OPEN ACCESS
VOL. 7, NO. 2, PP.163-172, MAY 2025

ISSN : 2622-8106 (ONLINE)
DOI: 10.20895/INISTA.V7I2.1814

Journal of Informatics, Information System, Software Engineering and Applications (INISTA)

Extracting Post-Disaster Health Impact Information from News Reports Using Named Entity Recognition

Nalar Istiqomah *1, Fanny Novika 2

1,2Actuary, Trisakti School of Insurance, Indonesia
1, 2Jl. Jenderal Ahmad Yani No.85, RT.15/RW.7, Kayu Putih, Kec. Pulo Gadung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13210, Indonesia

1 nalar.istiqomah23@gmail.com
2 novikafanny@gmail.com

Received on 09-04-2025, revised on 02-05-2025, accepted on 24 05-2025

Abstract

Natural disasters have a significant impact on public health, giving rise to various post-disaster illnesses. This study presents an automated information-extraction framework based on Named Entity Recognition (NER), leveraging the IndoBERT model to identify disaster types, health impacts, and affected locations from online news reports. Data were gathered via web scraping from multiple reputable news portals and subsequently processed through tokenization, stop-word removal, and lemmatization. Extracted entities were visualized via bar charts and word clouds to reveal disease patterns associated with each disaster type. Results indicate that floods have a significant public health impact, with skin diseases being the most prevalent, followed by diarrhea, fever, influenza, and Acute Respiratory Infections (ARIs). Volcanic eruptions are linked to health conditions such as ARI, hypertension, diarrhea, and influenza, whereas earthquakes show strong correlations with diarrhea, ARI, skin diseases, and fever. Droughts and landslides are closely associated with diarrheal outbreaks due to compromised sanitation resulting from limited access to clean water. Although less frequently reported, tsunamis also exhibit a notable association with cases of diarrhea. The proposed method achieves 90 % accuracy and an 88 % F1-score. These findings confirm the effectiveness of our NER-based approach in detecting causal relationships between disasters and health outcomes, providing valuable insights for policymakers and healthcare professionals in designing targeted post-disaster mitigation and response strategies.

Keywords: Health Impacts; IndoBERT; Named Entity Recognition; Natural Disasters; Text Mining

Corresponding Author:
* Nalar Istiqomah
Actuary, Trisakti School of Insurance,
Jl. Jenderal Ahmad Yani No.85, RT.15/RW.7, Kayu Putih, Kec. Pulo Gadung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13210, Indonesia
Email: nalar.istiqomah23@gmail.com

I. INTRODUCTION

Natural disasters such as floods, earthquakes, forest fires, and tsunamis often inflict significant health burdens on affected communities [1]. These include communicable diseases, such as diarrhea resulting from poor sanitation after floods [2], acute respiratory infections caused by air pollution from wildfires [3], and psychological disorders like Post-Traumatic Stress Disorder (PTSD) [4] following traumatic events [5-7]. In Indonesia, one of the world's most disaster-prone countries [8],[9], post-disaster health surveillance still relies on manual reports from various agencies. Such unstructured and delayed reporting hampers timely decision-making by government and health authorities, often leading to suboptimal public health responses [10].

Online news media, by contrast, provide a rich and timely stream of disaster-related information [11]. However, manually extracting health-impact data from thousands of articles, each written in diverse formats, is laborious and inconsistent. To address this, automated Natural Language Processing (NLP) methods are needed to efficiently identify and analyze health impacts directly from news texts [12].

Previous work has demonstrated the potential of NLP in disaster contexts. Named Entity Recognition (NER) has been used to detect disaster types and geographic mentions in tweets [13], while deep-learning models, such as the XLNet-BiLSTM-CRF model, can recognize natural hazards effectively [14]. In the public health domain, NLP is used to detect disease outbreaks from social media and news feeds [15]. Additionally, research has shown that NLP can be used to automate the classification of online sources for infectious disease occurrences [16], demonstrating its potential for real-time public health surveillance [17]. Some studies have utilized NLP to classify disaster news topics without extracting health-specific entities [18].

In contrast, others have applied NER for sentiment analysis on post-disaster tweets without addressing disease detection. More recent work has also shown how fine-tuned transformer models significantly improve health-related information extraction in low-resource languages [19] [20]. Building on these studies, the present research aims to develop an IndoBERT-based NER framework for automatic extraction of disaster types, health impacts, and affected locations from online news.

