

# EXTRACTING INFORMATION FROM TWITTER DATA TO IDENTIFY TYPES OF ASSISTANCE FOR VICTIMS OF NATURAL DISASTERS: AN INDONESIAN CASE STUDY

**Deden Ade Nurdeni, Universitas Indonesia**  
**Indra Budi\*, Universitas Indonesia**  
**Ariana Yunita, Universitas Pertamina**

## ABSTRACT

*The number of persons exposed to disaster risk in Indonesia, with a total potential life of 255 million people, is shown in the disaster risk research by the Indonesian National Board for Disaster Management. Furthermore, calamities significantly impact Indonesia, such as loss of life, property loss, and public facilities damage. To limit risks, the response system is critical, especially during the emergency reaction time. However, assisting catastrophe victims is impeded by several factors, including delays in assisting, a lack of information about the victims' whereabouts, and uneven aid distribution.*

*To provide fast and trustworthy information, several information systems were developed by the Indonesian National Board for Disaster Management, such as DIBI, InAware, Geospatial, Petabencana.id, and InaRisk. The existing systems, on the other hand, do not show the disaster site in real time, nor do they tell what kind of assistance the victim requires at any particular time. To solve these challenges, this study develops a model that can categorize text data from Twitter linked to the type of support needed by disaster victims in real-time. The location of the actual victim is also displayed on a dashboard in the form of a map-based application.*

*This research employs text mining techniques to analyze Twitter data using a multi-label classification strategy using the Stanford NER method to extract geographical information. Naive Bayes, Support Vector Machine, and Logistic Regression using OneVsRest, Binary Relevance, Power-set Label, and Classifier Chain are the algorithms employed. N-Grams with TF-IDF weighting are used to represent text. With 82 percent precision, 70 percent recall, and 75 percent F1-score, this study's best model for multi-label classification is a combination of Support Vector Machine and Classifier Chain with UniGram+BiGram features. For location categorization, which is the input for geocoding algorithms, Stanford NER produces an F1-score of 83 percent. In a map-based dashboard, geocoding results in the form of spatial information is displayed. The practical implications of this study are showing the best model for extracting information from citizens' Twitter data what the suitable assistance for victims, and visualizing the real-time locations. This study can be helpful for several stakeholders, such as the government. The Indonesian government, particularly in natural disaster management, will become more efficient if the assistance delivered to victims is more accurate.*

**Keywords:** Natural disaster management; Twitter data; Text mining; Multi label classification; Geocoding.

## INTRODUCTION

Geographically, Indonesia is located at the confluence of four tectonic plates, namely the Asian Continent, the Australian Continent, the Indian Ocean plate, and the Pacific Ocean. The southern and eastern parts of Indonesia have a volcanic belt in volcanic mountains and lowlands. Indonesia also has two seasons, namely hot and rainy seasons, with extreme changes in weather, temperature, and wind direction. The condition of Indonesia is very potential and prone to natural disasters such as volcanic eruptions, earthquakes, tsunamis, floods, landslides, forest fires, and droughts (BNPB, 2017).

According to the findings of a disaster risk study conducted in Indonesia in 2015 by the Indonesian National Board for Disaster Management, the number of people exposed to multi-threat disaster risk is dispersed across the country, with a total potential of more than 255 million people exposed to moderate to high disasters, and the value of affected assets totaling more than Rp. 650 trillion (BNPB, 2016). Disasters have a huge impact on Indonesia, because compared to the country's overall population of roughly 270 million people, about 95 percent of the Indonesian population is at danger of being affected by disasters. As a result, disaster management breakthroughs are required, particularly during the emergency response period, to reduce the dangers that may materialize.

In natural disaster management, the role of community activity is critical (Simon et al., 2015). One of the roles of the community in disaster management is the dissemination of information during the emergency response period related to information on disaster events, evacuation needs, and information on basic needs assistance (Sakurai & Murayama, 2019). According to an internet survey report conducted by the Indonesian Internet Service Providers Association (APJII), from 2019 to 2020, there was an increase in internet usage in Indonesia, with 73.7 percent of the Indonesian population using it, and approximately 51.5 percent of internet users using it for social media activities (APJII, 2020).

In many cases, the first information about disaster events and emergency response situations is now available on social media channels, owing to the ease of access and speed with which information is disseminated (Kim & Hastak, 2018). Twitter is one of the popular social media platforms that is able to provide real-time information. Not only for advertising purposes, but the public also uses Twitter to share ideas, information, and opinions (Garcia & Berton, 2021). In addition, people are increasingly using social media platforms such as Twitter and Facebook during disasters and emergencies to post updates on situations, including reports of injured or dead people, infrastructure damage, requests for urgent needs, and the like (Imran et al., 2020). Social media, especially Twitter, is considered a potential source of information for further observation and action on requests for assistance for the basic needs of disaster-affected communities (Vieweg et al., 2014).

