INDOTECH
INDONESIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCE

Indonesian Journal of Education And Computer Science

Vol. 4, No. 2, Agustus 2026
Hal 55-64

E-ISSN : 2987-2650
P-ISSN : 2987-7644

Site : https://jurnal.intekom.id/index.php/indotech/article/view/161
DOI: https://doi.org/10.60076/indotech.v4i2.2090

Performa Multinomial *Naive Bayes* dalam Klasifikasi Misinformasi *Megathrust* pada Komentar TikTok

Performance of Multinomial Naive Bayes in Classifying Megathrust Misinformation in TikTok Comments

Febri Yalda Sulistia[1*], Karina Nurfebia, Nurbeti Sinulingga, Nuzul Aini Ramadhani, Khairul
[1] Fakultas Sains dan Teknologi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Article Info

Article history:

Received Mei 27, 2026
Revised Agustus 30, 2026
Accepted Agustus 31, 2026
Published Agustus 31, 2026

Kata Kunci:

Misinformasi *Megathrust*;
Komentar TikTok;
Multinomial Naive Bayes;
TF-IDF;
Misinformasi Kebencanaan

Keywords:

Megathrust Misinformation;
TikTok Comments;
Multinomial Naive Bayes;
TF-IDF;
Disaster-Related Misinformation

ABSTRAK

Penyebaran misinformasi terkait bencana *megathrust* di media sosial TikTok berpotensi memicu kepanikan publik dan mengganggu efektivitas komunikasi risiko bencana. Penelitian ini bertujuan menganalisis model klasifikasi informasi valid dan misinformasi terkait *megathrust* menggunakan pendekatan *Natural Language Processing* (NLP), algoritma *Multinomial Naive Bayes*, dan *TF-IDF*. Data penelitian berupa 2.000 komentar TikTok berbahasa Indonesia yang dikumpulkan secara *purposive sampling* dari konten relevan dan dibagi ke dalam dua kelas: informasi valid dan misinformasi. *Preprocessing* meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming* menggunakan *library* Sastrawi. Evaluasi pada 400 data uji menghasilkan akurasi 52%, *macro-precision* 50,24%, *macro-recall* 50,20%, dan *macro-F1* 48,93%. Akurasi tersebut lebih rendah daripada *majority-class baseline* 54,5%, sehingga model belum memberikan peningkatan dibanding prediksi sederhana yang selalu memilih kelas mayoritas. Kesalahan paling banyak terjadi pada komentar valid yang diprediksi sebagai misinformasi. Hasil ini menunjukkan bahwa variasi bahasa informal dan keterbatasan *TF-IDF* dalam menangkap konteks masih menjadi tantangan utama pada klasifikasi komentar TikTok.

ABSTRACT

The spread of megathrust-related misinformation on TikTok can trigger public panic and interfere with disaster risk communication. This study evaluates a classification model for distinguishing valid information from misinformation in Indonesian TikTok comments using Natural Language Processing (NLP), Multinomial Naive Bayes, and TF-IDF. The dataset contains 2,000 comments collected through purposive sampling from relevant public content and assigned to two classes: valid information and misinformation. Text preprocessing included case folding, cleaning, tokenizing, stopword removal, and stemming with the Sastrawi library. Evaluation on 400 test comments produced 52% accuracy, 50.24% macro-precision, 50.20% macro-recall, and 48.93% macro-F1. The model accuracy was lower than the 54.5% majority-class baseline, indicating that it did not outperform the simple strategy of always predicting the majority class. Most errors occurred when valid comments were classified as misinformation. The results indicate that informal language variation and the limited contextual representation of TF-IDF remain major challenges for classifying TikTok comments.

Corresponding Author:

Febri Yalda Sulistia
Program Studi Teknologi Informasi, Universitas Pembangunan Panca Budi
Medan, Indonesia
Email: febriyaldasulistia7@gmail.com

1. PENDAHULUAN

Indonesia merupakan negara yang berada pada kawasan tektonik aktif akibat pertemuan tiga lempeng utama dunia, yaitu Indo-Australia, Eurasia, dan Pasifik. Kondisi geologis ini menjadikan Indonesia, termasuk wilayah Provinsi Sumatera Utara, memiliki tingkat kerentanan tinggi terhadap gempa bumi berskala besar, khususnya gempa *megathrust* yang berpotensi menimbulkan dampak destruktif dalam waktu singkat. Dalam situasi krisis semacam ini, ketersediaan informasi yang akurat dan dapat dipertanggungjawabkan memiliki peran krusial dalam mendukung upaya mitigasi risiko serta pengendalian kepanikan publik, sementara informasi yang keliru justru berpotensi menghambat efektivitas respons kedaruratan dan menimbulkan distorsi persepsi masyarakat terhadap tingkat keparahan bencana yang sesungguhnya [1].

