

Perbandingan Kinerja Model NER IndoBERT dan IndoLEM dalam Ekstraksi Informasi Kesehatan Pascabencana dari Berita Daring di Indonesia

Comparative Performance of IndoBERT and IndoLEM Baseline Models for Post-Disaster Health Information Extraction from Indonesian Online News

Nalar Istiqomah^{*1}, Fanny Novika²

^{1,2}Aktuaria, Sekolah Tinggi Manajemen Asuransi Trisakti

E-mail: ¹nalar.istiqomah23@gmail.com, ²novikafanny@gmail.com

Received: June 06, 2025 | Revised: June 28, 2025 | Accepted: July 05, 2025

Abstrak

Bencana alam sering kali menyebabkan dampak signifikan terhadap kesehatan masyarakat, namun pemantauan penyakit pascabencana di Indonesia masih belum sistematis. Penelitian ini membandingkan kinerja dua model *Named Entity Recognition* (NER) dalam mengekstraksi informasi penyakit, lokasi terdampak, dan jenis bencana dari artikel berita daring berbahasa Indonesia. Model pertama adalah IndoBERT yang telah dilatih ulang (*fine-tuned*) menggunakan 1.137 artikel berita pascabencana, sedangkan model kedua merupakan *baseline* dari *benchmark* IndoLEM, yakni mBERT dan XLM-RoBERTa tanpa pelatihan domain spesifik. Hasil evaluasi menunjukkan bahwa IndoBERT mencapai akurasi 90,00% dan F1-Score 88,26%, jauh lebih tinggi dibanding mBERT (72,93%) dan XLM-R (76,44%). Analisis lebih lanjut terhadap entitas hasil ekstraksi menunjukkan tren spasial dan temporal penyakit: banjir di Pulau Jawa secara konsisten dikaitkan dengan diare dan penyakit kulit, sementara erupsi gunung di wilayah timur Indonesia berkorelasi dengan infeksi saluran pernapasan dan hipertensi. Temuan ini menekankan pentingnya pemilihan model yang sesuai untuk mendukung sistem pemantauan kesehatan berbasis data di wilayah rawan bencana.

Kata kunci: *Named Entity Recognition*, IndoBERT, IndoLEM, penyakit pascabencana, berita daring, bencana alam, pemodelan spasial-temporal

Abstract

Natural disasters often have significant impacts on public health, yet systematic monitoring of post-disaster diseases in Indonesia remains limited. This study compares the performance of two *Named Entity Recognition* (NER) models in extracting health impacts, affected locations, and disaster types from Indonesian-language online news articles. The first model is IndoBERT, fine-tuned using 1,137 manually validated disaster-related news articles. The second comprises baseline models from the IndoLEM benchmark, namely mBERT and XLM-RoBERTa, without domain-specific training. Evaluation results show that IndoBERT outperforms the baseline models, achieving 90.00% accuracy and an F1-score of 88.26%, compared to mBERT (72.93%) and XLM-R (76.44%). Further analysis of the extracted entities reveals spatial and temporal disease trends: floods in Java are consistently associated with diarrhea and skin diseases, while volcanic eruptions in eastern Indonesia are linked to respiratory infections and hypertension.

These findings highlight the importance of selecting appropriate models to support data-driven public health monitoring systems in disaster-prone regions.

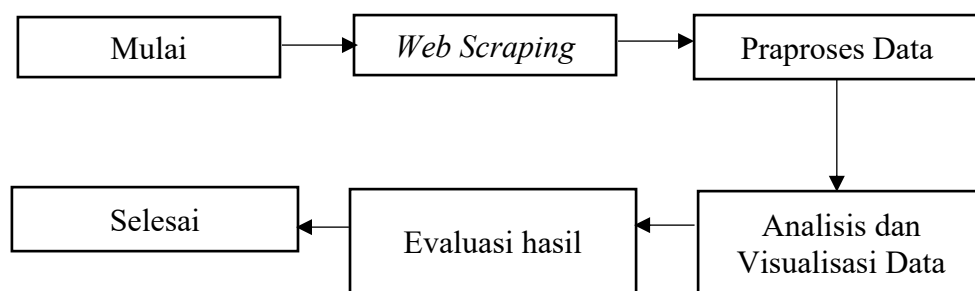
Keywords: Named Entity Recognition, IndoBERT, IndoLEM, post-disaster disease, online news, natural disaster, spatial-temporal modeling

1. PENDAHULUAN

Indonesia merupakan negara dengan tingkat risiko bencana alam yang sangat tinggi. Ratusan kejadian bencana terjadi setiap tahunnya dan berdampak pada jutaan penduduk [1]. Selain kerugian infrastruktur dan ekonomi, bencana juga memunculkan permasalahan kesehatan serius seperti penyakit infeksi, gangguan pernapasan, serta stres pascatrauma. Padahal, laporan dari United Nations Office for Disaster Risk Reduction menekankan bahwa keterlambatan respons kesehatan dapat memperburuk dampak bagi masyarakat terdampak [2]. Pemantauan penyakit pascabencana di Indonesia sebagian besar masih mengandalkan laporan manual dari fasilitas kesehatan atau instansi lokal, yang kerap mengalami keterlambatan dan ketidakkonsistenan distribusi data [3] [4] [5]. Kondisi ini melemahkan deteksi dini potensi wabah dan menghambat pengambilan keputusan berbasis bukti. Meningkatnya ketersediaan informasi digital menjadikan media berita daring sumber data alternatif yang menjanjikan. Berita lokal sering melaporkan kondisi lapangan lebih cepat dan rinci, termasuk keluhan kesehatan serta lokasi terdampak. Studi R. Rocca *et al.* (2023) dan G. F. Shidik *et al.* (2024) menunjukkan bahwa teks berita dapat dimanfaatkan untuk ekstraksi informasi darurat menggunakan *Natural Language Processing* (NLP) [6] [7]. Namun, ekstraksi manual dari ribuan artikel tidak efisien dan rawan bias [8]. Di sinilah teknik *Named Entity Recognition* (NER) berperan penting. Yadav & Bethard (2018) menunjukkan pendekatan berbasis deep-learning (BiLSTM, BERT) berperforma tinggi untuk NER di berbagai domain, termasuk kesehatan [9]. Beberapa penelitian telah dilakukan, di antaranya penggunaan BERT dapat mendeteksi gangguan mental [10] [11] ataupun dapat juga mendeteksi respon terhadap bencana [12]. *Benchmark IndoLEM*, yang pertama kali diperkenalkan oleh Koto *et al.* (2020), menyediakan korpus evaluasi multi-tugas beserta model *baseline* (mBERT, XLM-RoBERTa, dan varian IndoBERT) untuk menstandarkan pengujian NLP Bahasa Indonesia [13]. Pembaruan berikutnya dari Cahyawijaya *et al.*, 2021 memperluas cakupan *benchmark* tersebut. Walaupun pendekatan NER telah terbukti efektif, masih terbatas penelitian yang secara langsung membandingkan kinerja antar-model NER (IndoBERT *fine-tuned* vs. *baseline* IndoLEM) dalam konteks ekstraksi penyakit pascabencana dari berita daring [14]. Oleh karena itu, penelitian ini bertujuan mengevaluasi dan membandingkan performa IndoBERT yang telah dilatih ulang dengan *baseline* IndoLEM (mBERT dan XLM-RoBERTa) pada tugas ekstraksi entitas penyakit, lokasi, dan jenis bencana. Selain itu, penelitian ini juga menganalisis hasil ekstraksi untuk mengidentifikasi tren spasial dan temporal penyakit pascabencana di Indonesia. Menggunakan 1.137 artikel berita terverifikasi yang mencakup periode 2005–2025, studi ini diharapkan berkontribusi pada pengembangan sistem pemantauan kesehatan berbasis data guna mendukung kesiapsiagaan dan respons darurat di wilayah rawan bencana.

