Data in Brief 61 (2025) 111688

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Annotated data for semantic role labeling of crisis events in Indonesian Tweets

Amelia Devi Putri Ariyanto[a,b], Diana Purwitasari[a,*], Bilqis Amaliah[a], Chastine Fatichah[a], Muhammad Ghifari Taqiuddin[a], Haikal[c]

[a] *Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia*
[b] *Information Technology and Systems Study Program, Widya Husada University, Semarang 50146, Indonesia*
[c] *Department of Public Health, Faculty of Health Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia*

ARTICLE INFO

Article history:
Received 17 January 2025
Revised 16 April 2025
Accepted 13 May 2025
Available online 18 May 2025

Dataset link: Semantic Role Labeling Datasets for Crisis Event (Original data)

Keywords:
Semantic role labeling
Named entity recognition
Crisis event
Twitter data
Low-resource languages
Indonesian Tweets

ABSTRACT

Social media platforms like Twitter provide essential real-time information about crisis events. Although the text data generated is rich, its vast volume and unstructured format make manual analysis challenging. Information extraction technologies such as Semantic Role Labeling (SRL) are needed to identify a sentence's semantic roles, such as who is the victim, what happened, when and where the event occurred, and what objects are affected in the crisis text to speed up and facilitate the emergency response process. However, the availability of public SRL datasets, especially for Indonesian, still considered a low-resource language, is very limited. We aim to develop an Indonesian-language SRL dataset based on Twitter text focusing on crisis events. This dataset includes entity labels for Named Entity Recognition (NER), another information extraction technique besides SRL. Text data was obtained through a crawling process on Twitter using specific keywords from 2018–2023, then preprocessed to obtain clean and relevant data for crisis events in Indonesia. The cleaned text data was then manually annotated by two experts based on guidelines designed to maintain consistency, resulting in 99,206 tokens labeled with SRL and NER. The high inter-annotator agreement value (Cohen's Kappa >0.90) indicates reliable data quality. This dataset is designed to support the

* Corresponding author.
E-mail address: diana@its.ac.id (D. Purwitasari).

https://doi.org/10.1016/j.dib.2025.111688

development of automated models for information extraction, such as SRL and NER. The results of this extraction will be used for disaster impact analysis, mapping affected areas, and planning for crisis mitigation. By providing this dataset, the research opens up new opportunities for developing Natural Language Processing (NLP) in Indonesian, especially in crisis event analysis.

Specifications Table

Subject	Computer Science, Artificial Intelligence, Natural Language Processing
Specific subject area	Information Extraction on Tweets, Semantic Role Labeling for Crisis Events
Type of data	Tables containing text (*.csv format)
Data collection	This dataset was collected using the Twitter API for crisis events texts in Indonesia, covering events such as floods, accidents, earthquakes, and fires. The search keywords used included "*kebakaran* (fire)", "*terbakar* (burning)", "*kecelakaan* (accident)", "*banjir* (flood)" and "*gempa bumi* (earthquake)", with a period from 2018 to 2023. After the data was collected, preprocessing was carried out using techniques such as case folding, removing symbols, normalization, removing duplicate tweets, tokenization, and geolocation filtering. After going through the geolocation filtering process, only tweet texts that were relevant to crisis events in Indonesia were included. The final result was around 4,150 processed tweet texts, totaling 99,206 tokens. This data was manually labeled for two main tasks: entities (Named Entity Recognition) and semantic role arguments (Semantic Role Labeling). The labeling process was carried out by two annotators, with the level of inter-annotator agreement (IAA) for both types of labels (entities and semantic role arguments) reaching a value above 0.90. This value indicates a high level of consistency between annotators, thus ensuring the quality and accuracy of the dataset to support various NLP analyses.
Data source location	Institution: Institut Teknologi Sepuluh Nopember City: Surabaya Country: Indonesia
Data accessibility	Repository name: Semantic Role Labeling Datasets for Crisis Event (Mendeley Data) Data identification number: 10.17632/r76v5sjyv2.2 Direct URL to data: https://data.mendeley.com/datasets/r76v5sjyv2/2
Related research article	none

1. Value of the Data

- This dataset is very valuable because it offers insights into crisis events (such as fires, accidents, floods, and earthquakes) in Indonesia extracted from Twitter text, which can show how people respond and provide information such as location reports, time, and descriptions of the impact of the crisis event.
- This dataset is equipped with argument labels for the Semantic Role Labeling (SRL) task and entity labels for the Named Entity Recognition (NER) task, which provide added value in crisis event analysis. NER labels are designed to recognize entities, while SRL labels add in-depth information by classifying semantic roles, which not only record the presence of victims but also reveal the relationships and impacts among entities involved in the event. Datasets that combine entity and argument annotations are still scarce, especially in Indonesian, making this dataset a valuable reference to advance NLP research in the context of crisis event analysis. Beyond advancing NLP research, this dataset also presents strong

potential to support interdisciplinary collaboration between NLP researchers and disaster management experts. For example, information extracted through SRL—such as the number of deaths (DEATHVICTIM-ARG), injuries (WOUNDVICTIM-ARG), or damaged infrastructure (AFFECTEDOBJECTS-ARG)—can serve as core indicators for assessing disaster severity or prioritizing emergency response. These extracted roles can be integrated into decision-support systems that help authorities perform rapid situation assessments, identify affected locations, and allocate resources more effectively. Such collaboration bridges the gap between text-based social media analysis and real-world crisis event management needs.
- This dataset is instrumental for researchers to build information extraction models based on Named Entity Recognition (NER) and Semantic Role Labeling (SRL), which can be used for further analysis such as determining the severity of a disaster, mapping affected areas, or supporting disaster mitigation efforts.

2. Background

Twitter is one of the social media platforms frequently used to share real-time information, particularly during crises such as natural disasters or accidents [1]. However, the high volume and rapid flow of information on Twitter make manual extraction inefficient. This challenge can be addressed using information extraction techniques such as the SRL task, which can identify semantic roles in a sentence, such as who is the victim, what happened, when and where the event occurred, and what objects were affected more quickly [2].

Despite its potential, publicly available SRL datasets are still limited—particularly for Indonesian, a language considered low-resource in NLP research [3]. In particular, there is a lack of SRL datasets derived from social media platforms like Twitter that focus on crisis-related content in the Indonesian context, which hinders the development of systems capable of processing informal, real-time data during emergencies. This limitation is especially critical given Indonesia's high exposure to natural and non-natural disasters. Natural events such as floods and earthquakes are common due to their tropical climate and tectonic location [4], while urban density and infrastructure issues contribute to frequent fires and accidents. These events often cause severe impacts on human life and the environment [5].

3. Data Description

Our public dataset is already stored in Mendeley data [6], including labeled arguments for SRL, consisting of 99,206 tokens. This dataset also has entity labels for NER, which contains 99,206 tokens. We have created version 2 of the dataset published in Mendeley by adding English translations for each data row. The main structure of version 2 of our dataset includes folders for raw, preprocessed, and labeled data, as shown in Figure 1, with detailed descriptions in the following subsections.

a. Raw Data (Folder Name: 1-Raw Data)

Raw data collection was carried out using search keywords such as "*kebakaran* (fire)", "*terbakar* (burning)", "*kecelakaan* (accident)", "*gempa bumi* (earthquake)", and "*banjir* (flood)" on social media Twitter (now referred to as X). The data collection period starts from January 2018 to December 2023. The tweet text scraping results are then stored in .csv files according to the type of crisis event, with the naming format "type_of_crisis_event-with-english-translation.csv", for example, "fire-with-english-translation.csv". Explanations of the columns for each .csv file in the raw data folder are detailed in Table 1. The total number of tweets successfully obtained is around 269,652 texts, with details for each crisis event accident (39,030 texts), fire (149,308 texts), earthquake (38,286 texts), and flood (43,028 texts).

