
**INDOBERTWEET DENGAN TEMPORAL ATTENTION MECHANISM
UNTUK DETEKSI ISU BENCANA DINAMIS MULTIPLATFORM****Nurhayati¹, Tanti², Nuraina³, Arisman⁴, Felix⁵****Universitas Mikroskil, Sumatera Utara**E-mail: ¹nurhayati@mikroskil.ac.id, ²tanti@mikroskil.ac.id, ³nuraina@mikroskil.ac.id,
⁴arisman@mikroskil.ac.id, ⁵felix.pandi@mikroskil.ac.id

Abstract: *The development of disaster-related information in the digital space is very rapid and is often detected earlier through social media than through official government channels. This condition highlights the need for a system capable of detecting disaster-related issues quickly and dynamically across various digital platforms. This study aims to develop a dynamic disaster issue detection model based on IndoBERTweet with a Temporal Attention Mechanism using multiplatform big text data in Indonesia. The research methodology includes collecting textual data from various digital platforms such as Twitter, YouTube, TikTok, Quora, and Medium. The data are processed through preprocessing stages using Natural Language Processing (NLP) techniques, timestamp extraction to obtain temporal information, and the application of fine-grained labeling for more detailed classification of disaster-related issues. Subsequently, the IndoBERTweet model is trained with a Temporal Attention Mechanism to capture the relationship between textual context and temporal dynamics in the development of disaster-related issues. The expected results of this research are a model capable of dynamically detecting disaster-related issues by considering informal language contexts and temporal changes. This model is expected to support early warning systems and data-driven disaster management decision-making in Indonesia.*

Keywords: *Disaster Issue Detection; Social Media Text Analysis; Multiplatform Big Data; IndoBERTweet; Temporal Attention.*

Abstrak: Perkembangan informasi kebencanaan di ruang digital berlangsung sangat cepat dan sering kali lebih dahulu terdeteksi melalui media sosial dibandingkan melalui kanal resmi pemerintah. Kondisi ini menunjukkan perlunya sistem yang mampu mendeteksi isu bencana secara cepat dan dinamis dari berbagai platform digital. Penelitian ini bertujuan mengembangkan model deteksi isu bencana dinamis berbasis IndoBERTweet dengan Temporal Attention Mechanism pada big data teks multiplatform di Indonesia. Metode penelitian meliputi pengumpulan data teks dari berbagai platform digital seperti Twitter, YouTube, TikTok, Quora, dan Medium. Data diproses melalui tahapan preprocessing menggunakan teknik Natural Language Processing (NLP), ekstraksi timestamp untuk memperoleh informasi temporal, serta penerapan fine-grained labeling untuk klasifikasi isu bencana yang lebih rinci. Selanjutnya, model IndoBERTweet dilatih dengan Temporal Attention Mechanism untuk menangkap hubungan antara konteks teks dan dinamika waktu dalam perkembangan isu bencana. Hasil penelitian diharapkan menghasilkan model yang mampu mendeteksi isu bencana secara dinamis dengan mempertimbangkan konteks bahasa informal dan perubahan waktu. Model ini diharapkan mendukung sistem peringatan dini dan pengambilan kebijakan kebencanaan berbasis data di Indonesia.

Kata Kunci: Deteksi Isu Bencana; Analisis Teks Media Sosial; Big Data Multiplatform; IndoBERTweet; Temporal Attention.

mengubah cara masyarakat memperoleh dan menyebarkan informasi, termasuk informasi terkait bencana alam. Informasi mengenai kejadian bencana sering muncul lebih cepat melalui platform digital seperti Twitter, YouTube, TikTok, Quora, dan Medium dibandingkan kanal resmi pemerintah [1]. Hal ini menunjukkan bahwa media sosial memiliki potensi besar sebagai sumber data untuk mendeteksi isu kebencanaan secara cepat. Namun, tingginya volume data, keragaman bahasa, serta dinamika perkembangan informasi dari waktu ke waktu menjadi tantangan dalam analisis isu bencana secara efektif [2], [3].