II. RESEARCH METHOD

A. Research Data

This study employs online news articles that report on disaster events and their related health impacts. These articles were gathered via web scraping from four reputable news outlets: Kompas, CNN Indonesia, BBC Indonesia, and Tempo. For each collected article, several key fields were extracted, including the title, publication date, full text, disaster category (such as flood, earthquake, forest fire, or tsunami), affected location, and any mentioned health impacts. The extracted data were then preprocessed through a series of steps, including tokenization, stop-word removal, and lemmatization, to prepare the text for Named Entity Recognition (NER) using the IndoBERT model.

B. Research Instruments

Data collection and analysis were conducted using Python, along with several libraries and tools. For web scraping, BeautifulSoup and Selenium were employed to parse HTML and handle dynamic content, while the Requests library was used to access APIs when available. The data collected was stored in structured formats, specifically CSV and JSON, to facilitate further processing. Text preprocessing involved the use of spaCy for tokenization and Sastrawi for lemmatization and stop-word removal in the Indonesian language. For Named Entity Recognition (NER), two models were utilized: spaCy's pre-trained NER model served as a baseline, and a fine-tuned IndoBERT model was used for more accurate, domain-specific extraction of disaster types, health impacts, and location entities.

C. Data Collection Procedures

The overall data collection methodology is presented in Figure 1.

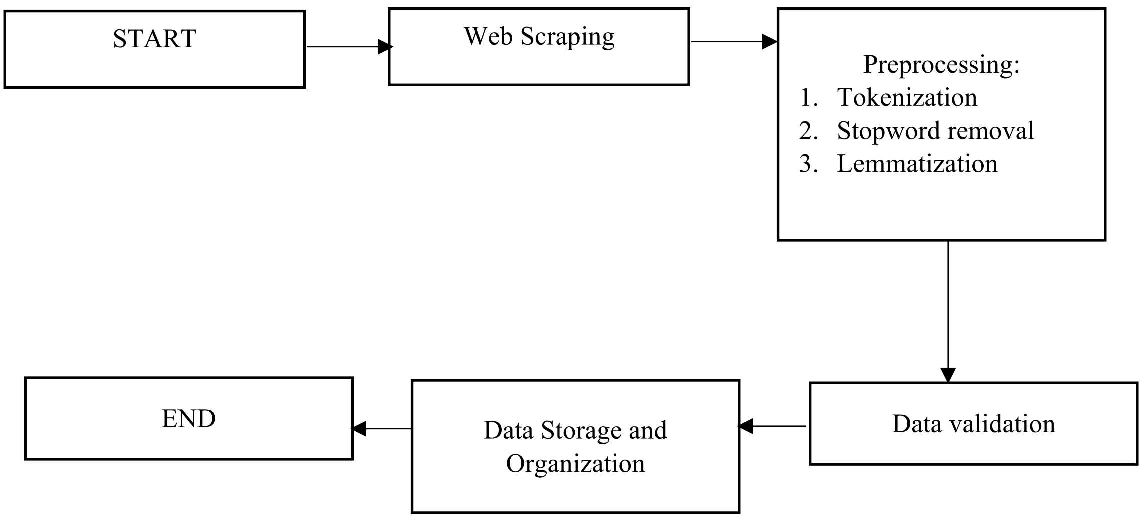

Fig. 1. Research procedure

1. Web Scraping of Online News

To collect news-text data, we employed web scraping, an automated technique for extracting information from websites. The scraping process was implemented using Python, with BeautifulSoup, Selenium, and the Requests library. From each article, the following fields were extracted: the title, publication date, full article text, disaster type (e.g., flood, earthquake, wildfire), affected location, and any mentioned health impacts.

2. Data Validation and Cleaning

Following the initial web scraping process, a thorough data cleaning procedure was conducted to ensure that the dataset consisted solely of articles relevant to the research objectives. The first step involved duplicate removal, in which articles with identical or highly similar content—often resulting from syndicated news or cross-posting between media outlets—were identified and excluded. This was followed by a relevance filtering stage, where articles lacking explicit mentions of health impacts related to the reported disasters were discarded. For example, articles that focused exclusively on economic losses or infrastructure damage were removed from the dataset. This step is essential in disaster-related NLP studies, as irrelevant data can significantly reduce the performance of downstream models such as Named Entity Recognition (NER) [21],[22]. To further ensure the reliability and quality of the data, a manual validation process was conducted. In this step, a randomly selected subset of articles was reviewed by human annotators to confirm thematic relevance and verify the accuracy and consistency of the extracted information, in accordance with best practices in corpus curation for machine learning applications [23].