Although social media platforms, particularly Twitter, offer numerous benefits and simplicity in exchanging information, processing Twitter text data (tweets) is challenging due to the lack of formal language in tweets, both in writing and in the usage of formal terminology (Imran et al., 2020). As a result, a Twitter data processing method is required to obtain a classification of important information that can be used by related parties to take actions and make decisions (Singh & Kumari, 2016).

The objectives of this study are to build a system model that is able to classify Twitter social media text data into the types of assistance requested by disaster victims during the

emergency response period, as well as the victims' location. Furthermore, based on the previously built classification model and location of the victims create a data visualization or dashboard for quick response actions to help disaster victims. Several research questions that guide this study are:

*RQ 1: What is the best model for classifying the types of aid in disaster-related tweets? RQ 1a: Which features have the greatest influence on the classification process?*

*RQ 1b: Which multi-label transformation method is the most influential?*

*RQ 2: How can disaster-related Twitter data be used to determine the location of disaster victims in need of assistance?*

*RQ 3: To what extent can classification and location determination results be presented?*

## RELATED WORKS

Extracting essential information with text classification on Twitter is very popular for researchers in the field of text mining because tweets are open and easily accessible without privacy restrictions (D'Andrea et al., 2019). Several previous studies have extracted important information on Twitter data related to disasters using text mining techniques (Benitez et al., 2018; Kabir & Madria, 2019; Devaraj et al., 2020; Loynes et al., 2020; Zhang et al., 2021). The results of these studies are the most optimal classification model based on experiments for the classification of tweet text in several categories or labels. The classification model is built using machine learning algorithms that are tested in training data and testing data.

Kabir and Madria (2019) aims to classify Twitter data related to the Hurricane Harvey and Hurricane Irma disasters into the types of assistance needed and extract information on disaster victims' location. Furthermore, a victim rescue scheduling system was also developed based on the location data of the victims for reliable disaster management. They compare classification models from machine learning algorithms such as Support Vector Machine (SVM) and Logistic Regression (LR) with deep learning techniques, namely Convolution Neural Network (CNN), for the classification of the types of needs of disaster victims. This study shows that the CNN technique has the most optimal performance compared to other algorithms.

Similarly, Benitez et al. (2019) and Devaraj et al. (2020) also classified Tweets. Benitez et al. (2019) analyzed Twitter data related to disasters for the classification of relevance, types of disasters, and types of disaster information using Naive Bayes (NB) and Genetic Algorithm (GA) models. Another study, Devaraj et al. (2020), classified tweets related to the 2017 Hurricane Harvey disaster in the US using a model built from the NB, Decision Tree (DT), AdaBoost, SVM, LR, Ridge Regression, and CNN algorithms. This study resulted in a model with SVM and CNN that produced the best performance with an F1-Score evaluation value above 0.86.

On the other hand, Loynes et al. (2020) not only analyzed social media data, especially Twitter, but also detected tweets that inform disaster events along with their locations. They detected the location of the closest Non-Governmental Organizations (NGO) communities to the disaster. The Support Vector Classifier (SCV) is used to recognize man-made and natural catastrophes, while Density-Based Spatial Clustering of Application with Noise (DBSCAN) is used to determine the disaster's geographic location. The model used in this study was built from the Random Forest (RF), LR, NB, and SVM algorithms. The SVM algorithm has the highest

performance with an accuracy evaluation value of 0.992.

Differently, Zhang et al. (2021) used topic modeling techniques to analyze social media data related to the need for humanitarian supplies for disaster victims. The analysis of Twitter data relating to the 2013 Typhoon Haiyan disaster in the Philippines is the topic of this study. Topic modeling techniques were used to create a flexible vocabulary of types of requests, which were then linked with information on the victims' locations.

There is a wide range of literature on social media analysis for natural disaster management. However, studies focusing specifically on the Indonesian context are limited. Furthermore, rather than focusing on the real-time location, most academics focus on establishing a model for classifying Twitter content in general. This study aims to fill this gap by integrating prior research findings to the context of Indonesia, classifying Tweets into sorts of aids for victims and developing a prototype to indicate the victims' real-time position using NER and geocoding to extract the area.

## RESEARCH METHODOLOGY

There are eight main research phases: Tweet collecting, corpus annotation, pre-processing, text representation, classification of data relevance and sorts of aid for disaster victims, extraction of disaster victim location information, assessment and model selection, and visualization.

### Tweets Collection

The collection of tweets related to disasters used the Crawling technique by utilizing the Twitter API and the Tweepy library. The script for crawling was built using Python with the following query parameters “disaster OR flood OR earthquake OR tsunami OR landslide OR (volcanic eruption) OR eruption OR (cyclone) OR (rainstorm) OR (strong wind) OR (forest fire) OR (tidal wave) OR abrasion-filter: links-filter: images-filter: videos-is: retweet lang: in”. The tweet collection period is from January 15, 2021, to February 28, 2021; the selection of this period is because there have been several catastrophic events such as earthquakes, floods, landslides, cyclones, and volcanic events eruptions in Indonesia. Tweet data taken was tweet text data, user account data, and user location. Before the annotation process was carried out, three steps were conducted: 1) deletion of retweets, 2) initial pre-processing such as removing special characters, hashtags, and links to simplify the annotation process, and 3) eliminating duplication.