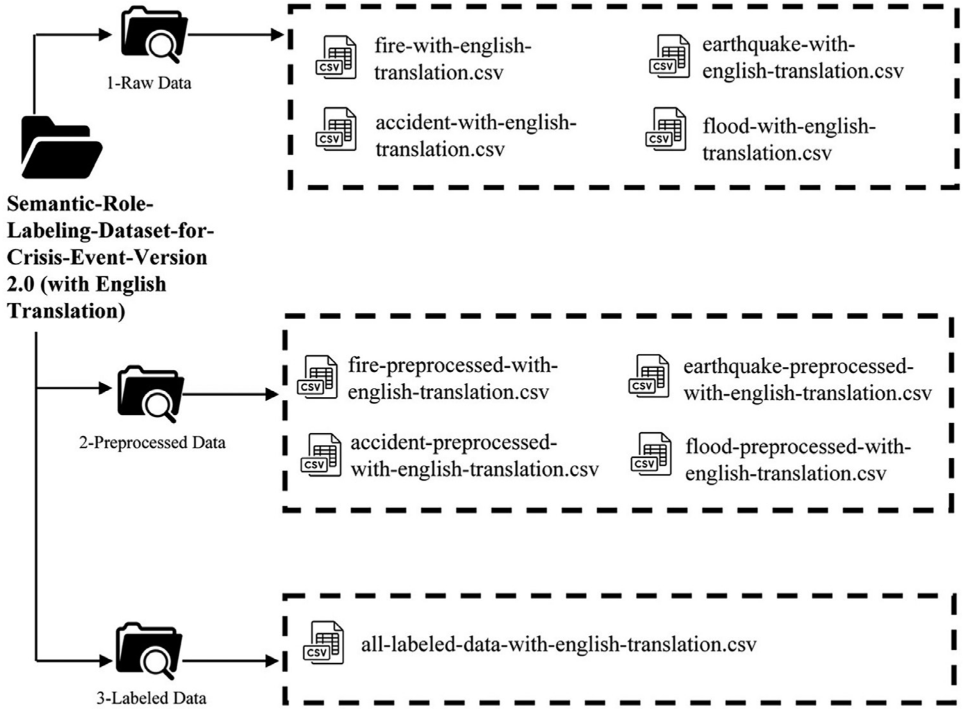

Fig. 1. Directory structure of semantic role labeling dataset for crisis event.

Table 1
Column details for each file in raw data and preprocessed data.

Column	Details
created_at	• Description: The time and date when the tweet was published. The time format follows Twitter's standard time format, which is a string that includes the year, month, day, and time. • Example: 2018-01-01 21:08:23+00:00
id_str	• Description: The unique ID value that Twitter assigns to each tweet. • Example: 947937572961112065
full_text_in_Indonesian	• Description: The complete text of the tweet posted by the user in Indonesian. • Example: "*imbas kecelakaan di jembatan Ciherang sebelum pom bensin…* (impact of the accident on the Ciherang bridge before the gas station…)"
full_text_in_English	• Description: The full text of the tweet written by the user in English. • Example: "impact of the accident on the Ciherang bridge before the gas station…"

b. Preprocessed Data (Folder Name: 2-Preprocessed Data)

This preprocessed data folder contains files derived from raw tweet data that have been cleaned and standardized through several preprocessing steps. First, all text was converted to lowercase (case folding) to ensure consistency in token representation, such as converting "*BANJIR* (FLOOD)" to "*banjir* (flood)." Various non-informative elements were then removed using regular expressions, including URLs, Twitter mentions (e.g., @username), hashtags (e.g., *#ban-*

Table 2
An example of raw text with the results after preprocessed data.

No	Raw Text (Indonesian)	Raw Text (English)	Preprocessed Text (Indonesian)	Preprocessed Text (English)
1	Gempa Bumi 6 4 SR di Banten Terasa Sampai Bandung https://t.co/jM7b1s4IBL #infoBDG https://t.co/OK9DHhgJZy	The 6.4 magnitude earthquake in Banten was felt as far away as Bandung https://t.co/jM7b1s4IBL #infoBDG https://t.co/OK9DHhgJZy	gempa bumi 6 4 sr di banten terasa sampai bandung	the 6.4 magnitude earthquake in banten was felt as far away as bandung
2	Pemotor Tewas Kecelakaan di Jalan Kapuk Raya Cengkareng #TauCepatTanpaBatas #BeritaTerkini #BeritaTerkini #NewsUpdate . https://t.co/zP722CuTUS	Motorcyclist dies in accident on Jalan Kapuk Raya Cengkareng #KnowFastWithoutLimits #NewsUpdate #NewsUpdate #NewsUpdate . https://t.co/zP722CuTUS	pemotor tewas kecelakaan di jalan kapuk raya cengkareng	motorcyclist dies in accident on jalan kapuk raya cengkareng
3	Kebakaran rumah di Jln Karet Pedurenan, Gang H. Sidik Rt 4/06, Kuningan, Jaksel dan api dalam proses pendinginan. @beritakebakaran #ListenToElshinta https://t.co/PVDHjsiAFW	House fire on Jln Karet Pedurenan, Gang H. Sidik Neighborhood unit 4/06, Kuningan, South Jakarta and the fire is in the process of cooling down. @beritakebakaran #ListenToElshinta https://t.co/PVDHjsiAFW	kebakaran rumah di jalan karet pedurenan gang h sidik rt 4/06 kuningan jaksel dan api dalam proses pendinginan	house fire on jalan karet pedurenan gang h sidik neighborhood unit 4/06 kuningan south Jakarta, and the fire was in the process of cooling
4	@infoGRESIK @e100ss @RTMCJatim banjir di perempatan kebomas arah lamongan https://t.co/iKR1tmIc9D	@infoGRESIK @e100ss @RTMCJatim flooding at kebomas intersection towards Lamongan https://t.co/iKR1tmIc9D	banjir di perempatan kebomas arah lamongan	flooding at the kebomas intersection towards lamongan

jir (#flood)), emojis, punctuation (commas, periods), newline characters, and redundant spaces. These steps helped reduce noise and standardize the structure of the text. Duplicate tweets were eliminated to avoid redundancy, and tweets that became empty or contained fewer than four words after cleaning were discarded, as they were unlikely to contain meaningful content for analysis.

Normalization was conducted using a dictionary-based method that replaced informal or abbreviated terms with their formal counterparts. This process utilized a publicly available Indonesian colloquial lexicon [7]. For instance, "*kec*" was replaced with "*kecamatan* (subdistrict)," "*jl*" with "*jalan* (street)," and "*prov*" with "*provinsi* (province)." Finally, tweets that did not mention crisis events occurring in Indonesia were excluded from the dataset to maintain the geographic relevance of the information. These preprocessing procedures ensured the resulting dataset was clean, consistent, and well-suited for annotation and further modeling tasks. Table 2 shows an example of raw text with the results after preprocessed data.

The preprocessing results are then saved in the form of a .csv file according to the type of crisis event with the naming format "type_of_crisis_event-preprocessed-with-english-translation.csv," such as "fire-preprocessed-with-english-translation.csv". The description of the columns for each .csv file in the preprocessed data folder is the same as the explanation in Table 1. The total preprocessed data is around 56,135 tweet texts, with details for each crisis event flood (7,321 texts), earthquake (7,408 texts), fire (33,251 texts), and accident (8,155 texts).

c. Labeled Data (Folder Name: 3-Labeled Data)

Several samples were taken from the preprocessed data to create a labeled dataset. The labeling process on these samples was carried out manually on approximately 4,150 tweet texts by two annotators. Both annotators are fluent native Indonesian speakers. One annotator is an expert in disaster management, and the other is a PhD student supervised by two experts in computer science. The decision to employ two annotators was based on combining domain-specific insight with technical annotation capability while ensuring manageable coordination and efficient iteration during the annotation process. Since SRL and NER work at the token level, the tokenization process was carried out before labeling. From the 4,150 tokenized tweets, a total of 99,206 tokens were generated, which needed to be labeled with arguments and entities by each annotator.

