

Klasifikasi Sentimen Pengguna Platform X Terhadap Isu Geopolitik Israel–Palestina Menggunakan BERT

Cellia Auzia Nugraha¹, Joan Angelina Widians², Nataniel Dengen³, Akhmad Irsyad⁴

¹Informatika, Universitas Mulawarman, celliaauzianugraha@gmail.com

²Magister Informatika, Universitas Mulawarman, angelwidians@unmul.ac.id

³Informatika, Universitas Mulawarman, ndengen@unmul.ac.id

⁴Magister Informatika, Universitas Mulawarman, akhmadirsyad@unmul.ac.id

Keywords:

Sentiment analysis,
BERT,
Deep learning,
Natural language processing,
VADER

ABSTRACT

The geopolitical issue between Israel and Palestine has sparked a wide range of public reactions on social media, particularly on the platform X. Sentiment analysis is a widely discussed topic in Natural Language Processing. This study develops a sentiment classification model for that issue using the BERT Base Uncased architecture, with data labeled automatically through the VADER approach. Data were collected via scraping with Tweet-Harvest, yielding 3,172 tweets, later reduced to 2,362 after preprocessing stages including cleaning, case folding, tokenizing, stopword removal, and stemming. VADER labeling produced a distribution of 73.7% negative sentiment and 26.3% positive sentiment. The model achieved 87.34% accuracy, with precision, recall, and F1-Score for the negative class of 92%, 90%, and 91%, respectively, while the positive class obtained 74%, 79%, and 77%. The macro-average F1-score reached 84%, reflecting the model's performance across both classes with equal weighting despite the imbalanced data distribution. These results indicate that BERT Base Uncased can classify VADER-generated labels with fairly good performance in the context of international geopolitical issues.

Kata Kunci:

Analisis Sentimen,
BERT,
Deep Learning,
Natural Language Processing,
VADER

ABSTRAK

Isu geopolitik antara Israel dan Palestina memunculkan beragam tanggapan publik yang beredar di media sosial, terutama platform X. Analisis sentimen sendiri menjadi salah satu bahasan yang sering dibicarakan dalam bidang *Natural Language Processing*. Penelitian ini membangun model klasifikasi sentimen untuk isu tersebut menggunakan arsitektur BERT Base Uncased, dengan pelabelan data yang dilakukan secara otomatis melalui pendekatan VADER. Data dikumpulkan menggunakan teknik *scraping* melalui Tweet-Harvest, sehingga diperoleh 3.172 tweet yang kemudian berkurang menjadi 2.362 tweet setelah melalui tahap *preprocessing* seperti *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil pelabelan VADER menunjukkan pembagian sentimen negatif sebesar 73,7% dan sentimen positif sebesar 26,3%. Model yang dibangun mencapai akurasi sebesar 87,34% berdasarkan hasil evaluasi dengan nilai *precision*, *recall*, dan *F1-score* pada kelas negatif masing-masing sebesar 92%, 90% dan 91%, sedangkan kelas positif memperoleh 74%, 79% dan 77%. Nilai *macro-average* F1-score yang dihasilkan sebesar 84%, yang mencerminkan performa model pada kedua kelas dengan bobot yang setara meskipun distribusi data tidak seimbang. Capaian ini menunjukkan bahwa BERT Base Uncased mampu mengklasifikasikan label yang dihasilkan VADER dengan performa yang cukup baik pada konteks isu geopolitik internasional.

Korespondensi Penulis:

Joan Angelina Widians,
Universitas Mulawarman, Samarinda
Telepon: +6282144911888
Email: angelwidians@unmul.ac.id

Submitted : 10-09-2026; Accepted : 07-10-2026;
Published : 09-10-2026

Copyright (c) 2026 The Author (s) This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

1. PENDAHULUAN

Konflik Israel-Palestina merupakan isu geopolitik yang telah berlangsung selama puluhan tahun dan bersifat rumit, terlihat dari adanya pendudukan wilayah, benturan kepentingan, serta upaya masing-masing pihak untuk memperoleh hak menentukan nasib sendiri. Akar permasalahan ini dipicu oleh perbedaan pandangan historis, keyakinan agama dan klaim atas wilayah, yang pada praktiknya turut menyeret keterlibatan sejumlah pihak, mulai

dari negara-negara terkait, kelompok berbasis keagamaan, hingga lembaga internasional. Dampak dari persoalan ini cukup besar terhadap kondisi geopolitik di kawasan tersebut, menjadikannya salah satu pemicu ketidakstabilan yang bertahan hingga beberapa dekade belakangan [1]. Konflik ini kini berkembang menjadi konflik antara Hamas dan Israel, dan dampaknya meluas melampaui kawasan Timur Tengah hingga memengaruhi geopolitik global [2]. Ketegangan yang terus berlanjut tersebut turut menarik perhatian publik di berbagai negara, salah satunya tercermin dari perdebatan yang berlangsung di jejaring sosial [3].

X merupakan platform media sosial yang cukup diandalkan pengguna di Indonesia sebagai sarana berinteraksi sekaligus memperoleh informasi dengan cepat [4]. Publik memanfaatkan X sebagai wadah untuk menyampaikan pendapat terhadap berbagai topik yang sedang dibicarakan. Interaksi antar pengguna pada platform tersebut menghasilkan kumpulan tulisan tweet yang dapat diolah dan dianalisis guna mengetahui sentimen pengguna X terhadap peristiwa yang sedang berlangsung [5]. Opini publik di X juga dipengaruhi oleh faktor eksternal, seperti peristiwa besar dan jatuhnya korban akibat konflik, sehingga diperlukan pendekatan yang dapat menangkap makna bahasa secara lebih menyeluruh sehingga hasil klasifikasi sentimen yang diperoleh benar-benar akurat dan objektif [6].

Analisis sentimen merupakan cabang kajian Natural Language Processing (NLP) yang berfokus pada mengenali dan mengelompokkan sentimen dalam data teks. Analisis ini digunakan untuk mengetahui pandangan atau sikap yang muncul dalam teks. Sentimen tersebut bisa dikategorikan sebagai sentimen positif atau negatif [7]. Analisis sentimen saat ini, yang merupakan bagian dari teks mining, yaitu proses penggalian pola dari informasi yang sebagian besar tidak terstruktur [8], telah menjadi topik yang banyak diminati para peneliti dengan tujuan menyediakan informasi dari dataset yang tidak terstruktur [9]. Relevansi pendekatan ini semakin diperkuat oleh penelitian yang menegaskan bahwa cara ini efektif untuk menilai opini masyarakat di X, khususnya melihat kecenderungan suatu persoalan ke arah positif atau negatif [10].