Media sosial, terutama TikTok, telah menjadi kanal utama penyebaran informasi kebencanaan karena sifatnya yang *real-time*, terbuka, dan berjangkauan luas. Namun di sisi lain, kanal ini memiliki kerentanan tinggi terhadap penyebaran misinformasi, sebagaimana ditunjukkan oleh berbagai kajian kebencanaan lima tahun terakhir. Mızrak [2] mensurvei masyarakat pascagempa Kahramanmaraş di Turki dan menemukan bahwa meskipun mayoritas responden memanfaatkan media sosial secara aktif untuk berbagi informasi, sebagian besar dari mereka turut menyoroti sisi negatifnya, termasuk beredarnya informasi yang tidak terverifikasi. Dallo et al. [3] menganalisis dinamika misinformasi terkait prediksi gempa bumi di Twitter menggunakan model *RoBERTa* dan menemukan bahwa unggahan misinformasi cenderung meningkat tajam setelah terjadinya gempa yang dirasakan masyarakat, sehingga institusi resmi kebencanaan perlu secara berkelanjutan mengklarifikasi klaim-klaim prediksi yang tidak berdasar. Sejalan dengan itu, Zhai et al. [4] menemukan bahwa penyebaran misinformasi bencana serta proses koreksinya di media sosial sangat dipengaruhi oleh kedekatan *spatiotemporal* peristiwa dan penularan sentimen antarpengguna. Tinjauan naratif oleh Hilberts et al. [1] turut menegaskan bahwa pola tersebut konsisten ditemukan lintas jenis bencana alam, sehingga mendorong perlunya teknologi deteksi misinformasi otomatis sebagai bagian dari kesiapsiagaan bencana.

Penanganan misinformasi secara manual menghadapi keterbatasan signifikan dari sisi kapasitas dan kecepatan respons, sehingga *Natural Language Processing* (NLP) hadir sebagai solusi komputasional untuk menganalisis dan mengklasifikasikan teks secara otomatis dalam skala besar. Truică dan Apostol [5] menunjukkan bahwa representasi berbasis *document embedding* mampu meningkatkan akurasi deteksi berita palsu dibandingkan representasi fitur konvensional seperti *TF-IDF* semata. Mouratidis et al. [6] membandingkan performa pengklasifikasi *machine learning* klasik (*Naïve Bayes*, SVM, *Random Forest*) dengan model *deep learning* berbasis *BERT* untuk deteksi berita palsu lintas platform, dan mendapati bahwa model *transformer* secara konsisten unggul, meskipun pengklasifikasi klasik tetap relevan sebagai *baseline* karena kebutuhan komputasinya yang jauh lebih ringan. Temuan serupa dilaporkan oleh Raza et al. [7] yang mengevaluasi model berbasis *BERT* dan *large language model* dengan data anotasi generatif, sekaligus menekankan bahwa pengklasifikasi ringan masih relevan pada konteks sumber daya terbatas.

Pada konteks nasional, sejumlah penelitian telah menerapkan *Naïve Bayes* untuk klasifikasi hoaks berbahasa Indonesia. Agustina et al. [8] mengimplementasikan *Naïve Bayes Classifier* berbasis *TF-IDF* untuk mendeteksi berita palsu di media sosial dan memperoleh akurasi yang cukup tinggi sebagai

baseline. Karo Karo et al. [9] menerapkan pendekatan serupa pada cuitan Twitter berbahasa Indonesia, sementara Prasetiyo et al. [10] mengombinasikan *Naïve Bayes* dengan metode *forward selection* untuk meningkatkan performa klasifikasi hoaks. Fernando et al. [11] menggunakan algoritma yang sama untuk mengklasifikasikan berita hoaks kampanye Pemilu 2024, sedangkan Adrian et al. [12] membandingkan *Multinomial Naïve Bayes* dengan *Passive Aggressive Classifier* berbasis *TF-IDF Vectorizer*. Di sisi lain, pendekatan berbasis *transformer* turut berkembang pesat: Geni et al. [13] memanfaatkan *IndoBERT* untuk analisis sentimen cuitan menjelang Pemilu 2024, Ridho dan Yulianti [14] membandingkan *IndoBERT* dengan *CNN-LSTM*, regresi logistik dan *Naïve Bayes* disertai teknik *oversampling* SMOTE untuk menangani data tidak seimbang, serta Kunaefi [15] melakukan *fine-tuning IndoBERT* dengan pendekatan *focal loss* untuk kasus serupa. Rangkaian penelitian ini secara konsisten menunjukkan bahwa *Naïve Bayes* tetap menjadi *baseline* yang efisien dan mudah diimplementasikan, meskipun performanya umumnya masih berada di bawah model berbasis *transformer*, terutama pada data yang bersih dan terstruktur seperti judul berita atau cuitan formal.

