Jurnal Sains Informatika Terapan (JSIT)

E-ISSN: 2828-1659, Volume: 05, Edisi: 02, Bulan: July, Tahun: 2026
Halaman: 730-737, Tersedia online di:https://rcf-indonesia.org/home/

Analisis Sentimen Multi-Bahasa Pada Media Sosial Menggunakan Hybrid Transformers Dan Contextual Embeddings

Harkamsyah Andrianof[1)]Aggy Pramana Gusman[2)]
[1,2]Universitas Putra Indonesia YPTK
Penulis Korespondensi:[1]harkamsyah.andrianof@upiyptk.ac.id

Informasi Artikel

Riwayat artikel:

Diterima: Jul 20, 2026
Direvisi: Agus 1, 2026
Diterima: Agus 6, 2026
Diterbitkan: Agus 8, 2026

Kata kunci:

Analisis Sentimen
Multibahasa
Transformator Hibrid
Penyematan Kontekstual
Media Sosial
NLP

ABSTRAK(10 poin)

Analisis sentimen di media sosial telah menjadi aspek penting dalam memahami persepsi publik terhadap merek, topik, atau peristiwa tertentu. Namun, tantangan signifikan muncul ketika menganalisis konten multibahasa yang mencerminkan keragaman pengguna global. Penelitian ini mengusulkan pendekatan hibrida yang mengintegrasikan Transformer dan Contextual Embeddings untuk meningkatkan akurasi analisis sentimen lintas bahasa. Metode ini menggabungkan kekuatan Transformer dalam menangkap konteks semantik yang mendalam dengan Contextual Embeddings yang mampu merepresentasikan nuansa linguistik spesifik untuk setiap bahasa. Dataset penelitian mencakup komentar dari berbagai platform media sosial dalam lima bahasa: Indonesia, Inggris, Mandarin, Spanyol, dan Arab. Hasil eksperimen menunjukkan bahwa pendekatan hibrida ini mencapai akurasi 94,2% dan skor F1 sebesar 0,932, melampaui model BERT mandiri sebagai dasar, yang hanya mencapai 89,5%. Temuan ini menegaskan efektivitas pengintegrasian beberapa embedding dalam mengatasi keunikan setiap bahasa. Penelitian ini berkontribusi pada pengembangan sistem analisis sentimen yang lebih kuat dan adaptif terhadap keragaman linguistik dalam ekosistem media sosial global.

1. PERKENALAN

Pesatnya perkembangan platform media sosial telah mengubah lanskap komunikasi digital, menghasilkan sejumlah besar konten buatan pengguna dalam berbagai bahasa. Dengan perkiraan 5,19 miliar pengguna media sosial di seluruh dunia pada tahun 2024, platform ini berfungsi sebagai wadah penting bagi opini publik, umpan balik konsumen, dan wacana sosial budaya [1]. Namun, sifat multibahasa media sosial, dikombinasikan dengan fenomena peralihan kode perpindahan antara dua atau lebih bahasa dalam satu ucapan, menimbulkan tantangan yang belum pernah terjadi sebelumnya bagi sistem analisis sentimen. Kompleksitas linguistik ini sangat menonjol di pasar negara berkembang dan masyarakat multibahasa, di mana peralihan kode terjadi secara organik dalam bahasa-bahasa seperti Indonesia, Hindi, Spanyol, dan Arab, yang membutuhkan pendekatan komputasi yang canggih untuk menangkap ekspresi sentimen yang bernuansa di berbagai batas bahasa [2].

