

# Relevance Classification of Flood-related Twitter Posts via Multiple Transformers

Wisal Mukhtiar<sup>1,†</sup>, Waliya Rizwan<sup>1,†</sup>, Aneela Habib<sup>1,†</sup>, Yasir Saleem Afridi<sup>1</sup>,  
Laiq Hasan<sup>1</sup> and Kashif Ahmad<sup>2</sup>

<sup>1</sup>Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar, Pakistan.

<sup>2</sup>Department of Computer Science, Munsters Technological University, Cork, Ireland.

## Abstract

In recent years, social media has been widely explored as a potential source of communication and information in disasters and emergency situations. Several interesting works and case studies of disaster analytics exploring different aspects of natural disasters have been already conducted. Along with the great potential, disaster analytics comes with several challenges mainly due to the nature of social media content. In this paper, we explore one such challenge and propose a text classification framework to deal with Twitter noisy data. More specifically, we employed several transformers both individually and in combination, so as to differentiate between relevant and non-relevant Twitter posts, achieving the highest F1-score of 0.87.

## 1. Introduction

Natural disasters, which are hazardous events and occur frequently in different parts of the world, can have devastating effects on society. Depending on the severity of the disaster, it may result in significant damage to the infrastructure and human lives. Rapid response to natural disasters may help in mitigating their adverse impact on society. In disasters and emergency situations, access to relevant and timely information is key to a rapid and effective response. However, the literature reports several situations where access to relevant and timely information may not be possible due to several factors [1].

In recent years, social media outlets, such as Twitter, Facebook, and Instagram, have been explored as a source of communication and information dissemination in emergency situations [2]. The literature already reports the feasibility and effectiveness of social media for a diversified list of tasks in disaster analytics. For instance, Ahmad et al. [3] explored social media outlets as a source of information collection and dissemination during natural disasters by proposing a system that is able to collect and analyze disaster-related multimedia content from social media. Similarly, social media content has also been utilized for disaster severity and damage assessment [4, 5].

Despite being very effective in disaster analytics, social media data also come with several limitations. For instance, social media content contains a lot of noise and irrelevant information. This paper targets one of such challenges by proposing several solutions for the Relevance Classification of Twitter Posts (RCTP), sub-task introduced in DisasterMM challenge of MediaEval 2022 [6]. The task aims at automatically analyzing and classifying flood-related tweets into relevant and non-relevant tweets.

---

*MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online*


\*Corresponding author.

†These authors contributed equally.

✉ kashif.ahmad@mtu.ie (K. Ahmad)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Related Work

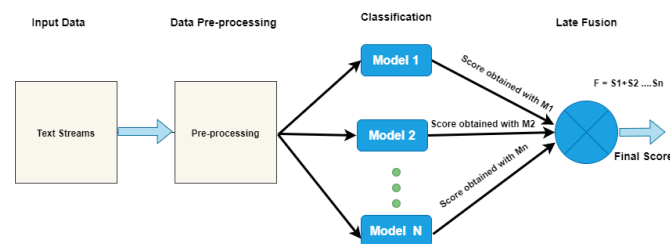
Disaster analysis in social media content has been one of the active topics of research in the domain over the last few years [2]. During this time, different aspects and applications of disaster analytics in social media content have been explored [7]. Some key applications include communication/information dissemination, damage assessment, response management, sentiment analysis, and identification of the needs of affected individuals. The literature already reports several interesting works on these applications. For instance, Nguyen et al. [8] utilized social media content for damage assessment by analyzing disaster-related visual media posts. Ahmad et al. [9] analyzed social media imagery for monitoring road conditions after floods. Moreover, a vast majority of the literature demonstrates how social media outlets can be used as means of communication in disasters and emergency situations [10, 1].

In the literature, different types of disasters including natural disasters, such as earthquakes, landslides, droughts, wildfires, and floods, as well as man-made disasters, such as accidents, have been explored [1, 11]. However, the majority of the works have targeted floods, being one of the most common natural disasters. The literature reports several interesting works on flood analysis in social media content for different tasks. For instance, Ahmad et al. [9] proposed a late fusion-based framework for the automatic detection of passable roads after a flood. For this purpose, several deep learning models are trained on flood-related images from social media. Alam et al. [4], on the other hand, employed social media imagery for post floods damage severity assessment.

Flood detection and analysis in social content have also been a part of the MediaEval benchmark initiative as a shared task for several years. Each time a separate aspect of flood analysis has been explored. For instance, in MediaEval 2017 the task aimed at the retrieval of flood-related images from social media. The task mainly involved analyzing the water level in different areas to differentiate between floods and regular water reservoirs, such as lakes [12]. In MediaEval 2018, the task was slightly modified by asking the participants to propose multi-modal classification frameworks for flood-related multimedia content [13]. In MediaEval 2019 and 2020, the tasks aimed at analyzing flood severity and flood events recognition in social media posts.