Berdasarkan uraian tersebut, penelitian mengenai misinformasi kebencanaan sebelumnya lebih banyak menggunakan Twitter/X, berita daring, dan konteks bencana di luar Indonesia. Sementara itu, penelitian yang secara khusus memanfaatkan komentar TikTok berbahasa Indonesia pada isu *megathrust* Sumatera Utara masih terbatas. Kebaruan penelitian ini terletak pada penerapan *Multinomial Naive Bayes* berbasis *TF-IDF* untuk mengklasifikasikan komentar TikTok ke dalam kategori informasi valid dan misinformasi. Penggunaan komentar TikTok sebagai sumber data juga memberikan konteks yang berbeda karena karakteristik bahasanya cenderung informal, singkat, dan beragam. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran mengenai kemampuan metode tersebut dalam mengidentifikasi misinformasi kebencanaan pada media sosial berbasis video pendek.

Penelitian ini bertujuan menganalisis performa model klasifikasi informasi valid dan misinformasi terkait *megathrust* pada komentar TikTok berbahasa Indonesia menggunakan pendekatan *Natural Language Processing* (NLP), algoritma *Multinomial Naive Bayes*, dan ekstraksi fitur *TF-IDF*, serta mengevaluasinya berdasarkan akurasi, *precision*, *recall*, dan *F1-score*.

2. METODE

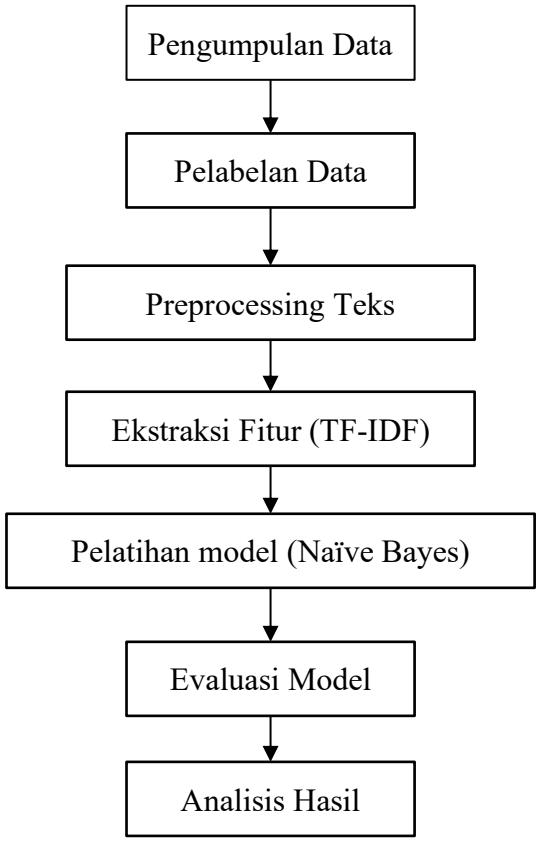
Penelitian ini merupakan penelitian eksperimental berbasis komputasi yang menggunakan pendekatan *Natural Language Processing* (NLP) untuk mengklasifikasikan komentar TikTok terkait isu *megathrust* Sumatera Utara ke dalam dua kelas, yaitu informasi valid dan misinformasi. Tahapan penelitian mencakup pengumpulan data, pelabelan manual, *preprocessing* teks, ekstraksi fitur *TF-IDF*, pelatihan model *Multinomial Naive Bayes*, dan evaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, serta *F1-score*. Penelitian dilaksanakan pada Januari-Februari 2026. Dokumentasi eksperimen yang tersedia belum mencatat secara lengkap konfigurasi *TF-IDF* (*ngram_range*, *min_df*, *max_df* atau *max_features*), nilai *alpha* MultinomialNB, dan versi pustaka yang digunakan. Karena rincian tersebut tidak dapat diverifikasi dari berkas yang tersedia, nilai parameter tidak ditambahkan berdasarkan asumsi; hal ini menjadi keterbatasan reproduksibilitas yang perlu diperbaiki pada replikasi berikutnya.

Data penelitian berupa 2.000 komentar TikTok berbahasa Indonesia yang dikumpulkan secara *purposive sampling* dari konten relevan dengan topik *megathrust* Sumatera Utara. Setiap komentar dilabeli ke dalam dua kelas: informasi valid dan misinformasi. *Dataset* dibagi dengan rasio 80:20 seperti ditunjukkan pada Tabel 1. Dokumentasi yang tersedia hanya mencatat rasio pembagian data; nilai *random_state*, penggunaan *stratify*, dan urutan *fit TF-IDF* terhadap proses *split* belum tercatat. Oleh karena itu, parameter tersebut tidak dituliskan dengan nilai asumsi. Pada replikasi berikutnya, pembagian data sebaiknya dilakukan secara terstratifikasi, *TF-IDF* di-*fit* hanya pada data latih sebelum mentransformasi data uji untuk mencegah *data leakage*, dan evaluasi dapat dilengkapi dengan *stratified k-fold* agar estimasi performa lebih stabil.