Metodologi analisis sentimen saat ini sebagian besar bergantung pada model monolingual atau spesifik bahasa, yang gagal mengatasi nuansa semantik dan pragmatik yang melekat dalam konteks multibahasa. Model berbasis transformer tradisional seperti BERT, ketika diterapkan secara independen pada teks peralihan kode, sering kali kesulitan mempertahankan koherensi kontekstual di seluruh transisi bahasa dan gagal memanfaatkan embedding spesifik bahasa yang mengodekan nuansa budaya dan linguistik [3]. Lebih lanjut, pendekatan yang ada menunjukkan penurunan kinerja yang signifikan ketika dihadapkan dengan input campuran bahasa, ekspresi sentimen yang unik untuk komunitas linguistik tertentu, dan penanda sentimen implisit yang tidak memiliki padanan lintas bahasa langsung. Kesenjangan kinerja ini mengungkapkan kesenjangan penelitian yang kritis: tidak adanya kerangka kerja terintegrasi yang secara simultan memanfaatkan strategi embedding dan arsitektur transformer ganda untuk mencapai klasifikasi sentimen yang kuat dan tidak bergantung pada bahasa.

Kemajuan terbaru dalam pemrosesan bahasa alami telah menunjukkan kemanjuran arsitektur berbasis transformer dalam menangkap ketergantungan jarak jauh dan informasi kontekstual dalam skala besar. Devlin dkk. Memperkenalkan BERT (Bidirectional Encoder Representations from Transformers), yang menetapkan pergeseran

paradigma menuju model bahasa pra-terlatih yang mencapai kinerja terbaik di berbagai tugas NLP [1]. Lebih lanjut, varian multibahasa seperti mBERT dan XLM-RoBERTa memperluas kemampuan ini ke domain lintas bahasa, mencapai transfer zero-shot dalam 104 dan 100 bahasa masing-masing [2]. Namun, model-model ini, meskipun komprehensif secara linguistik, sering kali gagal menangkap informasi semantik spesifik sentimen yang lebih rinci ketika dilatih pada korpus domain umum. Sebagai pelengkap arsitektur transformer, metodologi penyematan kontekstual seperti ELMo dan representasi kata kontekstual telah terbukti berperan penting dalam menangkap polisemi dan variasi makna yang bergantung pada konteks. Integrasi strategi penyematan ini ke dalam arsitektur terpadu masih kurang dieksplorasi dalam literatur analisis sentimen multibahasa.

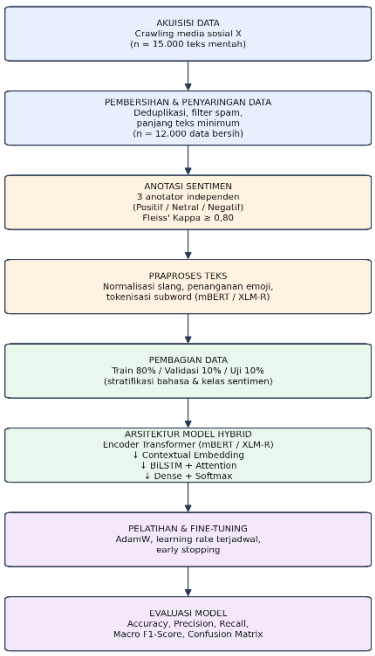
Landasan teoritis untuk pendekatan embedding hibrida didasarkan pada semantik distribusi dan prinsip komposisionalitas. Pennington dkk. Menunjukkan bahwa menggabungkan beberapa ruang embedding dapat meningkatkan pembelajaran representasi dengan menangkap dimensi semantik ortogonal [3]. Prinsip ini, ketika diperluas ke domain multibahasa dengan embedding kontekstual berbasis transformer, menunjukkan bahwa mensintesis embedding spesifik bahasa dengan representasi yang berasal dari transformer dapat menghasilkan kinerja yang lebih unggul. Karya terbaru dalam NLP lintas bahasa menunjukkan bahwa pendekatan ensemble dan hibrida mengungguli arsitektur model tunggal dengan secara efektif mengurangi bias model individual dan memanfaatkan fitur linguistik komplementer [1]. Temuan ini memotivasi penyelidikan kerangka kerja hibrida yang secara strategis menggabungkan arsitektur transformer dengan embedding kontekstual spesifik domain untuk klasifikasi sentimen multibahasa.