### Corpus Annotation

The corpus annotation process is the process of labeling tweet text data with a predetermined label category. It aims to classify the relevance of tweets (relevant and irrelevant) and directly to labeling the type of assistance needed (rescue and evacuation, temporary shelter, food assistance, non-food assistance, clothing, clean water and sanitation, and health services). The annotation technique was conducted by three annotators, two of whom work separately (Annotator A and Annotator B), and one person was responsible for the final justification. Two separate annotators from humanitarian social groups annotated this study, with the author serving as the final annotator. An annotation guideline exists to ensure that annotators have the same perspective of each label as determined previously. The annotation results from all the annotators were evaluated using Cohen's Kappa (Pérez et al., 2020).

## Pre-Processing

Case folding, document filtering, stop word removal, normalization, and stemming were some of the techniques utilized in pre-processing. The TALA stopword dataset (Tala, 2003) and the stop word corpus from the Indonesian NLTK Python library were used in this study.

Furthermore, normalization aims to change unofficial or non-standard words into standard words based on the Big Indonesian Dictionary. A normalization dictionary was being created by the authors. This research used the Python Sastrawi library (Pypi.org., 2021), which was created with Indonesian basic terms in view. All of the pre-processing techniques were done with Python scripts.

## Text Representation

This study used the Bag of Word feature with various forms of N-Grams combination with words parameter to evaluate the model's performance. The vector weighting uses the TF-IDF method. Results of this stage were the text features used in the classification process using machine learning algorithms. Feature testing was done by comparing several single features, namely Unigram, Bigram, and Trigram with combined features, namely Unigram+BiGram, UniGram+TriGram, and BiGram+TriGram.

## Classification of Data Relevance and Types of Assistance for Disaster Victims

Based on the text features generated in the previous stage, modeling was conducted using a supervised classification technique using the Naïve Bayes classification algorithm, Support Vector Machine, and Logistic Regression. The multi label data transformation method was also implemented to solve multi label classification problems. This stage produced a classification model that was used to predict and classify tweet text into one or more relevant or irrelevant labels, and the type of disaster assistance desired by disaster victims. The following is a list of categories of types of assistance: (1) Rescue and evacuation, (2) Temporary shelter, (3) Food assistance, (4) Non-food assistance, (5) Clothing, (6) Clean water and sanitation, and (7) Health services. Evaluation of training and testing data for this research using cross-validation with a value of  $k = 10$ , 10-fold cross-validation is widely used for classification problems (Goyal et al., 2018).

## Extraction of Disaster Victim Location Information

At this stage, the first step was to extract location information from the relevant labeled dataset using the Named Entity Recognizer (NER) method. The model has previously been trained using a research dataset (Alfina et al., 2017) which consisted of 599,600 tokens with class labels Person, Place, Organization, and O for other. Then to evaluate the model, this researcher uses the gold standard dataset generated in the study (Luthfi et al., 2014) which has 14,430 tokens with the labels Person, Place, Organization, and O. Furthermore, the results of the extraction of the NER location were converted into spatial location information data using the geocoding method. The result of this process was the spatial data of the relevant tweet locations, namely disaster victims who need assistance.

## Evaluation and Model Selection

The evaluation procedure was based on the models created in the first two stages of classification, as well as the findings of the preceding stage's location information extraction. The accuracy, precision, recall, and F1-Score evaluation metrics were used in this study as the measuring method. The model with the highest score performs better than the others and can be chosen and used in the following procedure.

## Visualization

A web-based dashboard was developed to provide information on where disaster victims might get help and what type of assistance they required. The classification model was used as an input, and the location extraction that has been chosen was tested using operational data. The Folium Library and OpenStreetMap tools were used to present the classification results, consisting of spatial data (latitude and longitude).

## RESULTS AND ANALYSES

### Addressing RQ1

Initially, data from crawling are 511.950. After deleting retweets and duplication, the data are 399.389. Furthermore, data were labeled and resulting in 9.859 data and corpus for the classification model. After labeling, there were 9.053 without label and 805 with minimal one label. The tweets without label mean that it is not relevant to the request for assistance for disaster victims, meanwhile, tweets with label indicate it is relevant.

The annotation results show that there are no tweets asking for non-food and sanitation assistance, so these two categories of disaster assistance can be omitted. In addition, if the relevant tweet does not mention one of the predefined types of assistance, then the tweet is included in the general type of assistance category. Table 1 describes the distribution of labeled data, which consists of the relevance and distribution of the type of disaster assistance requested.

The dataset is separated into two datasets for experimental purposes based on the five pre-processing processes, as follows: (1) Dataset 1, which consists of raw data that has been pre-processed to the stemming stage or without deleting stop words, and (2) Dataset 2, which consists of raw data that has been pre-processed till stop words have been removed. A multi-label classification approach is used in both of these datasets to determine what forms of disaster help are required by catastrophe victims on Twitter.