Deep Learning menjadi salah satu metode yang lazim dipakai dalam analisis sentimen karena beroperasi melalui jaringan saraf tiruan (*neural network*) berlapis untuk mengekstraksi fitur yang kompleks serta merepresentasikan data input secara hierarkis [11]. Selama pelatihan, model ini terus menyesuaikan dirinya guna menekan tingkat kesalahan prediksi. Di sisi lain, opini publik di media sosial terlihat kompleks, seperti yang tampak pada pola penyebaran sentimen pengguna platform X tentang konflik Israel-Palestina yang berbeda-beda menurut waktu dan tempat. Hal ini menunjukkan perlunya pendekatan yang lebih mendalam untuk memahami dinamika bahasa [12]. BERT atau *Bidirectional Encoder Representations from Transformers*, sebagai model transformer, kini lebih diminati karena bisa menangkap makna berdasarkan konteks secara detail [13]. BERT sebagai model bahasa kontekstual memakai pembelajaran dua arah. Model ini mampu mempelajari konteks tulisan dari kiri ke kanan dan sebaliknya secara bersamaan. Dengan cara ini, makna tiap kata jadi lebih tepat tertangkap [14].

Penelitian mengenai analisis sentimen menggunakan BERT pada data X telah berkembang cukup pesat. Namun, masih terdapat beberapa aspek yang belum banyak dibahas. Kebanyakan penelitian terdahulu lebih berpusat di topik domestik, seperti kebijakan pemerintah dan pemilihan umum. Sementara itu, penggunaan BERT untuk menganalisis sentimen publik terhadap isu geopolitik internasional, khususnya konflik Israel–Palestina, masih relatif sedikit diteliti. Meskipun demikian, penelitian lain menerapkan IndoBERT dengan pelabelan otomatis berbasis RoBERTa untuk menganalisis opini publik Indonesia terhadap konflik Israel-Iran dan memperoleh akurasi 75,41% dan F1-score 75,21% dengan sentimen negatif mendominasi sebesar 40,3% [15]. Penelitian lain membandingkan IndoBERT dan mBERT pada 10.140 tweet terkait pemilihan presiden 2024, di mana IndoBERT lebih unggul dengan memperoleh tingkat akurasi 84%, presisi 75%, recall 80%, dan F1-score 78%, sementara mBERT mencatat tingkat akurasi 81% disertai F1-score 73% [16]. Sementara itu, penelitian lain membandingkan metode pelabelan TextBlob, VADER, dan Flair yang diuji menggunakan IndoBERT pada 5.283 data opini publik pasca pelantikan presiden, dengan hasil TextBlob mencapai akurasi 73,35%, VADER mencapai akurasi 74,86%, dan Flair menghasilkan akurasi tertinggi sebesar 81,29% [17]. Penelitian lain juga membandingkan BERT-base-uncased, DistilBERT, dan BERTweet pada 30.000 ulasan aplikasi Instagram, dengan BERT-base-uncased memperoleh F1-score tertinggi 84,75% dan BERTweet unggul dalam presisi sebesar 86,19% [18].

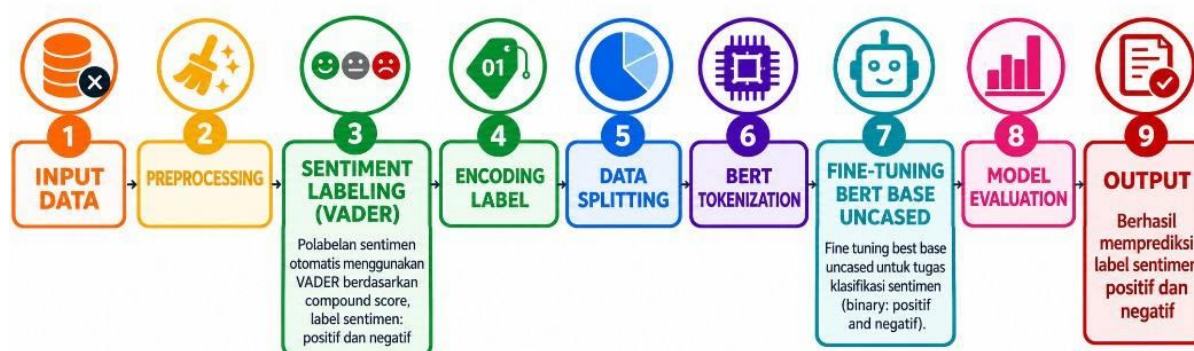
Di sisi lain, sejumlah penelitian mengenai konflik Israel-Palestina masih mengandalkan pendekatan machine learning klasik. Riset yang menganalisis sentimen publik terhadap konflik Gaza pada 2.175 tweet menggunakan ekstraksi fitur TF-IDF membandingkan *Multi-Layer Perceptron*, *XGBoost*, dan *Logistic Regression*, dengan hasil terbaik diperoleh oleh *Logistic Regression* pada akurasi 73,17%, serta secara eksplisit merekomendasikan penggunaan arsitektur berbasis transformer seperti BERT pada penelitian selanjutnya karena TF-IDF dinilai belum mampu menangkap informasi kontekstual pada teks media sosial [19]. Riset lainnya menganalisis sentimen masyarakat Indonesia terhadap konflik Palestina pada 2.960 tweet menggunakan pelabelan *lexicon-based* serta membandingkan KNN dan kombinasi KNN serta SVM, dengan hasil terbaik diperoleh gabungan KNN dan SVM pada akurasi 82%, presisi 83%, recall 82%, dan F1-score 82%, di mana sentimen negatif mendominasi sebesar 38,9% [20]. Sejumlah penelitian tersebut mengindikasikan bahwa model yang memakai BERT terbukti mumpuni dalam analisis sentimen pada berbagai domain dan platform media sosial. Namun, penelitian mengenai konflik Israel–Palestina masih didominasi oleh penggunaan IndoBERT yang dirancang untuk tulisan berbahasa Indonesia, sehingga belum sesuai untuk menganalisis tweet berbahasa Inggris yang mendominasi diskusi global. Selain itu, penelitian yang

mengombinasikan pelabelan otomatis menggunakan VADER dengan model BERT Base untuk klasifikasi sentimen pada isu konflik Israel–Palestina masih belum banyak ditemukan.