The final decision on each label was made based on an agreement between two annotators. When a disagreement occurred during labeling, an adjudication process was initiated. Both annotators revisited the data in question, exchanged interpretations concerning the established guidelines, and jointly determined the final label. For example, in the sentence "*gubernur tinjau korban banjir di simpanggambir warga setuju pindah dari bantaran sungai* (Governor inspects flood victims at Simpangambir, residents agree to move from riverbanks)", the word "*banjir* (flood)" was initially labeled as FALSE-EVENT by one annotator and FLOOD-EVENT by the other. Through adjudication, it was concluded that "*banjir* (flood)" referred to an actual flood that prompted a governmental response, including the relocation of residents. Since the event described was real and had observable consequences, the more appropriate label was determined to be FLOOD-EVENT rather than FALSE-EVENT, which is reserved for misleading or inaccurate information. To assess the consistency and objectivity of the annotation process, inter-annotator agreement (IAA) was calculated using Cohen's Kappa, a statistical measure specifically suited for evaluating agreement between two raters. Cohen's Kappa produces a score between -1 and 1, where a value of 1 indicates perfect agreement, 0 represents agreement equivalent to chance, and negative values suggest systematic disagreement [8]. The results showed high agreement: 0.9236 for argument labels and 0.9117 for entity labels. These values confirm that the annotation process resulted in highly consistent and reliable labeled data.

This labeled data folder only contains one .csv file that has been labeled with entities and arguments and consists of a combination of all types of crisis events named "all-labeled-data-with-english-translation.csv," not separate .csv files according to the type of crisis event as in the raw data or preprocessed data folder. The description of the columns in the .csv labeled data file is detailed in Table 3. The breakdown of the number of tokens in each crisis event is 28,780 tokens (equivalent to 1,250 tweets) for flood events, then 17,230 tokens (or equal to 866 tweets) for accident events, around 30,388 tokens (or same as 1,128 tweets) for fire events, and 22,808 tokens (or around 906 tweets) for earthquake events.

Our proposed dataset has a different labeling approach from the traditional SRL approach, which generally follows the Proposition Bank (PropBank) rules [9]. In the PropBank rules, semantic role labels are assigned using numbered arguments (ARG0-ARG5), where the meaning of labels ARG2 to ARG5 often varies depending on the predicate, which can cause ambiguity and confusion in determining the appropriate label [10]. In addition, the ARG0-ARG5 labeling is not specifically tailored for a particular domain, such as a crisis event, so it can less capture important domain-specific information, such as the number of victims or objects affected. To overcome this limitation, we use semantic role labels that are more specific to the crisis event, with fifteen SRL argument labels and detailed specifications for each label, as described in Table 4. This approach is designed to avoid ambiguity and ensure the relevance of the labels to the context of the crisis event. Thus, this dataset is more effective in supporting the analysis of critical information relevant to emergency response organizations.

The overall distribution of argument labels is shown in Table 5, which highlights the imbalanced proportion of labels in the dataset. A detailed breakdown of the argument label distribution for each type of crisis event is available in Table 6. Notably, the FALSE-EVENT label exhibits a spread distribution in all crisis events, emphasizing the importance of distinguishing between relevant and irrelevant contexts. Labels such as PLACE-ARG, TIME-ARG, and AFFECTEDOBJECTS-ARG also appear in all crisis events, indicating elements commonly used in crisis descriptions (place, time, affected objects). In one type of crisis event, there is a specific label according to its type of crisis event, such as the FLOOD-EVENT label, which is dominant in the flood crisis event type. However, some crisis events may contain labels that refer to other crisis events. This overlapping crisis event indicates that there are tweets that are relevant to more than one type

Table 3
Column details for each file in labeled data.

<table>
<tr><th>Column</th><th>Details</th></tr>
<tr><td>text_id</td><td>Description: A unique ID for each tweet in the dataset consisting of a crisis event type code and a sequence number.
Format:
A combination of crisis event code and number, such as BAN-00001.
Crisis event codes such as “BAN” for “banjir (flood)”, “KEC” for “kecelakaan (accident)”, “GEM” for “gempa bumi (earthquake)”, and “KEB” for “kebakaran (fire)”.
Example: Each tweet is assigned a sequence number, such as BAN-00001, the first tweet about flooding.</td></tr>
<tr><td>Id</td><td>Description: A token ID that indicates the sequence of tokens in a tweet based on text_id.
Format: Combination of text_id and token sequence number, separated by a dot (.), for example BAN-00001.002
Example: For tweet BAN-00001 with the sentence “13 janji akan mencari solusi banjir jakarta... (thirteen promises to find solutions to Jakarta floods...)”, each word will have a token ID like:
BAN-00001.001 for “13 (thirteen)”
BAN-00001.002 for “janji (promises)”
BAN-00001.003 for “akan (to)”
and so on.</td></tr>
<tr><td>token_in_Indonesian</td><td>Description: This column contains Indonesian tokens taken from the input text. Each row represents one token (word or punctuation mark) separated from the original text.
For example:
“13 (thirteen)”
“janji (promises)”
“akan (to)”</td></tr>
<tr><td>token_in_English</td><td>Description: This column contains the translation of the tokens in the token_in_Indonesian column into English
For example:
"thirteen"
“promises”
“to”</td></tr>
<tr><td>Entity</td><td>Description: Entity label that indicates the type of entity at the token level, especially for NER task.
Format: Labels use the BIO-tag scheme, where:
B- indicates the beginning of an entity.
indicates the continuation of the entity.
means the token is not part of an entity.
Example:
B-LOC for the first token of a place name. For example, from the phrase “North Jakarta”, the token “North” will be labeled B-LOC.
I-LOC for following tokens in a place name. For example, from the phrase “North Jakarta,” then the token “Jakarta” will be labeled I-LOC.</td></tr>
<tr><td>Argument</td><td>Description: This label indicates the specific function or argument of the token in the context of the event.
Format: Labels that represent specific roles in the context of the event and are given the suffix “-ARG“, such as DEATHVICTIM-ARG, WOUNDVICTIM-ARG, PLACE-ARG, CAUSE-ARG.
Example:
DEATHVICTIM-ARG for tokens representing deceased victims.
PLACE-ARG for tokens indicating the location of the incident.</td></tr>
</table>

Table 4
Details of label arguments for the semantic role labeling task and examples of each label.

Label Argument	Description	Example Phrase/Word	Original Tweet Text
DEATHVICTIM-ARG	Words or phrases indicating the presence of dead people	"*seorang kakek tewas* (a grandfather died)"	**In Indonesian:** "*terjebak di tengah kobaran api, seorang kakek tewas terbakar*" **Translated to English:** "trapped in the middle of the flames, a grandfather died in flames"
WOUNDVICTIM-ARG	Words or phrases indicating the presence of injured people	"*satu orang luka* (one person injured)"	**In Indonesian:** "*peristiwa kandang terbakar di sragen kembali terjadi: satu orang luka, api dari sisa pembakaran kotoran*" **Translated to English:** "the incident of a cage burning in Sragen happened again: one person injured, fire from remains of burning manure"
REASON-ARG	Reason for the occurrence of the incident	"*Sisa Pembakaran Kotoran* (Remains of Burning Manure)"	
AFFECTEDOBJECTS-ARG	Words or phrases indicating the presence of objects affected by the incident	"*Kandang* (Cage)"	
PLACE-ARG	Geopolitical location of a place, including villages, sub-districts, regencies, or cities	"*Sragen* (a city called Sragen)"	
O	Words or phrases that are not included in any semantic role	"*Kembali terjadi* (happened again)", and ":"	
TIME-ARG	Time of the incident	"*selasa (17/10)* (Tuesday (17/10))"	**In Indonesian:** "*sebuah rumah dan kendaraan roda empat di serang, banten, hangus terbakar, selasa (17/10). kebakaran diduga akibat bensin eceran tersulut puntung rokok*" **Translated to English:** "a house and four-wheeled vehicle were attacked, Banten, burned to the ground, Tuesday (17/10). The fire is suspected to have been caused by retail gasoline igniting a cigarette butt"
STREET-ARG	Name of the street where the incident occurred	"*jalan wr. supratman* (wr. supratman street)"	**In Indonesian:** "*toko bahan bangunan buana indah yang berlokasi di jalan wr. supratman, kelurahan penarukan, buleleng kebakaran*" **Translated to English:** "buana indah building materials store located on wr. supratman street, penarukan village, buleleng, fire"

(*continued on next page*)

Table 4 (*continued*)

