

ANALISIS KOMPARASI METODE SVM DAN MODERNBERT DALAM KLASIFIKASI SENTIMEN OPINI PUBLIK DI PLATFORM TIKTOK MENGENAI BANJIR BANDANG DI SUMATRA

Sabila Ayu Febriani, Virgania Sari

Matematika, Universitas Negeri Semarang

Jl. Raya Banaran, Sekaran, Kec. Gn. Pati, Kota Semarang, Jawa Tengah

Sabilaayufebriani418@gmail.com

ABSTRAK

Peristiwa banjir bandang di Sumatra pada tahun 2025 memicu berbagai reaksi masyarakat yang banyak disampaikan melalui media sosial, khususnya TikTok. Namun, tingginya volume komentar yang tidak terstruktur menyulitkan dalam memahami kecenderungan opini publik secara sistematis. Oleh karena itu, diperlukan pendekatan analisis sentimen untuk mengidentifikasi polaritas opini masyarakat secara otomatis. Penelitian ini bertujuan untuk menganalisis distribusi sentimen serta membandingkan kinerja metode Support Vector Machine (SVM), ModernBERT dan BERT sebagai metode pembandingan tambahan dalam mengklasifikasikan sentimen komentar TikTok terkait banjir bandang di Sumatra. Metode yang digunakan adalah pendekatan kuantitatif komparatif dengan dataset sebanyak 6.999 komentar yang dikumpulkan melalui teknik web scraping. Data diproses melalui tahapan *preprocessing*, kemudian diklasifikasikan menggunakan SVM berbasis TF-IDF serta model BERT dan ModernBERT melalui proses *fine-tuning*. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa ModernBERT memiliki performa terbaik dengan akurasi sekitar 93%, diikuti SVM sebesar 89,43%, dan BERT sebesar 88%. Temuan ini menunjukkan bahwa model berbasis Transformer yang telah dioptimasi mampu memahami konteks bahasa informal media sosial dengan lebih baik dibandingkan metode lainnya.

Kata kunci : Analisis Sentimen, TikTok, Support Vector Machine, ModernBERT, Machine Learning

1. PENDAHULUAN

Indonesia merupakan negara yang memiliki tingkat kerentanan tinggi terhadap berbagai bencana hidrometeorologi seperti banjir, tanah longsor, dan cuaca ekstrem. Peningkatan frekuensi bencana tersebut dipengaruhi oleh berbagai faktor, antara lain perubahan iklim global, kerusakan lingkungan, serta perubahan tata guna lahan yang tidak terkendali [1].

Salah satu peristiwa yang menjadi perhatian publik adalah banjir bandang yang terjadi di beberapa wilayah Sumatra pada akhir tahun 2025. Peristiwa ini tidak hanya menimbulkan dampak fisik, tetapi juga memicu berbagai reaksi masyarakat yang banyak disampaikan melalui media sosial.

Media sosial saat ini tidak hanya berfungsi sebagai sarana komunikasi dan hiburan, tetapi juga menjadi ruang publik digital bagi masyarakat untuk menyampaikan opini, kritik, serta persepsi terhadap suatu peristiwa sosial [2]. Salah satu platform yang mengalami perkembangan pesat adalah TikTok, yang memungkinkan pengguna untuk berbagi konten video pendek serta memberikan tanggapan melalui fitur komentar [3]. Banyaknya komentar yang dihasilkan membentuk kumpulan data teks yang besar dan tidak terstruktur, sehingga sulit dianalisis secara manual untuk mengetahui kecenderungan opini publik.

Permasalahan utama dalam penelitian ini adalah bagaimana mengolah dan menganalisis data komentar TikTok yang tidak terstruktur untuk mengidentifikasi sentimen opini masyarakat secara akurat, serta bagaimana membandingkan performa metode klasifikasi yang berbeda dalam memahami

karakteristik bahasa media sosial yang cenderung informal dan bervariasi. Analisis sentimen sebagai bagian dari Natural Language Processing (NLP) dapat digunakan untuk mengklasifikasikan opini ke dalam kategori positif, negatif, dan netral [4].

Metode machine learning seperti Support Vector Machine (SVM) telah banyak digunakan karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan model yang stabil [5]. Namun, metode ini memiliki keterbatasan dalam memahami konteks semantik yang kompleks.

Seiring dengan perkembangan teknologi deep learning, model berbasis arsitektur Transformer seperti BERT mampu memahami konteks bahasa secara lebih mendalam melalui representasi kontekstual [6]. Pengembangan lebih lanjut dari model tersebut adalah ModernBERT yang menawarkan peningkatan efisiensi komputasi serta kemampuan pemrosesan konteks teks yang lebih panjang [7].

Oleh karena itu, diperlukan penelitian yang membandingkan performa metode berbasis machine learning dan deep learning dalam analisis sentimen pada data media sosial.

Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna TikTok terkait banjir bandang di Sumatra serta membandingkan kinerja metode Support Vector Machine (SVM), BERT, dan ModernBERT dalam melakukan klasifikasi sentimen. Rencana penyelesaian masalah dilakukan melalui pengumpulan data komentar TikTok menggunakan teknik web scraping, dilanjutkan dengan tahap pra-pemrosesan data, pelatihan model klasifikasi

menggunakan SVM berbasis TF-IDF serta model BERT dan ModernBERT melalui proses fine-tuning, serta evaluasi kinerja model menggunakan metrik akurasi, precision, recall, dan F1-score. Dengan pendekatan tersebut, diharapkan dapat diperoleh metode terbaik dalam mengklasifikasikan sentimen pada data media sosial yang bersifat tidak terstruktur.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang dari Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi, mengekstraksi, serta mengklasifikasikan opini atau emosi yang terkandung dalam suatu teks. Analisis sentimen bertujuan untuk menentukan polaritas suatu pernyataan, apakah termasuk ke dalam kategori positif, negatif, atau netral [4].

Metode ini banyak digunakan dalam berbagai penelitian untuk memahami persepsi masyarakat terhadap suatu produk, layanan, maupun isu sosial yang berkembang di masyarakat.

Perkembangan media sosial telah menghasilkan volume data teks yang sangat besar dan bersifat tidak terstruktur. Oleh karena itu, analisis sentimen menjadi pendekatan yang efektif untuk menganalisis opini publik secara otomatis dengan memanfaatkan teknik pembelajaran mesin dan pemrosesan bahasa alami [8].

Data yang berasal dari media sosial biasanya memiliki karakteristik bahasa yang informal, seperti penggunaan singkatan, slang, serta variasi ekspresi emosi yang beragam.

Dalam beberapa penelitian, analisis sentimen telah diterapkan pada berbagai platform media sosial seperti Twitter, Facebook, dan Instagram untuk mengetahui kecenderungan opini masyarakat terhadap suatu isu tertentu [2].

Namun demikian, penelitian yang memanfaatkan data dari platform TikTok masih relatif terbatas, padahal platform ini memiliki tingkat interaksi pengguna yang sangat tinggi. Oleh karena itu, penelitian mengenai analisis sentimen pada komentar TikTok menjadi penting untuk memahami dinamika opini publik yang berkembang di media sosial.

2.2. Media Sosial TikTok

Media sosial merupakan platform komunikasi digital yang memungkinkan pengguna untuk berinteraksi, berbagi informasi, serta menghasilkan konten secara daring. Salah satu platform media sosial yang mengalami perkembangan pesat dalam beberapa tahun terakhir adalah TikTok. TikTok merupakan aplikasi berbasis video pendek yang memungkinkan pengguna untuk membuat, mengedit, serta membagikan konten video dengan durasi singkat kepada pengguna lainnya [3].

Popularitas TikTok terus meningkat secara global, termasuk di Indonesia. Platform ini tidak hanya digunakan sebagai media hiburan, tetapi juga menjadi sarana bagi masyarakat untuk menyampaikan opini,

kritik, maupun tanggapan terhadap berbagai isu sosial yang sedang terjadi. Melalui fitur komentar yang tersedia pada setiap video, pengguna dapat memberikan tanggapan yang mencerminkan persepsi mereka terhadap suatu peristiwa atau fenomena tertentu [2].

Karakteristik bahasa yang digunakan dalam komentar TikTok umumnya bersifat informal, spontan, dan seringkali mengandung unsur emosi yang kuat. Hal ini menyebabkan data komentar pada TikTok menjadi sumber data yang menarik untuk dianalisis menggunakan pendekatan analisis sentimen. Dengan menganalisis sentimen komentar pengguna, peneliti dapat memperoleh gambaran mengenai kecenderungan opini publik terhadap suatu isu yang sedang berkembang di masyarakat.

2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu algoritma pembelajaran terawasi (supervised learning) yang banyak digunakan dalam permasalahan klasifikasi dan regresi. Algoritma ini bekerja dengan menemukan *hyperplane optimal* dalam ruang fitur berdimensi tinggi untuk memisahkan data ke dalam kelas-kelas yang berbeda [5].

Prinsip utama dari SVM adalah memaksimalkan jarak atau margin antara hyperplane dengan titik data terdekat dari masing-masing kelas yang disebut sebagai *support vectors*.

Secara matematis, untuk data yang dapat dipisahkan secara linier, persamaan hyperplane dapat dinyatakan sebagai berikut:

$$w \cdot x + b = 0 \quad (1)$$

di mana w merupakan vektor normal terhadap hyperplane, x adalah vektor fitur dari data, dan b adalah nilai bias.