Mengacu pada kesenjangan tersebut, riset ini membangun model klasifikasi sentimen terhadap isu geopolitik Israel-Palestina menggunakan model BERT Base Uncased dengan pelabelan otomatis berbasis VADER pada data tweet berbahasa Inggris dari platform X. Riset ini berfokus pada evaluasi kinerja model saat mengelompokkan sentimen menjadi kelas positif dan negatif, sekaligus mengukur performanya melalui indikator evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

2. METODE PENELITIAN

Bidirectional Encoder Representations from Transformers (BERT) Base Uncased menjadi model yang digunakan pada penelitian ini untuk menganalisis sentimen tweet berbahasa Inggris terkait konflik Israel–Palestina pada platform X. BERT dipilih sebagai model klasifikasi sentimen sebab kapasitasnya dalam menangkap makna kata dari dua arah (*bidirectional*) lewat arsitektur Transformer, adapun pelabelan sentimen dilakukan dengan menggunakan *Valence Aware Dictionary and sEntiment Reasoner* (VADER), yaitu pendekatan berbasis leksikon yang dikembangkan untuk mengidentifikasi sentimen dalam tulisan informal serta mengenali ekspresi emosional dalam tulisan, khususnya media sosial [21], [22], [23]. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengambilan Data

Pengumpulan data dilakukan melalui scraping pada platform X menggunakan library Tweet-Harvest v2.6.1, dengan kata kunci "Israel", "Palestine", "Gaza", dan "Hamass" digabungkan dalam satu string pencarian menggunakan operator AND implisit, sehingga tiap tweet harus memuat keempat kata kunci sekaligus; retweet dikecualikan menggunakan *-filter:retweets* dan bahasa dibatasi pada bahasa Inggris menggunakan *lang:en*. Karena pencarian menggunakan satu string gabungan, tidak ada proses pencarian terpisah per kata kunci yang berisiko menimbulkan duplikasi lintas-keyword; duplikasi yang tetap muncul (misalnya diakibatkan oleh tumpang tindih antarperiode) ditangani melalui *drop_duplicates* berdasarkan teks tweet. Data diambil dalam delapan periode enam-bulanan (10 Mei 2021–10 Mei 2025) dengan parameter pengambilan sekitar 400 tweet per periode, menghasilkan 3.172 tweet sebelum preprocessing. Distribusi antarperiode relatif stabil (400, 411, 439, 382, 340, 400, 400, 400), dengan penurunan hanya pada periode Mei–November 2023 (340 tweet).

2.2 Text Preprocessing

Tahap *preprocessing* bertujuan mengubah data mentah yang masih memuat elemen yang tidak diperlukan menjadi data yang lebih teratur, konsisten serta tersusun dengan baik sehingga lebih praktis ketika diproses dalam analisis sentimen. Tahap *preprocessing* tersebut meliputi:

1. Cleaning

Proses cleaning dijalankan guna menghilangkan elemen yang tidak dibutuhkan, contohnya URL, simbol, mention, tagar (hashtag), data kosong, dan data duplikat sehingga data menjadi lebih terstruktur [24].

2. Case Folding

Proses menyeragamkan seluruh karakter ke dalam huruf kecil (lowercase) sehingga menghasilkan format tulisan yang seragam dan mendukung proses klasifikasi agar bekerja secara lebih optimal [25].

3. Tokenizing

Proses memecah sebuah kalimat menjadi kata, istilah, simbol, atau elemen-elemen lain yang memiliki makna berbeda, yang disebut sebagai token [26].

4. Stopword Removal

Proses menghilangkan kata-kata yang tidak relevan [27].

5. Stemming

Tahapan mengubah kata menjadi bentuk dasarnya [28]. Pada penelitian ini, hasil stemming digunakan sebagai masukan dalam proses pelabelan sentimen menggunakan VADER dan proses klasifikasi dengan BERT. Penggunaan stemming dipertahankan untuk menyederhanakan variasi kata pada data tweet yang bersifat informal dan tidak terstruktur. Meskipun BERT menggunakan tokenizer berbasis subword dan mempertimbangkan konteks kalimat, sedangkan VADER mempertimbangkan ekspresi linguistik dalam teks, stemming digunakan secara konsisten sebagai bagian dari pipeline prapemrosesan penelitian.

2.3 Pelabelan Data

Tahapan ini menggunakan pelabelan VADER, yaitu perangkat analisis sentimen yang mengandalkan leksikon serta sejumlah aturan (*lexicon and rule-based*) yang dirancang khusus untuk menyesuaikan dengan sentimen yang diekspresikan di media sosial [29]. Pelabelan sentimen pada penelitian ini terbatas pada kelas positif dan negatif.

2.4 Encoding Label

Hasil pelabelan sentimen yang masih berupa tulisan diubah menjadi nilai numerik menggunakan metode *Encoding Label* dengan bantuan fungsi Label Encoder dari pustaka *scikit-learn* agar dapat diproses oleh model BERT. Pada tahap ini, label *negative* diberi nilai 0 sedangkan label *positive* diberi nilai 1. Tahapan tersebut memungkinkan label dalam bentuk tulisan dikonversi ke dalam bentuk numerik sehingga dapat digunakan oleh algoritma klasifikasi selama proses pelatihan dan evaluasi model [30].

2.5 Data Splitting

Dataset yang sudah diberi label kemudian dipisahkan menjadi tiga kelompok melalui metode hold-out, yaitu data latih (*training*) sebesar 80% sebanyak 1.888 data, data validasi (*validation*) sebesar 10% sebanyak 237 data, dan data uji (*testing*) sebesar 10% sebanyak 237 data. Pembagian dilakukan secara *stratified* untuk mempertahankan proporsi kelas sentimen pada setiap kelompok data, dengan menggunakan *random_state* sebesar 42 agar proses pembagian dapat direproduksi. Sebelum proses pembagian, tweet duplikat berdasarkan teks (*full_text*) telah dihapus menggunakan fungsi *drop_duplicates*, sehingga tweet yang identik tidak tersebar ke kelompok *training*, *validation*, dan *testing*. Penanganan *near-duplicate* secara khusus tidak diterapkan dalam penelitian ini. Data latih digunakan untuk melatih model dalam mengenali pola pada dataset, data validasi digunakan untuk memantau kinerja model selama proses pelatihan guna mengidentifikasi potensi *overfitting*, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data yang belum dikenali sebelumnya [31].