2. METODE PENELITIAN

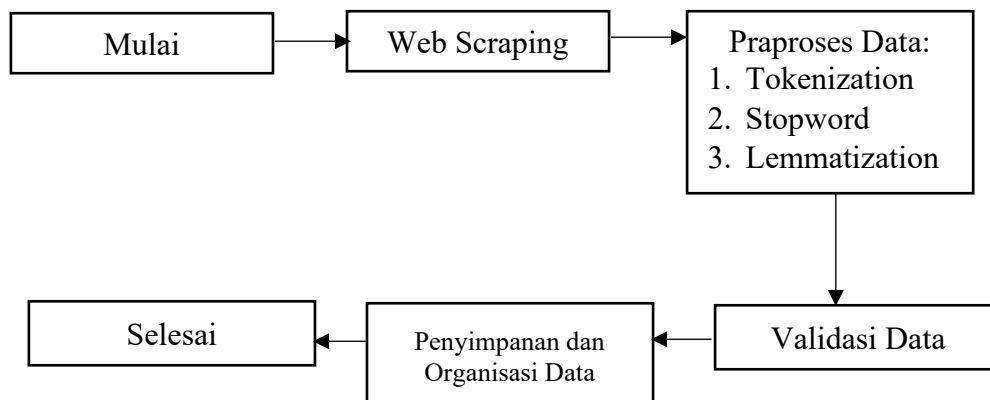
Penelitian ini menggunakan pendekatan kuantitatif berbasis data sekunder yang diperoleh secara otomatis melalui proses web scraping terhadap berita daring berbahasa Indonesia. Tujuan utama dari penelitian ini adalah mengekstraksi entitas yang berkaitan dengan jenis bencana, dampak kesehatan, dan lokasi terdampak menggunakan metode *Named Entity Recognition* (NER). Dalam konteks ini, penelitian juga membandingkan performa model IndoBERT yang telah dilatih ulang (*fine-tuned*) secara spesifik untuk domain bencana dan kesehatan, dengan dua model *baseline* dari *benchmark* IndoLEM, yaitu *Multilingual BERT* (mBERT) dan XLM-RoBERTa. Rangkaian tahapan penelitian secara umum meliputi pengumpulan data, validasi dan pembersihan, pra-pemrosesan, ekstraksi entitas, evaluasi model, serta analisis spasial dan temporal terhadap hasil ekstraksi. Tahapan penelitian ini disusun seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Penelitian ini menggunakan data berita daring dari portal berita yang kredibel. Gambar 2 adalah langkah pengumpulan data.



Gambar 2. Tahapan Pengumpulan Data

Penelitian ini diawali dengan pengumpulan data sekunder dari berita daring yang diperoleh melalui teknik *web scraping* menggunakan bahasa pemrograman Python. Portal berita yang dijadikan sumber adalah Kompas, CNN Indonesia, BBC Indonesia, dan Badan Nasional Penanggulangan Bencana (BNPB), yang dinilai memiliki kredibilitas tinggi dan jangkauan nasional. Proses *scraping* dilakukan dengan pustaka BeautifulSoup untuk parsing HTML,

Selenium untuk menangani interaksi dinamis, serta Requests untuk pengambilan data melalui API jika tersedia [15]. Data yang dikumpulkan mencakup judul, tanggal publikasi, isi berita, jenis bencana, lokasi kejadian, dan dampak kesehatan yang disebutkan dalam teks. Setelah data terkumpul, dilakukan proses validasi dan pembersihan. Duplikasi artikel yang berasal dari agregator atau media berbeda dihapus, dan berita yang tidak memuat informasi dampak kesehatan disaring. Validasi manual juga dilakukan pada sebagian data untuk memastikan bahwa isi berita memang relevan dengan tujuan penelitian. Pendekatan ini mengacu pada praktik validasi *corpus* dalam studi NLP bencana sebelumnya [16]. Langkah selanjutnya adalah *preprocessing* teks, yaitu proses pengolahan awal untuk menyiapkan data dalam format yang bisa diolah oleh model NLP. Tahapan ini meliputi tokenisasi, penghapusan *stopwords*, dan *lemmatization*. Tokenisasi dilakukan dengan pustaka spaCy [17], sementara *lemmatization* dan penghapusan kata umum dilakukan menggunakan spaCy dan Sastrawi, pustaka lemmatizer Bahasa Indonesia yang telah digunakan dalam beberapa studi klasifikasi teks [18]. Hasil *preprocessing* kemudian disimpan dalam format CSV dan JSON, yang masing-masing dipilih karena fleksibilitasnya dalam pengolahan tabular maupun berbasis objek untuk keperluan NLP [19].