Tabel 1. Pembagian *Dataset*

Jenis Data	Jumlah
Data Latih (80%)	1.600
Data Uji (20%)	400
Total	2.000

Alur penelitian ditunjukkan pada Gambar 1. Proses dimulai dari pengumpulan data, pelabelan manual, *preprocessing* teks, ekstraksi fitur *TF-IDF*, pelatihan model *Naive Bayes*, evaluasi model, hingga analisis hasil.



Gambar 1. Alur Penelitian

Data penelitian dikumpulkan dari 20 video publik menggunakan kata kunci dan *hashtag* seperti *megathrust* Sumatera, gempa Sumatera Utara, tsunami Aceh, #*megathrust*, #gempasumatera, dan #bmkg. Pengambilan komentar dilakukan melalui layanan Apify TikTok *Scraper* terhadap komentar yang tersedia secara publik selama periode penelitian Januari-Februari 2026. Komentar yang duplikat, kosong, hanya berisi emoji atau simbol, *spam*, dan tidak berkaitan dengan topik penelitian dieliminasi sehingga diperoleh 2.000 komentar unik. Identitas akun pengguna tidak ditampilkan dalam pelaporan dan dianonimkan. Rentang tanggal publikasi video dan komentar tidak tercatat pada dokumentasi sumber, sehingga tidak ditambahkan sebagai informasi yang tidak dapat diverifikasi.

Pelabelan dilakukan secara manual oleh dua *annotator* ke dalam dua kelas, yaitu informasi valid dan misinformasi. Informasi valid mengacu pada komentar yang sesuai dengan informasi resmi BMKG, BNPB, atau sumber ilmiah, sedangkan misinformasi mencakup klaim yang tidak terverifikasi atau bertentangan dengan sumber resmi. Untuk komentar berbentuk pertanyaan atau opini yang tidak memuat klaim faktual secara langsung, keputusan label didasarkan pada ada atau tidaknya klaim yang dapat diverifikasi. Komentar sarkastik atau ambigu ditelaah berdasarkan konteks kalimat; apabila kedua

annotator memberikan label berbeda, keputusan akhir dicapai melalui diskusi. Nilai kesepakatan awal antarannotator, seperti *Cohen's Kappa*, belum dihitung pada dokumentasi eksperimen sehingga reliabilitas anotasi belum dapat dilaporkan secara kuantitatif. Penelitian ini menggunakan klasifikasi biner, dan model *Multinomial Naive Bayes* dilatih untuk mempelajari pola kata pada kedua kelas tersebut.

Tahap *preprocessing* bertujuan membersihkan dan menormalkan teks komentar yang tidak baku. Lima tahapan dilakukan secara berurutan seperti pada Tabel 2: (1) *case folding* menyeragamkan huruf menjadi *lowercase*; (2) *cleaning* menghapus URL, emoji, dan simbol; (3) *tokenizing* memecah kalimat menjadi *token*; (4) *stopword removal* menghapus kata yang tidak informatif menggunakan kamus stopword bahasa Indonesia; dan (5) *stemming* mengubah kata ke bentuk dasar menggunakan *library* Sastrawi. Tidak dilakukan tahap normalisasi khusus untuk *slang*, singkatan, *typo*, atau variasi kata tidak baku di luar *cleaning* dan *stemming*. Karena itu, variasi bentuk kata tersebut masih dapat muncul sebagai fitur yang berbeda dan menjadi salah satu sumber kesulitan model.

Tabel 2. Tahapan Pre-Processing Data

No	Tahap	Contoh *Input*	Contoh *Output*
1	*Case Folding*	"GEMPA Besar"	"gempa besar"
2	*Cleaning*	"gempa!!! ^"	"gempa"
3	*Tokenizing*	"gempa besar"	["gempa","besar"]
4	*Stopword Removal*	"akan terjadi gempa"	"terjadi gempa"
5	*Stemming*	"terjadinya"	"jadi"