Penelitian dalam bidang NLP telah banyak digunakan untuk menganalisis teks media sosial guna mengidentifikasi isu publik. Model berbasis Transformer seperti BERT dan RoBERTa menunjukkan performa yang baik dalam tugas klasifikasi teks dan analisis informasi pada data berskala besar [4], [5]. Beberapa penelitian juga memanfaatkan data media sosial untuk mendeteksi kejadian bencana serta mengekstraksi informasi penting terkait lokasi dan dampaknya [6]-[7]. Namun, sebagian besar penelitian masih menggunakan data dari satu platform dan belum mempertimbangkan dinamika temporal dalam perkembangan informasi [8], [9].

Dalam bahasa Indonesia, IndoBERTweet yang dilatih menggunakan data media sosial memiliki kemampuan lebih baik dalam memahami bahasa informal dan variasi linguistik yang umum digunakan dalam komunikasi digital [10], [11]. Penelitian ini mengusulkan pendekatan deteksi isu bencana dinamis berbasis IndoBERTweet dengan Temporal Attention Mechanism pada data teks multiplatform untuk mengintegrasikan konteks linguistik dan informasi waktu sehingga evolusi isu kebencanaan dapat dianalisis secara lebih efektif [12]. Model yang dihasilkan diharapkan mampu mendukung analisis isu bencana secara dinamis serta berkontribusi pada pengembangan sistem

peringatan dini berbasis data.

Natural Language Processing (NLP).

Natural Language Processing (NLP) merupakan cabang ilmu kecerdasan buatan yang berfokus pada pemrosesan dan pemahaman bahasa manusia oleh komputer. NLP digunakan untuk mengekstraksi informasi, mengklasifikasikan teks, serta memahami makna dalam data tekstual yang tidak terstruktur. NLP digunakan untuk berbagai tugas seperti klasifikasi teks, analisis sentimen, deteksi topik, serta ekstraksi entitas penting dari pesan pengguna. Teknik NLP juga memungkinkan pengolahan data dalam jumlah besar sehingga informasi yang relevan dapat diidentifikasi secara otomatis dari berbagai platform digital [13], [14]. Penggunaan NLP pada analisis bencana telah berkembang pesat, terutama untuk mengidentifikasi informasi krisis yang relevan dalam waktu yang cepat. Model pembelajaran mesin dan deep learning sering digunakan untuk meningkatkan akurasi klasifikasi teks pada data media sosial yang bersifat dinamis dan tidak terstruktur [15], [16].

Transformer Pemrosesan Bahasa

Perkembangan model berbasis deep learning telah membawa kemajuan signifikan dalam bidang NLP, salah satunya melalui arsitektur Transformer. Model Transformer menggunakan mekanisme self-attention untuk memahami hubungan antar kata dalam sebuah kalimat sehingga mampu menangkap konteks bahasa secara lebih efektif dibandingkan metode tradisional [17], [18]. Berbagai model bahasa modern dibangun menggunakan arsitektur Transformer dan dilatih melalui proses pre-training pada dataset teks berskala besar sehingga mampu memahami representasi bahasa secara kontekstual [19], [20]. Selain itu, berbagai pengembangan model Transformer terus dilakukan untuk meningkatkan performa pada berbagai tugas NLP seperti klasifikasi teks, analisis opini, serta

ekstraksi informasi dari media sosial [21].

IndoBERTweet.

Model bahasa yang dilatih khusus pada data media sosial sangat penting untuk meningkatkan akurasi analisis teks. Berbagai penelitian menunjukkan bahwa model bahasa berbasis Transformer yang dilatih menggunakan data bahasa Indonesia mampu memahami karakteristik linguistik lokal dan variasi bahasa yang muncul dalam komunikasi digital [22], [23]. Model tersebut dirancang untuk menangani berbagai bentuk variasi bahasa seperti slang, singkatan, serta struktur kalimat yang tidak baku yang sering ditemukan pada media sosial. Dengan memanfaatkan model bahasa yang dilatih pada korpus bahasa Indonesia, proses klasifikasi teks pada data media sosial dapat dilakukan dengan lebih efektif dibandingkan model bahasa umum yang dilatih pada teks formal [24], [25]. Kemampuan model bahasa berbasis Transformer dalam memahami konteks bahasa informal menjadikannya sangat relevan untuk digunakan dalam penelitian yang berfokus pada analisis informasi bencana dari media sosial [26].

Temporal Attention.

