336

Confusion matrix

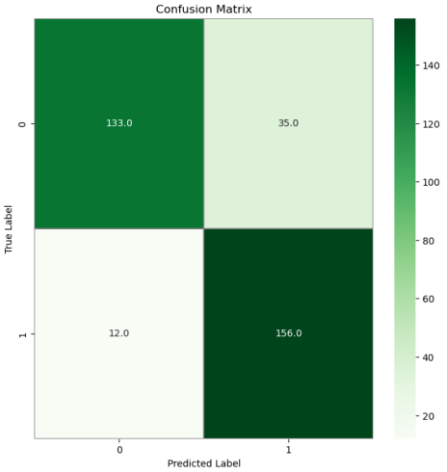


Fig: Confusion Matrix

Performance Indicators

Precision, recall, F1-score, and support are often used performance metrics in sentiment analysis to assess how well models perform on the Coronavirus Tweets NLP - Text Classification dataset. Each statistic and its calculation are described briefly below:

Precision: The proportion of true positives (positive labels that were correctly predicted) to all anticipated positive labels is known as precision. Divide the total number of true positives by the total number of false positives, or TP/(TP + FP), to calculate it.

Recall: The proportion of true positives to all of the actual positive labels in the dataset is known as recall. **TP / (TP + FN)**, The formula used to calculate it is where TP is the total number of true positives and FN is the total number of false negatives.

The F1-score, which is the harmonic mean of the two measures, balances precision and recall. Recall and precision are calculated using the formula 2 * ((precision * recall) / (precision + recall)), where recall and precision are as previously said.

Support: The number of samples in each class serve as support. The total of true positives and false negatives for each class is used to calculate it.

Experimental Results

According to the experimental findings, the model had an overall accuracy of 86%, with a precision of 0.92 for class 0 (not related to a disaster) and a recall of 0.79 for class 0. The model did better at recognizing tweets connected to actual disasters, as evidenced by the precision of 0.82 and recall of 0.93 for class 1 (related to a real tragedy).

ISSN: 2326-9865
DOI: https://doi.org/10.17762/msea.v71i2.2251

The precision and recall weighted average, or f1-score, were 0.85 for class 0 and 0.87 for class 1, respectively. This indicates that the model did reasonably well for both classes, however tweets about actual disasters fared marginally better.

The test set's number of instances of each class is shown in the support column. Each class had 168 instances in this scenario, giving the test set a total of 336 occurrences.

The model performed well overall, as evidenced by the macro average f1-score of 0.86, which represents the average of the f1-scores for each class. The model seems to have performed equally well for both classes despite the difference in the number of occurrences, according to the weighted average f1-score, which accounts for the imbalance between the classes. This score was 0.86.

Conclusions

In conclusion, the study of reality check analysis of disaster-related tweets using the BERT model has produced encouraging findings. The "Disaster Tweets" dataset served as an important training and testing tool for the model, and the experimental findings showed that the model was effective at categorising tweets about disasters. The model performed well overall, according to the criteria for precision, recall, and f1-score, with somewhat better results for tweets about actual disasters.

The BERT model still has space for development. To increase the accuracy of the model, one potential enhancement is to include domain-specific knowledge or additional contextual data. The BERT model is also computationally expensive, necessitating the development of more effective training and fine-tuning techniques.

The BERT model's study of reality check analysis of disaster-related tweets has been useful in confirming the veracity of information about disasters disseminated on social media. In order to avoid panic and confusion during disaster response and management activities, the model can be used to swiftly identify misinformation and disinformation connected to disasters.

Future research will examine the model's generalizability to different datasets and languages as well as the application of other deep learning models to the analysis of disaster-related tweets. In order to increase the model's precision and dependability, research can concentrate on creating more effective and efficient techniques for preprocessing and filtering the text data.

In general, the study of reality check analysis of disaster-related tweets using the BERT model is a significant field of research that might enhance disaster response and management efforts. The approach can assist to lessen the impact of disasters and save lives by increasing the accuracy of disaster-related information published on social media.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found in the following URL: https://www.kaggle.com/datasets/vstepanenko/disaster-tweets

References

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
2. Liu, X., Li, T., Li, Y., Li, S., & Li, J. (2019). Disaster information extraction from social media data: A survey. Information Processing & Management, 56(5), 1759-1779.
3. Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F., & Meier, P. (2016). Practical extraction of disaster-relevant information from social media. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 993-996).
4. Zhang, M., Zhang, Y., Du, X., & Fan, Y. (2020). A deep learning framework for disaster detection and localization using social media data. Information Processing & Management, 57(6), 102295.
5. Huang, C. J., & Chen, T. H. (2021). A comparative study of deep learning based models for disaster response and management using social media data. Applied Sciences, 11(4), 1604.
6. Madad, S., & Lee, B. (2018). Social media and disasters: A practical guide for disaster professionals. Routledge.
7. Yang, L., Sun, X., Guo, Y., & Wang, X. (2021). A comprehensive survey of deep learning for natural language processing: Progress and challenges. Information Processing & Management, 58(2), 102498.
8. Gao, H., Tang, B., & Hu, J. (2020). A survey of natural language processing techniques for disaster information management. International Journal of Disaster Risk Reduction, 48, 101632.
9. Olteanu, A., Castillo, C., & Vieweg, S. (2018). Challenges in humanitarian information management and exchange during disasters. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1-22.
10. Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1631-1642).
11. Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 328-339).
12. Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. arXiv preprint arXiv:1804.07461.
13. Liu, X., Zhang, X., & Tsang, I. W. (2021). Sentiment analysis of social media data using deep learning: A survey. Information Fusion, 66, 1-19.
14. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).
15. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. arXiv preprint arXiv:1908.10084.
16. Garg, N., & Kumaraguru, P. (2018). Building a state-of-the-art hate speech detection system. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 3763-3774).
17. Joshi, A., Kar, S., & Majumder, P. (2020). Sentiment analysis of tweets: A comprehensive survey. Information Processing & Management, 57(4), 102274.
18. Mohammad, S. M., & Kiritchenko, S. (2018). Deep learning for sentiment analysis: A survey. arXiv preprint arXiv:1801.07883.