2.2 Ekstraksi dan Analisis Entitas

Proses analisis utama dimulai dengan ekstraksi entitas menggunakan metode *Named Entity Recognition* (NER). Dalam penelitian ini, tiga model dievaluasi dan dibandingkan: (1) model IndoBERT yang telah dilatih ulang (*fine-tuned*) dengan data berita bencana berbahasa Indonesia, (2) model mBERT sebagai *baseline* multibahasa dari *benchmark* IndoLEM, dan (3) model XLM-RoBERTa sebagai *baseline* lintas bahasa lainnya dari *benchmark* yang sama [13].

Ketiga model digunakan untuk mengenali tiga kategori entitas: jenis bencana (misalnya banjir, gempa), dampak kesehatan (diare, ISPA, stres), dan lokasi geografis (nama kota, provinsi, atau wilayah).

2.3 Evaluasi Kinerja Model

Setelah proses ekstraksi selesai, dilakukan evaluasi untuk menilai sejauh mana model mampu mengenali entitas dengan benar. Evaluasi dilakukan dengan membandingkan hasil prediksi model dengan dataset anotasi manual yang berfungsi sebagai *gold standard*. Metrik evaluasi yang digunakan adalah akurasi, *precision*, *recall*, dan F1-Score, sebagaimana disarankan dalam studi NLP berbasis entitas [20] [21]

Perhitungan metrik dilakukan menggunakan pustaka *seqeval*, yang dirancang untuk evaluasi tugas *sequence labeling* seperti NER. Evaluasi ini menjadi dasar perbandingan kuantitatif antara performa model IndoBERT (*fine-tuned*) dan *baseline* IndoLEM.

2.4 Visualisasi dan Interpretasi Data

Hasil ekstraksi yang telah dievaluasi selanjutnya divisualisasikan untuk analisis spasial dan temporal. Visualisasi *word cloud* digunakan untuk menampilkan kata-kata yang paling sering muncul dalam konteks kesehatan. Grafik batang digunakan untuk menunjukkan hubungan antara jenis bencana dan penyakit yang paling umum. Selain itu, dilakukan pemetaan spasial dan analisis tren tahunan untuk mengidentifikasi pola kemunculan penyakit pascabencana, misalnya keterkaitan banjir dengan diare, atau kebakaran hutan dengan ISPA.. Visualisasi ini diharapkan dapat memperkaya hasil penelitian dan mendukung pengembangan sistem pemantauan kesehatan pascabencana yang berbasis data.

3. HASIL DAN PEMBAHASAN

2.1 Hasil Pengumpulan dan Pra-proses Data

Data dalam penelitian ini dikumpulkan secara otomatis menggunakan Python dengan metode *web scraping* dari berbagai portal berita daring yang kredibel, seperti Kompas, CNN Indonesia, Detik, Tempo, dan Tribunnews. Pengumpulan dilakukan dengan menggunakan kombinasi kata kunci yang terdiri dari jenis bencana alam (misalnya banjir, gempa, longsor) dan penyakit pascabencana (misalnya diare, ISPA, gatal-gatal). Kata kunci bencana mencakup tujuh kategori umum, seperti banjir, gempa bumi, longsor, kekeringan, tsunami, dan erupsi gunung berapi, sementara kata kunci penyakit mencakup lebih dari 20 jenis penyakit fisik dan psikologis yang umum ditemukan dalam laporan pascabencana. Contohnya, flu, sakit kepala, hipertensi, DBD, dehidrasi, demam, diare/disentri, gizi buruk, ISK, ISPA, kolera, penyakit kulit, malaria, nyeri sendi, typhoid, hepatitis, leptospirosis dan PTSD. Tabel 1 menunjukkan contoh hasil *scrapping* berita online.

Tabel 1. Contoh hasil *scrapping* berita

Tanggal	Link	Sumber	Judul	Isi Berita
06/08/ 2022	https://www.kompas.id/baca/nusantara/2022/08/06/warga-terdampak-bencana-kekeringan-di-lanny-jaya-capai-2740-orang	Kompas	Warga Terdampak Bencana Kekeringan di Lanny Jaya Bertambah Jadi 2.740 Orang	Warga di Distrik Kuyawage, Kabupaten Lanny Jaya, Papua, yang terdampak musibah kekeringan bertambah menjadi 2.740 orang. Masyarakat yang terdampak kekeringan itu membutuhkan bantuan makanan dan pelayanan kesehatan karena mereka rentan terserang diare. Manajer Pusat Pengendalian Operasi Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Papua Jonathan Koirewoa, Sabtu (6/8/2022), mengatakan, 2.740 warga terdampak itu berasal dari empat kampung yang dilanda kekeringan akibat fenomena

26/11/2019	https://regional.kompas.com/read/2019/11/26/20510431/korban-banjir-dan-longsor-solok-selatan-mulai-terserang-penyakit	Kompas	Korban Banjir dan Longsor Solok Selatan Mulai Terserang Penyakit	embun beku. Empat kampung ini adalah Uwone, Luare, Yugunomba, dan Tumbubur. Korban banjir dan longsor di Kabupaten Solok Selatan, Sumatera Barat (Sumbar), mulai terserang penyakit seperti Infeksi Saluran Pernafasan Akut (ISPA), demam, batuk, penyakit kulit dan diare. Empat puskesmas di daerah bencana disiagakan selama 24 jam. Bahkan 10 petugas kesehatan standby di dua lokasi pengungsian. "Selama masa tanggap darurat, empat puskesmas di daerah bencana disiagakan. Kita berikan pelayanan kepada korban secara gratis," kata Kepala Dinas Kesehatan Solok Selatan Novirman kepada Kompas.com, Selasa (26/11/2019). Baca juga: Evakuasi Korban Banjir, Basarnas dan Polisi Bergantian Gendong Kakek Sejauh 2 Km Novirman mengatakan berdasarkan data dari empat puskesmas yang disiagakan, terdapat lonjakan korban bencana sejak
------------	---	--------	--	---