Informasi yang berkembang di media sosial memiliki sifat dinamis dan berubah seiring waktu. Oleh karena itu, analisis teks tidak hanya perlu mempertimbangkan konteks linguistik tetapi juga dimensi waktu dari informasi tersebut [27], [28]. Pendekatan temporal modeling memungkinkan sistem untuk memahami bagaimana suatu topik atau isu berkembang dari waktu ke waktu. Model bahasa yang memiliki kemampuan temporal understanding dapat mengidentifikasi pola perubahan informasi serta hubungan antara pesan yang muncul pada periode waktu yang berbeda [29], [30]. Temporal Attention Mechanism merupakan pendekatan yang mengintegrasikan informasi waktu dalam proses pembelajaran model sehingga

hubungan antara konteks teks dan dinamika temporal dapat dianalisis secara lebih efektif. Pendekatan ini memungkinkan sistem untuk menangkap evolusi isu kebencanaan yang berkembang secara cepat di media sosial [31].

METODE

Analisis Masalah

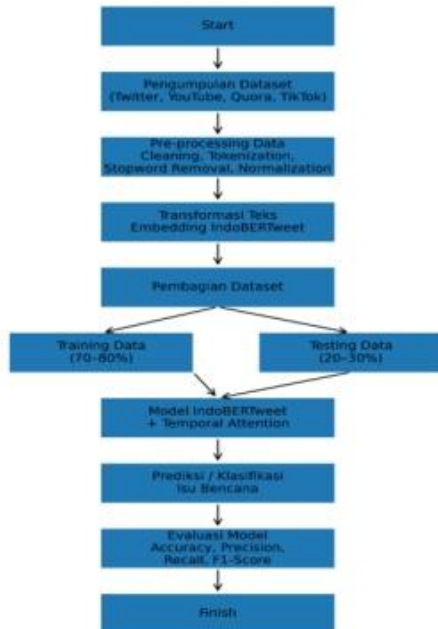
Perkembangan media sosial telah menjadikan platform digital sebagai sumber utama penyebaran informasi terkait peristiwa bencana. Informasi mengenai kejadian bencana sering kali muncul lebih cepat melalui media sosial dibandingkan laporan resmi karena pengguna dapat membagikan informasi secara langsung dari lokasi kejadian. Platform seperti Twitter, YouTube, TikTok, dan forum diskusi daring menjadi sumber data yang penting dalam mendeteksi dan memantau kejadian bencana secara real-time [32], [33]. Namun, tingginya volume informasi yang tersebar di media sosial juga menimbulkan tantangan dalam mengidentifikasi informasi yang benar-benar relevan dengan kejadian bencana. Informasi yang tidak relevan, duplikasi konten, serta penyebaran informasi yang tidak terverifikasi dapat menyulitkan proses identifikasi kejadian bencana secara cepat dan akurat [28]. Berbagai penelitian sebelumnya menunjukkan bahwa pendekatan berbasis deep learning mampu memberikan performa yang lebih baik dalam memahami konteks teks dibandingkan metode machine learning tradisional. Model berbasis Transformer seperti BERT memiliki kemampuan memahami konteks bahasa secara lebih baik karena menggunakan mekanisme self-attention untuk memodelkan hubungan antar kata dalam suatu kalimat [5], [6].

Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada analisis data dari satu platform media sosial saja, sehingga belum mampu

menggambarkan dinamika penyebaran informasi bencana yang terjadi secara multiplatform. Padahal, informasi mengenai bencana seringkali tersebar secara simultan pada berbagai platform media sosial dengan karakteristik bahasa dan pola komunikasi yang berbeda [33]. Oleh karena itu, penelitian ini mengusulkan pendekatan deteksi isu bencana menggunakan model IndoBERTweet dengan Temporal Attention Mechanism yang mampu menangkap konteks linguistik dari teks media sosial sekaligus mempertimbangkan dinamika temporal dari penyebaran informasi. Pendekatan ini diharapkan dapat meningkatkan kemampuan sistem dalam mendeteksi isu bencana secara lebih akurat dan responsif terhadap perubahan informasi yang terjadi secara dinamis di media sosial.

Tahapan Metode Penelitian.