Evaluasi model menggunakan *confusion matrix* dan empat metrik standar. *Accuracy* = (TP+TN)/(TP+TN+FP+FN) mengukur proporsi prediksi benar secara keseluruhan. *Precision* = TP/(TP+FP) mengukur ketepatan prediksi positif, *recall* = TP/(TP+FN) mengukur kemampuan mendeteksi kelas positif, sedangkan *F1-score* = 2 × (*precision* × *recall*)/(*precision* + *recall*) merupakan rata-rata harmonik *precision* dan *recall*. Karena penelitian menggunakan dua kelas dengan distribusi yang tidak sepenuhnya seimbang, nilai *precision*, *recall*, dan *F1-score* tunggal yang dilaporkan pada abstrak dan Tabel 4 menggunakan *macro average* sehingga kedua kelas diberi bobot yang sama. Nilai per kelas, *macro average*, dan *weighted average* ditampilkan pada Tabel 5.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Dataset penelitian terdiri dari 2.000 komentar TikTok berbahasa Indonesia terkait *megathrust* Sumatera Utara. Distribusi kelas pada Tabel 3 menunjukkan 45,7% informasi valid dan 54,3% misinformasi. Dominasi misinformasi memperlihatkan bahwa isu kebencanaan di media sosial memuat cukup banyak klaim yang belum terverifikasi. Karakteristik komentar TikTok yang menggunakan *slang*, singkatan, variasi ejaan, dan penulisan tidak baku menyebabkan representasi *TF-IDF* kurang konsisten. Tidak adanya normalisasi khusus terhadap *slang*, singkatan, dan *typo* membuat bentuk kata yang sebenarnya serupa dapat terbaca sebagai fitur yang berbeda. Selain itu, kosakata pada komentar valid dan misinformasi sering tumpang tindih, misalnya kata gempa, tsunami, BMKG, dan waspada, sehingga pendekatan *bag-of-words* sulit menangkap konteks, negasi, atau ironi. Temuan ini sejalan dengan Mouratidis et al. [6] dan Raza et al. [7] yang menunjukkan keunggulan representasi kontekstual dibandingkan fitur berbasis frekuensi. Dibandingkan Agustina et al. [8] yang menggunakan data berita

dan Karo Karo et al. [9] yang menggunakan *tweet*, penelitian ini berada pada *domain* dan korpus yang berbeda sehingga angka performanya tidak dapat dibandingkan secara langsung.

Tabel 3. Distribusi *Dataset*

No	Kelas	Jumlah	Persentase
1	Informasi Valid	914	45,7%
2	Misinformasi	1.086	54,3%
Total		2.000	100%

Hasil evaluasi model *Multinomial Naive Bayes* pada 400 data uji ditunjukkan pada Tabel 4. Model menghasilkan akurasi 52%, *macro-precision* 50,24%, *macro-recall* 50,20%, dan *macro-F1* 48,93%. Pada data uji, kelas mayoritas adalah misinformasi sebanyak 218 dari 400 data, sehingga *majority-class baseline* mencapai 54,5%. Akurasi model sebesar 52% berada di bawah *baseline* tersebut. Dengan demikian, model belum memberikan peningkatan dibandingkan strategi sederhana yang selalu memprediksi kelas mayoritas dan masih memerlukan perbaikan sebelum digunakan untuk klasifikasi praktis.

Tabel 4. Hasil Performa Model *Naive Bayes*

Metrik	Nilai
Accuracy	52%
Precision (*Macro Avg*)	50,24%
Recall (*Macro Avg*)	50,20%
F1-Score (*Macro Avg*)	48,93%

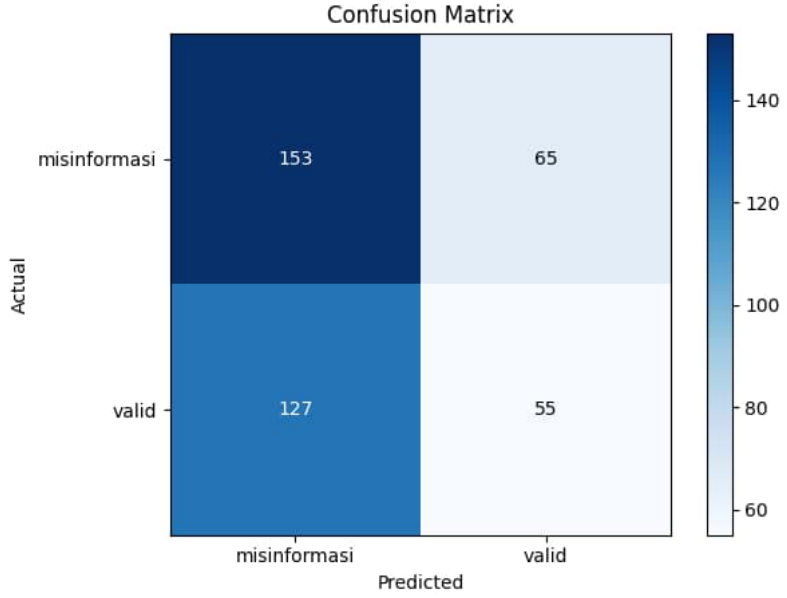
Berdasarkan hasil evaluasi, kelas misinformasi memiliki nilai *recall* sebesar 0,70, sedangkan kelas valid hanya 0,30. Hal ini menunjukkan bahwa model lebih mampu mengenali komentar yang mengandung misinformasi dibandingkan komentar yang berisi informasi valid. Kondisi tersebut mengindikasikan adanya kecenderungan model untuk memprediksi data ke dalam kelas misinformasi, sehingga sebagian komentar valid masih sering salah diklasifikasikan.