mulai masa
tanggap
darurat.
"Selama 23-25
November total
99 korban
berobat dengan
jenis penyakit
mulai diare,
kulit, ISPA,
batuk dan
demam," kata
dia. Sementara
untuk di posko
pengungsian di
Balai Adat
Nagari Pakaan
Rabaa Timur
dan Simpang
Salak, Dinas
Kesehatan telah
menyiagakan
10 petugas
mulai dokter,
bidan, hingga
tenaga
sanitarian. "Di
posko
pengungsian
ada sekitar 330
warga. Kita
siagakan
petugas
sebanyak 10
orang,"
jelasnya. Baca
juga: BNPB:
340 Orang
Mengungsi
Akibat Banjir
di Solok
Selatan,
Sumbar
Diberitakan
sebelumnya,
Pemerintah
Kabupaten
Solok Selatan,
Sumatera Barat,
akhirnya
menetapkan
status tanggap
darurat akibat
bencana banjir
selama 14 hari,
terhitung mulai
Jumat
(22/11/2019).
Dampak banjir
yang
diakibatkan
cukup parah.
Dari data Sabtu
(23/11/2019)
siang, tercatat
ada empat
kecamatan yang
terdampai yakni
Koto Parik

30/11/2022	https://www.cnnindonesia.com/nasional/20221130184936-20-881003/krisis-air-bersih-pengungsi-gempa-cianjur-mulai-terserang-diare	CNN Indonesi a	Krisis Air Bersih, Pengungsi Gempa Cianjur Mulai Terserang Diare	Gadang Diateh, Sungai Pagu dan Sangir, Pauh Duo ""Warga terdampak gempa Cianjur yang mengungsi di posko pengungsian mulai terserang berbagai penyakit seperti flu, demam, batuk hingga diare. Nurdin (28) warga Desa Rancagoong, Kecamatan Cilaku, Cianjur yang merupakan salah seorang pengungsi mengatakan, posko pengungsian yang jauh dari layak huni serta krisis air bersih diduga menjadi penyebab penyebab banyak warga yang terserang penyakit. Faktor cuaca di lokasi yang kerap dilanda hujan memperburuk kondisi kesehatan warga. ""Faktornya sih selain dari trauma atau pikiran dan emosi yang tidak stabil. Banyak posko yang berdiri di atas sawah, lapangan becek, hujan, cuaca angin besar, dan sanitasi yang kurang baik,"" ujar
------------	---	----------------------	---	--

Nurdin, Rabu
(30/11).
Tak hanya
warga korban
gempa, para
relawan yang
berada di lokasi
juga mulai
terserang
penyakit.
""Jangankan
korban atau
warga yang
terdampak,
beberapa
relawan juga
ada yang mulai
berjatuh
sakit,""
katanya.

Setelah proses *scraping* selesai, terkumpul 6.464 artikel berita dari tahun 2005 hingga Maret 2025. Namun, setelah dilakukan proses validasi dan pembersihan, hanya 1.137 berita (17,59%) yang benar-benar relevan, yakni memuat informasi tentang penyakit yang terjadi sebagai akibat langsung dari bencana alam di Indonesia. Validasi dilakukan secara manual untuk menyaring artikel yang meskipun mengandung kata kunci, namun tidak sesuai konteks, seperti berita tentang penyakit pada hewan, perumpamaan, atau istilah metaforis (contoh: “tsunami Covid-19”).

Data yang lolos validasi kemudian diproses lebih lanjut melalui tahapan *preprocessing* yang mencakup *case folding*, *tokenizing*, penghapusan *stopwords*, dan *stemming* dengan menggunakan pustaka Python seperti spaCy dan Sastrawi (Explosion AI, 2023; Siswanto & Dani, 2021). Hasil akhir berupa dataset siap pakai dalam format CSV dan JSON, yang digunakan dalam proses pelabelan manual dan pelatihan model NER.

2.2 Analisis Data

Analisis data dalam penelitian ini dilakukan dengan menerapkan tiga model *Named Entity Recognition* (NER) untuk mengekstrak tiga kategori entitas utama dari berita daring yaitu jenis bencana, dampak kesehatan, dan lokasi terdampak. Ketiga model yang dibandingkan adalah:

- (1) IndoBERT yang telah dilatih ulang (*fine-tuned*) menggunakan dataset berita pascabencana,
- (2) mBERT (*Multilingual BERT*) sebagai *baseline* multibahasa dari IndoLEM, dan
- (3) XLM-RoBERTa, *baseline* lintas bahasa dari IndoLEM.

Setelah melalui tahap pelabelan manual, pelatihan model, dan penerapan NER terhadap 1.137 artikel berita, diperoleh hasil ekstraksi yang kemudian dianalisis lebih lanjut untuk melihat kemampuan masing-masing model dalam mengenali entitas penting yang berkaitan dengan bencana dan dampak kesehatannya. Tabel 2 merupakan contoh hasil ekstraksi entitas NER.

Tabel 2. Contoh hasil ekstraksi entitas masing-masing model NER

No	Kalimat Berita	Model	Jenis Bencana	Dampak Kesehatan	Lokasi Terdampak
1	Banjir bandang yang terjadi di Kabupaten Luwu Utara menyebabkan puluhan warga mengalami infeksi saluran pernapasan akut.	IndoBERT	banjir bandang	infeksi saluran pernapasan akut	Kabupaten Luwu Utara
		mBERT	banjir	infeksi saluran pernapasan	Luwu
		XLM-R	banjir bandang	-	Luwu Utara
2	Gempa bumi berkekuatan 5,6 SR mengguncang Kota Mataram, menyebabkan beberapa warga mengalami luka ringan dan trauma.	IndoBERT	gempa bumi	luka ringan	Kota Mataram
		mBERT	gempa	-	Mataram
		XLM-R	gempa bumi	-	—
3	Kabut asap akibat kebakaran hutan di Kalimantan Tengah mengakibatkan meningkatnya kasus asma dan iritasi mata di kalangan anak-anak.	IndoBERT	kebakaran hutan	asma, iritasi mata	Kalimantan Tengah
		mBERT	kebakaran	-	Kalimantan
		XLM-R	kebakaran hutan	-	Kalimantan Tengah
4	Usai banjir melanda wilayah Indramayu, sejumlah warga terserang diare akibat air bersih yang tercemar.	IndoBERT	banjir	diare	Indramayu
		mBERT	banjir	-	Indramayu
		XLM-R	banjir	-	-

Pada bagian ini, dijabarkan secara rinci hasil dari setiap tahapan analisis, mulai dari efektivitas model dalam mengenali entitas, perbedaan performa antar model, hingga contoh kesalahan yang umum terjadi.

a. Hasil Ekstraksi Entitas

Model IndoBERT yang telah *define-tune* menunjukkan performa terbaik dalam mengenali entitas secara lengkap dan akurat. IndoBERT berhasil mengekstrak entitas seperti “diare”, “infeksi saluran pernapasan akut (ISPA)”, serta lokasi-lokasi lokal seperti “Kabupaten Cianjur” atau “Kupang” dengan sangat baik. Model ini juga unggul dalam menangkap frasa penyakit yang panjang atau majemuk, seperti “penyakit kulit akibat genangan air” dan “gizi buruk pascabencana”.