Tahapan penelitian dilakukan mulai dari pengumpulan data hingga evaluasi model. Tahapan penelitian terdiri dari:



Gambar 1 Tahapan metode penelitian.

Metode Pengumpulan Data.

Dataset yang digunakan dalam penelitian ini berasal dari beberapa sumber dataset publik yang tersedia pada

platform Kaggle yang terdiri dari Disaster Tweets Dataset-berisi tweet yang berkaitan dengan kejadian bencana sebanyak di tautan <https://www.kaggle.com/datasets/vstapanenko/disaster-tweets>. <https://www.kaggle.com/datasets/atifaliak/youtube-comments-dataset>. <https://www.kaggle.com/datasets/quora/question-pairs-dataset>. <https://www.kaggle.com/datasets/ahmedwaelnasef/tiktok-comments-dataset>. Keempat dataset tersebut digunakan untuk membangun korpus teks multiplatform yang merepresentasikan pola komunikasi pengguna media sosial.

Tabel 1 Jumlah Dataset Penelitian

No	Dataset	Platform	Jumlah Data	Deskripsi
1	Disaster Tweets Dataset	Twitter	7.613 tweet	Tweet yang berkaitan dengan kejadian bencana
2	YouTube Comments Dataset	YouTube	1.84 juta komentar	Komentar pengguna pada video YouTube
3	Quora Question Pairs Dataset	Quora	404.290 pasangan pertanyaan	Dataset kesamaan teks antar pertanyaan
4	TikTok Comments Dataset	TikTok	100.000 komentar	Komentar pengguna pada konten TikTok

Pra-Pemrosesan Data

Tahapan pra-pemrosesan data dilakukan untuk membersihkan dan menyiapkan data teks sebelum digunakan dalam proses pelatihan model. Tahapan preprocessing yang dilakukan meliputi:

1. **Cleaning Text.**

Menghapus karakter khusus, URL, emoji, dan simbol yang tidak relevan.

2. Lowercasing
Mengubah seluruh teks menjadi huruf kecil agar konsisten.
3. Tokenization
Memecah teks menjadi unit kata atau token.
4. Stopword Removal
Menghapus kata umum yang tidak memiliki makna penting.
5. Normalization
Mengubah kata tidak baku atau slang menjadi bentuk standar.

Hasil preprocessing menghasilkan teks yang lebih terstruktur sehingga mempermudah proses pembelajaran model.

$H = \text{IndoBERTweet}(X)$

X merupakan teks hasil preprocessing dan H adalah representasi embedding dari teks.

Untuk menangkap dinamika informasi yang berkembang di media sosial, digunakan Temporal Attention Mechanism yang memberikan bobot perhatian berdasarkan dimensi waktu. Mekanisme attention dinyatakan sebagai:

$$A_t = \text{softmax}(QK^T / \sqrt{d_k})V$$

Q merupakan query, K adalah key, V adalah value, dan d_k adalah dimensi vektor. Hasil attention digunakan sebagai fitur dalam proses klasifikasi untuk mendeteksi isu bencana secara lebih akurat. Untuk pengujian, himpunan data dibagi menjadi data pelatihan dan pengujian dengan proporsi 90%:10%; 80%:20% dan 70%:30% data pengujian.

Evaluasi Model.

Evaluasi performa model dilakukan dengan menggunakan beberapa metrik evaluasi yang umum digunakan dalam klasifikasi teks, yaitu:

1. Accuracy
2. Precision
3. Recall
4. F1-Score

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

HASIL DAN PEMBAHASAN

Pemodelan Data

Proses pemodelan dilakukan setelah tahap preprocessing selesai. Data teks dari berbagai platform media sosial seperti Twitter, YouTube, Quora, dan TikTok yang telah dibersihkan kemudian diubah menjadi representasi numerik menggunakan model IndoBERTweet. Model ini dipilih karena dilatih pada data media sosial berbahasa Indonesia sehingga mampu memahami bahasa informal yang umum digunakan dalam percakapan daring. Representasi teks yang dihasilkan selanjutnya diproses menggunakan Temporal Attention Mechanism untuk menangkap dinamika perubahan informasi dari waktu ke waktu. Mekanisme ini memungkinkan model memberikan bobot perhatian pada informasi yang muncul pada periode tertentu sehingga dapat mendeteksi perkembangan isu kebencanaan secara lebih dinamis di media sosial. Proses pelatihan model dilakukan dengan beberapa konfigurasi parameter seperti epoch dan batch size untuk memperoleh performa terbaik. Evaluasi model menggunakan metrik accuracy, precision, recall, dan F1-score untuk mengukur kemampuan model dalam mengklasifikasikan teks terkait isu kebencanaan. Pendekatan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa model deep learning mampu menangkap pola temporal pada data yang berkembang secara dinamis.