<b>Label</b>	<b>Number Of Tweets</b>
Relevant	805
General	365
Evacuation	338
Food	182
Clothes	81
Health	27
Shelter	10

The multi-label classification in these two datasets compares the text extraction features, namely Unigram, Bigram, Trigram, UniBigram, UniTrigram, and BiTrigram with the implementation of four multi-label transformation methods (BR, LP, and CC) in each classification algorithm (NB, SVM, and LR). After all, the combination of the dataset, classification algorithms, multi-label data transformation, and feature extraction was conducted. The results can be seen in Table 2 and Table 3 for Dataset 1 and Dataset 2 respectively.

**TABLE 2**  
**DATASET 1 EVALUATION RESULTS**

Algorithm/ Transformation		Dataset 1																	
		Fitur N-Grams Word Vectorizer, TF-IDF, Cross-Validation k = 10																	
		UniGram			BiGram			TriGram			UniBiGram			UniTriGram			BiTriGram		
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
NB	BR	0.567	0.083	0.142	0.651	0.222	0.318	0.650	0.086	0.150	0.626	0.258	0.356	0.685	0.274	0.372	0.675	0.262	0.360
	LP	0.373	0.014	0.026	0.702	0.084	0.147	0.579	0.020	0.038	0.723	0.103	0.176	0.755	0.123	0.207	0.686	0.119	0.200
	CC	0.643	0.101	0.172	0.711	0.267	0.377	0.720	0.108	0.183	0.768	0.314	0.425	0.772	0.335	0.445	0.741	0.319	0.430
SVM	BR	0.807	0.636	0.703	0.805	0.495	0.597	0.689	0.314	0.423	0.826	0.630	0.703	0.826	0.633	0.705	0.801	0.491	0.595
	LP	0.836	0.595	0.685	0.798	0.462	0.570	0.695	0.292	0.406	0.846	0.594	0.682	0.848	0.590	0.681	0.793	0.458	0.567
	CC	0.798	0.692	0.737	0.784	0.536	0.622	0.662	0.335	0.439	0.820	0.704	0.751	0.820	0.701	0.749	0.766	0.534	0.620
LR	BR	0.809	0.412	0.527	0.770	0.214	0.322	0.470	0.025	0.047	0.809	0.372	0.489	0.809	0.372	0.489	0.781	0.209	0.315
	LP	0.782	0.418	0.532	0.748	0.232	0.351	0.731	0.044	0.082	0.756	0.379	0.496	0.769	0.385	0.502	0.765	0.234	0.354
	CC	0.821	0.512	0.616	0.791	0.289	0.416	0.672	0.041	0.076	0.835	0.478	0.590	0.833	0.474	0.587	0.784	0.289	0.416

**TABLE 3**  
**DATASET 2 EVALUATION RESULTS**

Algorithm/ Transformation		Dataset 2																	
		Fitur N-Grams Word Vectorizer, TF-IDF, Cross-Validation k = 10																	
		UniGram			BiGram			TriGram			UniBiGram			UniTriGram			BiTriGram		
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
NB	BR	0.592	0.105	0.174	0.600	0.152	0.240	0.666	0.056	0.102	0.610	0.225	0.322	0.660	0.247	0.347	0.631	0.182	0.276
	LP	0.442	0.017	0.033	0.651	0.060	0.108	0.577	0.014	0.028	0.654	0.089	0.153	0.716	0.115	0.194	0.726	0.095	0.165
	CC	0.638	0.126	0.208	0.694	0.171	0.268	0.657	0.059	0.107	0.761	0.266	0.374	0.765	0.286	0.396	0.783	0.215	0.323
SVM	BR	0.794	0.633	0.697	0.755	0.445	0.547	0.644	0.240	0.343	0.797	0.642	0.703	0.802	0.638	0.703	0.734	0.432	0.531
	LP	0.807	0.582	0.670	0.764	0.398	0.511	0.623	0.228	0.330	0.831	0.594	0.682	0.833	0.594	0.682	0.757	0.387	0.498
	CC	0.783	0.681	0.724	0.754	0.488	0.579	0.619	0.247	0.347	0.802	0.698	0.742	0.808	0.696	0.743	0.722	0.477	0.565
LR	BR	0.804	0.414	0.531	0.756	0.200	0.306	0.487	0.026	0.048	0.784	0.382	0.496	0.777	0.378	0.491	0.747	0.188	0.289
	LP	0.769	0.405	0.518	0.746	0.226	0.341	0.657	0.042	0.078	0.742	0.384	0.496	0.758	0.386	0.500	0.737	0.209	0.319
	CC	0.829	0.504	0.614	0.761	0.259	0.378	0.734	0.044	0.082	0.831	0.483	0.595	0.825	0.475	0.587	0.767	0.251	0.370

Table 2 and Table 3 above only show the value of P for precision, R for recall, and F for F1-score, while the accuracy value for each combination produces a value above 90%. This means that the model has succeeded in classifying tweets into relevant and irrelevant categories correctly in large numbers rather than misclassifying in all datasets. However, high accuracy does not guarantee that the model's performance is good because the characteristics of the research dataset is imbalance between tweets that have labels (relevant) and tweets that do not have labels (irrelevant).