3. Data Preprocessing

Before applying Named Entity Recognition (NER), the collected text data underwent a series of standard natural language processing (NLP) preprocessing steps to improve the quality and consistency of the input. First, tokenization was performed to split the text into individual words or meaningful phrases, enabling more granular analysis. This was followed by stop-word removal, which involved eliminating commonly used words such as “and,” “or,” and “yang” that carry limited semantic value and can introduce noise into the model. Lastly, lemmatization was applied to reduce each word to its base or dictionary form, helping to normalize variations and improve entity recognition performance. The preprocessing was carried out using two key libraries: spaCy [24], which handled the tokenization process [25], and Sastrawi, which provided lemmatization and stop-word removal capabilities tailored for the Indonesian language [26].

4. Data Storage and Organization

The cleaned and preprocessed data were stored in formats that facilitated efficient downstream analysis and model integration. CSV (Comma-Separated Values) files were used for storing data in a structured, tabular format, which allowed for straightforward import into statistical and data analysis tools. In parallel, JSON (JavaScript Object Notation) format was employed for storing hierarchical and nested data structures, making it particularly suitable for use with natural language processing (NLP) libraries that require flexible and richly annotated inputs. This structured dataset is now ready for further processing using an IndoBERT-based NER pipeline to extract health-impact entities, locations, and disaster types from the news articles.

D. Data Analysis Procedure

In this study, data analysis is conducted through several stages, including text data processing, entity extraction using the IndoBERT-based Named Entity Recognition (NER) model, model performance evaluation, and visualization of the extracted results to analyze post-disaster health impacts. The detailed techniques are as follows:

1. Entity Extraction using Named Entity Recognition (NER)

After preprocessing, the primary step in the data analysis phase involves entity extraction using Named Entity Recognition (NER) models. NER is a subtask of information extraction that aims to locate and classify named entities in text into predefined categories, such as person names, organizations, locations, expressions of time, quantities, monetary values, percentages, and more. In this study, the focus is extended to disaster-related entities, including disaster types, health impacts, and affected locations. Mathematically, a NER model can be viewed as a sequence labeling function that maps a sequence of tokens $X = \{x_1, x_2, \dots, x_n\}$ to a sequence of labels $Y = \{y_1, y_2, \dots, y_n\}$, where each $y_i \in Entity\ Labels$, such as B-DISASTER, I-LOCATION, O, etc. This can be represented in Equation (1):

$$f_{NER}: X \rightarrow Y \quad (1)$$

Where the function f is typically modeled using deep learning architectures such as BiLSTM-CRF or transformer-based models like BERT [27]. To illustrate this, consider the example sentence: *"Sebanyak 350 warga menderita ISPA akibat kebakaran hutan di Palangkaraya."*

NER will tag:
1. "ISPA" as B-HEALTH
2. "kebakaran hutan" as B-DISASTER
3. "Palangkaraya" as B-LOCATION

Two NER models were utilized in this study:
1. spaCy Pre-trained Model: Serves as a baseline, trained on general-purpose corpora. While effective for standard entities such as PERSON, GPE, and DATE, it may underperform for domain-specific entities, particularly those related to health crises or disasters in Indonesian texts [24].
2. Fine-tuned IndoBERT: This model was adapted further using domain-specific corpora focused on Indonesian disaster news. Fine-tuning allows the transformer-based model to learn contextual representations that are highly sensitive to subtle variations and local terms in disaster reports, thus improving recall and precision for specialized entities [26].

The entity extraction process in this study focused on three main categories. The first is disaster type detection, which involves identifying terms that refer to specific types of disasters such as banjir (flood), gempa (earthquake), kabut asap (haze), and tsunami. The second category is health impact detection, which includes both physical health conditions like diare (diarrhea) and ISPA (acute respiratory infections), as well as psychological effects such as trauma and depresi (depression). The third category is affected location detection, which entails extracting the names of places, administrative regions, or other geographical indicators—such as cities or provinces—that are mentioned as being impacted in the news articles.