2.6 Model BERT

Tahap pemodelan menggunakan BERT Base Uncased sebagai model klasifikasi sentimen. Setiap tweet diproses terlebih dahulu menggunakan BERT Tokenizer sebelum memasuki tahap pelatihan, di mana proses ini secara otomatis melakukan tokenisasi tulisan, menambahkan token spesifik [CLS] pada bagian awal kalimat dan [SEP] pada akhir kalimat, lalu melakukan padding agar panjang setiap masukan seragam, serta membentuk *attention mask* agar dapat membedakan token asli dengan token padding. Hasil tokenisasi berupa *input_ids* dan *attention_mask* tersebut kemudian dimanfaatkan sebagai *input* pada model BERT. Model BERT yang sudah menyelesaikan tahap *pre-training* kemudian menjalani proses *fine-tuning* yang bermaksud melakukan penyesuaian terhadap parameter model berdasarkan karakteristik data yang digunakan agar model bisa memahami konteks dan mengklasifikasikan sentimen khususnya pada isu sosial-politik [32]. Penelitian ini melatih model menggunakan 80% data latih, 10% data validasi, dan 10% data uji. Proses pelatihan berlangsung dalam 3 *epoch* dengan *learning rate* sebesar $0,00003 (3 \times 10^{-5})$, *batch size* 16 untuk data latih maupun data validasi, serta *weight decay* sebesar 0,02 guna membantu mengurangi *overfitting*. Data validasi digunakan untuk memantau perkembangan model selama tahapan pelatihan berlangsung, lewat nilai *training loss*, *validation loss* serta *validation accuracy* pada setiap *epoch*. Model hasil *fine-tuning* tersebut selanjutnya dimanfaatkan untuk melakukan klasifikasi sentimen tweet ke dalam kelas positif dan negatif.

2.7 Evaluasi Model

Kinerja model dievaluasi dengan menerapkan *confusion matrix*, yaitu metode yang menampilkan hasil klasifikasi dalam bentuk tabel yang menunjukkan perbandingan antara nilai prediksi model terhadap nilai aktual. *Confusion matrix* memiliki empat kategori utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Komponen-komponen menjadi dasar dalam menghitung berbagai metrik evaluasi, seperti *accuracy*, *precision*, *recall*, serta *F1-score*, guna mengukur performa model klasifikasi [33]. Rumus untuk masing-masing metrik evaluasi tersebut disajikan dalam persamaan (1), (2), (3) dan (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengambilan Data

Proses pengambilan data dilakukan menggunakan teknik scraping yang dijalankan melalui Google Colab, dengan memanfaatkan bahasa pemrograman Python, pustaka Tweet-Harvest versi 2.6.1, serta X *Authentication Token*. Langkah pertama adalah mengakses platform X menggunakan *auth_token* sebagai autentikasi, kemudian melakukan pencarian dengan kata kunci 'Israel', 'Palestine', 'Gaza', dan 'Hamas'. Pengambilan data dilakukan pada rentang waktu 10 Mei 2021 hingga 10 Mei 2025 dan dibatasi hanya pada tweet berbahasa Inggris dengan menerapkan parameter '*lang:en*'. Proses *scraping* selanjutnya dilakukan untuk mengekstrak data tweet yang sesuai dengan kata kunci tersebut.

3.2 Text Preprocessing

Tahap *preprocessing* dijalankan guna menyiapkan data sebelum memasuki tahap pemodelan. Pada tahap ini, data mentah dibersihkan serta diubah agar memiliki format yang lebih konsisten dan sesuai untuk diproses oleh model klasifikasi sentimen. Tahap *preprocessing* yang diterapkan dalam penelitian ini meliputi beberapa prosedur, yaitu *Cleaning*, *Case folding*, *Tokenizing*, *Stopword removal*, dan *Stemming*. Setelah dilakukan tahapan *text preprocessing*, diperoleh perubahan pada isi tulisan sehingga data menjadi lebih teratur, konsisten, dan siap digunakan pada tahapan pelabelan sentimen dan pemodelan menggunakan BERT. Perubahan jumlah data setelah proses *preprocessing* disajikan melalui Tabel 1.

Tabel 1. Jumlah Data *Preprocessing*

Tahapan	Jumlah Data
Sebelum <i>Preprocessing</i>	3.172 data
Sesudah <i>Preprocessing</i>	2.362 data

3.3 Pelabelan Data

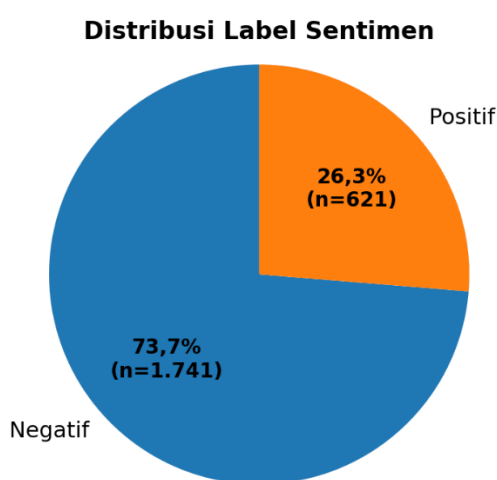
Proses pelabelan data dijalankan secara otomatis dengan menerapkan metode VADER. Penentuan label pada penelitian ini didasarkan pada nilai *compound score* dengan menerapkan ambang batas (nilai *threshold*) 0,05. Tweet yang memiliki nilai *compound score* $\geq 0,05$ dikelompokkan menjadi sentimen positif, sedangkan tweet yang mempunyai nilai *compound score* $\leq -0,05$ dikelompokkan menjadi sentimen negatif. Sementara itu, tweet dengan nilai *compound score* di antara $> -0,05$ dan $< 0,05$ dikategorikan sebagai sentimen netral dan tidak digunakan dalam proses klasifikasi biner. Dengan demikian, skor yang berada di sekitar titik netral tidak dipaksakan ke dalam kelas positif maupun negatif. Penghilangan kelas netral dilakukan karena penelitian ini berfokus pada klasifikasi biner untuk membedakan sentimen positif dan negatif. Setelah proses pelabelan, data dengan label netral dihapus dari dataset menggunakan fungsi *dropna()*. Setiap tweet kemudian diberi label sesuai dengan nilai *compound score* yang dihasilkan. Hasil pelabelan data dapat dilihat melalui Tabel 2.