Sebaliknya, model mBERT sering mengalami kesulitan dalam mengenali istilah lokal atau singkatan, seperti “ISPA” atau “DBD”, serta keliru dalam memisahkan entitas geolokasi dan penyakit. Dalam beberapa kasus, mBERT mengklasifikasikan kata “banjir” sebagai nama tempat atau entitas netral. Model ini juga kerap gagal mengenali entitas penyakit jika tidak berada dalam kalimat yang sangat umum atau eksplisit.

XLM-RoBERTa menunjukkan performa lebih baik dibanding mBERT dalam mengenali entitas lokasi, terutama nama provinsi dan kota besar. Namun, XLM-R juga sering mengelompokkan kata secara tidak tepat, misalnya menggabungkan “sakit kepala dan demam” menjadi satu entitas yang keliru (“sakit kepala dan demam” → satu tag entitas tunggal, padahal seharusnya dua). Selain itu, XLM-R mengalami penurunan akurasi pada frasa penyakit yang tidak umum atau medis spesifik, seperti “leptospirosis” atau “disentri”.

4. Perbedaan dalam Jumlah Entitas yang Terdeteksi

Tabel 3 menunjukkan jumlah entitas yang berhasil dikenali oleh masing-masing model berdasarkan estimasi terhadap total entitas aktual dalam data uji. IndoBERT (yang telah *define-tune*) berhasil mengenali hampir seluruh entitas secara tepat, terutama pada kategori penyakit

dan lokasi terdampak, yang sering kali menggunakan istilah lokal atau domain-spesifik. Sebaliknya, mBERT menunjukkan performa paling rendah dalam mengenali penyakit karena keterbatasannya dalam memahami struktur dan istilah dalam Bahasa Indonesia, apalagi yang bersifat domain medis. XLM-RoBERTa menunjukkan kinerja menengah dan cenderung lebih baik dalam mendeteksi nama lokasi dibanding mBERT, namun masih tertinggal dari IndoBERT dalam mengenali entitas penyakit kompleks atau majemuk

Tabel 3. Perbandingan Jumlah Entitas yang Berhasil Diekstraksi oleh Masing-Masing Model

Kategori Entitas	IndoBERT (<i>fine-tuned</i>)	mBERT (<i>baseline</i>)	XLM-RoBERTa (<i>baseline</i>)
Jenis Bencana	100% (semua dikenali)	±88%	±90%
Penyakit	Sangat baik (90–95%)	Rendah (70–75%)	Menengah (75–80%)
Lokasi	Sangat baik (95%)	Cenderung rendah (70–80%)	Baik (85–90%)
Terdampak			

5. Contoh Kesalahan Model (*Error Analysis*)

Tabel 4 menampilkan beberapa contoh kalimat dari korpus berita yang digunakan untuk menguji model. Dari hasil ini terlihat bahwa model IndoBERT mampu mengidentifikasi entitas secara tepat, bahkan ketika entitas muncul dalam frasa majemuk atau tidak eksplisit. Sebaliknya, mBERT cenderung gagal mengenali entitas jika tidak muncul dalam bentuk baku atau jika konteksnya terlalu lokal. XLM-RoBERTa cenderung mengalami kesalahan dalam segmentasi (*chunking*), yaitu menggabungkan beberapa kata menjadi satu entitas secara keliru, atau melewatkan sebagian entitas yang sebenarnya relevan.

Tabel 4. Contoh Kesalahan Model dalam Mengenali Entitas pada Kalimat Berita

Kalimat Asli	IndoBERT	mBERT	XLM-RoBERTa
“Puluhan anak pengungsi di Kupang terserang diare dan demam setelah banjir.”	diare, demam, Kupang	demam saja (diare tidak dikenali)	demam dan Kupang digabung
“ISPA menyebar di tenda-tenda pengungsian akibat abu vulkanik.”	ISPA	Tidak dikenali sama sekali	“ISPA menyebar” jadi satu entitas, padahal harusnya hanya ISPA yang terdeteksi
“Trauma dan hipertensi dilaporkan meningkat di lokasi pengungsian.”	trauma, hipertensi	trauma saja (hipertensi tidak terdeteksi)	trauma saja (hipertensi tidak terdeteksi)
“Ratusan warga terserang penyakit kulit dan diare usai banjir bandang.”	penyakit kulit, diare	hanya diare (penyakit kulit tidak terdeteksi)	penyakit kulit terdeteksi, diare tidak terdeteksi

Tabel 5 menyajikan nilai F1-score untuk tiga kategori entitas utama yang diekstraksi menggunakan tiga model berbeda: IndoBERT (yang telah dilatih ulang dengan data domain spesifik), mBERT, dan XLM-RoBERTa sebagai *baseline* dari *benchmark* IndoLEM. Tiga

kategori entitas utama tersebut adalah B-DISASTER (jenis bencana), B-HEALTH (penyakit atau dampak kesehatan), dan B-LOC (lokasi terdampak).