Dalam kondisi data nyata, seringkali terdapat data yang tidak dapat dipisahkan secara sempurna akibat adanya noise atau overlap antar kelas. Oleh karena itu, SVM menggunakan pendekatan *Soft Margin* yang memungkinkan terjadinya kesalahan klasifikasi pada tingkat tertentu dengan memperkenalkan variabel slack. Fungsi tujuan dari Soft Margin SVM dapat dirumuskan sebagai berikut [9]:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (2)$$

di mana ξ_i merupakan *variabel slack* yang mengizinkan kesalahan klasifikasi pada data ke- i , sedangkan C merupakan parameter regulasi yang digunakan untuk menyeimbangkan antara lebar margin dan tingkat kesalahan pelatihan.

Metode SVM banyak digunakan dalam analisis teks karena kemampuannya dalam menangani data berdimensi tinggi serta menghasilkan model klasifikasi yang stabil dengan tingkat generalisasi yang baik [5].

2.4. Term Frequency–Inverse Document Frequency (TF-IDF)

Dalam proses klasifikasi teks, data teks perlu diubah terlebih dahulu ke dalam bentuk numerik agar dapat diproses oleh algoritma machine learning. Salah satu metode yang umum digunakan untuk merepresentasikan teks adalah *Term Frequency–Inverse Document Frequency (TF-IDF)*. Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam suatu dokumen relatif terhadap keseluruhan dokumen dalam korpus [10].

Rumus umum TF-IDF dapat dinyatakan sebagai berikut:

$$TFIDF(t, d) = TF(t, d) \times \log\left(\frac{N}{df(t)}\right) \quad (3)$$

di mana:

$TF(t, d)$ = frekuensi kemunculan kata t dalam dokumen d

N = jumlah total dokumen dalam korpus

$df(t)$ = jumlah dokumen yang mengandung kata t

Melalui pembobotan TF-IDF, kata-kata yang sering muncul dalam suatu dokumen tetapi jarang muncul pada dokumen lain akan memperoleh bobot yang lebih tinggi. Hal ini memungkinkan model klasifikasi untuk lebih fokus pada kata-kata yang memiliki informasi penting dalam proses analisis sentimen.

2.5. ModernBERT dan Arsitektur Transformer

ModernBERT merupakan pengembangan terbaru dari model *Bidirectional Encoder Representations from Transformers (BERT)* yang dirancang untuk meningkatkan efisiensi komputasi serta kemampuan pemrosesan konteks yang lebih panjang. Model ini termasuk dalam kategori *encoder-only transformer models* yang digunakan secara luas dalam berbagai tugas pemrosesan bahasa alami (Natural Language Processing/NLP) seperti klasifikasi teks, analisis sentimen, dan pemahaman bahasa alami [7].

ModernBERT mengintegrasikan beberapa inovasi teknis, salah satunya adalah penggunaan *Rotary Positional Embeddings (RoPE)* yang memungkinkan model untuk memproses konteks teks hingga panjang yang lebih besar secara efisien. Pendekatan ini mendukung pemrosesan hingga ribuan token dalam satu input tanpa mengalami penurunan performa yang signifikan [11].

Mekanisme utama dalam arsitektur Transformer adalah *self-attention*, yang memungkinkan model untuk menimbang hubungan antar kata dalam suatu kalimat secara bidirectional. Mekanisme tersebut dapat dirumuskan sebagai berikut:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

di mana:

Q = Query K = Key

V = Value d_k = dimensi vektor

Melalui mekanisme self-attention, model mampu memahami hubungan semantik antar kata dalam suatu kalimat secara lebih mendalam dibandingkan metode berbasis frekuensi kata seperti TF-IDF.

2.6. Perbandingan Karakteristik Model

Tabel 1. Perbandingan Karakteristik Model

Karakteristik	SVM	ModernBERT
Paradigma	Statistika Klasik	Deep Learning
Representasi	Frekuensi Leksikal	Contextual Embedding
Kebutuhan GPU	Tidak wajib	Wajib

Dari Tabel 1 dapat dilihat bahwa kedua metode memiliki karakteristik yang berbeda dalam proses representasi data serta kebutuhan komputasi. Metode SVM termasuk dalam pendekatan machine learning klasik yang menggunakan representasi berbasis frekuensi kata seperti TF-IDF, sehingga tidak memerlukan perangkat komputasi yang terlalu besar. Sebaliknya, ModernBERT merupakan model *deep learning* berbasis Transformer yang menggunakan *contextual embedding* untuk memahami hubungan antar kata dalam suatu teks secara kontekstual. Oleh karena itu, model ini umumnya membutuhkan sumber daya komputasi yang lebih besar seperti GPU untuk proses pelatihan model.

2.7. Penelitian Terdahulu

Beberapa penelitian terdahulu telah mengkaji perbandingan metode Support Vector Machine (SVM) dan model berbasis Transformer seperti BERT dalam analisis sentimen pada media sosial. Penelitian yang dilakukan oleh Adiwinata dan Afyati menunjukkan bahwa SVM memiliki akurasi yang sedikit lebih tinggi sebesar 89,74% dibandingkan BERT sebesar 87,90%, khususnya dalam klasifikasi sentimen positif. Namun demikian, model BERT terbukti lebih unggul dalam menangani sentimen negatif serta memahami nuansa bahasa yang kompleks seperti penggunaan slang dan sarkasme [12].