Tabel 5. Matriks Evaluasi Per Kelas

Kelas	*Precision*	*Recall*	*F1-Score*	*Support*
Misinformasi	0,55	0,70	0,61	218
Valid	0,46	0,30	0,36	182
Accuracy			0,52	400
Macro Avg	0,50	0,50	0,49	400
Weighted Avg	0,51	0,52	0,50	400

Visualisasi *confusion matrix* pada Gambar 2 menunjukkan bahwa dari 218 data misinformasi aktual, sebanyak 153 berhasil diklasifikasikan dengan benar sebagai misinformasi dan 65 salah diklasifikasikan sebagai valid. Dari 182 data valid aktual, sebanyak 55 berhasil diklasifikasikan dengan

benar sebagai valid dan 127 salah diklasifikasikan sebagai misinformasi. Tingginya kesalahan pada kelas valid menunjukkan kecenderungan model memprediksi komentar ke kelas misinformasi. Kondisi ini berkaitan dengan kemiripan kosakata antarkelas serta karakteristik komentar TikTok yang tidak baku, mengandung singkatan, *slang*, dan variasi penulisan yang tinggi.



Gambar 2. Visualisasi *Confusion Matrix*

Perbandingan dengan penelitian terdahulu perlu dibaca secara hati-hati karena sumber data dan skema eksperimennya berbeda. Agustina et al. [8] menggunakan data berita, sedangkan Karo Karo et al. [9] menggunakan *tweet* berbahasa Indonesia. Penelitian ini menggunakan komentar TikTok yang memiliki variasi ejaan, *slang*, singkatan, dan konteks percakapan yang tinggi. Karena itu, selisih performa tidak dapat dijelaskan hanya dari tingkat formalitas teks, tetapi juga dipengaruhi oleh *domain* data, distribusi kelas, tahapan *preprocessing*, representasi fitur, dan prosedur evaluasi. *Confusion matrix* menunjukkan bahwa masalah utama berada pada kelas valid: 127 dari 182 komentar valid diprediksi sebagai misinformasi. Temuan ini menguatkan bahwa kemiripan kosakata antarkelas dan keterbatasan *TF-IDF* dalam menangkap konteks menjadi tantangan utama pada data yang digunakan.

3.2 Pembahasan

Hasil penelitian menunjukkan bahwa *Multinomial Naive Bayes* berbasis *TF-IDF* belum memberikan performa yang memadai untuk klasifikasi misinformasi *megathrust* pada komentar TikTok. Akurasi sebesar 52% tidak hanya relatif rendah, tetapi juga berada di bawah *majority-class baseline* sebesar 54,5%. Nilai *macro-precision* 50,24%, *macro-recall* 50,20%, dan *macro-F1* 48,93% memperkuat temuan bahwa kemampuan model dalam membedakan kedua kelas masih terbatas ketika setiap kelas dinilai dengan bobot yang sama. Dengan demikian, pada konfigurasi data dan tahapan pemrosesan yang digunakan dalam penelitian ini, model belum dapat dianggap lebih efektif daripada strategi prediksi sederhana yang selalu memilih kelas mayoritas.

Keterbatasan tersebut terlihat lebih jelas pada perbedaan *recall* antarkelas. Model mampu mengenali 153 dari 218 komentar misinformasi, tetapi hanya mengenali 55 dari 182 komentar valid. *Recall* kelas misinformasi sebesar 0,70 berbanding jauh dengan *recall* kelas valid sebesar 0,30, sehingga 127 komentar valid salah diprediksi sebagai misinformasi. Dalam konteks komunikasi kebencanaan, pola kesalahan ini penting karena komentar yang sebenarnya valid berpotensi ditandai sebagai informasi keliru. Kecenderungan tersebut menunjukkan bahwa model lebih sensitif terhadap pola kata yang

diasosiasikan dengan misinformasi, tetapi belum cukup mampu membedakan penggunaan kosakata yang sama ketika muncul dalam konteks informasi yang benar.

Salah satu faktor yang dapat menjelaskan kondisi tersebut adalah karakteristik komentar TikTok yang singkat, informal, dan memiliki variasi penulisan yang tinggi. *Slang*, singkatan, *typo*, serta penggunaan kata yang sama pada komentar valid dan misinformasi dapat menghasilkan representasi *TF-IDF* yang saling berdekatan. Pendekatan *bag-of-words* pada *TF-IDF* menekankan kemunculan dan bobot kata, tetapi tidak secara langsung merepresentasikan hubungan kontekstual, negasi, sarkasme, atau maksud suatu komentar. Temuan ini relevan dengan penelitian Truică dan Apostol [5], Mouratidis et al. [6], serta Raza et al. [7] yang menunjukkan bahwa representasi berbasis *embedding* dan model kontekstual memiliki keunggulan dalam menangkap informasi semantik dibandingkan representasi berbasis frekuensi semata.