## 3. Approach

Figure 1 provides the block diagram of the proposed framework for the RCTP task. The framework is composed of three main components namely (i) Pre-processing, (ii) Training and Classification, and (iii) Fusion. In the first step, different pre-processing techniques are employed to clean the dataset. Three different transformers are then trained on the data to obtain classification scores. In the final step, the classification scores of the individual models are combined in a late fusion scheme. The details of these steps are provided below.



**Figure 1:** Block diagram of the proposed approach.

### 3.1. Pre-processing

In the pre-processing step, we employed different techniques for cleaning the dataset. More specifically, we removed unnecessary information, such as user names, URLs, emojis, punctuation marks, stop words, etc. Besides this, we also performed the necessary pre-processing tasks that are required to transform the raw text into a form that is suitable for the transformers. To achieve this, we used the TF.text library<sup>1</sup>.

### 3.2. Classification via Transformers

After cleaning and pre-processing the data, we trained three different models, namely BERT [14], RoBERTa [15], and XLNet [16]. The selection of these models for the task is motivated by their proven performance on similar tasks [17]. A brief overview of these models is provided below.

- **BERT:** Bidirectional Encoder Representations from Transformers (BERT) is one of the state-of-the-art NLP algorithms for text processing. The model is pre-trained on a large collection of unlabeled text and can be fine-tuned for different text-analysis applications. The key attributes of the model include its bi-directional nature, pre-training with Masked Language Modeling (MLM), and Next Structure Prediction (NSP) objectives. In the experiments with BERT, we used the Adam optimizer with a learning rate of 0.001 and a batch size of 8 for 3 epochs.
- **RoBERTa:** Robustly Optimized BERT is a modified version of the BERT model with an improved training mechanism. More specifically, in RoBERTa the NSP capabilities are removed. Moreover, dynamic masking is introduced. In addition, a larger batch size and a larger amount of training data were used in the training process. In this work, we used a learning rate of 0.001, batch size of 20, and 10 epochs during the fine-tuning of the model for the desired task.
- **XLNet:** XLNet is another state-of-the-art NLP algorithm. Similar to BERT, XLNet is also a bidirectional transformer and uses an improved training approach. In contrast to BERT and traditional NLP algorithms, XLNet relies on Permutation Language Modeling (PLM) by predicting all the tokens in random order. This allows XLNet to handle dependencies and bidirectional relationships in a better way. In this work, we used a learning rate of 0.002, a batch size of 32, and 4 epochs during the fine-tuning of the model for the desired task.

We obtained the results in the form of posterior probabilities from these models, which are then used in the fusion scheme to obtain the final predicted labels. The fusion method used in this work is described in the next section.

### 3.3. Fusion

Our fusion method is based on late fusion, where we combined the classification scores obtained with the individual models for the final classification decision as shown in Equ. 1. In the equation,  $S_{final}$  represents the final classification score while  $s_n$  is the score obtained with the  $n$ th model. We note that in the current implementation, we used a simple fusion method by treating all the models equally (i.e., simple aggregation of the individual scores).

$$S_{final} = S_1 + S_2 + s_3 + \dots + S_n \quad (1)$$

<sup>1</sup>[https://www.tensorflow.org/text/guide/bert\\_preprocessing\\_guide#text\\_preprocessing\\_with\\_tftext#](https://www.tensorflow.org/text/guide/bert_preprocessing_guide#text_preprocessing_with_tftext#)

## 4. Results and Analysis

Table 1 provides the experimental results of the proposed solutions on the development set. As can be seen in the table, overall better results are obtained with the BERT model, and surprisingly, a lower F1-score is observed for RoBERTa. In the future, we will further investigate the potential causes of the lower performance of RoBERTa by exploring different implementations and hyperparameter settings for it. As far as the performance of the fusion methods is concerned, overall better results are obtained with the pair of XLNet and BERT. One of the potential reasons for the lower performance of the fusion of all the models is the less accurate prediction of RoBERTa, as also evident from the performance of the individual models.

**Table 1**

Experimental results of the proposed solutions on the development set.