Tabel 2. Hasil Pelabelan Sentimen menggunakan VADER

No.	Tweet	Compound Score	Label Sentimen
0.	hey idget hama alway want peac terrorist israel wont let mayb need get head actual see what go palestin israel gaza freepalestin	-0,6597	<i>negative</i>
1.	trump soninlaw want build condo gaza that israel continu attack palestin even though hama opinion longer real threat right dont even think theyr still palestin	-0,7351	<i>negative</i>
2.	news gaza westbank palestin part ii middleeasi aka westasia qud news network palestinian girl razan muhaisen pass away isra occup deni permiss travel outsid gaza lifesav treatment suffer critic	-0,6808	<i>negative</i>
3.	actual dutch fake import suicid bomber fuck hama fuck gaza fuck palestin done antisemit thank remind us talk goe nowher cheer israel	-0,8316	<i>negative</i>

4.	end hama war posit benefit israel palestine gaza know propos condit im wait hear russia readi speak ukraine want deal end conflict make ukraine happy put	-0,4404	<i>negative</i>
5.	least attack target school report gaza strip last month averag nearli two incid per day accord data partner israel palestine hama	0,3724	<i>positive</i>

Tabel 2 menunjukkan bahwa setiap tweet telah memperoleh label sentimen berdasarkan nilai *compound score* yang dihasilkan oleh VADER. Tweet dengan nilai *compound score* negatif diberi label *negative*, sedangkan tweet dengan nilai *compound score* positif diberi label *positive*. Meskipun demikian, terdapat contoh tweet yang memperoleh label *positive* dengan *compound score* sebesar 0,3724, meskipun secara konteks membahas serangan terhadap sekolah di Gaza. Hal ini menunjukkan bahwa VADER memiliki keterbatasan dalam memahami konteks secara menyeluruh. Setelah seluruh data berhasil diberi label, diperoleh sebaran jumlah data untuk setiap kelas sentimen yang ditunjukkan melalui Gambar 2.



Gambar 2. Distribusi Label Sentimen

Gambar 2 memperlihatkan distribusi hasil pelabelan sentimen menggunakan metode VADER. Dari total 2.362 tweet, sebanyak 1.741 tweet (73,7%) dikategorikan sebagai sentimen *negative*, sedangkan 621 tweet (26,3%) termasuk sentimen *positive*. Temuan tersebut memperlihatkan bahwa mayoritas tweet dalam dataset menunjukkan kecenderungan sentimen negatif terhadap isu konflik Israel–Palestina. Selain itu, distribusi data pada kedua kelas belum seimbang (*class imbalance*), karena jumlah tweet berlabel *negative* lebih banyak dibandingkan *positive*. Kondisi tersebut perlu menjadi perhatian dalam proses pemodelan karena dapat mempengaruhi kemampuan model dalam mengenali kelas yang memiliki jumlah data yang lebih sedikit. Oleh sebab itu, kinerja model selanjutnya dinilai menggunakan sejumlah metrik meliputi *accuracy*, *precision*, *recall* dan *F1-score* agar hasil klasifikasi dapat dinilai dengan lebih menyeluruh.

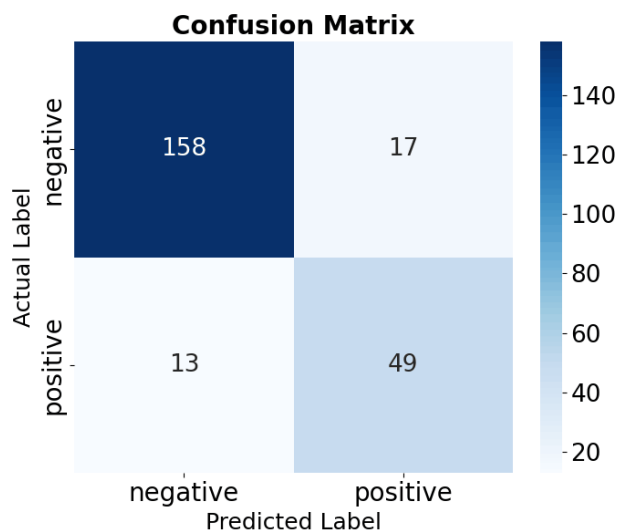
3.4 Hasil Model BERT

Pelatihan model dijalankan dengan menerapkan BERT Base Uncased dengan data latih yang telah melalui tahapan *preprocessing*, pelabelan sentimen, *label encoding*, dan pembagian data. Kinerja model kemudian dievaluasi menggunakan data uji dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui kinerja model dalam mengklasifikasikan sentimen. Hasil evaluasi model disajikan melalui Tabel 3.

Tabel 3. Hasil Evaluasi Model BERT

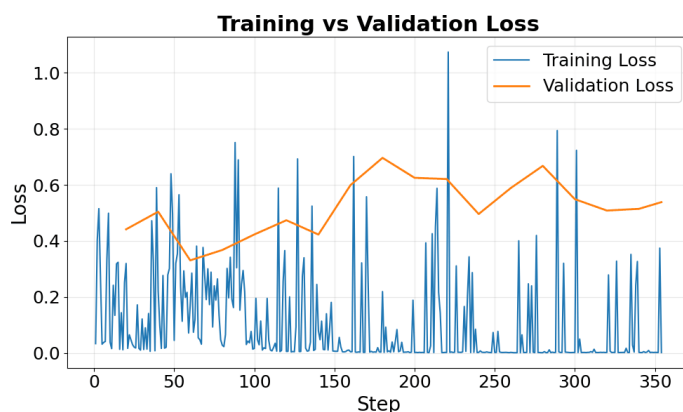
Kelas/Metrik	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Negative</i>	0,92	0,90	0,91	175
<i>Positive</i>	0,74	0,79	0,77	62
<i>accuracy</i>	-	-	0,87	237
<i>macro avg</i>	0,83	0,85	0,84	237
<i>weighted avg</i>	0,88	0,87	0,87	237