Penelitian lain oleh Sabrina juga melakukan komparasi antara SVM dan BERT dalam klasifikasi sentimen dan deteksi sarkasme pada platform X. Hasil penelitian tersebut menunjukkan bahwa BERT secara umum memiliki performa yang lebih baik dibandingkan SVM, terutama dalam menangani kompleksitas bahasa yang tidak dapat direpresentasikan secara optimal oleh metode berbasis frekuensi kata [13].

Selanjutnya, Wijaya dalam penelitiannya terkait deteksi kesehatan mental di media sosial menunjukkan bahwa model BERT memiliki keunggulan dalam memahami ekspresi emosional yang kompleks dibandingkan SVM. Hal ini mengindikasikan bahwa model berbasis Transformer lebih efektif dalam menangkap konteks semantik dibandingkan metode machine learning tradisional [14]. Berdasarkan

berbagai penelitian tersebut, dapat disimpulkan bahwa metode SVM dan BERT memiliki keunggulan masing-masing, di mana SVM cenderung unggul dalam stabilitas dan efisiensi komputasi, sedangkan BERT lebih unggul dalam memahami konteks bahasa yang kompleks.

Namun demikian, sebagian besar penelitian masih berfokus pada perbandingan antara SVM dan BERT konvensional, dan belum banyak yang mengkaji penggunaan model terbaru seperti ModernBERT. Selain itu, hingga saat ini belum ditemukan penelitian yang secara khusus melakukan komparasi antara SVM dan ModernBERT, terutama pada data media sosial berbasis TikTok yang memiliki karakteristik bahasa yang sangat informal dan dinamis. Oleh karena itu, penelitian ini memiliki kontribusi dalam mengisi celah penelitian (*research gap*) dengan membandingkan kinerja SVM dan ModernBERT dalam klasifikasi sentimen komentar TikTok terkait banjir bandang di Sumatra.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komparatif yang bertujuan untuk membandingkan kinerja dua algoritma klasifikasi sentimen, yaitu Support Vector Machine (SVM) dan ModernBERT. Proses penelitian dilakukan melalui beberapa tahapan utama yang meliputi pengumpulan data, pra-pemrosesan data, penanganan ketidakseimbangan data, pelatihan model, serta evaluasi kinerja model.

Alur penelitian dimulai dari proses pengumpulan data komentar pengguna TikTok terkait banjir bandang di Sumatra. Data yang telah diperoleh kemudian melalui tahap pra-pemrosesan untuk membersihkan dan menormalisasi teks. Selanjutnya data digunakan dalam proses pelatihan model klasifikasi menggunakan algoritma SVM dan ModernBERT. Tahap terakhir adalah melakukan evaluasi terhadap performa model menggunakan beberapa metrik evaluasi seperti akurasi, precision, recall, dan F1-score.

3.1. Pengumpulan dan Pra-pemrosesan Data

Data yang digunakan dalam penelitian ini berupa komentar pengguna TikTok yang berkaitan dengan peristiwa banjir bandang di wilayah Sumatra. Proses pengumpulan data dilakukan menggunakan teknik web scraping dengan memanfaatkan bahasa pemrograman Python untuk mengambil komentar dari video yang relevan dengan topik penelitian. Melalui proses tersebut diperoleh sebanyak 6.999 komentar yang kemudian digunakan sebagai dataset penelitian.

Setelah proses pengumpulan data, tahap berikutnya adalah pra-pemrosesan data (data preprocessing) yang bertujuan untuk meningkatkan kualitas data teks sebelum dilakukan proses klasifikasi. Tahapan pra-pemrosesan yang diterapkan pada model Support Vector Machine (SVM) meliputi beberapa langkah berikut:

- Case Folding**
Mengubah seluruh huruf dalam teks menjadi huruf kecil untuk menghindari perbedaan makna akibat penggunaan huruf kapital.
- Cleansing**
Menghapus karakter khusus, tanda baca, URL, angka, serta simbol yang tidak memiliki makna linguistik.
- Normalisasi Kata Slang**
Mengubah kata-kata tidak baku atau slang yang umum digunakan pada media sosial menjadi bentuk kata baku.
- Stopword Removal**
Menghapus kata-kata umum yang tidak memiliki kontribusi signifikan terhadap makna kalimat.

Sementara itu, pada model ModernBERT, proses pra-pemrosesan dilakukan secara lebih minimalis. Hal ini bertujuan untuk mempertahankan integritas semantik kalimat, karena model berbasis Transformer telah memiliki mekanisme pemahaman konteks bahasa yang lebih kompleks melalui proses tokenisasi internal [7].