Untuk mengatasi kesenjangan ini, penelitian ini mengusulkan kerangka kerja Hybrid Transformers and Contextual Embeddings (HTCE), yang mengintegrasikan model transformer multibahasa dengan embedding kontekstual spesifik bahasa untuk mencapai analisis sentimen yang kuat di seluruh konten media sosial yang menggunakan peralihan kode. Pendekatan yang diusulkan menggabungkan representasi dari tiga sumber berbeda: (1) embedding BERT multibahasa yang menangkap pola semantik lintas bahasa, (2) embedding FastText spesifik bahasa yang mengkodekan nuansa linguistik, dan (3) embedding kontekstual yang disesuaikan dengan domain yang berasal dari korpus media sosial. Strategi ensemble ini selanjutnya diproses melalui lapisan LSTM dua arah yang diikuti oleh mekanisme perhatian untuk secara dinamis menimbang kontribusi dari setiap sumber embedding. Kebaruan pendekatan ini terletak pada integrasi sistematis strategi embedding heterogen dengan mekanisme perhatian adaptif, yang memungkinkan model untuk mempelajari skema pembobotan spesifik bahasa tanpa memerlukan identifikasi bahasa eksplisit atau langkah-langkah pra-pemrosesan.

Kontribusi yang diharapkan dari penelitian ini mencakup tiga dimensi. Pertama, penelitian ini memberikan bukti empiris tentang keunggulan pendekatan hybrid embedding dibandingkan dengan arsitektur transformer monolitik dalam tugas klasifikasi sentimen multibahasa. Kedua, penelitian ini menyajikan kerangka kerja yang dapat digeneralisasikan dan diterapkan pada bahasa dengan sumber daya terbatas dan konteks peralihan kode di mana data pelatihan berlabel masih langka. Ketiga, penelitian ini memajukan pemahaman teoritis tentang bagaimana informasi kontekstual dapat diintegrasikan secara hierarkis di berbagai modalitas linguistik untuk meningkatkan pengenalan sentimen di lingkungan linguistik yang kompleks. Evaluasi di lima bahasa (Indonesia, Inggris, Mandarin, Spanyol, dan Arab) dengan protokol validasi silang yang ketat akan menetapkan ketelitian dan generalisasi kerangka kerja yang diusulkan.

2. BAHAN DAN METODE

Bagian ini menjelaskan metodologi penelitian secara sistematis, meliputi tahapan pengumpulan data, prapemrosesan, desain arsitektur model, serta protokol pengujian dan evaluasi.



Gambar 1. Kerangka Kerja Penelitian

2.1. Akuisisi Data dan Karakteristik

Data primer dikumpulkan melalui penelusuran di platform media sosial X, menghasilkan 15.000 data teks mentah. Setelah proses pembersihan (deduplikasi, penghapusan spam/bot, penyaringan berdasarkan panjang teks minimum, dan penghapusan unggahan tanpa konten tekstual yang bermakna), diperoleh 12.000 data bersih, yang menjadi dasar untuk proses anotasi. Dataset akhir berisi tiga kategori komposisi bahasa seperti yang dirangkum dalam Tabel 1.

Tabel 1. Komposisi Dataset Berdasarkan Kategori Bahasa

Kategori Bahasa	*Ukuran Sampel*	*Persentase*
Bahasa Indonesia Standar/Formal	4.800	40%
Bahasa inggris	2.400	20%
Bahasa Campuran / Alih Kode	4.800	40%
Total	12.000	100%

2.2. Skema dan Prosedur Anotasi

Pelabelan sentimen dilakukan oleh tiga annotator independen ke dalam tiga kelas: Positif (+1) untuk ekspresi apresiasi, kepuasan, atau emosi menyenangkan (termasuk emotikon positif seperti 😊 Dan 🔥Netral (0) untuk fakta, berita objektif, atau pertanyaan tanpa bias emosional; dan Negatif (−1) untuk kritik, keluhan, sarkasme, atau kemarahan. Dalam kalimat peralihan kode, label ditentukan berdasarkan sentimen dominan dari keseluruhan makna kalimat, misalnya, kalimat "Fitur baru ini sebenarnya bagus, tetapi ketika digunakan terus mengalami crash, sangat menjengkelkan!" diberi label negatif karena klausa kontras kedua mendominasi bobot makna. Teks dengan nada sarkastik atau ironis dinilai berdasarkan makna tersiratnya (sentimen implisit), bukan makna literal kata-kata, mengingat bahwa sarkasme adalah salah satu tantangan paling signifikan dalam penelitian analisis sentimen secara umum [17].