2. Model Performance Evaluation

To assess the effectiveness of the NER models, performance evaluation was conducted using two widely adopted metrics: accuracy and F1-score. Accuracy in Equation (2) measures the proportion of correct predictions made by the model relative to the total number of predictions, providing a general sense of overall performance. It is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives. However, because NER tasks often involve imbalanced data and subtle distinctions between entity types, the F1-score, which is the harmonic mean of precision and recall, offers a more balanced and informative measure of model performance in this context. The F1-score is computed using the following formulas in Equation (3)-(5),

$$\text{Precision} = \frac{\sum_{i=1}^{k} \mathrm{TP}_i}{\sum_{i=1}^{k} TP_i + FP_i} \quad (3)$$

$$\text{Recall} = \frac{\sum_{i=1}^{k} \mathrm{TP}_i}{\sum_{i=1}^{k} \mathrm{TP}_i + \mathrm{FN}_i} \quad (4)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

These formulas are widely used in classification tasks, especially in NLP evaluations, as outlined by [28] and [29]. The evaluation process involved comparing the outputs of each NER model with a manually annotated dataset that served as the ground truth. This comparison ensured that the extracted entities were both relevant and correctly identified, thereby validating the reliability of the models in capturing health-related and disaster-specific information from Indonesian news articles.

3. Data Visualization

Following the entity extraction process, the extracted information was visualized to facilitate data interpretation and reveal meaningful patterns. Visualization techniques were used to explore trends and relationships within the dataset. A word cloud was created to represent the most frequently mentioned health-related terms, such as infectious diseases and psychological conditions, found in disaster-related news coverage. This visualization provides a concise and intuitive overview of the most prevalent health topics in the corpus. Additionally, a chart was developed to illustrate the relationship between disaster types, such as floods, earthquakes, and forest fires, and the corresponding health impacts they tend to trigger. These visualizations help highlight recurring patterns, such as which diseases are commonly associated with specific disaster events, and offer valuable insights for public health preparedness and response planning.

4. Analysis of Disease Trends Following Disasters

In addition to entity-based extraction, this study also examines the trends of diseases that emerge in the aftermath of disasters, based on information extracted from news articles. Using disaster type and location data, the analysis aims to uncover recurring patterns in which specific diseases tend to follow certain types of disaster events. For example, the study examines whether an increase in cases of diarrhea or skin-related illnesses frequently follows floods, and whether forest fires are associated with a rise in Acute Respiratory Infections (ARIs) due to smoke-induced air pollution. By identifying these patterns, the analysis contributes to a better understanding of potential health threats in post-disaster contexts. It provides valuable insights for anticipating and mitigating similar risks in future disaster events.

III. RESULTS AND DISCUSSION

A. Data Collection Results

The data used in this study consists of online news articles collected using Python programming. These articles were sourced from various reputable online news portals, including Kompas, CNN Indonesia, detik.com, Tempo, Tribunnews, and others. Data scraping was conducted using a set of keywords that combined terms related to natural disasters and post-disaster diseases. The keywords related to natural disasters included flood, eruption, earthquake, volcanic eruption, drought, landslide, and tsunami.

Meanwhile, the health-related keywords used to identify post-disaster disease impacts encompassed a wide range of conditions, such as flu (cough, runny nose, chills), headache (cephalgia, dizziness, vertigo), hypertension or high blood pressure, dengue fever, dehydration, fever, diarrhea or dysentery, stomach ache (dyspepsia, nausea, gastritis, vomiting), malnutrition, urinary tract infection, respiratory infections (ARI), cholera, skin diseases (athlete's foot, itching), malaria, myalgia, pneumonia, typhoid, psychological illness (including trauma, depression, anxiety, and stress), hepatitis, and leptospirosis. These keyword combinations were designed to capture comprehensive coverage of disaster events and their associated health impacts across different geographic and temporal contexts.

After the scraping process, a total of 6,464 news articles containing both disaster-related and disease-related keywords were collected, spanning the period from 2005 to March 2025. These online news articles were then subjected to a rigorous validation and preprocessing procedure. The first step involved filtering relevant content, retaining only articles that explicitly reported health impacts following natural disasters in Indonesia. Despite containing relevant keywords, many articles were excluded because they were not directly relevant to the actual context of post-disaster health. Examples of such excluded content include titles like “8 Facts About the Tsunami of Covid-19 Cases in China,” “Importance of Tetanus Vaccine for Earthquake Victims and Volunteers,” “Indonesia at Risk of Cataract Tsunami in 2030,” “Jakarta's 2020 Flood: How to Stay Healthy?” “Post-Earthquake Trauma: Cats in Majene Vomit and Have Diarrhea,” and

"447 Pigs Dead from Cholera in Karo." These articles were omitted because they did not directly reflect the human health impacts caused by disaster events in Indonesia.