The Naive Bayes algorithm produces the lowest F1-Score value in all datasets in the

Power-set (LP) transformation method. The lowest value is found in the combination of Dataset 1 and the UniGram feature, which is 0.026 or about 3%. F1-Score is the average proportional value of precision and recall. This is because the True Positive value is very small, meaning that the model does not classify correctly each positive label to a positive label.

The selection of the model with the best performance in this study does not take into account the value of accuracy, because it considers the characteristics of the imbalanced dataset. The calculation of the recall value is preferred in this study because it wants to reduce the occurrence of False Negatives. A good model is prioritized to be able to reduce the identification of relevant tweets and need help into irrelevant categories. If this happens, it is feared that disaster victims who need assistance will not be detected properly. The value of False Positive is not that much of a concern because, in the implementation of the tweet monitoring system related to disasters, evaluation and confirmation are carried out first by parties related to the validity of information before taking action or distributing aid. Based on the above considerations, the best model in this research experiment is the model in Dataset 1 with the Support Vector Machine (SVM) algorithm, the multi-label Classifier Chain (CC) transformation method, and a single feature combination, namely UniBiGram, because it has the highest recall value, namely 0.704 or 70% and F1-score 0.751 or 75%.

The Wilcoxon statistical test was conducted to answer what features and transformation methods have the most influence on the model. The evaluation of significant features was carried out by comparing the results of the F1-score of the features used in the experiment on a combination of dataset 1, the SVM algorithm, and the CC transformation method. The choice of this method combination is because it is a combination that provides the highest model performance. Meanwhile, the evaluation of the multi-label transformation method was carried out on dataset 1 with the SVM algorithm and on the UniBiGram feature.

	UniGram	BiGram	TriGram	UniTriGram	BiTriGram
UniBiGram	0.123	0.005	0.005	0.515	0.005

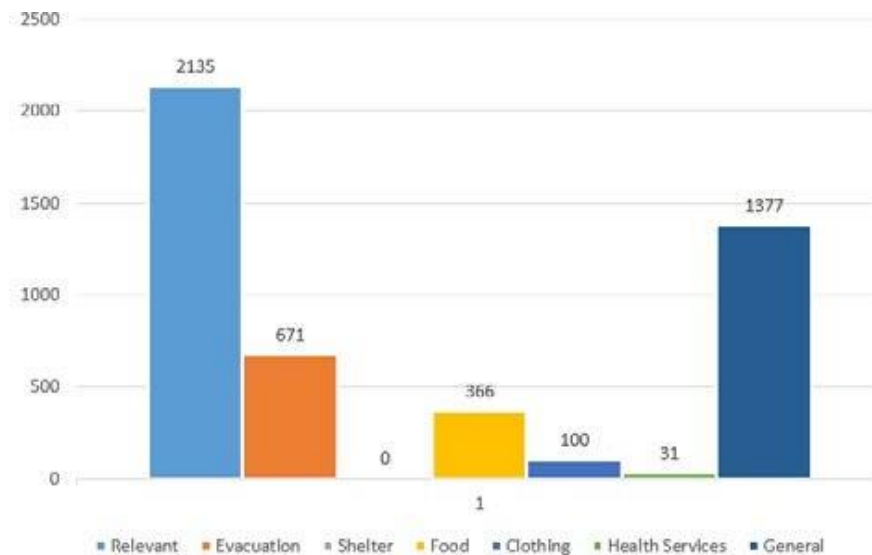
Table 4 shows that the UniBiGram feature has a p-value of less than 0.05 (5% significance level) for the BiGram, TriGram, and BiTriGram features, which is 0.005. However, it has a value above 0.05 for the UniGram and UniTriGram features which are 0.123 and 0.515, respectively. Results with a p-value of less than 0.05 mean that there is a statistical difference between the proposed approach and other approaches (Shi et al., 2021). Thus, it can be concluded that UniBiGram has an effect and significantly better performance than the BiGram, TriGram, and BiTriGram features, but there is no significant difference between the UniGram and UniTriGram features.

	Binary Relevance	Label Power-set
Classifier Chain	0.005	0.005



Table 5 provides evidence that the Classifier Chain transformation method has a significant difference in effect on other methods because it has a p-value below 0.05. Thus, Classifier Chain is the best multi-label transformation method in the classification model for the identification of the type of aid for disaster victims.

The label prediction stage is performed by the implementation of the best model for classifying datasets related to natural disasters in Indonesia. A dataset consisting of 399,389 tweets is predicted for the identification of relevant tweets about requests for assistance from disaster victims.