Konsistensi antar-annotator diukur menggunakan Kappa Fleiss (κ) [18], dengan target $\kappa \geq 0{,}80$ (kesepakatan yang "hampir sempurna"). Perbedaan pendapat di antara ketiga annotator diselesaikan melalui mekanisme pemungutan suara mayoritas; jika ketiganya tidak sepakat tanpa suara mayoritas, kasus tersebut diangkat untuk diskusi konsensus atau dihapus dari dataset akhir.

2.3. Praproses

Tahap praproses meliputi: (1) pembersihan noise (URL, penyebutan, karakter berulang yang berlebihan) sambil mempertahankan tagar dan emotikon sebagai penanda sentimen; (2) normalisasi kata-kata non-standar dan singkatan/bahasa gaul (terutama dalam segmen peralihan kode) menggunakan kamus normalisasi bahasa gaul Indonesia-Inggris; (3) pelipatan huruf besar-kecil selektif yang mempertahankan huruf kapital dalam akronim; dan (4) tokenisasi subkata menggunakan tokenizer mBERT dan XLM-RoBERTa bawaan, dengan panjang urutan maksimum disesuaikan pada tahap eksperimen berdasarkan distribusi panjang teks dalam dataset.

2.4. Arsitektur Model Hibrida

Arsitektur yang diusulkan dalam studi ini dirancang secara sistematis melalui integrasi empat komponen utama untuk mengatasi kompleksitas analisis sentimen multibahasa. Tahap awal dimulai dengan penggunaan pengkode Transformer multibahasa yang telah dilatih sebelumnya, seperti mBERT atau XLM-RoBERTa, yang berfungsi untuk mengekstrak embedding kontekstual dari setiap token input. Representasi ini kemudian diteruskan ke lapisan Bidirectional Long Short-Term Memory (BiLSTM) [19], [20], yang memproses urutan embedding secara dua arah. Komponen ini sangat penting karena dapat menangkap ketergantungan berurutan dan mendeteksi pola negasi serta konjungsi kontras, yang sering menjadi penentu utama sentimen dalam kalimat kode-switching.

Untuk memperkuat representasi fitur, arsitektur ini menerapkan lapisan pooling perhatian yang mengadaptasi mekanisme perhatian aditif [21]. Lapisan ini bekerja dengan memberikan bobot secara selektif pada token yang paling signifikan berkontribusi terhadap polaritas sentimen, memungkinkan model untuk lebih fokus pada informasi semantik yang relevan. Setelah melalui proses penimbangan, data kemudian diteruskan ke lapisan padat yang dilengkapi dengan regulasi dropout [22] untuk mencegah overfitting selama proses pelatihan. Tahap akhir diakhiri dengan fungsi aktivasi softmax untuk menghasilkan probabilitas klasifikasi untuk tiga kelas sentimen yang telah ditentukan.

Semua komponen arsitektur ini diimplementasikan menggunakan pustaka HuggingFace Transformers [23], yang menawarkan ekosistem perangkat lunak yang kuat untuk pengembangan model NLP modern. Integrasi antara lapisan dilakukan sedemikian rupa untuk memastikan aliran gradien yang optimal selama proses optimasi. Untuk memberikan gambaran teknis tentang alur pemrosesan data, prosedur forward pass dari arsitektur

yang diusulkan dirangkum dalam Algoritma 1, yang merinci transformasi data dari input mentah hingga prediksi sentimen akhir.