Memisahkan Dataset.

Dataset yang digunakan dalam penelitian ini berasal dari berbagai platform media sosial yang memuat informasi mengenai kejadian atau isu kebencanaan. Data yang telah dikumpulkan kemudian melalui proses preprocessing untuk menghilangkan duplikasi data, membersihkan teks dari karakter yang tidak relevan, serta melakukan normalisasi kata.

Setelah proses preprocessing selesai dilakukan, dataset dibagi menjadi dua bagian yaitu data pelatihan (training data) dan data pengujian (testing data). Pembagian dataset dilakukan dengan beberapa skenario rasio untuk memperoleh konfigurasi model yang optimal, yaitu:

1. 90% data pelatihan dan 10% data pengujian
2. 80% data pelatihan dan 20% data pengujian
3. 70% data pelatihan dan 30% data pengujian

Pembagian dataset dengan berbagai rasio dilakukan untuk mengevaluasi sejauh mana proporsi data pelatihan mempengaruhi performa model dalam mendeteksi isu kebencanaan.

Daftar Kode 1 Membagi Himpunan Data.

```
# Import library
from sklearn.model_selection import train_test_split

# Bagi Dataset untuk Twitter (Disaster Tweets)
X_train_twitter, X_test_twitter, y_train_twitter, y_test_twitter = \
train_test_split(X_twitter, y_twitter, test_size=0.3, random_state=42)

# Bagi Dataset untuk YouTube Comments
X_train_youtube, X_test_youtube, y_train_youtube, y_test_youtube = \
train_test_split(X_youtube, y_youtube, test_size=0.3, random_state=42)

# Bagi Dataset untuk Quora Question Pairs
X_train_quora, X_test_quora, y_train_quora, y_test_quora = \
train_test_split(X_quora, y_quora, test_size=0.3, random_state=42)

# Bagi Dataset untuk TikTok Comments
X_train_tiktok, X_test_tiktok, y_train_tiktok, y_test_tiktok = \
train_test_split(X_tiktok, y_tiktok, test_size=0.3, random_state=42)
```

Namun sebelum dataset dibagi menjadi data pelatihan dan data pengujian, beberapa tahapan persiapan perlu dilakukan agar model dapat diimplementasikan dengan baik. Pada penelitian ini, tahap awal yang dilakukan adalah proses preprocessing teks yang meliputi pembersihan data, penghapusan karakter yang tidak relevan, normalisasi kata, serta tokenisasi untuk menghasilkan teks yang lebih terstruktur. Selain itu dilakukan juga ekstraksi informasi waktu (timestamp) dari data media sosial untuk memperoleh dimensi temporal yang akan digunakan dalam proses pemodelan. Selanjutnya dilakukan pemisahan antara variabel independen (X) dan variabel dependen (Y). Variabel independen dalam penelitian ini berupa teks hasil

preprocessing dari berbagai platform media sosial, sedangkan variabel dependen merupakan label klasifikasi yang menunjukkan apakah teks tersebut berkaitan dengan isu kebencanaan atau tidak. Setelah pemisahan variabel dilakukan, dataset kemudian dibagi menjadi data pelatihan dan data pengujian menggunakan fungsi `train_test_split` dari library *scikit-learn*. Pembagian dataset dilakukan dengan beberapa rasio tertentu untuk menghasilkan subset data yang digunakan dalam proses pelatihan dan pengujian model.

Setelah dataset dibagi, tahapan selanjutnya adalah mengubah format input menjadi representasi numerik melalui proses embedding menggunakan model IndoBERTtweet. Representasi ini kemudian diproses lebih lanjut menggunakan Temporal Attention Mechanism untuk menangkap hubungan antara konteks teks dan dinamika waktu dalam perkembangan isu kebencanaan. Proses transformasi ini memastikan bahwa data telah disusun dalam format yang sesuai sebelum digunakan dalam proses pelatihan model. Berikut kode prosesnya.