After this filtering step, only 1,137 articles, which represent approximately 17.59 percent of the original dataset, were retained for further analysis. These remaining articles underwent preprocessing using standard natural language processing techniques. The title and full-text content of each article were processed through case folding, tokenization, stop-word removal, and stemming, utilizing Python libraries specifically designed for handling Indonesian-language text. This preprocessing ensured that the data were clean, normalized, and ready for subsequent Named Entity Recognition (NER) and analysis steps.

B. Data Analysis Results

The analysis was conducted using Python. Named Entity Recognition (NER) based on IndoBERT was employed to identify frequent keywords associated with post-disaster diseases. The analysis revealed that floods were the most frequently reported disaster linked to diseases. This is presented in Table 1.

TABLE I. OCCURRENCES OF NATURAL DISASTERS ARE MENTIONED IN ONLINE NEWS

Types of disasters	Count
Flood	526
Flash floods	35
Tidal flood	8
Eruption	197
Earthquake	185
Drought	64
Landslide	83
Tsunami	38

In addition, the analysis reveals that the most frequently mentioned post-disaster diseases in news articles are diarrhea, skin diseases, acute respiratory infections, flu, and fever. Detailed information on the frequency of occurrence of each disease is presented in Table 2.

TABLE II. OCCURRENCES OF POST-DISASTER DISEASES REPORTED IN NEWS ARTICLES

Types of disasters	Count
Diarrhea	223
Skin Disease	223
Acute Respiratory Infection	133
Flu	107
Fever	103
Hypertension	98
Stomachache	43
Myalgia	33
Headache	30
Psychological Disorders	27
Dengue Fever	25
Hepatitis	17
Leptospirosis	16
Malaria	16
Pneumonia	9
Typhoid	8
Cholera	6
Dehydration	4
Pharyngitis	4
Pulmonary Tuberculosis	3
Malnutrition	2
Bronchitis	1
Dysentery	1
Urinary Tract Infection	1
Gastritis	1
Meningitis	1
Constipation	1

An analysis of disease occurrences following disasters was also conducted, revealing several key findings from the dataset. Floods were most frequently associated with skin diseases and diarrhea, mainly due to poor sanitation conditions in evacuation shelters. Earthquakes were linked to diarrhea, acute respiratory infections (ARI), skin diseases, and fever, with ARI resulting from dust exposure and inadequate shelter ventilation. Volcanic eruptions were associated with an increase in respiratory illnesses, including asthma and bronchitis. Droughts were commonly associated with hepatitis, often stemming from the lack of clean water and improper food handling. In addition, hypertension frequently emerges after disasters, triggered by stress, shock, or trauma experienced by evacuees. The data also indicated that the number of disease-related news articles tended to rise within one to two weeks following a disaster event. Areas affected by poor sanitation consistently showed higher incidences of disease. Furthermore, food safety in shelters was identified as a contributing factor to illness; in some cases, diarrhea resulted from food poisoning due to spoiled or expired relief aid, as documented in news reports from Kupang and Nganjuk. Meanwhile, psychological illnesses were rarely mentioned in the news, likely because such conditions require clinical diagnosis and are not immediately observable in the aftermath of a disaster.

Cases of prolonged evacuation also increased the risk of malnutrition, often caused by limited access to nutritious food, uneven distribution of aid, and poor sanitation that impairs nutrient absorption. These conditions were reported in tsunami-hit Mentawai and earthquake-affected Cianjur. A province-level analysis of post-disaster disease patterns, detailed in Appendix 2, reveals, for example, that volcanic eruptions in East Nusa Tenggara (NTT) were associated with ARI, diarrhea, and hypertension. At the same time, floods in East and West Java were linked to skin diseases. Earthquakes in West Java were linked to skin diseases, diarrhea, and ARI. Additionally, temporal analysis based on the year of news articles indicates that in 2023, widespread droughts were linked to diarrhea outbreaks. In 2024, frequent flooding events were associated with a broader set of health impacts, including skin diseases, diarrhea, fever, flu, and hypertension. That same year, several volcanic eruptions were reported to cause ARI, diarrhea, hypertension, and flu. In early 2025, repeated flooding events continued to result in skin diseases and diarrhea.

C. *Data Visualization*

The relationship between disasters and diseases was visualized using two main forms: bar charts and word clouds. These visualizations help in understanding the pattern of disease occurrence resulting from natural disasters.