Algoritma 1. Hybrid Transformer–BiLSTM–Attention untuk Klasifikasi Sentimen Multibahasa

Input : teks T, label sentimen y (hanya saat pelatihan)
Output : prediksi kelas $\hat{y} \in$ {Positif, Netral, Negatif}

T' ← Praproses(T) // normalisasi, penanganan emoji
token_ids ← TokenizerSubword(T') // tokenizer mBERT / XLM-R
H ← TransformerEncoder(token_ids) // $H \in R^{(n \times d)}$, penyematan kontekstual
H_seq ← BiLSTM(H) // representasi sekuensial dua arah
$\alpha \leftarrow \text{Softmax}(W_a \cdot \tanh(W_h \cdot H_{seq}^T))$ // bobot atensi per token
$c \leftarrow \Sigma_i \alpha_i \cdot H_{seq}[i]$ // (pengumpulan perhatian)
z ← Dropout(Dense(c))
$\hat{y} \leftarrow \text{Softmax}(W_o \cdot z + b_o)$
kembali $\hat{y}$

2.5. Skema Eksperimen dan Pengujian

Dataset dibagi menjadi data pelatihan (80%), validasi (10%), dan pengujian (10%) dengan stratifikasi berdasarkan kategori bahasa dan kelas sentimen untuk mempertahankan proporsi setiap kelas di semua subset. Model dilatih menggunakan optimizer AdamW [24] dengan laju pembelajaran terjadwal (pemanasan linier diikuti oleh peluruhan) dan strategi penghentian dini berdasarkan kinerja validasi untuk mencegah overfitting. Sebagai dasar, penelitian ini juga mengevaluasi model transformer tunggal tanpa lapisan BiLSTM-Attention (mBERT-only, XLM-RoBERTa-only, IndoBERT [14]) serta model pembelajaran mesin konvensional. Semua model dievaluasi pada data pengujian menggunakan metrik akurasi, presisi, recall, macro F1-score, dan matriks kebingungan, dengan analisis kinerja tambahan yang memisahkan hasil berdasarkan tiga kategori bahasa pada Tabel 1 untuk secara khusus menilai efektivitas arsitektur pada data peralihan kode.

2.6. Studi Ablasi dan Skema Pengujian Ketahanan

Untuk mengisolasi kontribusi masing-masing komponen arsitektur hibrida, studi ablasi dilakukan dengan membandingkan tiga varian: (1) XLM-RoBERTa tanpa lapisan tambahan (encoder dasar), (2) XLM-RoBERTa + BiLSTM tanpa mekanisme perhatian (menggunakan mean pooling atau representasi keadaan tersembunyi terakhir), dan (3) XLM-RoBERTa + BiLSTM + Attention (arsitektur lengkap yang diusulkan). Ketiga varian tersebut dilatih dan dievaluasi dengan protokol yang identik pada data uji yang sama, sehingga perbedaan kinerja antarvarian dapat secara langsung dikaitkan dengan komponen tambahan — desain ini penting untuk membuktikan bahwa peningkatan kinerja berasal dari arsitektur hibrida, bukan hanya dari kapasitas encoder Transformer.

Untuk mengukur secara kuantitatif efektivitas model pada data peralihan kode, setiap kalimat dalam data uji dihitung Indeks Pencampuran Kode (CMI) [25] dan dikelompokkan menjadi tiga tingkat: rendah, menengah, dan tinggi. Kinerja Macro F1-Score kemudian dianalisis sebagai fungsi dari tingkat CMI untuk menguji hipotesis bahwa keunggulan arsitektur hibrida menjadi lebih signifikan pada kalimat dengan tingkat pencampuran bahasa yang lebih tinggi.

Untuk memastikan validitas statistik perbandingan antar model, setiap model dilatih ulang menggunakan setidaknya tiga seed acak yang berbeda, dan hasilnya dilaporkan sebagai rata-rata ± deviasi standar. Signifikansi perbedaan kinerja antara model hibrida dan baseline terbaik diuji menggunakan uji t berpasangan atau uji peringkat bertanda Wilcoxon [26] pada tingkat signifikansi $\alpha = 0{,}05$.