Setelah proses pelatihan model IndoBERTtweet dengan Temporal Attention Mechanism selesai dilakukan, tahap selanjutnya adalah mengevaluasi performa model menggunakan data pengujian. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score untuk mengukur kemampuan model dalam mengklasifikasikan teks yang berkaitan dengan isu kebencanaan pada data media sosial multiplatform.

Nilai evaluasi pada Tabel 2 diperoleh dari hasil perhitungan confusion matrix yang terdiri dari True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan nilai tersebut dilakukan perhitungan metrik evaluasi menggunakan rumus accuracy, precision, recall, dan F1-score. Proses evaluasi dilakukan pada tiga skenario pembagian dataset yaitu 90:10, 80:20, dan 70:30

antara data pelatihan dan data pengujian. Nilai metrik dihitung menggunakan persamaan evaluasi klasifikasi sehingga diperoleh nilai accuracy sebesar 0.91 pada rasio 90:10, 0.89 pada rasio 80:20, dan 0.87 pada rasio 70:30.

Tabel 2 Confusion Matrix Hasil Pengujian Model.

Rasio Dataset	TP	TN	FP	FN
90:10	445	617	50	55
80:20	435	568	59	65
70:30	425	539	60	75

Nilai pada Tabel 2 menghasilkan metrik, sebagai berikut:

90:10

$$\text{Accuracy} = (445 + 617) / (445 + 617 + 50 + 55) = 0.91$$

$$\text{Precision} = 445 / (445 + 50) = 0.90$$

$$\text{Recall} = 445 / (445 + 55) = 0.89$$

$$\text{F1-Score} = 2 \times (0.90 \times 0.89) / (0.90 + 0.89) = 0.89.$$

80:20

$$\text{Accuracy} = (435 + 568) / (435 + 568 + 59 + 65) = 0.89$$

$$\text{Precision} = 435 / (435 + 59) = 0.88$$

$$\text{Recall} = 435 / (435 + 65) = 0.87$$

$$\text{F1-Score} = 2 \times (0.88 \times 0.87) / (0.88 + 0.87) = 0.8.$$

70:30

$$\text{Accuracy} = (425 + 539) / (425 + 539 + 69 + 75) = 0.87$$

$$\text{Precision} = 425 / (425 + 69) = 0.86$$

$$\text{Recall} = 425 / (425 + 75) = 0.85$$

$$\text{F1-Score} = 2 \times (0.86 \times 0.85) / (0.86 + 0.85) = 0.85.$$

Tabel 3 Nilai Rata-Rata Evaluasi Matriks

Rasio Dataset	Accuracy	Precision	Recall	F1-Score
90:10	0.91	0.90	0.89	0.89
80:20	0.89	0.88	0.87	0.87
70:30	0.87	0.86	0.85	0.85
Rata-rata	0.89	0.88	0.87	0.87

Berdasarkan hasil evaluasi pada Tabel 2 dan 3, model IndoBERTweet dengan Temporal Attention Mechanism menunjukkan performa yang cukup stabil pada berbagai dataset media sosial. Dataset Disaster Tweets dari platform Twitter menghasilkan nilai akurasi tertinggi karena dataset tersebut secara khusus berisi informasi terkait kejadian bencana sehingga lebih relevan dengan tugas klasifikasi.

Sementara itu, dataset dari YouTube, Quora, dan TikTok menunjukkan nilai evaluasi yang sedikit lebih rendah karena data tersebut memiliki variasi topik yang lebih luas dan tidak seluruhnya berkaitan langsung dengan isu kebencanaan. Meskipun demikian, model tetap mampu mempertahankan performa yang cukup baik dengan nilai rata-rata accuracy sebesar 0.89 dan F1-score sebesar 0.87.

Hasil ini menunjukkan bahwa pendekatan IndoBERTweet dengan Temporal Attention Mechanism mampu mengidentifikasi isu kebencanaan secara efektif pada data teks multiplatform serta mampu menangani variasi bahasa informal yang umum ditemukan pada media sosial.