Tabel 3 menyajikan hasil evaluasi model BERT yang menghasilkan *accuracy* sebesar 87,34% pada data uji. Kelas *negative* menghasilkan *precision* 92%, *recall* 90%, dan *F1-score* 91% di mana total data uji berjumlah 175, sedangkan kelas *positive* menghasilkan *precision* 74%, *recall* 79%, dan *F1-score* 77% dengan total data uji berjumlah 62. Nilai *macro average* sebesar 83% (*precision*), 85% (*recall*), dan 84% (*F1-score*) menggambarkan kinerja rata-rata model tanpa memperhitungkan bobot banyaknya data di tiap kelas, sementara nilai *weighted average* sebesar 88% (*precision*), 87% (*recall*), dan 87% (*F1-score*) mempertimbangkan proporsi data pada setiap kelas. Gambaran lebih rinci mengenai hasil prediksi model terhadap setiap kelas disajikan dalam *confusion matrix* di Gambar 3.



Gambar 3. Hasil *Confusion Matrix*

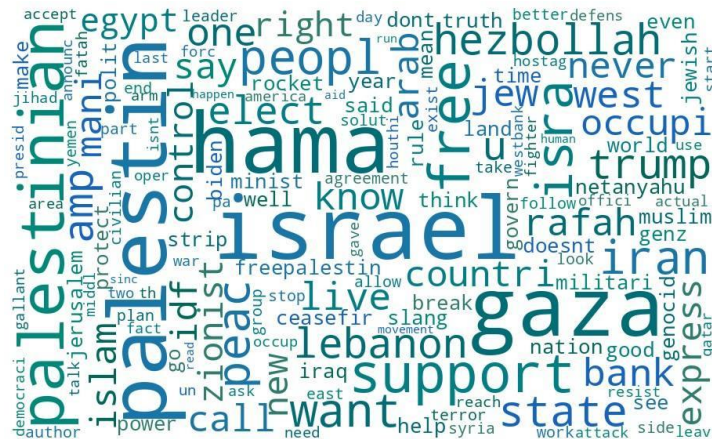
Gambar 3 menunjukkan bahwa dari 175 data aktual berlabel *negative*, model berhasil memprediksi 158 data dengan benar, sedangkan 17 data salah diklasifikasikan sebagai *positive*. Pada kelas *positive*, dari 62 data aktual, model memprediksi 49 data dengan benar, sementara 13 data salah diklasifikasikan sebagai *negative*. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam mengenali kelas *negative* dibandingkan kelas *positive*. Hal ini terlihat dari nilai *recall* kelas *negative* sebesar 90% (158/175), sedangkan *recall* kelas *positive* sebesar 79% (49/62). Perbedaan tersebut menunjukkan bahwa performa model antarkelas masih belum sepenuhnya seimbang, yang dapat berkaitan dengan distribusi data yang tidak setara. Proses pembelajaran model selanjutnya diamati melalui perubahan nilai *training loss* dan *validation loss* pada setiap *step*. Perubahan kedua nilai tersebut dapat dilihat melalui grafik pada Gambar 4.



Gambar 4. *Training vs Validation Loss*

Gambar 4 menunjukkan nilai *training loss* dan *validation loss* berubah selama proses pelatihan model. *Validation loss* berada di sekitar 0,44 pada *step* ke-20, kemudian naik menjadi sekitar 0,50 pada *step* ke-40, sebelum menurun hingga mencapai titik terendah sekitar 0,33 pada *step* ke-60. Setelah *step* ke-60, *validation loss* cenderung meningkat secara bertahap hingga sekitar 0,47 pada *step* ke-120, sempat turun ke sekitar 0,42 pada *step* ke-140, lalu naik tajam hingga mencapai puncaknya sekitar 0,70 pada *step* ke-180. Nilainya kemudian menurun hingga sekitar

0,50 pada *step* ke-240, sebelum kembali meningkat hingga sekitar 0,67 pada *step* ke-280. Setelah *step* ke-280, *validation loss* kembali menurun hingga sekitar 0,51 pada *step* ke-320 dan relatif stabil hingga akhir pelatihan pada *step* ke-355. Sementara itu, *training loss* berfluktuasi cukup tinggi dengan beberapa lonjakan tajam, dengan nilai tertinggi sekitar 1,1 pada *step* ke-220, tetapi cenderung berada di nilai rendah pada sebagian besar *step*, meski lonjakan serupa masih muncul hingga mendekati akhir pelatihan. Kondisi tersebut, disertai *validation loss* yang juga berfluktuasi tanpa tren penurunan yang konsisten, mengindikasikan adanya *overfitting*, yaitu model semakin menyesuaikan diri terhadap data latih, tetapi belum menunjukkan generalisasi yang stabil terhadap data validasi. Meskipun demikian, model dapat melakukan klasifikasi sentimen terhadap isu geopolitik Israel-Palestina secara akurat, sebagaimana terlihat pada visualisasi *Word Cloud* di Gambar 5.



Gambar 5. Word Cloud Sentimen Positif

Gambar 5 menunjukkan visualisasi kata pada sentimen positif seperti kata *Israel*, *Palestine*, *Gaza*, *Hamas*, *support*, *peace*, *free*, *people*, dan *Iran*. Kata-kata yang muncul mengindikasikan frekuensi kata yang paling sering digunakan pada sentimen positif, seperti nama pihak yang terlibat dalam konflik, serta ungkapan umum seperti “support”, “free” dan “peace”.



Gambar 6. Word Cloud Sentimen Negatif

Gambar 6 menunjukkan visualisasi kata pada sentimen negatif seperti kata *Israel*, *Palestine*, *Gaza*, *Hamas*, *attack*, *war*, *terrorist*, *Iran*, *Hezbollah* dan *kill*. Kata-kata yang muncul mengindikasikan frekuensi kata yang paling sering digunakan pada sentimen negatif, seperti nama pihak yang terlibat dalam konflik, serta ungkapan umum seperti “attack”, “war”, “terrorist”, dan “kill”.