1. *Bar Chart*

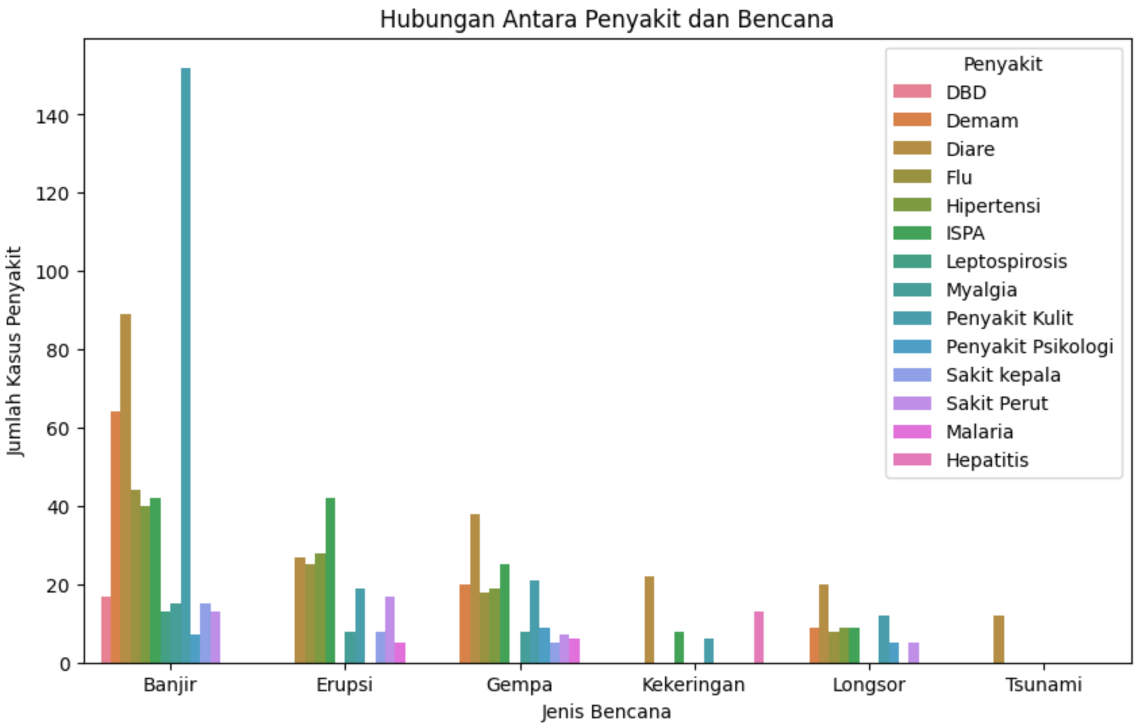

Fig. 2. Bar Chart of the Relationship Between Diseases and Natural Disasters

The bar chart, as presented in Figure 2, displays the number of disease cases categorized by the type of disaster, offering insight into the relationship between specific disasters and public health outcomes. Floods were found to have the most significant impact on public health, with skin diseases being the most frequently reported condition, followed by diarrhea, fever, flu, and respiratory infections. Volcanic eruptions are strongly associated with respiratory ailments, particularly respiratory infections, and are also linked to hypertension, diarrhea, and flu. Earthquakes show a high correlation with diarrhea, respiratory infections, skin diseases, and fever, reflecting the combined effects of displacement and poor shelter conditions. Droughts are closely related to diarrhea, primarily due to inadequate sanitation and limited access to clean water. Similarly, landslides are often linked to diarrhea, as affected populations frequently face poor hygiene in temporary shelters. In contrast, tsunamis are less commonly reported in the dataset. However, based on the available news, tsunamis are most associated with diarrhea, suggesting the persistence of sanitation-related health issues even in less-reported disaster events.

2. Word Cloud

Fig. 3. Word Cloud of Online News on Post-Disaster Diseases

The word cloud is used to illustrate the relationship between disasters and diseases in textual form, with the size of each word reflecting its frequency of occurrence in the dataset. From the visualization, it is evident that the terms “Flood”, “Landslide”, “Tsunami”, “Earthquake”, and “Drought” appear in larger sizes, indicating that these five types of disasters are frequently mentioned and have significant health impacts. In addition, diseases commonly associated with disaster events—such as acute respiratory infections (ARIs), stomach pain, diarrhea, skin diseases, leptospirosis, and hypertension—are prominently featured in the word cloud, suggesting that they often occur in post-disaster contexts. The presence of several medical terms such as pharyngitis, bronchitis, dysentery, and meningitis further highlights the elevated risk of serious infections following certain types of disasters. Visual representation of these findings is presented in Figure 3.