Label Argument	Description	Example Phrase/Word	Original Tweet Text
INFORMATION-ARG	Additional information or details related to an incident	"*kerugian ratusan juta* (losses of hundreds of millions)"	**In Indonesian:** "*kebakaran di sibolga: 24 rumah terbakar, kerugian ratusan juta*" **Translated to English:** "fire in sibolga: 24 houses burned, losses of hundreds of millions"
OFFICER-ARG	People (officers, officials, individuals) or units directly involved in disaster management and rescue efforts	"*6 unit mobil pemadam kebakaran* (6 fire engines)"	**In Indonesian:** "*kebakaran satu unit rumah tinggal terjadi di kelapa gading, Jakarta utara, sebanyak 6 unit mobil pemadam kebakaran dikerahkan ke lokasi*" **Translated to English:** "fire of one residential unit occurred in kelapa gading, north jakarta, as many as six fire trucks were deployed to the location"
FALSE-EVENT	• Phrases or sentences that include words related to the four crisis events, fire, accident, earthquake, and flood, without describing the incidents. • This label indicates that the text only contains general statements, plans, hopes, personal stories, or promises that do not indicate actual events.	**• Example 1:** "*banjir* (flood)" **• Example 2:** "*kebakaran* (fire)"	**• Example 1:** **In Indonesian:** *"13 janji akan mencari solusi banjir Jakarta dan kemacetan Jakarta ketika jadi presiden bahkan statemen di 2014 akan menggunakan cara lain ..."* **Translated to English:** "13 promises to find solutions to Jakarta floods and Jakarta traffic jams when he became president even a statement in 2014 would use other methods..." **• Example 2:** **In Indonesian:** "*kmren pas ke kudus terheran2 krn ni pasar matahari udh lama kebakaran, tp blm di renov juga*" **Translated to English:** "yesterday, when I went to kudus, I was surprised because this matahari market had been on fire for a long time, but it had not been renovated yet"
EARTHQUAKE-EVENT	• Phrases or sentences that describe actual earthquake events. • This label indicates that an earthquake is	"*gempa bumi* (earthquake)"	**In Indonesian:** "*guncangan gempa bumi terasa hingga di seluruh kecamatan kubu sekitar pukul 8.12 malam*

(*continued on next page*)

Table 4 (*continued*)

Label Argument	Description	Example Phrase/Word	Original Tweet Text
	occurring or has just occurred, with a description that includes the scale of the earthquake, location, impact, or time of the incident.		*guncangannya sedikit lemah tapi dirasakan oleh warga senin 02/01/2018"* **Translated to English:** "earthquake tremors were felt throughout the kubu sub-district at around 8.12 pm. The tremors were weak but were felt by residents on Monday, 02/01/2018"
FLOOD-EVENT	• Phrases or sentences that describe actual flooding. • These labels are marked with information about flooding at a specific location, including the extent of the flooding, location of the event, damage caused, or impact on the community.	"*banjir* (flood)"	**In Indonesian:** "*banjir di wilayah tlogopojok perempatan petrokimian gresik"* **Translated to English:** "floods in the tlogopojojo area, petrochemical intersection, gresik"
ACCIDENT-EVENT	• Phrases or sentences describing actual accidents, such as traffic, work, or other accidents. • These labels indicate an accident and include details such as location, time, type of accident, and impact on victims or property.	"*kecelakaan* (accident)"	**In Indonesian:** "*telah terjadi kecelakaan antara dua sepeda motor di desa sumberjo kecamatan kandat"* **Translated to English:** "there was an accident between two motorbikes in sumberjo village, kandat sub-district"
FIRE-EVENT	• Phrases or sentences that describe actual fires. • These labels indicate that a fire has occurred, with related information such as location, impact of the fire, cause, or number of victims affected.	"*kebakaran* (fire)"	**In Indonesian:** "*kebakaran di kampung pulo jakarta timur, suami istri tewas"* **Translated to English:** "fire in kampung pulo east jakarta, husband and wife died"

of disaster. For instance, in the case of a flood crisis event, tweets may not only contain the label FLOOD-EVENT but also the label EARTHQUAKE-EVENT. An example tweet could be: "*habis diguncang gempa 5,3 SR pada pagi, sorenya diterjang banjir di kecamatan Kempo* (After experiencing a 5.3 magnitude earthquake in the morning, Kempo sub-district was hit by a flood in the afternoon)". In the context of accident crisis events, tweets may include the ACCIDENT-EVENT label and the FIRE-EVENT label. For example, a tweet read, "*tol jombang Mojokerto-update kecelakaan yang menyebabkan kendaraan terbakar* (Jombang Mojokerto toll road update: an accident has caused a vehicle fire)", indicates that the accident on the toll road has resulted in a fire originating from the victim's vehicle.

Table 5
Overall frequency distribution of label arguments for semantic role labeling task.

Label Argument	Frequency	Percentage (%)
O	68,753	69.303268
PLACE-ARG	10,229	10.310868
FALSE-EVENT	3448	3.475596
STREET-ARG	3154	3.179243
AFFECTEDOBJECTS-ARG	2789	2.811322
TIME-ARG	2424	2.443401
OFFICER-ARG	2113	2.129911
INFORMATION-ARG	1977	1.992823
FIRE-EVENT	834	0.840675
REASON-ARG	759	0.765075
ACCIDENT-EVENT	683	0.688466
DEATHVICTIM-ARG	632	0.637058
EARTHQUAKE-EVENT	579	0.583634
FLOOD-EVENT	470	0.473762
WOUNDVICTIM-ARG	362	0.364897

Table 6
Frequency distribution of label arguments for each type of crisis event for semantic role labeling task.

Label Argument	Frequency for Each Crisis Event			
	Flood	Accident	Fire	Earthquake
O	22,045	11,090	19,287	16,331
PLACE-ARG	2,894	1,236	3,848	2,251
FALSE-EVENT	1,125	321	551	1,451
STREET-ARG	205	1,537	1,396	16
AFFECTEDOBJECTS-ARG	260	799	1,639	91
TIME-ARG	354	331	1,218	521
OFFICER-ARG	527	138	897	551
INFORMATION-ARG	385	514	183	895
FIRE-EVENT	0	1	833	0
REASON-ARG	323	167	207	62
ACCIDENT-EVENT	0	683	0	0
DEATHVICTIM-ARG	93	254	235	50
EARTHQUAKE-EVENT	3	0	0	576
FLOOD-EVENT	467	0	3	0
WOUNDVICTIM-ARG	99	159	91	13

Visualization of the most frequently uttered words by the public on Twitter, depicted in a word cloud for each type of crisis event, as in Figs. 2 and 3 for the AFFECTEDOBJECTS-ARG label and the REASON-ARG label, respectively. The visualization was done by taking the top 500 Indonesian words that most frequently appear for each type of crisis event on specific argument labels such as AFFECTEDOBJECTS-ARG and REASON-ARG. Then, the 500 words were translated into English with the help of the Google library in Python programming. The purpose of taking the top 500 words was to shorten the translation time into English, and it was considered quite representative of public tweets about a type of crisis event.

Fig. 3 shows the word cloud for the REASON-ARG label for each type of crisis event. The causes for each type of crisis event vary. Words that appear larger in the word clouds indicate a higher frequency of occurrence, suggesting that these represent the primary cause of a crisis event. For example, in the type of crisis event accident, the words “driver”, “fall”, “experience” and “hit” appear in large sizes, indicating that accidents often involve drivers, falls, and collisions. There are also supporting words such as “barrier” and “sleepy” which indicate that accidents are caused by collisions with barriers on the road or sleepy drivers.