3.2. Penanganan Ketidakseimbangan Data

Dalam analisis sentimen, seringkali ditemukan kondisi ketidakseimbangan distribusi kelas pada dataset, di mana jumlah data pada satu kelas lebih dominan dibandingkan kelas lainnya. Ketidakseimbangan ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas sehingga menurunkan performa prediksi pada kelas minoritas.

Untuk mengatasi permasalahan tersebut, pada model *Support Vector Machine* (SVM) diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Metode SMOTE bekerja dengan menghasilkan data sintetis baru pada kelas minoritas melalui proses interpolasi linier antara sampel data yang berdekatan dalam ruang fitur [15].

Secara matematis, proses pembangkitan sampel sintetis pada SMOTE dapat dirumuskan sebagai berikut:

$$x_{new} = x_i + \delta(x_{zi} - x_i) \quad (5)$$

di mana:

x_i = sampel data minoritas

x_{zi} = salah satu tetangga terdekat dari x_i

δ = nilai acak antara 0 dan 1

Melalui pendekatan ini, jumlah data pada kelas minoritas dapat ditingkatkan sehingga distribusi kelas dalam dataset menjadi lebih seimbang.

Sementara itu, pada model ModernBERT, penanganan ketidakseimbangan data dilakukan secara internal melalui penggunaan fungsi kerugian *Weighted Cross-Entropy* pada tahap pelatihan model. Metode ini memberikan penalti yang lebih besar terhadap kesalahan klasifikasi pada kelas minoritas sehingga model dapat belajar secara lebih seimbang terhadap seluruh kelas sentimen.

3.3. Pelatihan Model Klasifikasi

Setelah tahap preprocessing dan penanganan ketidakseimbangan data selesai dilakukan, langkah selanjutnya adalah proses pelatihan model klasifikasi menggunakan algoritma SVM dan ModernBERT.

Pada metode SVM, data teks terlebih dahulu diubah menjadi representasi numerik menggunakan teknik TF-IDF sebelum dilakukan proses pelatihan model. Model kemudian dilatih untuk menemukan hyperplane optimal yang dapat memisahkan kelas sentimen secara maksimal.

Sedangkan pada metode ModernBERT, proses pelatihan dilakukan dengan pendekatan fine-tuning terhadap model *pre-trained* yang telah dilatih sebelumnya pada korpus teks berskala besar. Proses fine-tuning memungkinkan model menyesuaikan parameter internalnya dengan karakteristik dataset penelitian.

3.4. Evaluasi Model

Evaluasi model dilakukan untuk mengetahui tingkat kinerja algoritma klasifikasi yang digunakan dalam penelitian ini. Proses evaluasi dilakukan menggunakan *confusion matrix* yang terdiri dari empat komponen utama yaitu:

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)

Berdasarkan nilai tersebut kemudian dihitung beberapa metrik evaluasi sebagai berikut:

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Precision

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Recall

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

Penggunaan beberapa metrik evaluasi tersebut bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai performa model klasifikasi dalam mengidentifikasi sentimen pada komentar pengguna TikTok.

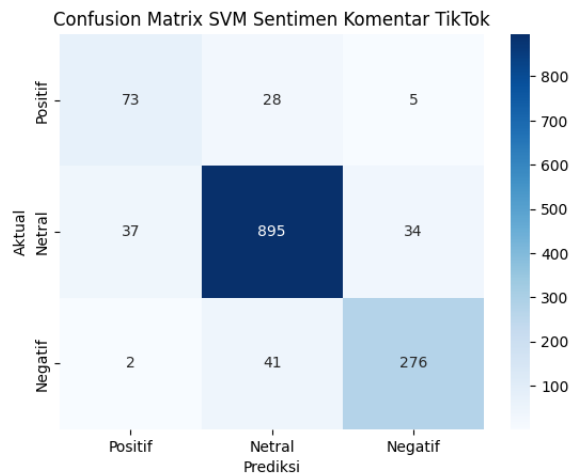
4. HASIL DAN PEMBAHASAN

4.1. Performa Klasifikasi Support Vector Machine (SVM)

Model Support Vector Machine (SVM) yang dikombinasikan dengan metode pembobotan TF-IDF menunjukkan performa yang cukup stabil dalam melakukan klasifikasi sentimen pada data komentar TikTok. Metode ini memanfaatkan representasi frekuensi kata untuk mengidentifikasi pola sentimen yang terdapat dalam teks pendek pada media sosial.

Tabel 2. Hasil classification report untuk model SVM

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.88	0.87	0.87	319
Netral	0.93	0.93	0.93	966
Positif	0.65	0.69	0.67	106
Accuracy			0.89	1391



Gambar 1. Confusion matrix SVM sentimen komentar TikTok

Berdasarkan Tabel 2 dan Gambar 1, model SVM berhasil mencapai akurasi sebesar 89,43%. *Confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar, terutama pada kelas Netral yang memiliki jumlah prediksi benar sebanyak 895 sampel. Hal ini menunjukkan bahwa model SVM cukup efektif dalam mengenali komentar yang bersifat informatif atau netral.