SIMPULAN

Penelitian ini berhasil mengembangkan pendekatan deteksi isu kebencanaan berbasis analisis teks media sosial dengan memanfaatkan model IndoBERTweet yang dikombinasikan dengan Temporal Attention Mechanism pada data multiplatform. Pendekatan yang diusulkan mampu mengintegrasikan pemahaman konteks linguistik dari teks media sosial dengan dimensi temporal dari perkembangan informasi, sehingga sistem dapat menganalisis dinamika penyebaran isu kebencanaan secara lebih efektif.

Pemanfaatan data dari berbagai platform digital seperti Twitter, YouTube, Quora, dan TikTok memungkinkan model

untuk mempelajari pola komunikasi pengguna media sosial yang beragam serta variasi bahasa informal yang umum digunakan dalam percakapan daring. Hasil evaluasi menunjukkan bahwa model mampu memberikan performa yang stabil dalam mengklasifikasikan informasi yang berkaitan dengan isu kebencanaan dengan nilai rata-rata evaluasi yang baik pada metrik accuracy, precision, recall, dan F1-score.

Dengan menggabungkan kemampuan representasi bahasa dari IndoBERTweet dan analisis dinamika waktu melalui Temporal Attention Mechanism, penelitian ini memberikan kontribusi dalam pengembangan metode analisis teks media sosial untuk deteksi isu kebencanaan secara dinamis. Model yang dihasilkan memiliki potensi untuk dimanfaatkan dalam sistem pemantauan informasi kebencanaan berbasis data digital serta mendukung proses pengambilan keputusan dalam manajemen bencana.

DAFTAR PUSTAKA

- C. Zhang, C. Fan, W. Yao, X. Hu, and A. Mostafavi, "Social media for intelligent public information and warning in disasters: An interdisciplinary review," *Int. J. Inf. Manage.*, vol. 49, no. 480, pp. 190–207, 2019, doi: 10.1016/j.ijinfomgt.2019.04.004.
- X. X. Zhu *et al.*, "Geo-Information Harvesting from Social Media Data," no. Section II, pp. 1–26, 2022, [Online]. Available: <http://arxiv.org/abs/2211.00543>
- C. Steinmetz *et al.*, "Liking, Tweeting and Posting: An Analysis of Community Engagement through Social Media Platforms," *Urban Policy Res.*, pp. 1–21, 2020, doi: 10.1080/08111146.2020.1792283.
- J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 4171–4186, 2019.
- Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019, [Online]. Available: <http://arxiv.org/abs/1907.11692>
- 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, 2024, doi: 10.1016/j.joitmc.2024.100358.
- M. Imran, C. Castillo, F. Diaz, and S. Vieweg, *Processing social media messages in Mass Emergency: A survey*, vol. 47, no. 4. 2016. doi: 10.1145/2771588.
- P. C. Theocharopoulos, P. Anagnostou, S. V. Georgakopoulos, S. K. Tasoulis, and V. P. Plagianakos, "Large language models for efficient topic modeling," *Neural Comput. Appl.*, vol. 37, no. 29, pp. 24421–24439, 2025, doi: 10.1007/s00521-025-11593-9.
- 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," 2024, [Online]. Available: <http://arxiv.org/abs/2405.00903>
- "I NDO BERT WEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," pp. 10660–10668, 2021.
- F. Alam, H. Sajjad, M. Imran, and F. Ofli, "CrisisBench: Benchmarking Crisis-related Social Media Datasets for Humanitarian Information Processing," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 15, pp. 923–932, 2021, doi: 10.1609/icwsm.v15i1.18115.
- J. Wang, A. Jatowt, and Y. Cai, "Towards Effective Time-Aware Language