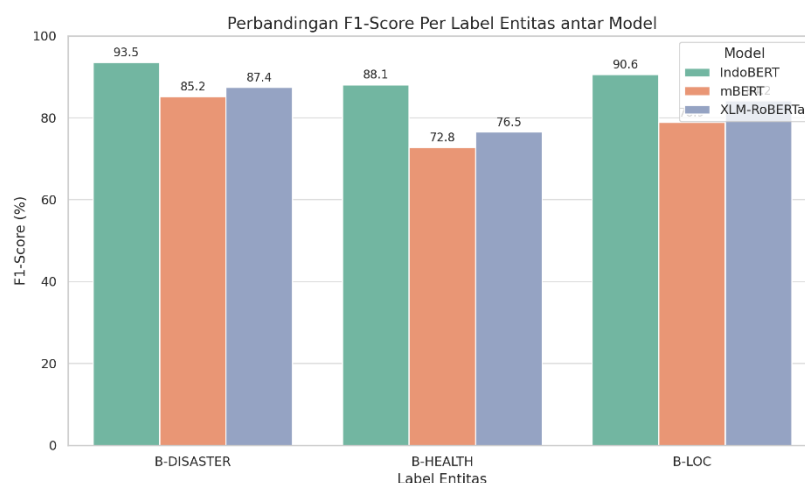
Berdasarkan tabel tersebut, IndoBERT menunjukkan kinerja tertinggi secara konsisten di ketiga kategori entitas. Untuk label B-DISASTER, model ini mencapai *F1-score* sebesar 93,5%, jauh lebih unggul dibandingkan mBERT (85,2%) dan XLM-RoBERTa (87,4%). Ini menunjukkan bahwa IndoBERT mampu mengenali frasa atau kata kunci terkait bencana secara akurat, meskipun disampaikan dengan variasi linguistik khas media lokal.

Tabel 5. F1-Score per Label Entitas untuk Masing-Masing Model

Label Entitas	IndoBERT (<i>fine-tuned</i>)	mBERT (<i>baseline</i>)	XLM-RoBERTa (<i>baseline</i>)
B-DISASTER	93.5%	85.2%	87.4%
B-HEALTH	88.1%	72.8%	76.5%
B-LOC	90.6%	78.9%	84.2%
Rata-rata (Macro F1)	90.7%	78.9%	82.7%

Kesenjangan performa paling mencolok terjadi pada label B-HEALTH. Di sini, IndoBERT meraih *F1-score* sebesar 88,1%, sementara mBERT hanya mencapai 72,8%, dan XLM-RoBERTa sebesar 76,5%. Nilai yang relatif rendah pada dua model *baseline* menunjukkan bahwa mereka kesulitan mengenali entitas penyakit atau dampak kesehatan, khususnya yang menggunakan istilah non-baku, singkatan medis, atau frasa panjang yang tidak umum dalam korpus multibahasa. Hal ini mendukung hipotesis bahwa *fine-tuning* model pada data domain-spesifik dapat secara signifikan meningkatkan akurasi ekstraksi pada konteks *low-resource* seperti ini.

Untuk label B-LOC, ketiga model tampil relatif baik, namun IndoBERT tetap unggul dengan skor 90,6%, dibanding XLM-RoBERTa (84,2%) dan mBERT (78,9%). Performa lebih tinggi pada XLM-R untuk label ini dapat dijelaskan oleh kekuatan arsitektur RoBERTa dalam mengenali nama entitas geografis lintas bahasa, namun IndoBERT tetap lebih unggul karena sensitivitasnya terhadap struktur kalimat dan istilah lokal.



Gambar 3. F1-Score per Label Entitas untuk Masing-Masing Model

Visualisasi hasil ini diperkuat dalam Gambar 3, yang menampilkan grafik batang perbandingan F1-score antar model pada masing-masing label entitas. Grafik tersebut secara visual menggarisbawahi keunggulan IndoBERT di seluruh kategori. Selain menampilkan nilai absolut, grafik juga membantu menyoroti bahwa *gap* performa paling besar terjadi pada label B-HEALTH, yang merupakan fokus utama dalam studi ini.

Secara keseluruhan, baik data tabel maupun grafik memperkuat temuan bahwa IndoBERT yang dilatih secara khusus untuk tugas NER pada teks berita pascabencana di Indonesia jauh lebih efektif dibanding model *baseline*. Ini juga menunjukkan pentingnya pendekatan *fine-tuning* untuk tugas NER dalam domain spesifik, terutama di bahasa dengan sumber daya terbatas seperti Bahasa Indonesia.

Analisis kesalahan ini penting dilakukan karena menunjukkan bagaimana pelatihan ulang model dengan data yang spesifik secara domain (*fine-tuning*) dapat secara signifikan meningkatkan sensitivitas dan presisi model dalam konteks *low-resource* seperti Bahasa Indonesia dan subdomain pascabencana. Temuan ini konsisten dengan studi sebelumnya dalam *benchmark* IndoNLU dan IndoLEM (Wilie et al., 2020; Koto et al., 2020).

6. Temuan Umum

Berdasarkan hasil evaluasi dan analisis ekstraksi entitas, dapat disimpulkan bahwa model IndoBERT menunjukkan performa terbaik dibandingkan dua model *baseline* lainnya. Keunggulan IndoBERT terutama terletak pada kemampuannya mengenali entitas secara akurat dalam konteks Bahasa Indonesia yang bersifat domain-spesifik. Karena telah melalui proses *fine-tuning* menggunakan data yang relevan dengan bencana dan kesehatan, model ini lebih sensitif terhadap istilah lokal, singkatan umum seperti “ISPA” atau “DBD”, serta frasa penyakit yang kompleks atau majemuk. Hal ini membuat IndoBERT sangat andal dalam mengekstraksi informasi dari teks berita yang sering kali tidak menggunakan terminologi medis baku.

Sementara itu, mBERT menunjukkan keterbatasan dalam memahami struktur kalimat berbahasa Indonesia, terutama dalam mengenali entitas yang tidak eksplisit atau bersifat lokal. Sebagai model multibahasa yang tidak dilatih khusus untuk Bahasa Indonesia, mBERT cenderung melewati entitas penyakit dan terkadang salah mengklasifikasikan entitas, seperti menganggap nama bencana sebagai nama tempat. Hal ini menunjukkan bahwa model ini kurang optimal untuk konteks berita lokal Indonesia, terutama yang berkaitan dengan isu-isu kesehatan pascabencana.