Namun demikian, performa model pada kelas Positif masih relatif lebih rendah dibandingkan kelas lainnya. Hal ini terlihat dari nilai F1-score sebesar 0,67 pada kelas tersebut. Kondisi ini mengindikasikan bahwa model mengalami kesulitan dalam membedakan kata-kata yang mengandung sentimen positif dengan kalimat yang bersifat informatif atau netral. Permasalahan tersebut terjadi karena metode TF-IDF hanya mempertimbangkan frekuensi kemunculan kata, sehingga hubungan konteks antar kata dalam kalimat tidak dapat dipahami secara mendalam.

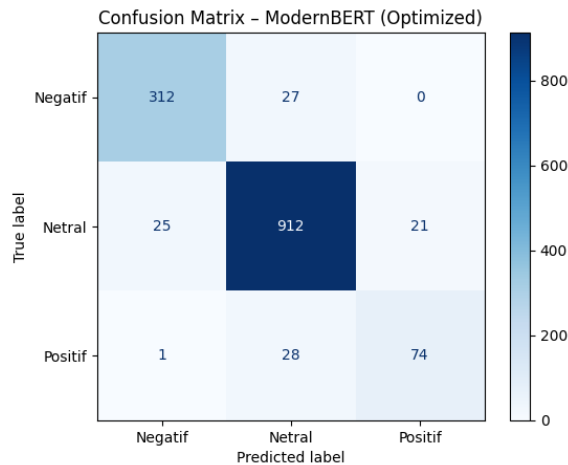
Dengan demikian, meskipun SVM mampu memberikan performa yang cukup baik dalam klasifikasi sentimen, metode ini masih memiliki keterbatasan dalam memahami struktur semantik yang kompleks pada teks media sosial.

4.2. Performa Klasifikasi ModernBERT

Model ModernBERT yang digunakan dalam penelitian ini telah melalui proses *fine-tuning* dengan teknik optimasi seperti *early stopping* untuk mencegah terjadinya overfitting selama proses pelatihan model.

Tabel 3. Hasil classification report Metode ModernBERT (Optimized)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.92	0.92	0.92	339
Netral	0.94	0.95	0.95	958
Positif	0.78	0.72	0.75	103
Accuracy			0.93	1400



Gambar 2. Confusion matrix ModernBERT (Optimized)

Berdasarkan Tabel 3 dan Gambar 2, model ModernBERT menunjukkan peningkatan performa yang signifikan dibandingkan metode SVM. Model ini berhasil mencapai akurasi global sebesar 93%, yang menunjukkan kemampuan klasifikasi yang lebih baik dalam mengidentifikasi sentimen pada komentar TikTok.

Nilai F1-score pada kelas negatif mencapai 0,92, sedangkan pada kelas netral mencapai 0,95, yang menunjukkan bahwa model mampu mengenali pola sentimen dengan tingkat presisi dan recall yang tinggi. Selain itu, performa pada kelas positif juga mengalami peningkatan, dengan nilai F1-score sebesar 0,75, lebih tinggi dibandingkan model SVM.

Berdasarkan analisis confusion matrix, terlihat bahwa ModernBERT mampu mengurangi kesalahan klasifikasi antar kelas secara signifikan. Hal ini disebabkan oleh penggunaan mekanisme self-attention pada arsitektur Transformer yang memungkinkan model memahami hubungan kontekstual antar kata dalam suatu kalimat.

Selain itu, penggunaan *Rotary Positional Embeddings* (RoPE) pada arsitektur ModernBERT memungkinkan model menangkap ketergantungan semantik jarak jauh dalam teks. Hal ini sangat penting dalam analisis komentar TikTok yang sering kali memiliki struktur kalimat yang tidak baku, singkat, dan mengandung berbagai ekspresi informal.

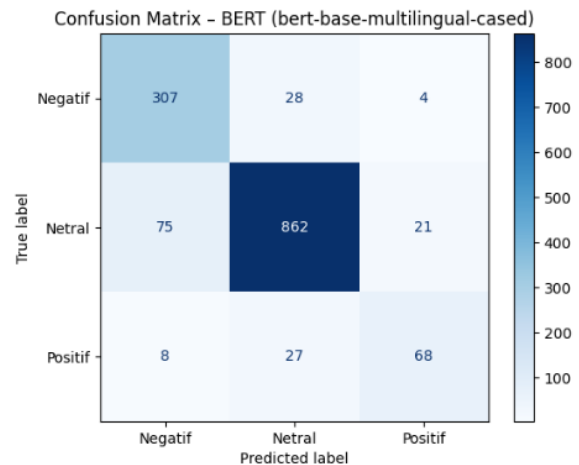
4.3. Performa Klasifikasi BERT

Sebagai pembandingan tambahan, penelitian ini juga melakukan pengujian menggunakan model BERT konvensional, yaitu *bert-base-multilingual-cased*, yang merupakan salah satu model transformer yang

banyak digunakan dalam berbagai penelitian *Natural Language Processing*.