D. Model Evaluation

To assess the performance of the NER model, an evaluation was conducted using manually annotated data as the gold standard. A total of 1,137 news articles were annotated to mark relevant entities, including disaster types, health impacts, and affected locations. The dataset was then divided into three parts: 70% (796 articles) for training, 15% (170 articles) for validation, and 15% (171 articles) for testing. The evaluation employed two commonly used metrics: accuracy and F1 Score. Accuracy reflects the proportion of correct predictions over the total number of predictions. At the same time, the F1-score, calculated as the harmonic mean of precision and recall, provides a balanced view of the model’s ability to detect relevant entities. The evaluation was conducted using the test set, where the model’s predictions were compared to manually annotated entities. The seqeval Python library was used to compute the metrics. The results show that the fine-tuned IndoBERT model achieved an accuracy of 90.00% and an F1-score of 88.26%, indicating strong performance with a good balance between false positives and correct detections. The detailed performance results of the NER model are presented in Table III.

TABLE III. MODEL PERFORMANCE

Metrics	Score
Accuracy	90.00 %
Precision	87.91 %
Recall	88.61 %
F1-Score	88.26 %

IV. CONCLUSION

This study demonstrates that a text mining algorithm based on Named Entity Recognition (NER), supported by the IndoBERT model, can effectively extract information on disaster types, health impacts, and affected locations from online news articles. The developed method successfully identifies causal relationships between various natural disasters and post-disaster diseases, achieving an accuracy of 90% and an F1 score of 88%. The analysis results, visualized through bar charts and word clouds, provide a comprehensive overview of post-disaster health patterns. Events such as floods, eruptions, earthquakes, droughts, landslides, and even tsunamis have been found to have different associations with the emergence of diseases, including skin infections, diarrhea, fever, influenza, and acute respiratory infections (ARI).

Recommendations for future research include expanding the data collection scope by incorporating news from local or regional news portals to improve the representativeness of disaster events across Indonesia. Furthermore, the development of this system into a web-based application accessible in real-time is expected to enhance its utility for policymakers, medical professionals, and the public in addressing post-disaster health impacts. Such a web application could be integrated with health data from hospitals or public health agencies. The development of an AI-based monitoring system for the early detection of post-disaster diseases is also an important direction for future work.

ACKNOWLEDGMENT

I extend my sincere gratitude to the leadership of the Trisakti School of Insurance for providing the opportunity to conduct this research. Appreciation is also extended to the Research and Community Service Center for facilitating this work through financial support and other essential resources.

REFERENCES

[1] M. Mavrouli, S. Mavroulis, E. Lekkas, and A. Tsakris, "The impact of earthquakes on public health: A narrative review of infectious diseases in the post-disaster period aiming to disaster risk reduction," *Microorganisms*, vol. 11, no. 2, p. 419, 2023.

[2] World Health Organization, "Public Health Risk Assessment and Interventions Tropical Cyclone PAM: Vanuatu," 2015. [Online]. Available: https://iris.who.int/bitstream/handle/10665/254640/9789290617495-eng.pdf;jsessionid=F64851D50EC3DBA92D5D2575E84269AC?sequence=1.

[3] Kementrian Kesehatan, "Dampak Karhutla Bagi Kesehatan Masyarakat," *Pusat Krisis Kesehatan*, 2022. https://pusatkrisis.kemkes.go.id/dampak-karhutla-bagi-kesehatan-masyarakat (accessed Feb. 05, 2025).

[4] B. Beaglehole, R. T. Mulder, C. M. Frampton, J. M. Boden, G. Newton-Howes, and C. J. Bell, "Psychological distress and psychiatric disorder after natural disasters: systematic review and meta-analysis," *Br. J. Psychiatry*, vol. 213, no. 6, pp. 716–722, 2018.

[5] T. Powell, K. M. Wegmann, and E. Backode, "Coping and post-traumatic stress in children and adolescents after an acute onset disaster: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 9, p. 4865, 2021.

[6] S. Tahernejad, S. Ghaffari, A. Ariza-Montes, U. Wesemann, H. Farahmandnia, and A. Sahebi, "Post-traumatic stress disorder in medical workers involved in earthquake response: a systematic review and meta-analysis," *Heliyon*, vol. 9, no. 1, 2023.

[7] E. Kaya, E. I. Onal, S. Fatih, and O. Güler, "Prevalence and predictors of post-traumatic stress disorder among survivors of the 2023 earthquakes in Türkiye: The case of a temporary camp," *Int. J. Disaster Risk Reduct.*, vol. 114, p. 104976, 2024.