The reasons for the occurrence of crisis event flood can be seen in Fig. 3, where the words “increase”, “broken”, “heavy”, “day”, “overflow”, and “area” appear in large sizes, indicating sev-

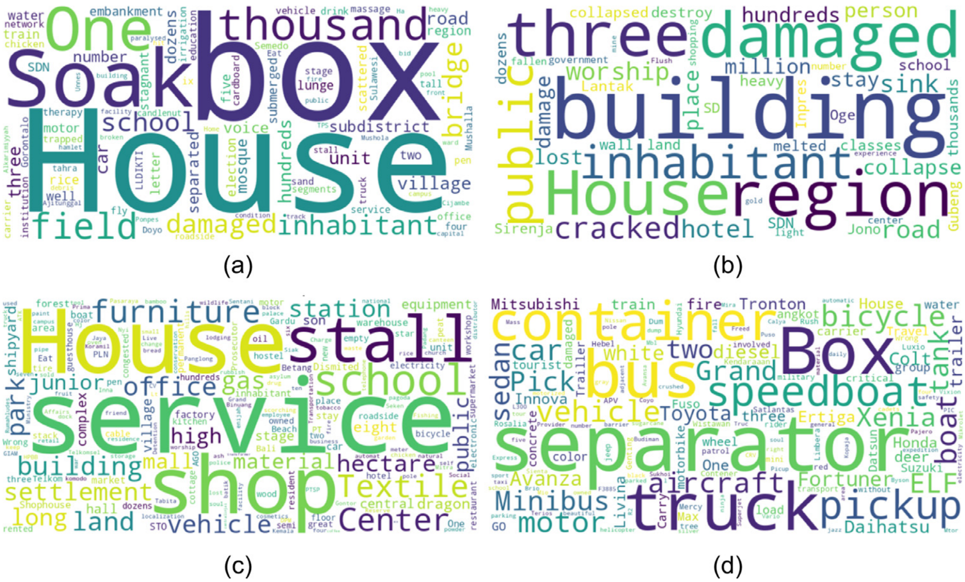

Fig 2. Visualization of the most frequently mentioned words on Twitter using word clouds for the AFFECTEDOBJECTS-ARG label related to crisis events: (a) flood, (b) earthquake, (c) fire, and (d) accident.

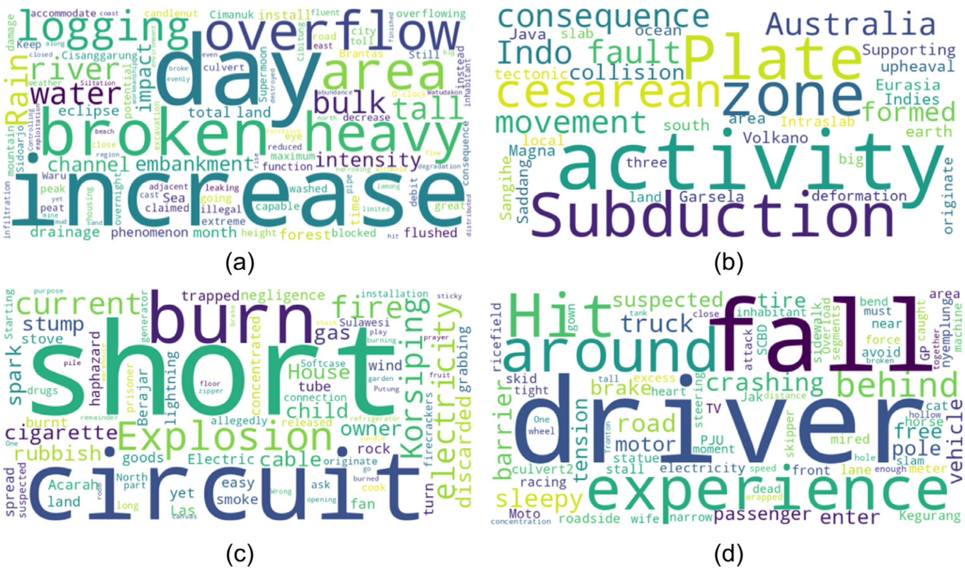

Fig 3. Visualization of the most commonly mentioned words on Twitter using word clouds for the REASON-ARG label during crisis events: (a) flood, (b) earthquake, (c) fire, (d) accident.

eral factors causing floodings such as increased water discharge, damaged channels, high rainfall, or overflowing water. Supporting words such as “logging” indicate deforestation activities that reduce the soil’s ability to absorb water, causing flooding. Words such as “circuit”, “short”, “burn”, “explosion”, “fire”, and “electricity” appear in large sizes in the crisis event fire type, indicating several leading causes of fires such as electrical short circuits or explosions due to gas cylinders. Fig. 3(c) also shows that supporting words such as “cigarette” can be the cause of small fires that have the potential to grow bigger. For the crisis event earthquake, it can be seen that the words “subduction”, “activity” and “plate” have large sizes, indicating that earthquakes can occur due to plate subduction activity, where two or more plates collide with each other.

Fig. 2 depicts word clouds for the AFFECTEDOBJECTS-ARG label. The words “house” and “box” in the crisis event flood have large sizes, indicating that houses and boxes (ballot boxes for the election of the people’s representative council) are objects that are often affected by floods. The words “building” and “house” for the crisis event fire and earthquake also have large sizes. Thus, building objects such as “houses” are the most frequently affected objects by crisis events such as floods, earthquakes, and fires. Meanwhile, vehicle objects such as “truck”, “bus”, and “container” are frequently affected objects in crisis event accidents. The word “Box” referred to in the crisis event accident is a vehicle such as a box car, which people in Indonesia often use to transport goods.

The addition of entity labels to this dataset uses the Begin Inside Outside (BIO) tag scheme, with detailed explanations and example sentences in Table 7. The overall distribution of each entity label in this dataset is illustrated in Table 8. Meanwhile, Table 9 provides a detailed breakdown of the frequency distribution of entity labels for each type of crisis event. The label "O" has the highest distribution in each type of crisis event because it shows the number of text sections that are not directly related to the entity, such as connecting words or general narratives. Based on Table 9, disaster events such as floods and earthquakes generally require detailed location information due to the high-frequency distribution of B-LOC labels, with 1,263 events for floods and 1,137 events for earthquakes. In contrast, fire crisis events tend to involve more information about affected objects, as indicated by the highest frequency of the B-ARG label among the various crisis events.

4. Experimental Design, Materials and Methods

The creation and evaluation of our dataset, shown in Fig. 4, consists of several main steps, such as data collection, preprocessing, annotation, and classification. A detailed explanation of each of these main steps will be explained in the subsections below.

4.1. Data collection

This semantic role labeling dataset was obtained by crawling text data on Twitter using the Python programming language, which consists of several processes, such as retrieving Twitter tokens with specific keywords, determining the date range in one month, collecting tweets in each date range, and setting the time interval. Some modules that need to be imported are calendar, datetime, and time [11]. The calendar module includes functions to calculate date ranges within a specific month. The DateTime module and the time delta function are essential for representing time differences. It can be used to add or subtract days or seconds, which allows for calculating the next day within a date range. The time module with sleep function temporarily stops the program’s execution for a certain number of seconds to help create a time interval so that it complies with Twitter’s API (Application Programming Interface) restrictions when collecting tweets. The sleep function is used if the tweet collection automation process fails due to the rate limit (the maximum number of requests that can be made in a specific time), then the program will wait for two minutes before trying again. If the tweet collection automation process is successful, the program rests for 20 seconds before continuing to the following date range. A total of 269,652 tweets were successfully collected and saved in CSV file format.

4.2. Data preprocessing

Once the data has been successfully crawled, it proceeds to the preprocessing process for cleaning. Some Python modules that need to be imported are pandas, string, re, and coun-

Table 7
Details of label entities for the named entity recognition task and examples of each label.

Label Entity	Description	Example Phrase/Word	Original Tweet Text
B-LOC	• This label marks the first word of the event's central location or specific location in the text. • The location with the LOC label is the most specific place of the event, such as the name of a village, hamlet, or place that is the center of the event.	"*dusun* (sub-district)"	**In Indonesian:** "*bencana banjir di dusun krajan desa winong kecamatan tugu*" **Translated to English:** "flood disaster in krajan sub-district, winong village, tugu district"
I-LOC	The I-LOC label marks the words that follow the first word of the central location or specific event labeled as LOC.	"*Krajan* (a sub-district called Krajan) "	
B-PLOC	The B-PLOC label marks the first word of a broader location, not the most specific location of the event, but an additional geographic context that provides general information about the larger administrative area.	"*desa* (village)", "*kecamatan* (district)"	
I-PLOC	• The I-PLOC label marks the words after the first word that complete the details of the PLOC (Pseudo-Location) label or the broader contextual location of the specific event location. • The words in this label are usually the names of sub-districts or regencies around the event's location.	"*Winong* (a village called Winong)", "*Tugu* (a district called Tugu)"	
B-ORG	The B-ORG label marks the first word of the name of an organization, institution, or group involved in handling a crisis event.	"26 (twenty six)"	**In Indonesian:** "*si jago merah menghanguskan 13 ruko di jl. oto Iskandar dinata kota bandung 26 unit mobil pemadam kebakaran dikerahkan ke lokasi kejadian*" **Translated to English:** "the red rooster burns down 13 shophouses on oto Iskandar dinata street, bandung city 26 fire engines deployed to the location"
I-ORG	The I-ORG label marks the next word after the first word, which completes the details of the B-ORG label and indicates the name of the organization or institution mentioned in the text.	"*unit mobil pemadam kebakaran* (fire engine units)"	