Tabel 4. Hasil classification report untuk model BERT (Multilingual)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.79	0.91	0.84	339
Netral	0.94	0.90	0.92	958
Positif	0.73	0.66	0.69	103
Accuracy			0.88	1400



Gambar 3. Confusion matrix BERT

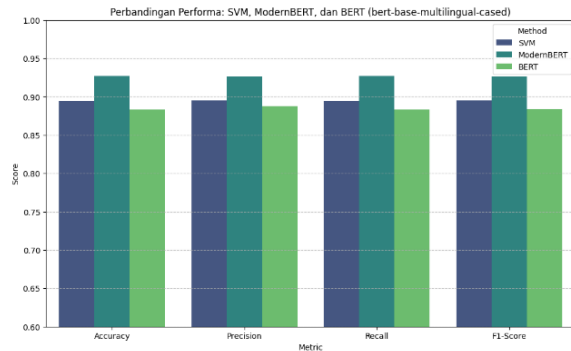
Berdasarkan Tabel 4 dan Gambar 3, model BERT memperoleh akurasi sebesar 88%. Meskipun model ini berbasis arsitektur Transformer, performanya ternyata sedikit berada di bawah model SVM dan cukup jauh di bawah ModernBERT.

Hasil ini menunjukkan bahwa model BERT konvensional tidak selalu memberikan performa terbaik ketika digunakan pada data media sosial yang memiliki karakteristik bahasa yang sangat informal dan bising. Hal ini dapat disebabkan oleh beberapa faktor, salah satunya adalah tokenisasi standar BERT yang kurang optimal dalam menangani variasi kata slang, singkatan, dan struktur kalimat yang tidak baku pada komentar TikTok.

Selain itu, model *bert-base-multilingual-cased* dirancang untuk menangani berbagai bahasa secara umum, sehingga kapasitas representasi bahasa spesifik pada dataset penelitian menjadi kurang optimal dibandingkan model yang telah dioptimasi secara khusus seperti ModernBERT.

4.4. Analisis Komparatif Tiga Metode

Untuk memperoleh gambaran yang lebih komprehensif mengenai performa ketiga metode klasifikasi yang digunakan, dilakukan visualisasi perbandingan performa model berdasarkan beberapa metrik evaluasi.



Gambar 4. Perbandingan performa: SVM, ModernBERT, dan BERT

Berdasarkan grafik pada Gambar 4, terlihat bahwa model ModernBERT secara konsisten mengungguli kedua metode lainnya, yaitu SVM dan BERT, pada seluruh metrik evaluasi yang digunakan, termasuk Accuracy, Precision, Recall, dan F1-Score.

ModernBERT memperoleh nilai akurasi tertinggi sebesar sekitar 0,93, diikuti oleh SVM dengan nilai sekitar 0,89, dan BERT dengan nilai sekitar 0,88. Hasil ini menunjukkan bahwa pendekatan berbasis Transformer yang telah dioptimasi secara arsitektural mampu memberikan performa yang lebih baik dalam memahami konteks bahasa pada data media sosial.

Temuan menarik dari penelitian ini adalah performa SVM yang mampu mengungguli model BERT konvensional. Hal ini menunjukkan bahwa pada data teks yang relatif singkat seperti komentar TikTok, pendekatan berbasis statistik seperti SVM dengan TF-IDF terkadang masih dapat memberikan performa yang kompetitif.

Namun demikian, ModernBERT tetap menjadi metode dengan performa terbaik karena kemampuannya dalam memahami konteks semantik kata melalui mekanisme self-attention serta optimasi pada arsitektur encoder. Dengan demikian, model ModernBERT terbukti lebih efektif dalam melakukan analisis sentimen pada data media sosial yang memiliki struktur bahasa yang kompleks dan tidak terstruktur.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode Support Vector Machine (SVM) dan ModernBERT dapat digunakan untuk melakukan klasifikasi sentimen pada komentar pengguna TikTok terkait banjir bandang di Sumatra. Hasil pengujian menunjukkan bahwa model ModernBERT memiliki performa yang lebih baik dibandingkan SVM dalam memahami konteks bahasa yang kompleks pada data media sosial. Hal ini menunjukkan bahwa model berbasis Transformer lebih efektif digunakan dalam analisis sentimen pada data teks yang bersifat tidak terstruktur. Untuk penelitian selanjutnya disarankan menggunakan dataset yang lebih besar serta mengembangkan model berbasis deep learning lainnya guna meningkatkan akurasi klasifikasi sentimen.