Jika dibandingkan dengan penelitian berbahasa Indonesia sebelumnya, hasil penelitian ini perlu dipahami berdasarkan perbedaan *domain* dan karakteristik data. Agustina et al. [8] menggunakan data berita, Karo Karo et al. [9] menggunakan *tweet*, Prasetiyo et al. [10] menerapkan *forward selection*, sedangkan Fernando et al. [11] dan Adrian et al. [12] menggunakan skenario klasifikasi berita hoaks dengan karakteristik teks dan prosedur eksperimen yang berbeda. Karena itu, performa yang lebih rendah pada penelitian ini tidak dapat diartikan semata-mata sebagai kelemahan algoritma *Naive Bayes*. Hasil tersebut lebih menunjukkan bahwa kombinasi *Multinomial Naive Bayes* dan *TF-IDF* yang digunakan belum cukup adaptif terhadap komentar TikTok yang lebih tidak terstruktur dan kontekstual dibandingkan teks berita atau korpus yang lebih formal.

Hasil penelitian juga memberikan arah pengembangan model berikutnya tanpa mengubah temuan utama penelitian ini. Normalisasi *slang*, singkatan, dan *typo* perlu dipertimbangkan agar variasi bentuk kata tidak membentuk fitur yang terpisah. Selain itu, konfigurasi *TF-IDF* dan parameter *Multinomial Naive Bayes* perlu didokumentasikan dan dioptimalkan, pembagian data perlu dilakukan secara terstratifikasi, dan evaluasi dapat diperkuat menggunakan *stratified k-fold*. Perbandingan dengan model kontekstual seperti *IndoBERT* juga relevan karena Geni et al. [13], Ridho dan Yulianti [14], serta Kunaefi et al. [15] menunjukkan potensi model berbasis *transformer* pada pengolahan teks berbahasa Indonesia. Dengan demikian, kontribusi penelitian ini terutama terletak pada penyediaan *baseline* empiris untuk komentar TikTok bertema *megathrust* sekaligus menunjukkan batas kemampuan pendekatan *TF-IDF* dan *Multinomial Naive Bayes* pada *domain* tersebut.

4. KESIMPULAN

Berdasarkan hasil evaluasi, *Multinomial Naive Bayes* berbasis *TF-IDF* masih memiliki keterbatasan dalam mengklasifikasikan misinformasi *megathrust* pada komentar TikTok berbahasa Indonesia. Dengan 2.000 komentar yang dibagi menjadi 1.600 data latih dan 400 data uji, model memperoleh akurasi 52%, *macro-precision* 50,24%, *macro-recall* 50,20%, dan *macro-F1* 48,93%. Akurasi tersebut berada di bawah *majority-class baseline* 54,5%, sehingga model belum melampaui prediksi sederhana yang selalu memilih kelas mayoritas. Kesalahan terutama terjadi pada kelas valid, yang sering diprediksi sebagai misinformasi. Hasil ini menunjukkan bahwa variasi bahasa informal, *slang*, singkatan, kemiripan kosakata antarkelas, serta keterbatasan representasi *bag-of-words* masih menjadi kendala utama. Penelitian ini dapat digunakan sebagai titik acuan awal untuk pengembangan deteksi misinformasi kebencanaan pada komentar media sosial Indonesia. Penelitian lanjutan perlu menambahkan normalisasi bahasa informal, mendokumentasikan parameter eksperimen secara lengkap, menggunakan evaluasi *stratified k-fold*, dan membandingkan model dengan representasi kontekstual seperti *IndoBERT* atau *transformer* lainnya.

PERNYATAAN PENGGUNAAN AI

Penulis menyatakan bahwa dalam penelitian ini tidak menggunakan AI

KONTRIBUSI PENULIS

Konseptualisasi: Febri Yalda Sulistia, Karina Nurfebia, Nurbeti Sinulingga, Nuzul Aini Ramadhani
Metodologi: Karina Nurfebia
Pengembangan perangkat lunak/sistem: -
Pengumpulan data: Nurbeti Sinulingga
Analisis data: Nurbeti Sinulingga, Nuzul Aini Ramadhani
Validasi: Nuzul Aini Ramadhani
Visualisasi: Nuzul Aini Ramadhani, Nurbeti Sinulingga
Penelusuran literatur: Karina Nurfebia
Penulisan draf awal: Febri Yalda Sulistia
Penelaahan dan penyuntingan: Khairul
Supervisi: -
Administrasi penelitian: Karina Nurfebia

PERNYATAAN PENDANAAN

Penelitian ini dilakukan secara mandiri oleh penulis dan tidak mendapatkan dukungan pendanaan dari institusi kampus, pihak eksternal, maupun program hibah penelitian mana pun.