XLM-RoBERTa memberikan hasil yang lebih baik dibandingkan mBERT, khususnya dalam mengenali nama-nama tempat atau lokasi geografis yang umum disebutkan dalam berita. Namun, model ini masih rentan terhadap kesalahan dalam segmentasi entitas (*chunking error*), terutama pada frasa penyakit yang panjang. Beberapa kesalahan umum yang ditemukan adalah penggabungan dua entitas menjadi satu atau pengenalan entitas yang tidak lengkap. Meskipun demikian, performa XLM-R masih berada di bawah IndoBERT, terutama dalam hal konsistensi dan ketepatan deteksi entitas pada kalimat-kalimat kompleks.

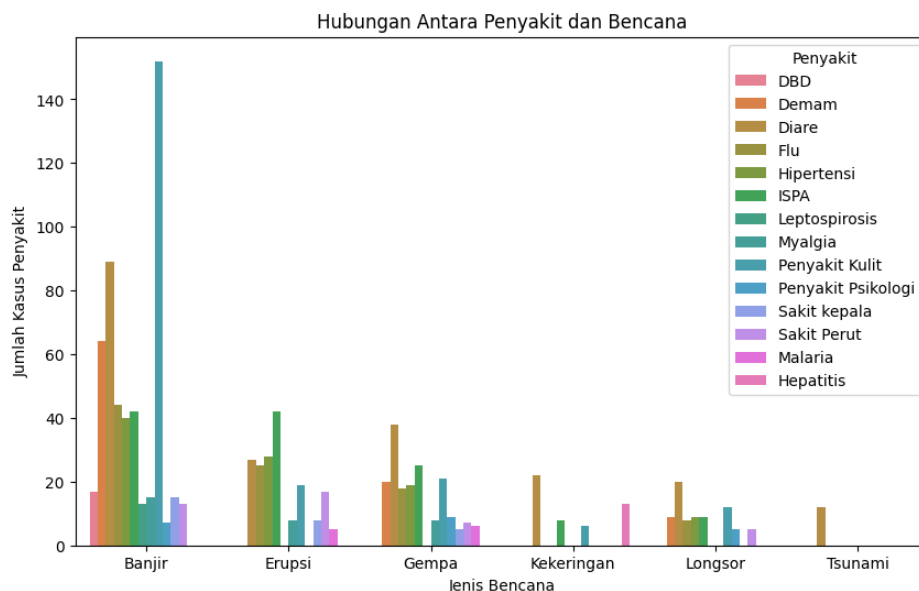
Secara keseluruhan, hasil ini memperkuat temuan bahwa pelatihan ulang model bahasa dengan data yang sesuai konteks domain sangat krusial, terutama dalam tugas-tugas ekstraksi informasi untuk bahasa dengan sumber daya terbatas seperti Bahasa Indonesia.

2.3 Visualisasi Data

Berdasarkan hasil evaluasi performa model NER, diketahui bahwa IndoBERT yang telah melalui proses *fine-tuning* memiliki kinerja tertinggi dibandingkan mBERT dan XLM-RoBERTa, dengan F1-score rata-rata sebesar 90,7%. Oleh karena itu, visualisasi pada bagian ini hanya didasarkan pada hasil ekstraksi entitas dari model IndoBERT, guna memperoleh representasi data yang paling akurat. Visualisasi digunakan untuk mengidentifikasi tren penyakit pascabencana di Indonesia selama dua dekade terakhir berdasarkan hubungan antara jenis bencana dan dampak kesehatan yang muncul.

Analisis hubungan antara penyakit dan bencana divisualisasikan dalam dua bentuk utama, yaitu grafik batang dan *word cloud*. Kedua visualisasi ini membantu dalam memahami pola kejadian penyakit akibat bencana alam.

1. Grafik Batang



Gambar 4. Grafik Batang Hubungan Antara Penyakit dan Bencana Alam

Grafik batang menampilkan jumlah kasus penyakit berdasarkan jenis bencana. Beberapa temuan penting dari grafik ini adalah:

1. Banjir memiliki dampak terbesar terhadap kesehatan masyarakat, dengan penyakit kulit menjadi yang paling dominan, diikuti oleh diare, demam, flu dan ISPA.
2. Erupsi gunung berapi berhubungan erat dengan penyakit pernapasan seperti ISPA, hipertensi, diare dan flu.
3. Gempa bumi menunjukkan korelasi tinggi dengan diare, ISPA, penyakit kulit dan demam.
4. Kekeringan berkaitan erat dengan diare akibat buruknya sanitasi yang disebabkan keterbatasan air bersih.
5. Longsor sering dikaitkan dengan diare, akibat sanitasi yang buruk di pengungsian.
6. Tidak seperti bencana lainnya, tsunami tidak banyak disebutkan di berita. Namun, dari berita, dapat diketahui bahwa tsunami erat kaitannya dengan diare.

2. Word Cloud

Word cloud digunakan untuk menggambarkan hubungan antara bencana dan penyakit dalam bentuk teks dengan ukuran yang mencerminkan frekuensi kemunculan suatu kata. Temuan utama dari word cloud ini meliputi:

1. Kata "Banjir", "Longsor", "Tsunami", "Gempa" dan "Kekeringan" muncul dalam ukuran besar, menunjukkan bahwa kelima bencana ini memiliki dampak kesehatan yang cukup signifikan.
2. Penyakit yang dominan akibat bencana, seperti ISPA, sakit perut, diare, penyakit kulit, leptospirosis dan hipertensi, tampak lebih besar dalam visualisasi, menandakan seringnya kasus tersebut terjadi pasca-bencana.

Beberapa istilah medis seperti faringitis, bronkitis, disentri dan meningitis juga muncul, menunjukkan adanya risiko infeksi serius setelah bencana tertentu.