Method	F1-Score
BERT	0.94
RoBERTa	0.78
XLNet	0.93
Fusion 1 (RoBERTa, BERT, XLNet)	0.75
Fusion 2 (BERT, XLNet)	0.93
Fusion 3 (RoBERTa, XLNet)	0.92

Table 2 provides the official results of the proposed solutions on the test set. In total, three different runs were submitted. The first run is based on the fusion of all three models used in this work. The remaining two runs are based on the fusion of the models in pairs of two. In run 2, BERT and XLNet are combined while in run 3 RoBERTa and XLNet are jointly used. As can be seen in the table, better results are obtained for the fusion of the models in pairs of two where the best performing pair of two models obtained an improvement of 20% over the fusion of all three models.

**Table 2**

Experimental results of the proposed solutions on the test set.

Run	Precision	Recall	F1-Score
1 (Fusion of BERT, RoBERTa, XLNet)	0.6738	0.5431	0.6014
2 (Fusion of BERT and XLNet)	0.8044	0.6948	0.7456
3 (Fusion of RoBERTa and XLNet)	0.8977	0.8598	0.8784

## 5. Conclusions

In this paper, we presented our solutions for the RCTP subtask of DisasterMM challenge posted in MediaEval 2022. We proposed a late fusion framework incorporating several state-of-the-art transformers for the task. In the current implementation, all the models are treated equally by assigning them equal weights (i.e., 1). In the future, we aim to employ merit-based fusion methods to further improve the final classification score.

## References

- [1] K. Ahmad, K. Pogorelov, M. Riegler, N. Conci, P. Halvorsen, Social media and satellites, *Multimedia Tools and Applications* 78 (2019) 2837–2875.
- [2] N. Said, K. Ahmad, M. Riegler, K. Pogorelov, L. Hassan, N. Ahmad, N. Conci, Natural disasters detection in social media and satellite imagery: a survey, *Multimedia Tools and Applications* 78 (2019) 31267–31302.
- [3] K. Ahmad, M. Riegler, A. Riaz, N. Conci, D.-T. Dang-Nguyen, P. Halvorsen, The jord system: Linking sky and social multimedia data to natural disasters, in: *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, 2017, pp. 461–465.
- [4] F. Alam, M. Imran, F. Ofli, Image4act: Online social media image processing for disaster response, in: *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining* 2017, 2017, pp. 601–604.
- [5] F. Alam, F. Ofli, M. Imran, Crisismmmd: Multimodal twitter datasets from natural disasters, in: *Twelfth international AAAI conference on web and social media*, 2018.
- [6] S. Andreadis, A. Bozas, I. Gialampoukidis, A. Moumtzidou, R. Fiorin, F. Lombardo, T. Mavropoulos, D. Norbiato, S. Vrochidis, M. Ferri, I. Kompatsiaris, DisasterMM: Multimedia Analysis of Disaster-Related Social Media Data Task at MediaEval 2022, in: *Proceedings of the MediaEval 2022 Workshop, Bergen, Norway and Online*, 2023.
- [7] F. Ofli, M. Imran, F. Alam, Using artificial intelligence and social media for disaster response and management: an overview, *AI and Robotics in Disaster Studies* (2020) 63–81.
- [8] D. T. Nguyen, F. Ofli, M. Imran, P. Mitra, Damage assessment from social media imagery data during disasters, in: *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining* 2017, 2017, pp. 569–576.
- [9] K. Ahmad, K. Pogorelov, M. Riegler, O. Ostroukhova, P. Halvorsen, N. Conci, R. Dahyot, Automatic detection of passable roads after floods in remote sensed and social media data, *Signal Processing: Image Communication* 74 (2019) 110–118.
- [10] L. Palen, A. L. Hughes, Social media in disaster communication, *Handbook of disaster research* (2018) 497–518.
- [11] K. Ahmad, A. Sohail, N. Conci, F. De Natale, A comparative study of global and deep features for the analysis of user-generated natural disaster related images, in: *2018 IEEE 13th image, video, and multidimensional signal processing workshop (IVMSP)*, IEEE, 2018, pp. 1–5.
- [12] B. Bischke, P. Helber, C. Schulze, V. Srinivasan, A. Dengel, D. Borth, The multimedia satellite task at mediaeval 2017., in: *MediaEval*, 2017.
- [13] B. Benjamin, H. Patrick, Z. Zhengyu, B. Damian, et al., The multimedia satellite task at mediaeval 2018: Emergency response for flooding events (2018).
- [14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).
- [15] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, *arXiv preprint arXiv:1907.11692* (2019).
- [16] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, *Advances in neural information processing systems* 32 (2019).
- [17] K. Ahmad, M. Ayub, J. Khan, N. Ahmad, A. Al-Fuqaha, Social media as an instant source of feedback on water quality, *IEEE Transactions on Technology and Society* (2022).