KONFLIK KEPENTINGAN

Para penulis menyatakan tidak terdapat konflik kepentingan dalam penelitian dan publikasi artikel ini.

DAFTAR PUSTAKA

[1] S. Hilberts, M. Govers, E. Petelos, and S. Evers, “The impact of misinformation on social media in the context of natural disasters: Narrative review,” *JMIR Infodemiology*, vol. 5, Art. no. e70413, 2025, doi: 10.2196/70413.

[2] S. Mızrak, “Public's social media use during the Kahramanmaraş earthquakes on 6 February 2023,” *International Journal of Disaster Risk Reduction*, vol. 108, Art. no. 104541, 2024, doi: 10.1016/j.ijdrr.2024.104541.

[3] I. Dallo, O. Elroy, L. Fallou, N. Komendantova, and A. Yosipof, “Dynamics and characteristics of misinformation related to earthquake predictions on Twitter,” *Scientific Reports*, vol. 13, Art. no. 13391, 2023, doi: 10.1038/s41598-023-40399-9.

[4] W. Zhai, H. Yu, and C. Y. Song, “Disaster misinformation and its corrections on social media: Spatiotemporal proximity, social network, and sentiment contagion,” *Annals of the American Association of Geographers*, vol. 114, no. 2, pp. 408–435, 2024, doi: 10.1080/24694452.2023.2271549.

[5] C.-O. Truică and E.-S. Apostol, “It’s all in the embedding! Fake news detection using document embeddings,” *Mathematics*, vol. 11, no. 3, Art. no. 508, 2023, doi: 10.3390/math11030508.

[6] D. Mouratidis, A. Kanavos, and K. Kermanidis, “From misinformation to insight: Machine learning strategies for fake news detection,” *Information*, vol. 16, no. 3, Art. no. 189, 2025, doi: 10.3390/info16030189.

[7] S. Raza, D. Paulen-Patterson, and C. Ding, “Fake news detection: Comparative evaluation of BERT-like models and large language models with generative AI-annotated data,” *Knowledge and Information Systems*, vol. 67, no. 4, pp. 3267–3292, 2025, doi: 10.1007/s10115-024-02321-1.

[8] N. Agustina, Adrian, and M. Hermawati, “Implementasi algoritma Naïve Bayes classifier untuk mendeteksi berita palsu pada sosial media,” *Faktor Exacta*, vol. 14, no. 4, pp. 206–213, 2021, doi: 10.30998/faktorexacta.v14i4.11259.

[9] I. M. Karo Karo, Romia, S. Dewi, and P. M. Fadilah, "Hoax detection on Indonesian tweets using Naïve Bayes classifier with TF-IDF," *Journal of Information System Research*, vol. 4, no. 3, pp. 914–919, 2023, doi: 10.47065/josh.v4i3.3317.

[10] D. B. C. Prasetiyo, P. N. Andono, and C. Supriyanto, "Metode Naive Bayes classifier dan forward selection untuk deteksi berita hoaks bahasa Indonesia," *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1541–1550, 2023, doi: 10.30865/mib.v7i3.6459.

[11] S. Fernando, A. Voutama, and A. A. Hendriadi, "Klasifikasi berita hoaks kampanye pemilihan umum (Pemilu) 2024 menggunakan algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2112–2115, 2024, doi: 10.36040/jati.v8i2.9400.

[12] R. Adrian, Musaddam, M. Ikhsan, and M. R. Pahlevi B., "Detection of hoax news using TF-IDF vectorizer and Multinomial Naïve Bayes and Passive Aggressive," *Media Journal of General Computer Science*, vol. 1, no. 2, pp. 54–61, 2024, doi: 10.62205/mjgcs.v1i2.24.

[13] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment analysis of tweets before the 2024 elections in Indonesia using IndoBERT language models," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 746–757, 2023, doi: 10.26555/jiteki.v9i3.26490.

[14] M. Y. Ridho and E. Yulianti, "From text to truth: Leveraging IndoBERT and machine learning models for hoax detection in Indonesian news," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 3, pp. 544–555, 2024, doi: 10.26555/jiteki.v10i3.29450.

[15] A. Kunaefi, Z. Abidin, and R. Kusumawati, "Klasifikasi berita hoaks bahasa Indonesia menggunakan IndoBERT fine-tuning dengan pendekatan focal loss pada data tidak seimbang," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1706–1714, 2025, doi: 10.29100/jipi.v10i2.7811.