Gambar 5. *Word cloud* berita online penyakit pasca bencana

Hasil visualisasi ini menunjukkan bahwa penyakit yang muncul setelah bencana bersifat spesifik terhadap jenis bencananya. Misalnya, banjir paling sering dikaitkan dengan penyakit kulit dan diare akibat sanitasi buruk, sementara erupsi gunung api cenderung berkaitan dengan ISPA dan hipertensi. Pola-pola ini memberikan wawasan penting yang dapat dimanfaatkan dalam sistem peringatan dini dan mitigasi risiko kesehatan pascabencana.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa pendekatan *Named Entity Recognition* (NER) berbasis model transformer, khususnya IndoBERT yang telah melalui proses *fine-tuning*, secara signifikan lebih unggul dibandingkan dua *baseline* model dari IndoLEM (mBERT dan XLM-RoBERTa) dalam mengekstraksi informasi penyakit pascabencana dari berita daring berbahasa Indonesia. IndoBERT mencatat F1-score rata-rata tertinggi dan menunjukkan keunggulan dalam mengenali entitas penyakit yang kompleks, istilah lokal, serta nama lokasi yang tidak selalu eksplisit, menjadikannya model yang paling andal untuk konteks domain bencana dan kesehatan. Hasil

ekstraksi menunjukkan bahwa banjir merupakan bencana yang paling sering dikaitkan dengan penyakit seperti diare, ISPA, dan penyakit kulit, sedangkan bencana seperti erupsi dan kekeringan cenderung terkait dengan gangguan pernapasan dan penyakit akibat sanitasi buruk. Temuan ini mengindikasikan bahwa berita daring dapat menjadi sumber data alternatif yang cepat dan informatif untuk pemantauan kesehatan pascabencana. Berdasarkan temuan tersebut, disarankan agar pelatihan ulang model NLP dilakukan menggunakan data lokal dan domain-spesifik untuk meningkatkan akurasi ekstraksi informasi. Selain itu, perlu dikembangkan sistem deteksi dini berbasis berita daring yang dapat membantu pemerintah dan otoritas terkait dalam merespons risiko kesehatan dengan lebih cepat dan tepat. Penelitian selanjutnya dapat memperluas cakupan dengan menggabungkan analisis hubungan antarentitas, pelibatan media sosial sebagai sumber data, serta integrasi hasil ekstraksi ke dalam dashboard visual interaktif yang dapat dimanfaatkan dalam pengambilan keputusan tanggap darurat berbasis data.

UCAPAN TERIMA KASIH

Saya mengucapkan terima kasih yang sebesar-besarnya kepada pimpinan Sekolah Tinggi Manajemen Asuransi Trisakti atas kesempatan yang diberikan untuk melaksanakan penelitian ini. Apresiasi juga saya sampaikan kepada Pusat Penelitian dan Pengabdian kepada Masyarakat atas dukungan finansial dan berbagai sumber daya penting lainnya yang telah memfasilitasi penelitian ini.

DAFTAR PUSTAKA

- [1] B. Warbung, K. Kusuma, B. Wahyudi, M. N. Gibran, and P. Widodo, "Strategi Penerapan Teknologi IOT dalam Sistem Komunikasi Kebencanaan di Indonesia," *Nusant. J. Ilmu Pengetah. Sos.*, vol. 11, no. 8, pp. 3108–3117, 2024.
- [2] GAR, *Global Assessment Report on Disaster Risk Reduction 2022: Our World at Risk: Transforming Governance for a Resilient Future*. 2022.
- [3] R. Febtrina *et al.*, "Mitigasi Bencana Banjir di Desa Palung Raya: Dampak Kesehatan dan Upaya Penanggulangannya," *SIGDIMAS*, vol. 2, no. 1, pp. 1–12, 2024.
- [4] I. D. M. Arsyad, A. Md, A. A. Arsyad, and M. A. F. I. Aslim, *Mitigasi Bencana di Lingkungan Kawasan Karst*. Indonesia Emas Group, 2025.
- [5] 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.
- [6] R. Rocca, N. Tamagnone, S. Fekih, X. Contla, and N. Rekabsaz, "Natural language processing for humanitarian action: Opportunities, challenges, and the path toward humanitarian NLP," *Front. big Data*, vol. 6, p. 1082787, 2023.
- [7] G. F. Shidik *et al.*, "Indonesian disaster named entity recognition from multi source information using bidirectional LSTM (BiLSTM)," *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 3, p. 100358, 2024.
- [8] 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.
- [9] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," *arXiv Prepr. arXiv1910.11470*, 2019.
- [10] J. Sunjaya, J. Ong, R. F. Ziliwu, H. Risni, and A. Pratama, "Applying BERT Model for Early Detection of Mental Disorders Based on Text Input," *J. Ilm. Tek. Inform. dan Komun.*, vol. 5, no. 2, pp. 660–670, 2025.

- [11] M. Košprdić, N. Prodanović, A. Ljajić, B. Bašaragin, and N. Milošević, "From zero to hero: harnessing transformers for biomedical named entity recognition in zero-and few-shot contexts," *Artif. Intell. Med.*, vol. 156, p. 102970, 2024.
- [12] A. O. Alharm and S. Naim, "Enhancing natural disaster response: A deep learning approach to disaster sentiment analysis using bert and lstm," *Available SSRN 4755638*.
- [13] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," *arXiv Prepr. arXiv2011.00677*, 2020.
- [14] S. Cahyawijaya *et al.*, "IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation," *arXiv Prepr. arXiv2104.08200*, 2021.
- [15] J. L. Arsianto and T. K. Gautama, "Pengembangan Aplikasi Web Scraping untuk Crawling Web Data dari Situs E-Commerce Properti," *J. Strateg. Maranatha*, vol. 7, no. 1, pp. 151–163, 2025.
- [16] 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.
- [17] P. Bhadekar, R. Gavali, R. Chavan, A. Karve, and P. Shelke, "Legal Document Summarizer using Spacy and BART.," *Grenze Int. J. Eng. Technol.*, vol. 10, 2024.
- [18] V. P. Vasani, S. C. Pawar, S. Ahamad, A. Sahu, and G. Talele, "Transformer Models for Enhanced Natural Language Processing in Medical Records Management," in *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 2024, pp. 1808–1814.
- [19] Z. Hu, W. Hou, and X. Liu, "Deep learning for named entity recognition: a survey," *Neural Comput. Appl.*, vol. 36, no. 16, pp. 8995–9022, 2024.
- [20] A. Salam and S. R. Sidiq, "SciBERT Optimisation for Named Entity Recognition on NCBI Disease Corpus with Hyperparameter Tuning," *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 432–441, 2025.
- [21] S. Kanbara, T. Ando, and R. Shaw, "Enhancing Planetary Health Through Data Visualization and Transdisciplinary Actions: A Case of Primary Healthcare and Climate Disaster Response in Japan," *Available SSRN 4519610*, 2023.