4. KESIMPULAN

Penelitian ini membangun model klasifikasi sentimen berbasis BERT Base Uncased untuk mengkaji isu geopolitik Israel-Palestina, dengan pelabelan otomatis menggunakan VADER pada data tweet hasil *scraping* menggunakan Tweet-Harvest, diperoleh 2.362 tweet setelah tahap *preprocessing*, dengan distribusi sentimen 73,7% *negative* dan 26,3% *positive*. Setelah melalui proses *fine-tuning* selama 3 *epoch* dengan *learning rate* 3×10^{-5} , *batch size* 16, dan *weight decay* 0,02, model berhasil mencapai *accuracy* sebesar 87,34% pada data uji, dengan *precision*,

recall, dan *F1-score* berturut-turut sebesar 92%, 90%, dan 91% untuk kelas *negative* serta 74%, 79%, dan 77% untuk kelas *positive*, di mana skor *macro average* dan *weighted average F1-score* berturut-turut sebesar 84% dan 87%. Hasil *confusion matrix* menunjukkan model berhasil memprediksi 158 dari 175 data *negative* dan 49 dari 62 data *positive* dengan benar, disertai pemantauan terhadap *training loss* dan *validation loss*. Hasil penelitian ini menunjukkan bahwa kombinasi pelabelan otomatis VADER dengan model BERT Base Uncased mampu menghasilkan performa klasifikasi sentimen yang cukup baik pada isu geopolitik Israel–Palestina. Namun, masih ada beberapa keterbatasan, terutama pada kelas dengan jumlah data yang lebih sedikit. Berdasarkan temuan tersebut, penelitian selanjutnya dapat menggunakan anotasi manual sebagai pembandingan pelabelan otomatis, mempertimbangkan penerapan teknik penyeimbangan data seperti *class weighting*, *oversampling*, pada data training, atau augmentasi teks, serta menambahkan kategori sentimen netral.

REFERENSI

- [1] A. Delia, F. A. Khatimah, H. S. Andien, R. Arissaputra, and M. Parhan, “Peran Muslim Dalam Konflik Geopolitik: Analisis Perang Israel-Palestina,” *Pendas: Jurnal Ilmiah Pendidikan Dasar*, vol. 9, no. 2, pp. 152–166, Jun. 2024, doi: 10.23969/jp.v9i2.15182.
- [2] R. Hasna Rahmawan and I. Wayan Midhio, “Efek Konflik Israel-Hamas di Indonesia: Opini Publik dan Upaya Diplomasi,” *Almikraj*, vol. 5, no. 01, pp. 1752–1762, Dec. 2024, doi: 10.37680/almikraj.v5i01.643.
- [3] W. Widya Ariestya *et al.*, “Reaksi Media Sosial Terhadap Konflik Israel-Palestina: Pendekatan Analisis Sentimen Dengan VADER Dan BERT,” *Jurnal Ilmu Komputer dan Informatika*, vol. 5, no. 2, pp. 150–157, Dec. 2024, doi: 10.24127/jiki.v5i2.7420.
- [4] M. H. Al-Areef and K. Saputra, “Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM,” *Jurnal Sains Manajemen Informatika dan Komputer*, vol. 22, no. 2, pp. 270–279, Aug. 2023, doi: 10.53513/jis.v22i2.8680.
- [5] B. J. Manurung, B. Parga Zen, Y. Setiya, R. Nur, R. C. Felle, and E. A. Firdaus, “Analisis Sentimen Pengguna Twitter terhadap Konflik Rusia-Ukraina Menggunakan Naïve Bayes dan Lexicon Based Features,” *Jurnal Informatika*, vol. 4, no. 1, pp. 8–16, Mar. 2025, doi: 10.57094/ji.v4i1.2574.
- [6] M. G. Al-Kadzim, R. Rasim, and H. Herbert, “Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT,” *Digital Transformation Technology*, vol. 4, no. 2, pp. 1167–1174, Jan. 2025, doi: 10.47709/digitech.v4i2.5312.
- [7] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, “Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review,” *Natural Language Processing Journal*, vol. 6, pp. 1–30, Mar. 2024, doi: 10.1016/j.nlp.2024.100059.
- [8] J. A. Widians, R. Wardoyo, and S. Hartati, “Feature selection based on chi-square and ant colony optimization for multi-label classification,” *International Journal of Electrical and Computer Engineering*, vol. 14, no. 3, pp. 3303–3312, Jun. 2024, doi: 10.11591/ijece.v14i3.pp3303-3312.
- [9] R. Azhar and M. F. Wijayanto, “Analisis Sentimen di Twitter: Mengungkap Persepsi dan Emosi Publik Seputar Konflik Palestina-Israel,” *STAINS (Seminar Nasional Teknologi & Sains)*, vol. 3, no. 1, pp. 118–124, Jan. 2024, doi: 10.29407/stains.v3i1.4132.
- [10] A. Amelia and R. Yusuf, “Analisis Sentimen Masyarakat Indonesia Pada Platform X Terhadap Isu Fufufafa Menggunakan Bidirectional Encoder Representations From Transformers,” *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 7, no. 1, pp. 72–80, Feb. 2025, doi: 10.51401/jinteks.v7i1.5160.
- [11] N. Lubis, Mhd. Z. Siambaton, and R. Aulia, “Implementasi Algoritma Deep Learning pada Aplikasi Speech to Text Online dengan Metode Recurrent Neural Network (RNN),” *Jurnal Teknik Informatika*, vol. 3, no. 3, pp. 113–126, Sep. 2024, doi: 10.56211/sudo.v3i3.583.
- [12] D. Wang *et al.*, “Unveiling the spatiotemporal propagation patterns of sentiments regarding the Israeli–Palestinian military conflict,” *Humanities and Social Sciences Communications*, vol. 12, no. 1, pp. 1–17, Dec. 2025, doi: 10.1057/s41599-025-05134-x.
- [13] M. Hasraddin Hasnan, R. Yuslana Bakti, M. Faisal, T. Khawa Abd Rahman, M. S. Syafaat Kuba, and T. Wahyuni, “Deteksi Berita Hoaks Berbahasa Indonesia yang Dapat Dijelaskan Menggunakan IndoBERT dan SHAP,” *Buletin Sistem Informasi dan Teknologi Islam*, vol. 7, no. 2, pp. 160–170, May 2026, doi: 10.33096/busiti.v7i2.3466.
- [14] L. Geni, E. Yulianti, and D. I. Sensuse, “Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using BERT Language Models,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 746–757, Aug. 2023, doi: 10.26555/jiteki.v9i3.26490.
- [15] A. Sehatman Saragih *et al.*, “Analisis Sentimen Publik Indonesia Terhadap Konflik Israel-Iran Di Media Sosial Menggunakan IndoBERT Dan Explainable AI (LIME),” *Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 20, no. 1, pp. 78–87, Jan. 2026, doi: 10.47111/jti.v20i1.24465.
- [16] R. Dwiyanaputra, S. Ika Murpratiwi, and A. Aranta, “Analisis Sentimen Pengguna Platform Media Sosial X Pada Topik Pemilihan Presiden 2024 Menggunakan Perbandingan Model Monolingual Dan Multilingual BERT,” *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 1, pp. 626–634, Feb. 2025, doi: 10.36040/jati.v9i1.12430.