(*continued on next page*)

Table 7 (*continued*)

Label Entity	Description	Example Phrase/Word	Original Tweet Text
B-ARG	• The B-ARG label marks the first word that has a role as a semantic argument of an event. • Arguments can include time, objects affected, street names, or other elements directly related to the event. • This does not include locations or organizations.	“13 (thirteen)”, “*jl* (street)”	
I-ARG	• The I-ARG label marks the following words after the first word of the semantic argument label (B-ARG). • The words in this label complete the details of the argument, such as time, objects affected, or street names.	“*ruko* (shophouses)”, “*oto Iskandar dinata*”	
B-EVE	• The B-EVE label marks the first word of an actual event or incident in the text, such as a disaster. • The words in this label provide core information about the type of event that occurred.	“*si* (the)”	
I-EVE	The I-EVE label marks words that complete the event details after the first word of the B-EVE label if the event name consists of more than one word.	“*jago merah* (red rooster)”	
O	The O label is used to mark words or phrases that do not belong to any entity category.	“*dikerahkan ke lokasi kejadian* (deployed to the location)”	

Table 8
Overall frequency distribution of entity labels for the named entity recognition task.

Label Entity	Frequency	Percentage (%)
O	68,724	69.274036
I-ARG	7285	7.343306
B-ARG	4838	4.876721
B-EVE	4746	4.783985
B-LOC	3955	3.986654
B-PLOC	2786	2.808298
I-LOC	2089	2.105719
I-PLOC	1465	1.476725
B-ORG	1315	1.325525
I-EVE	1264	1.274116
I-ORG	739	0.744915

Table 9
Frequency distribution of entity labels for each type of crisis event for the named entity recognition task.

Label Entity	Frequency for Each Crisis Event			
	Flood	Accident	Fire	Earthquake
O	22,032	11,085	19,291	16,316
I-ARG	1027	2386	2829	1,043
B-ARG	668	1388	2172	610
B-EVE	1427	911	1341	1067
B-LOC	1263	557	998	1137
B-PLOC	745	361	1163	517
I-LOC	619	228	809	433
I-PLOC	306	89	896	174
B-ORG	386	107	483	339
I-EVE	167	91	43	963
I-ORG	140	27	363	209

try_list [11]. The Pandas library offers a DataFrame structure to manage structured data. The string module provides constants and classes for string manipulation that can be used to remove punctuation or make letters all lowercase (case folding). The re module contains regular expression operations in Python that are useful for matching patterns and manipulating complex strings. The country_list module with the countries_for_language function returns a list of countries that match the given language.

Preprocessing includes case folding and the removal of URLs, mentions, and hashtags using the re and string modules. Then, tweet filtering removes Indonesian language tweets that tell important news events abroad. The first step is to take all the lists of country names (except Indonesia) using the country_list module with the countries_for_language function. The list of country names is used to filter rows in the DataFrame. If text is found in the DataFrame that mentions countries in the list, the text will be removed. The output is a DataFrame that only contains text that does not mention any country from the list.

After filtering the country names, the process of filtering the names of provinces and districts/cities in Indonesia continues. The first step is reading data from a CSV file containing a list of provinces and districts/cities taken from GitHub [12] based on information from the Central Bureau of Statistics Indonesia. The list of provinces and districts/cities is converted to all lowercase and combined into one combined list to check whether any tweet rows in the DataFrame mention one of the names. The output is tweet rows in the DataFrame that mention the provinces or districts/cities from the combined list.

Text normalization involves replacing slang words with their formal equivalents using a slang dictionary. This dictionary is a CSV file [7] that contains pairs of slang words and their corresponding formal Indonesian translations. The dictionary is read into a DataFrame, and a function named ‘fix_slang’ is created to process the text within that DataFrame. This function splits the text into a list of words, replaces each slang word found in the dictionary with its equivalent, and then reassembles the words into standard text.

Similar tweets are removed by calculating the similarity level using TF-IDF and cosine similarity. Each text is converted into a TF-IDF vector representation, and then a cosine similarity matrix is calculated to measure the similarity between the texts. If two texts are similar above a threshold of 0.8, one is removed from the dataset. Approximately 56,135 tweets were successfully preprocessed. The cleaned data was then tokenized to facilitate the data annotation because SRL and NER work at the token level. Tokenization is done by breaking each text into individual words using the split() function from the Python module, which separates text based on spaces.

4.3. Data annotation

The labeling guidelines document was created before the labeling process for two annotators (1 expert in disaster management and 1 PhD student in computer science). Table 10 explains

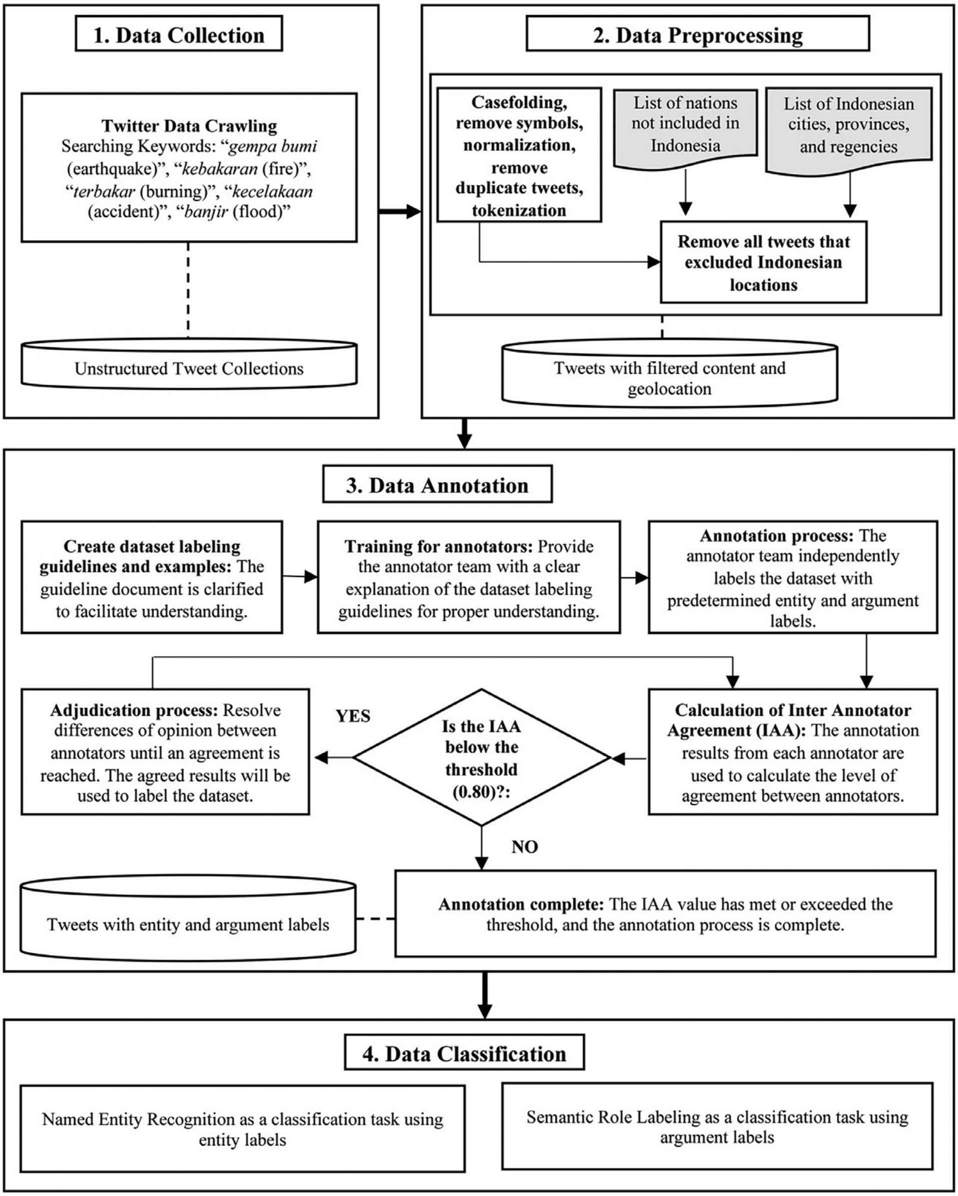

Fig 4. Flowchart for creation and evaluation of a semantic role labeling dataset for crisis events.

the labeling annotation guide to maintain data label consistency, especially those related to determining whether tweets are FALSE-EVENT and determining LOC and PLOC entity labels. The guide is also accompanied by labeling examples to provide an overview for annotators when labeling data. In addition, the definition of argument labels (Table 4) and entity labels (Table 7) also serve as additional guidelines for annotators in labeling arguments and entity labels.