**FIGURE 1**

### **DISTRIBUTION OF NUMBER OF TWEETS BY LABEL TYPE**

Figure 1 shows the distribution of the number of tweets that have at least one label. The number of tweets that have relevance to requests for assistance from disaster victims is 2,135 tweets. The General relief type label of 1,377 tweets was detected by the model because many tweets did not mention any of the other disaster type labels. Evacuation labels consist of 671 tweets, 366 tweets for food labels, 100 tweets for clothing labels, and 31 tweets for health labels. The model cannot predict tweets with shelter labels at all, this can happen because the term or the word "penampungan" or "shelter" is very small compared to the number of other data terms that often appear. In addition, this is also caused by the model does not have enough training data in the iteration of the cross-validation process.

### **Addressing RQ2**

Developing a model for the extraction of location information using the Stanford NER method. The research dataset used for evaluation was tweet data with annotated results and relevant labels. The dataset with relevant labels related to requests for disaster assistance consists of 9,308 words terms that have been annotated to the label O for other, organization, person, and place by the author.

The evaluation results in Table 6 show that the organization's precision and recall values are zero, indicating that the model cannot correctly predict any of the organization labels. For the person and place labels, the model yields a precision value of 75% for the person label and 92% for the place label. This research focuses on determining the location and paying close attention to the entity place evaluation results. Table 6 shows the evaluation results of the NER model, which is quite good for place labels because it has a precision value of 92 %, recall value of 77 percent, and F1-score of 83 %. Furthermore, location data is extracted from the relevant labeled dataset using this model.

Entity	Precision	Recall	F1-score	TP	FP	FN
Organization	0.000	0.000	0.000	0	0	9
Person	0.750	0.300	0.429	3	1	7
Place	0.924	0.765	0.837	208	17	64

The implementation of geocoding generates geographic information sourced from Openstreetmap. Using the Geopy Library, Table 7 provides examples of geographic information results from tweets that have location information as a result of the NER method.

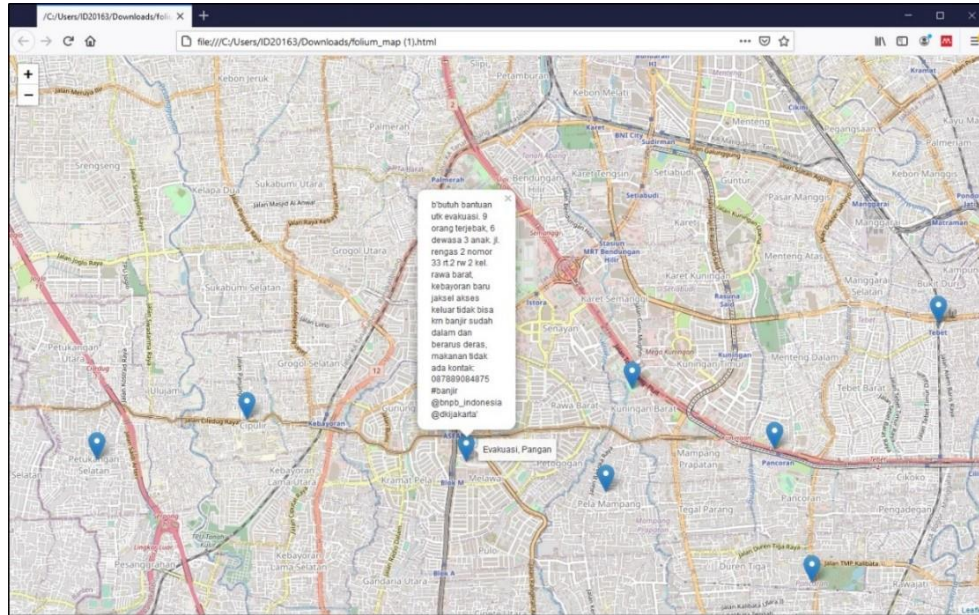
No	Tweet	Location Results using NER	Geocoding Results (Latitude, Longitude)
1	Butuh bantuan utk evakuasi. 9 orang terjebak, 6 dewasa 3 anak.\n\nJl. Rengas 2 Nomor 33 RT.2 RW 2 kel. Rawa barat, <Place>Kebayoran Baru</Place> Jaksel\n\nAkseskeluar tidak bisa krn banjir sudah dalam dan berarus deras,makanan tidak ada\n\nkontak: 087889084875\n\n#banjir @BNPB_Indonesia @DKIJakarta	KebayoranBaru	Kebayoran Baru, Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia (-6.24316375, 106.799850296448)
2	BUTUH PERAHU KARET AREA JABABEKA, <Place>Cikarang Baru</Place>. (Jalan Kasuari ) Kab <Place>Bekasi</Place>. Banjir masih belum surut. Untuk mobilisasi logistik (basic food supply etc )\n\n@ridwankamil @RadioElshinta @urbancikarang @QrJabar	Cikarang Baru; Bekasi	Jababeka Golf &Country Club, Cikarang baru, Jalan Haji UmarIsmail, Bekasi, Jawa Barat, 17581, Indonesia (-6.29884675, 107.175824716467)

Table 7 shows examples of geocoding interpretation results from the location information of a tweet. The geocoding process with the Geopy Library produces an interpretation of the location text based on the Openstreetmap along with spatial information, namely the latitude and longitude coordinates. The example of tweet number 1 generates Kebayoran Baru location information from NER results and then the geocoding process provides a complete location interpretation as Kebayoran Baru, South Jakarta, Special Capital Region of Jakarta, Indonesia and

with coordinates latitude -6.24316375, and longitude 106.79850296448.