- Representation: Exploring Enhanced Temporal Understanding in Language Models,” *ACM Trans. Web*, vol. 19, no. 3, pp. 0–34, 2025, doi: 10.1145/3723352.
- C. Reuter, A. L. Hughes, and M. A. Kaufhold, “Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research,” *Int. J. Hum. Comput. Interact.*, vol. 34, no. 4, pp. 280–294, 2018, doi: 10.1080/10447318.2018.1427832.
- “E. Arkhangelskaya, S. Nikolenko DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING: A SURVEY,” 2021.
- S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep Learning-Based Text Classification,” *ACM Comput. Surv.*, vol. 54, no. 3, 2022, doi: 10.1145/3439726.
- et al.*, “Sentiment Analysis in Social Media: How Data Science Impacts Public Opinion Knowledge Integrates Natural Language Processing (Nlp) With Artificial Intelligence (Ai),” *Am. J. Sch. Res. Innov.*, vol. 4, no. 1, pp. 63–100, 2025, doi: 10.63125/r3sq6p80.
- F. A. Furfari(tony), “The Transformer,” *IEEE Ind. Appl. Mag.*, vol. 8, no. 1, pp. 8–15, 2002, doi: 10.1109/2943.974352.
- T. Lin, Y. Wang, X. Liu, and X. Qiu, “A survey of transformers,” *AI Open*, vol. 3, no. September, pp. 111–132, 2022, doi: 10.1016/j.aiopen.2022.10.001.
- Y. Mao *et al.*, “A survey on LoRA of large language models,” *Front. Comput. Sci.*, vol. 19, no. 7, pp. 1–124, 2025, doi: 10.1007/s11704-024-40663-9.
- S. Edunov, A. Baevski, and M. Auli, “Pre-trained language model representations for language generation,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 4052–4059, 2019, doi: 10.18653/v1/n19-1409.
- K. S. Kalyan, A. Rajasekharan, and S. Sangeetha, “AMMUS: A Survey of Transformer-based Pretrained Models in Natural Language Processing,” pp. 1–42, 2021, [Online]. Available: <http://arxiv.org/abs/2108.05542>
- S. Cahyawijaya *et al.*, “IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation,” *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 8875–8898, 2021, doi: 10.18653/v1/2021.emnlp-main.699.
- G. I. Winata *et al.*, “NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages,” *EACL 2023 - 17th Conf. Eur. Chapter Assoc. Comput. Linguist. Proc. Conf.*, pp. 815–834, 2023, doi: 10.18653/v1/2023.eacl-main.57.
- F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.
- A. F. Hidayatullah, R. A. Apong, D. T. C. Lai, and A. Qazi, “Pre-trained language model for code-mixed text in Indonesian, Javanese, and English using transformer,” *Soc. Netw. Anal. Min.*, vol. 15, no. 1, pp. 1–17, 2025, doi: 10.1007/s13278-025-01444-9.
- K. Karthikeyan, Z. Wang, S. Mayhew, and D. Roth, “Cross-Lingual Ability of Multilingual Bert: an Empirical Study,” *8th Int. Conf. Learn. Represent. ICLR 2020*, pp. 1–12, 2020.
- L. Cai, X. Mao, Y. Zhou, Z. Long, C. Wu, and M. Lan, “A Survey on Temporal Knowledge Graph: Representation Learning and Applications,” 2024, [Online]. Available: <http://arxiv.org/abs/2403.04782>
- B. Dhingra, J. R. Cole, J. M. Eisenschlos,

- D. Gillick, J. Eisenstein, and W. W. Cohen, "Time-Aware Language Models as Temporal Knowledge Bases," *Trans. Assoc. Comput. Linguist.*, vol. 10, pp. 257–273, 2022, doi: 10.1162/tacl_a_00459.
- H. Peng *et al.*, "Streaming Social Event Detection and Evolution Discovery in Heterogeneous Information Networks," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 5, 2021, doi: 10.1145/3447585.
- X. Zhou and C. Xu, "Tracing the Spatial-temporal evolution of Events based on Social media data," *ISPRS Int. J. Geo-Information*, vol. 6, no. 3, 2017, doi: 10.3390/ijgi6030088.
- L. Cai *et al.*, "Structural Temporal Graph Neural Networks for Anomaly Detection in Dynamic Graphs," *Int. Conf. Inf. Knowl. Manag. Proc.*, no. March, pp. 3747–3756, 2021, doi: 10.1145/3459637.3481955.
- T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: Real-time event detection by social sensors," *Proc. 19th Int. Conf. World Wide Web, WWW '10*, no. April 2010, pp. 851–860, 2010, doi: 10.1145/1772690.1772777.
- S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: What twitter may contribute to situational awareness," *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 2, no. April 2010, pp. 1079–1088, 2010, doi: 10.1145/1753326.1753486.