When the labeling guidelines document has been created, the guide is conveyed clearly to the annotators. Each annotator labels arguments and entities. The level of agreement between

Table 10
Annotation guidelines on the semantic role labeling for crisis event dataset used by anonator.

No.	Details	Examples
1	• The context of each tweet is examined first to determine whether it qualifies as a FALSE-EVENT. • Tweets containing phrases or sentences related to one of the four crisis events (fire, accident, earthquake, or flood) but do not describe an actual event will be labeled as FALSE-EVENT. Additionally, tweets that only convey general statements, plans, hopes, personal stories, or promises without relating to an actual event will also be categorized as FALSE-EVENT. • If the tweets are not labeled as FALSE-EVENT, they will be classified as describing an actual event, which could be categorized as an EARTHQUAKE-EVENT, FIRE-EVENT, FLOOD-EVENT, or ACCIDENT-EVENT.	• A tweet reads, "alasan para pedagang bahwa kenaikan harga beras disebabkan banyaknya petani gagal panen akibat banjir dianggap tidak masuk akal.. (the reason traders say the increase in rice prices is due to many farmers failing to harvest due to flooding is considered unreasonable..)" ◦ The tweet is labeled FALSE-EVENT because it does not describe a flood that is currently or has occurred but instead talks about opinions or arguments that link the impact of flooding to rice prices. No information shows that a particular flood is currently taking place or has occurred. • Meanwhile, the tweets "banjir rendam 184 rumah di halmahera tengah (floods submerge 184 houses in Central Halmahera)" are labeled FLOOD-EVENT. ◦ The reason is that the information in the tweet shows the location, impact, and facts of the event that is currently or has occurred.
2	Street names and details, such as the name of the housing complex, house number, alley number, neighborhood number (neighborhood association/RT), and hamlet number (community association/RW), are labeled ARG entities and STREET-ARG argument labels.	• There is a tweet sound like this: "*kebakaran (konsleting listrik) di komplek cpi blok b no 100 rt 02 rw 13 desa cincin kecamatan soreang* (fire (electrical short circuit) in the cpi complex, block b, no. 100, rt 02, rw 13, cincin village, soreang district)". ◦ Therefore, the phrases "cpi complex", "block b", "no 100", "rt 02", and "rw 13" are labeled with the ARG entity and the STREET-ARG argument label.
3	• Determining the LOC and PLOC entity labels is used to distinguish the level of location specificity in a text. • The LOC label is given to the location of an event's main point or primary focus, such as the name of a hamlet, village, or place most relevant to the event mentioned. • Meanwhile, the PLOC label marks additional broader locations such as sub-districts, cities, regencies, provinces, or geographic areas that provide a more general location context but are not the center of the event.	• For example, the text is known to read " *kebakaran gudang milik pabrik tekstil di desa purwosuman, kecamatan sidoharjo, sragen, jawa tengah* (warehouse fire belonging to a textile factory in purwosuman village, sidoharjo district, sragen, central java)". ◦ Therefore, the phrase "purwosuman village" is labeled LOC because it is the central location where the flood occurred. ◦ The phrases "sidoharjo district", "sragen", and "central java" are labeled PLOC because all three are administrative areas that provide additional information about the location of the incident, but not the most specific location.

annotators/IAA is measured using Cohen's Kappa for both labels using sample data of around 1000 tweets, equivalent to 22,908 tokens.

Fig 4 shows that if the level of agreement between annotators on argument labels and entity labels is above 0.80, the labeling guidelines document is deemed appropriate, and labeling can continue until the entire dataset is completed. However, if the level of agreement is below 0.80, an adjudication process will be initiated to resolve conflicts where annotators assign different labels to the same data. The adjudication process involves exchanging opinions between annotators to understand each other's context and arguments. The label agreed upon through adjudication becomes the final label for the conflicting data, and revisions to the guidelines are made if necessary. This adjudication process aims to reduce individual annotator bias and produce a more reliable dataset. Once the adjudication process is complete, the IAA value is recalculated. If the IAA value is above 0.80, it indicates that the guidelines are sufficiently clear, and the la-

Table 11
Test results for the SRL task using argument labels as target labels, TF-IDF as the text representation, and various machine learning models.

Classifier	Ratio	Precision	Recall	F1-Score	Accuracy
Naïve Bayes	90:10	0.79	0.80	0.75	0.80
	80:20	0.79	0.81	0.75	0.81
	70:30	0.79	0.80	0.75	0.80
K-Nearest Neighbors	90:10	0.79	0.79	0.78	0.79
	80:20	0.78	0.78	0.76	0.78
	70:30	0.79	0.80	0.78	0.80
Logistic Regression	90:10	0.78	0.81	0.77	0.81
	80:20	0.80	0.82	0.80	0.82
	70:30	0.80	0.82	0.79	0.82
Decision Tree	90:10	0.75	0.77	0.71	0.77
	80:20	0.75	0.77	0.71	0.77
	70:30	0.75	0.77	0.71	0.77
Support Vector Machine	90:10	0.82	0.84	0.82	0.84
	80:20	0.82	0.84	0.82	0.84
	70:30	0.82	0.84	0.82	0.84
Average		0.79	0.80	0.77	0.80

Table 12
Test results for the NER task using entity labels as target labels, TF-IDF as the text representation, and various machine learning models.

Classifier	Ratio	Precision	Recall	F1-Score	Accuracy
Naïve Bayes	90:10	0.76	0.79	0.74	0.79
	80:20	0.77	0.80	0.74	0.80
	70:30	0.74	0.79	0.73	0.79
K-Nearest Neighbors	90:10	0.77	0.78	0.77	0.78
	80:20	0.77	0.76	0.76	0.76
	70:30	0.77	0.77	0.76	0.77
Logistic Regression	90:10	0.78	0.81	0.77	0.81
	80:20	0.78	0.81	0.77	0.81
	70:30	0.77	0.80	0.77	0.80
Decision Tree	90:10	0.74	0.78	0.71	0.78
	80:20	0.74	0.78	0.71	0.78
	70:30	0.74	0.77	0.71	0.77
Support Vector Machine	90:10	0.79	0.82	0.79	0.82
	80:20	0.79	0.82	0.79	0.82
	70:30	0.78	0.81	0.79	0.81
Average		0.77	0.79	0.75	0.79

beling process can proceed for the entire dataset. However, if the IAA value is still below 0.80, the adjudication process will be repeated until the IAA value meets or exceeds 0.80.

In our dataset, the inter-annotator agreement (IAA) value is 0.9236 for the argument label and 0.9117 for the entity label. These values indicate a very high level of agreement among the annotators, demonstrating that the labeled data offers strong consistency and low subjective bias [13].

4.4. Data classification

The experimental results using this dataset are used as a guide for future research related to crisis events. There are two experimental scenarios. In the first scenario, we extract information using the SRL task with argument labels as the target (shown in Table 11). In the second scenario, we extract information using the NER task with entity labels as the target (Table 12). In each experimental scenario, we conduct trials with several different training and testing ratios such as 90:10, 80:20, and 70:30. Term Frequency Inverse Document Frequency (TF-IDF) is a statistical-based technique used to represent text in the form of numbers so that machine learning models can process it. TF-IDF measures the importance of a word in a document relative to the collection of documents (corpus). The less frequently a word appears in the corpus, the higher its value in the TF-IDF representation [14].