DAFTAR PUSTAKA

- [1] M. B. Sulthan, I. Wahyudi, and L. Suhartini, "Analisis Sentimen Pada Bencana Alam Menggunakan Deep Neural Network dan Information Gain," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 2, no. 2, pp. 65–71, 2021, doi: 10.31102/jatim.v2i2.1273.
- [2] F. Fathoni, A. Ibrahim, F. Rizka Mumtaz, M. Azmi Zaky, M. Jodi Pratama, and I. Akbar Kurniawan, "Analisis Sentimen Public Twitter Terhadap Kebijakan Pemerintah Menggunakan Metode Svm (Studi Kasus : Ruu Tni)," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 4, pp. 6322–6329, 2025, doi: 10.36040/jati.v9i4.14036.
- [3] A. N. Fitri, L. B. Pertiwi, and M. P. Sary, "Pengaruh Media Sosial TikTok terhadap Kreativitas Mahasiswa Ilmu Komunikasi Universitas Negeri Jakarta Angkatan 2019," *Komuniti J. Komun. dan Teknol. Inf.*, vol. 13, no. 1, pp. 37–46, 2021, doi: 10.23917/komuniti.v13i1.13045.
- [4] E. Nofiyanti and E. M. O. N. Haryanto, "Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia," *J. Ilm. SINUS*, vol. 19, no. 2, p. 17, 2021, doi: 10.30646/sinus.v19i2.563.
- [5] B. U. Fahnun and S. A. S. Muhammad, "Perbandingan Model Machine Learning Dalam Analisis Sentimen Ulasan Pengguna Aplikasi E-Commerce," *J. Ilm. Teknol. dan Rekayasa*, vol. 30, no. 1, pp. 73–94, 2025, doi: 10.35760/tr.2025.v30i1.13530.
- [6] E. C. Sukmawati, L. Suryaningrum, D. Angelica, and N. G. Ramadhan, "Klasifikasi Berita Palsu Menggunakan Model Bidirectional Encoder Representations From Transformers (BERT)," *SisInfo*, vol. 6, no. 2, pp. 76–85, 2024, doi: 10.37278/sisinfo.v6i2.934.
- [7] B. Warner *et al.*, "Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference," vol. 1, pp. 2526–2547, 2024, doi: 10.18653/v1/2025.acl-long.127.
- [8] N. M. T. O. Adriana, I. M. A. D. Suarjaya, and D. P. Githa, "ANALISIS SENTIMEN PUBLIK TERHADAP AKSI DEMONSTRASI DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE DAN RANDOM FOREST," *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 236–245, 2023, [Online]. Available: <http://journal.umkendari.ac.id/index.php/decode>
- [9] B. Prasetyo, Ahmad Yusuf Al-Majid, and Suharjito, "A Comparative Analysis of MultinomialNB, SVM, and BERT on Garuda Indonesia Twitter Sentiment," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 12, no. 2, pp. 445–454, 2024, doi: 10.33558/piksel.v12i2.9966.
- [10] V. A. Shahputri and Y. Yamasari, "Analisis Sentimen Mengenai Pasca Bencana Alam Menggunakan Metode K-Nearest Neighbor (K-

- NN) dan Decision Tree,” *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 377–388, 2023, doi: 10.26740/jinacs.v5n03.p377-388.
- [11] W. Antoun, B. Sagot, and D. Seddah, “ModernBERT or DeBERTaV3? Examining Architecture and Data Influence on Transformer Encoder Models Performance,” vol. 2, pp. 1–11, 2025, [Online]. Available: <http://arxiv.org/abs/2504.08716>
- [12] C. Adiwinata and A. Afyati, “Sentiment Analysis of X Users Toward Electric Motorcycles Using SVM and BERT Algorithms,” *JUITA J. Inform.*, vol. 13, no. 2, pp. 119–126, 2025, doi: 10.30595/juita.v13i2.26152.
- [13] S. S. Sabrina, D. F. Shiddieq, and F. F. Roji, “Comparative Analysis of SVM and BERT for Sentiment and Sarcasm Detection in the Boycott of Israeli Products on Platform X,” *Sinkron*, vol. 9, no. 2, pp. 872–883, 2025, doi: 10.33395/sinkron.v9i2.14723.
- [14] H. Wijaya, M. F. Hadi, and N. Sulistianingsih, “Using Sentiment Analysis with BERT and SVM for Detect Mental Health Detection on Social Media,” *J. Digit. Heal. Innov. Med. Technol.*, vol. 1, no. 2, pp. 54–63, 2025.
- [15] D. I. Sumantiawan, J. E. Suseno, and W. A. Syaifei, “Sentiment Analysis of Customer Reviews Using Support Vector Machine and Smote-Tomek Links For Identify Customer Satisfaction,” *J. Sist. Info. Bisnis*, vol. 13, no. 1, pp. 1–9, 2023, doi: 10.21456/vol13iss1pp1-9.