### Addressing RQ3

The dashboard was developed with the Folium Framework and data sources from disaster-related tweets that are relevant in requesting assistance. The data used for dashboard development is the annotated dataset that have at least one category label for disaster aid. The tweet data source for this dashboard is filtered to include data with geographic information obtained through the NER process and geocoding.



**FIGURE 2**

### DISPLAY OF DISASTER ASSISTANCE TYPE INFORMATION AND TWEET TEXT

Figure 2 shows the location points in Kebayoran Baru showing information on what types of assistance are requested, along with the tweet text if the dashboard user clicks on the location point. At the location point in Kebayoran Baru, evacuation assistance and food are needed for flood disasters.

## CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

### Conclusions

This study develops a text classification model based on Twitter text data to identify the types of assistance required by victims of natural disasters. Tweet data is classified into multi-label information such as whether the tweet is relevant, whether it requires evacuation assistance, shelter assistance, food, clothing, health, or general assistance. The resulting model employs a multi-label classification strategy that combines classification algorithms, multi-label data transformation methods, and N-Grams feature extraction. This study also developed a model for identifying the

location of disaster victims in need of assistance based on tweet text that includes location information.

It was found that the model with the combination of the Support Vector Machine (SVM) classification algorithm, Classifier Chain and the combination of UniBiGram features produced the best performance than other models with an evaluation value of 82% precision, 70% recall and 75% F1-score using Dataset I, which consists of raw data that has been pre-processed to the stemming stage or without deleting stop words. This study also concludes that the best feature in the classification, namely UniBiGram, has a significant effect on the features of BiGram, TriGram, and BiTriGram. However, it does not provide a significant difference when compared to the UniGram and UniTriGram features. It can be concluded that the UniBiGram feature is the most influential feature when compared to the BiGram, TriGram, and BiTriGram features. Furthermore, the multi-label Classifier Chain transformation method has the most significant influence on the performance evaluation results of other methods. This is evidenced by the value of the Wilcoxon statistical test results which show the Classifier Chain has a p-value below 0.05 for the OneVsRest, Binary Relevance, and Label Power-set methods.

To determine the location of victims who need help this study produces a model for determining the location using the Stanford NER method. This model produces a precision of 92%, recall of 77%, and an F1-score of 83%. Furthermore, this research produces a prototype visualization of the results of tweet classification and location information extraction in the form of a simple web-based dashboard. The dashboard can display the location points of victims of natural disasters who need help with information on the type of assistance needed and details from tweets uploaded by Twitter users.

## Implications

The results of this study are a dashboard prototype that shows the location of disaster victims' requests for assistance, models, steps for modeling, and datasets that can be used for further research related to the same topic. Several stakeholders that potentially use this proposed system are

1. The government and disaster management parties, both central and local, to immediately take policies and take action quickly in the disaster management process. The government, in this case, the Indonesia National Board for Disaster Management, The National Search and Rescue Agency, or a combination of the Indonesian National Police and the Indonesian National Armed Forces can directly validate the information from this system, then can immediately establish a command post to carry out the process of evacuating victims if the source of the information is valid.

2. Non-Government Organizations (NGOs) or private humanitarian agencies can use this dashboard to start preparing programs for fundraising or logistics and volunteers to help victims of natural disasters quickly.

3. For the general public, this dashboard can be used as initial information to increase awareness and awareness of disaster events. Communities that were not affected by the disaster can seek initial assistance for disaster victims while waiting for the relevant authorities to do so. Furthermore, the public can provide detailed information about disaster victims to the larger community via social media to accelerate the dissemination of information on requests for assistance.

4. The media can use information from this dashboard as a starting point for their news

content. The mass media, both electronic and printed, can help disseminate information to the general public and the government about urging disaster victims to increase their speed in dealing with natural disasters, particularly in rescuing victims.

Research results can also be used as input for the development of a more complex and comprehensive disaster management information system. Starting from monitoring social media data to the process of communicating and disseminating information to related parties can be automatically carried out. This is to increase awareness and effectiveness of the disaster management process in Indonesia. It is important to note that the prediction model might still perform poorly, either False Negative or False Positive. As a result, a validation process is needed by relevant stakeholders to ensure that the disaster victim assistance request data is valid.

## Recommendations

Methods, research flows, models, and research datasets are expected to be further developed by further research. Recommendations for future research are, first, Twitter data collection related to disasters should include all types of natural disaster events in Indonesia to obtain comprehensive data. This study only takes Twitter data on the occurrence of several natural disasters such as floods, earthquakes, and volcanic eruptions. It needs to be developed for Twitter data related to other natural disasters such as tsunamis, landslides, hurricanes, and so forth.