[8] F. Novika, I. Maulidi, B. Marsanto, and A. N. Amalina, "Comparasion Model Analysis Time of Earthquake Occurrence in Indonesia based on Hazard Rate with Single Decrement Method," *J. Teor. dan Apl. Mat.*, vol. 6, no. 1, pp. 163–176, 2022.

[9] I. Maulidi, F. Novika, V. Apriliani, and M. Syazali, "The Estimation of the hazard function of earthquakes in aceh province with likelihood approach," *Desimal J. Mat.*, vol. 7, no. 3, pp. 557–566, 2024, doi: 10.24042/djm.

[10] G. Gunawan, "Disaster event, preparedness, and response in Indonesian coastal areas: Data mining

of official statistics," *Int. J. Comput. Digit. Syst.*, vol. 16, no. 1, pp. 249–264, 2024.
[11] M. Sinambela *et al.*, *Mitigasi dan manajemen bencana*. Yayasan Kita Menulis, 2021.
[12] C. Wu *et al.*, "Natural language processing for smart construction: Current status and future directions," *Autom. Constr.*, vol. 134, p. 104059, 2022.
[13] M. A. Sit, C. Koylu, and I. Demir, "Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma," in *Social Sensing and Big Data Computing for Disaster Management*, Routledge, 2020, pp. 8–32.
[14] J. Sun, Y. Liu, J. Cui, and H. He, "Deep learning-based methods for natural hazard named entity recognition," *Sci. Rep.*, vol. 12, no. 1, p. 4598, 2022.
[15] M. Chipapi, "Automated disease outbreak detection and reporting system," Manipal, 2024. doi: 10.13140/RG.2.2.24858.04807.
[16] M. Kim, K. Chae, S. Lee, H.-J. Jang, and S. Kim, "Automated classification of online sources for infectious disease occurrences using machine-learning-based natural language processing approaches," *Int. J. Environ. Res. Public Health*, vol. 17, no. 24, p. 9467, 2020.
[17] J. Mangoma and W. Sulistiadi, "Island Health Crisis: Bridging Gaps in Indonesia's Healthcare Deserts," *J. Indones. Heal. Policy Adm.*, vol. 9, no. 2, p. 5, 2024.
[18] A. Mehmood, M. T. Zamir, M. A. Ayub, N. Ahmad, and K. Ahmad, "A named entity recognition and topic modeling-based solution for locating and better assessment of natural disasters in social media," *arXiv Prepr. arXiv2405.00903*, 2024.
[19] N. Bui *et al.*, "Fine-tuning large language models for improved health communication in low-resource languages," *Comput. Methods Programs Biomed.*, vol. 263, p. 108655, 2025.
[20] S. Amin, "Learning entity and relation representation for low-resource medical language processing," 2024.
[21] R. S. Wilkho and N. G. Gharaibeh, "FF-NER: A named entity recognition model for harvesting web-based information about flash floods and related infrastructure impacts," *Int. J. Disaster Risk Reduct.*, p. 105604, 2025.
[22] R. Suwaileh, T. Elsayed, M. Imran, and H. Sajjad, "When a disaster happens, we are ready: Location mention recognition from crisis tweets," *Int. J. Disaster Risk Reduct.*, vol. 78, p. 103107, 2022.
[23] J.-C. Klie, R. E. de Castilho, and I. Gurevych, "Analyzing dataset annotation quality management in the wild," *Comput. Linguist.*, vol. 50, no. 3, pp. 817–866, 2024.
[24] V. Vennila, A. Rajivkannan, S. Savitha, G. J. Santhosh, R. Jeevanantham, and K. Kavin, "Integrated T5 Neural Network and Spacy-Based AI Framework for Advanced Grammar and Speech Analysis," in *International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024)*, 2025, pp. 741–754.
[25] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," *arXiv Prepr. arXiv1603.01360*, 2016.
[26] B. Wilie *et al.*, "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," *arXiv Prepr. arXiv2009.05387*, 2020.
[27] M. M. K. Dandu, S. Singiri, S. Nadukuru, S. Jain, R. Agarwal, and S. P. Singh, "Unsupervised Information Extraction with BERT," *Int. J. Res. Mod. Eng. Emerg. Technol. 9*, vol. 1, 2021.
[28] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
[29] W. Holmes, *Speech synthesis and recognition*. CRC press, 2002.