The numerical representation of the text using TF-IDF is then input into a machine learning model for SRL and NER tasks. Several classifiers are employed, including Naïve Bayes, K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree, and Support Vector Machine (SVM). The selection of several classifiers reflects a strategy to try various approaches based on different principles in machine learning, such as probability-based (with Naïve Bayes), distance (KNN), linearity (Logistic Regression), rules (Decision Tree), and margin (SVM). The Naïve Bayes classifier uses a probability approach that assumes all features (words that TF-IDF has represented) are independent [15]. The KNN classifier predicts a data class based on its nearest neighbors, using distance as an indicator of similarity [16]. A Logistic Regression classifier is a linear classification model based on a sigmoid/logistic function that converts the linear combination value of input features into target class probability [15]. In contrast, a Decision Tree classifier builds a series of rules to separate data into classes based on specific features such as resembling tree structure [16]. SVM looks for a dividing line in feature space that is used to separate data from different classes (hyperplane) with maximum margin [15].

We classify data for SRL and NER tasks using the scikit-learn library [17] with several types of classifiers in a Python programming language. Parameters used for the Naïve Bayes classifier are default parameters for multinomial Naïve Bayes, while our KNN classifier parameter uses n_neighbors is 2. We use a linear kernel for the SVM classifier to speed up the computation process. The parameters we use for the Decision Tree are criterion entropy, with a max depth of up to 40 to limit the depth of the tree because if it is too deep, it can make the tree learn irrelevant details from the training data. The parameters in the Logistic Regression classifier use the liblinear solver and max iter 100 times to avoid the training process being too long.

Tables 11 and 12 show that for the SRL and NER tasks, the SVM classifier obtains the best precision, recall, F1-score, and accuracy values among other classifiers because it can maximize the separation margin between classes in the feature space, even with variations in data distribution. The average precision, recall, and F1-score values exceeding 0.70 on various classifiers (Naïve Bayes, KNN, Logistic Regression, Decision Tree, and SVM) indicate our dataset has good quality. The TF-IDF feature representation effectively captures relevant patterns about crisis events from our dataset, enabling the model to distinguish classes accurately.

Limitations

This dataset has several limitations but does not reduce its value as a quality data source. First, this dataset is specifically designed for the Indonesian language, which is the main focus of this study. This limitation provides an advantage, considering the small number of public datasets available for Natural Language Processing (NLP) tasks in Indonesian. Second, this dataset only covers four types of crisis events: fires, accidents, floods, and earthquakes. This selection was made considering that these four categories are Indonesia's most frequent and impactful events. Despite its limited scope, this dataset can provide specific insights that are highly relevant for analyzing crisis events in Indonesia.

Ethics Statement

The dataset we developed fully complies with Twitter's 2023 developer policy [18]. The data collection process was conducted carefully and under all applicable regulations, including those regarding the use and distribution of data from Twitter. Notably, the identities of Twitter users are not included in the dataset, ensuring that no personal or sensitive information can be identified. These measures are implemented to protect users' privacy rights and prevent any potential data misuse.

Data Availability

Semantic Role Labeling Datasets for Crisis Event (Original data) (Mendeley Data)

CRediT Author Statement

Amelia Devi Putri Ariyanto: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing – original draft; **Diana Purwitasari:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition; **Bilqis Amaliah:** Data curation; **Chastine Fatichah:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision; **Muhammad Ghifari Taqiuddin:** Software, Resources, Data curation; **Haikal:** Validation, Data curation.

Acknowledgements

The authors acknowledge the financial support of the Ministry of Education and Culture of the Republic of Indonesia for Doctoral Dissertation Research scheme under Grant No. 038/E5/PG.02.00.PL/2024.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Zhou, et al., VictimFinder: harvesting rescue requests in disaster response from social media with BERT, Comput. Environ. Urban. Syst. 95 (2022), doi:10.1016/j.compenvurbsys.2022.101824.

[2] A.D.P. Ariyanto, D. Purwitasari, C. Fatichah, A systematic review on semantic role labeling for information extraction in low-resource data, IEEE Access. 12 (April) (2024) 57917–57946, doi:10.1109/ACCESS.2024.3392370.

[3] A.D.P. Ariyanto, C. Fatichah, D. Purwitasari, Semantic role labeling for information extraction on Indonesian texts: a literature review, in: 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA), IEEE, 2023, pp. 119–124, doi:10.1109/ISITIA59021.2023.10221008.

[4] S. Heo, S. Park, D.K. Lee, Multi-hazard exposure mapping under climate crisis using random forest algorithm for the Kalimantan Islands, Indonesia, Sci. Rep. 13 (1) (2023) 1–13, doi:10.1038/s41598-023-40106-8.

[5] P.K. Putra, D.B. Sencaki, G.P. Dinanta, F. Alhasanah, R. Ramadhan, Flood monitoring with information extraction approach from social media data, in: Proceeding - AGERS2020: IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology: Understanding the Interaction of Land, Ocean and Atmosphere: Disaster Mitigation and Regional Resillience, 2020, pp. 113–119, doi:10.1109/AGERS51788.2020.9452770.

[6] A. D. P. Ariyanto, D. Purwitasari, C. Fatichah, M. G. Taqiuddin, and H. Haikal, "Semantic role labeling datasets for crisis event," Mendeley Data. (Accessed 26 November 2024). [Online]. Available: https://data.mendeley.com/datasets/r76v5sjyv2/1.

[7] N. Aliyah Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, A. Jamal, Colloquial Indonesian Lexicon, in: 2018 International Conference on Asian Language Processing (IALP), 2018, pp. 226–229, doi:10.1109/IALP.2018.8629151.

[8] I.M.S. Bimantara, D. Purwitasari, N.A.S. ER, P.G.S. Natha, Balinese story texts dataset for narrative text analyses, Data Br. 56 (2024), doi:10.1016/j.dib.2024.110781.

[9] K. Xu, H. Wu, L. Song, H. Zhang, L. Song, D. Yu, Conversational semantic role labeling, IEEE/ACM Trans. Audio Speech. Lang. Process. 29 (2021) 2465–2475, doi:10.1109/TASLP.2021.3074014.

[10] Q. Zhou, Z. Jiang, F. Yang, Sentences similarity based on deep structured semantic model and semantic role labeling, in: 2020 International Conference on Asian Language Processing, IALP 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 40–44, doi:10.1109/IALP51396.2020.9310496.

[11] Python Software Foundation, "Documentation of The Python Standard Library." (Accessed 13 November 2024). [Online]. Available: https://docs.python.org/3/library/.

[12] E. Samuel, “Administrative Subdivisions of Indonesia (Provinces, Regencies/Cities, Districts, Villages).” (Accessed 17 November 2024). [Online]. Available: https://github.com/edwardsamuel/Wilayah-Administratif-Indonesia.
[13] M.T. Uliniansyah, et al., Twitter dataset on public sentiments towards biodiversity policy in Indonesia, Data Br. 52 (2024) 109890, doi:10.1016/j.dib.2023.109890.
[14] M. El Barachi, S.S. Mathew, M. Alkhatib, Combining named entity recognition and emotion analysis of tweets for early warning of violent actions, in: 2022 7th International Conference on Smart and Sustainable Technologies, SpliTech 2022, 2022, pp. 1–6, doi:10.23919/SpliTech55088.2022.9854231.
[15] R.S. Patil, S.R. Kolhe, Supervised classifiers with TF-IDF features for sentiment analysis of Marathi tweets, Soc. Netw. Anal. Min. 12 (1) (2022) 1–16, doi:10.1007/s13278-022-00877-w.
[16] J. Zhang, Z. Li, Z. Pu, C. Xu, Comparing prediction performance for crash injury severity among various machine learning and statistical methods, IEEE Access. 6 (2018) 60079–60087, doi:10.1109/ACCESS.2018.2874979.
[17] F. Pedregosa, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (85) (2011) 2825–2830 [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html .
[18] X. Corp, “Developer Policy.” (Accessed 21 November 2024). [Online]. Available: https://developer.x.com/en/developer-terms/policy.