- [17] K. Kusnawi and K. Anam, "Comparison of Sentiment Labeling Using Textblob, Vader, and Flair in Public Opinion Analysis Post-2024 Presidential Inauguration with IndoBERT," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, pp. 803–818, Apr. 2025, doi: 10.52436/1.jutif.2025.6.2.4015.
- [18] J. L. Halim, "BERT-base, DistilBERT, and BERTweet: A Comparative Study On App Review," *JIKO (Jurnal Informatika dan Komputer)*, vol. 10, no. 1, p. 45, Feb. 2026, doi: 10.26798/jiko.v10i1.2217.
- [19] A. I. Fadri, N. Futri, A. Jelita, D. Dimas Bagaskara, and R. Zahra, "Sentiment Analysis of Public Opinion on the Gaza Conflict Using Machine Learning," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 3, no. 2, pp. 138–149, Jan. 2026, doi: 10.57152/predatecs.v3i2.2088.
- [20] R. Fauzi, E. Iman, and H. Ujianto, "Sentiment Analysis of Indonesian Responses to the Conflict in Palestine Using KNN and SVM Methods," *Journal of Applied Informatics and Computing (JAIC)*, vol. 8, no. 2, pp. 542–549, Dec. 2024, doi: 10.30871/jaic.v8i2.8725.
- [21] A. Areshy and H. Mathkour, "Transfer Learning for Sentiment Classification Using Bidirectional Encoder Representations from Transformers (BERT) Model," *Sensors*, vol. 23, no. 11, pp. 1–18, Jun. 2023, doi: 10.3390/s23115232.
- [22] H. Tang and R. Zhang, "The Research on Aspect-Level Sentiment Analysis for Online Reviews Based on BERT-BiLSTM Model," *Frontiers in Artificial Intelligence and Applications*, vol. 373, pp. 928–936, Nov. 2023, doi: 10.3233/FAIA230903.
- [23] S. D. Wahyuni and H. Februariyanti, "Analisis Sentimen dan Rating Pengguna Aplikasi Traveloka Menggunakan Vader," *semantik Teknik Informasi*, vol. 12, no. 1, pp. 79–88, Apr. 2026, doi: 10.55679/semantik.v12i1.285.
- [24] K. Nugroho and F. N. Hasan, "Analisis Sentimen Masyarakat Mengenai RUU Perampasan Aset Di Twitter Menggunakan Metode Naïve Bayes," *SMATIKA JURNAL*, vol. 13, no. 02, pp. 273–283, Dec. 2023, doi: 10.32664/smatika.v13i02.899.
- [25] M. M. Dakwah, A. A. Firdaus, F. Furizal, and R. Faresta, "Sentiment Analysis on Marketplace in Indonesia using Support Vector Machine and Naïve Bayes Method," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 1, p. 39, Feb. 2024, doi: 10.26555/jiteki.v10i1.28070.
- [26] D. A. Kristiyanti and S. Hardani, "Sentiment Analysis of Public Acceptance of Covid-19 Vaccines Types in Indonesia using Naïve Bayes, Support Vector Machine, and Long Short-Term Memory (LSTM)," *Jurnal RESTI*, vol. 7, no. 3, pp. 722–732, Jun. 2023, doi: 10.29207/resti.v7i3.4737.
- [27] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, "Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naive Bayes (Studi Kasus: Kuliah Daring di Masa Pandemi)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 9, no. 4, pp. 717–724, Aug. 2022, doi: 10.25126/jtiik.202294707.
- [28] D. Ariatmanto and A. M. Rifai, "The Impact of Feature Extraction in Random Forest Classifier for Fake News Detection," *Jurnal RESTI*, vol. 8, no. 6, pp. 730–736, Dec. 2024, doi: 10.29207/resti.v8i6.6017.
- [29] S. Ernawati and R. Wati, "Evaluasi Performa Kernel SVM dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter dan VADER Lexicon," *Jurnal Buana Informatika*, vol. 15, no. 1, pp. 40–49, Apr. 2024, doi: 10.24002/jbi.v15i1.7925.
- [30] M. Jundi, R. Rahmadden, P. Utami, S. Perdana Arifin, and L. Sinaga, "Klasifikasi Pengaduan Kekerasan Berbasis Bidirectional Encoder Representations from Transformers (BERT) pada Kementerian Pemberdayaan Perempuan dan Perlindungan Anak Kementerian PPPA)," *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, vol. 5, no. 2, pp. 258–264, Aug. 2025, doi: 10.55382/jurnalpustakaai.v5i2.1109.
- [31] M. Ilham and B. Priambodo, "Analisis Sentimen Publik Terhadap Program Makan Siang Gratis Menggunakan BERT Neural Network Pada Platform X," *JEMSI: Jurnal Ekonomi Manajemen Sistem Informasi*, vol. 6, no. 2, pp. 1039–1047, Nov. 2024, doi: 10.38035/jemsi.v6i2.
- [32] F. M. Apriansyah, T. I. Ramadhan, C. R. Hidayat, and A. K. Wijaya, "Perbandingan IndoBERT dan IndoRoBERTa Untuk Analisis Sentimen Pada Film Dokumenter Dirty Vote," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 3, pp. 593–605, Jul. 2025, doi: 10.30591/jpit.v10i3.8607.
- [33] F. Rafiandi Andhika, W. Witanti, and P. N. Sabrina, "Analisis Sentimen Menggunakan Metode IndoBERT pada Ulasan Aplikasi Zoom Menggunakan Fitur Ekstraksi GloVe," *Metik Jurnal*, vol. 9, no. 2, pp. 439–448, Aug. 2025, doi: 10.47002/g38fxb78.