To increase the model performance, further research can use the best parameter trials from implementing the chosen algorithm method. This study only uses the standard parameters of each algorithm for model experiments. It is recommended that the time consumed by a combination of methods in completing the classification process be measured to assess the model's efficiency. It is also necessary to balance the dataset, such as using the Synthetic Minority Oversampling Technique (SMOTE). This is to avoid biased accuracy values on model performance due to unbalanced data.

Another recommendation is related to the NER method. The NER method produces a large number of False Negatives. This is because the structure of the tweet writing abbreviates a lot of regional locations such as South Kalimantan, Jababeka, TangSel, and so forth. In addition, the use of lowercase letters at the beginning of words or all capital letters causes the NER model to be unable to detect. Therefore, it is recommended to develop a NER classifier that is able to overcome this problem in future research. Further research is expected to develop a more interactive dashboard and provide more detailed location information to the minor point, namely housing, neighborhood association, or village. The addition of details information would be important to extract accurate location and information, thus would be very important for natural disaster management.

## ACKNOWLEDGEMENT

This research was funded by Hibah Publikasi Terindeks Internasional (PUTI) Q2, Universitas Indonesia, grant number NKB-4060/UN2.RST/HKP.05.00/2020.

## REFERENCES

Alfina, I., Savitri, S., & Fanany, M. I. (2017). Modified DBpedia entities expansion for tagging automatically NER dataset. *Proceedings of 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 216-221.

- APJII. (2020). *Internet Survey Report*.
- Benitez, I. P., Sison, A. M., & Medina, R. P. (2018). Implementation of GA-Based Feature Selection in the Classification and Mapping of Disaster-Related Tweets. *Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval*, 1–6.
- BNPB. (2016). *Risiko Bencana Indonesia*.
- BNPB. (2017). *Potensi Ancaman Bencana - BNPB*. <https://bnpb.go.id/potensi-ancaman-bencana>
- D’Andrea, E., Ducange, P., Bechini, A., Renda, A., & Marcelloni, F. (2019). Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, 116, 209-226.
- Devaraj, A., Murthy, D., & Dontula, A. (2020). Machine-learning methods for identifying social media-based requests for urgent help during hurricanes. *International Journal of Disaster Risk Reduction*, 51, 101757.
- Garcia, K., & Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing*, 101, 107057.
- Goyal, A., Gupta, V., & Kumar, M. (2018). Recent Named Entity Recognition and Classification techniques: A systematic review. *Computer Science Review*, 29, 21-43).
- Imran, M., Ofli, F., Caragea, D., & Torralba, A. (2020). Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions. In *Information Processing and Management*, 57(5), 102261.
- Kabir, M. Y., & Madria, S. (2019). A deep learning approach for tweet classification and rescue scheduling for effective disaster management. *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 269-278.
- Kim, J., & Hastak, M. (2018). Social network analysis: Characteristics of online social networks after a disaster. *International Journal of Information Management*, 38(1), 86-96.
- Loynes, C., Ouenniche, J., & De Smedt, J. (2020). The detection and location estimation of disasters using Twitter and the identification of Non-Governmental Organisations using crowdsourcing. *Annals of Operations Research*, 1–33.
- Luthfi, A., Distiawan, B., & Manurung, R. (2014). Building an Indonesian named entity recognizer using Wikipedia and DBPedia. *2014 International Conference on Asian Language Processing (IALP)*, 19–22.
- Pérez, J., Díaz, J., Garcia-Martin, J., & Tabuenca, B. (2020). Systematic literature reviews in software engineering—enhancement of the study selection process using Cohen’s Kappa statistic. *Journal of Systems and Software*, 168, 110657.
- PyPI.org. (2021). *Sastrawi PyPI*. <https://pypi.org/project/Sastrawi/>
- Sakurai, M., & Murayama, Y. (2019). Information technologies and disaster management Benefits and issues. *Progress in Disaster Science*, 2, 100012.
- Shi, R., Xu, X., Li, J., & Li, Y. (2021). Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Applied Soft Computing*, 109, 107538.
- Simon, T., Goldberg, A., & Adini, B. (2015). Socializing in emergencies - A review of the use of social media in emergency situations. *International Journal of Information Management*, 35(5), 609–619.
- Singh, T., & Kumari, M. (2016). Role of Text Pre-processing in Twitter Sentiment Analysis. *Procedia Computer Science*, 89, 549–554.
- Tala, F. Z. (2003). The impact of stemming on information retrieval in Bahasa Indonesia. *Proc. CLIN, the Netherlands*, 2003.
- Vieweg, S., Castillo, C., & Imran, M. (2014). Integrating social media communications into the rapid assessment of sudden onset disasters. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8851, 444–461.
- Zhang, T., Shen, S., Cheng, C., Su, K., & Zhang, X. (2021). A topic model based framework for identifying the distribution of demand for relief supplies using social media data. *International Journal of Geographical Information Science*, 35(11), 2216-2237.

\* Corresponding Author: [indra@cs.ui.ac.id](mailto:indra@cs.ui.ac.id)