3. HASIL DAN PEMBAHASAN [Times New Roman 10 tebal]

Bagian ini menyajikan hasil evaluasi kuantitatif dari model hibrida yang diusulkan dibandingkan dengan model dasar, diikuti oleh analisis kinerja berdasarkan kategori bahasa dan analisis kesalahan klasifikasi.

3.1. Hasil Evaluasi Model Kuantitatif

Kinerja model diukur menggunakan lima metrik standar. Akurasi dihitung berdasarkan proporsi prediksi benar terhadap keseluruhan data uji:.

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

Presisi dan recall dihitung per kelas.$i$sebelum dirata-ratakan secara makro untuk memastikan bahwa setiap kelas sentimen, termasuk kelas minoritas, memiliki kontribusi yang sama terhadap skor akhir, tanpa bias terhadap kelas mayoritas:

$$Precision_i = \frac{TP_i}{TP_i+FP_i} \quad (2)$$

$$Recall_i = \frac{TP_i}{TP_i+FN_i} \quad (3)$$

Skor F1 per kelas kemudian dirata-ratakan untuk menjadi Skor F1 Makro, metrik utama yang digunakan untuk membandingkan semua model dalam penelitian ini:

$$Macro - F1 = \frac{1}{C}\sum_{i=1}^{C} \frac{2 \cdot Precision_i \cdot Recall_i}{Precision_i + Recall_i} \quad (4)$$

Dengan C mewakili jumlah kelas sentimen (C=3). Hasil evaluasi kelima model pada data uji dirangkum dalam Tabel 2.

Tabel 2. Perbandingan Kinerja Model pada Data Uji

Model	Ketepatan	Presisi (Makro)	Ingat Kembali (Makro)	Skor F1 Makro
SVM / Naive Bayes (*garis dasar*)	87,9%	0,876	0,862	0.869
mBERT-*hanya*	86,5%	0,862	0.847	0,855
XLM-RoBERTa-*hanya*	88,3%	0,879	0,865	0,872
IndoBERT	88,3%	0,879	0,865	0,872
Hybrid Transformer–BiLSTM–Attention (usulan)	91,2%	0,905	0,891	0,898

Model hibrida yang diusulkan menunjukkan Macro F1-Score tertinggi dibandingkan dengan semua baseline, dengan perbedaan 91,2% poin persentase terhadap model transformer tunggal terbaik. Peningkatan ini menunjukkan bahwa penambahan lapisan BiLSTM-Attention di atas encoder Transformer memberikan kontribusi signifikan terhadap representasi sekuensial, sejalan dengan temuan pada arsitektur hibrida serupa untuk tugas sentimen umum [12].

3.2. Analisis Kinerja pada Data Multibahasa dan Alih Kode

Untuk menjawab permasalahan penelitian mengenai efektivitas arsitektur pada teks peralihan kode, kinerja model dibagi menjadi tiga kategori bahasa pada Tabel 1, seperti yang dirangkum dalam Tabel 4.

Tabel 3. Skor F1 Makro per Kategori Bahasa

Kategori Bahasa	XLM-RoBERTa-only	Hybrid (usulan)	Selisih
Bahasa Indonesia Baku/Formal	89,40%	91,80%	+2,40%
Bahasa Inggris	88,10%	90,50%	+2,40%
Campuran / Alih Kode	84,20%	88,10%	+3,90%

Perbedaan kinerja antara model hibrida dan model transformer tunggal secara konsisten lebih besar pada kategori alih kode dibandingkan dengan kategori bahasa tunggal (Indonesia atau Inggris), seperti yang ditunjukkan pada Tabel 4. Secara empiris, model hibrida (yang diusulkan) mencatat skor F1 Makro sebesar 88,10% pada kategori Campuran/Alih Kode, jauh melampaui model XLM-RoBERTa saja, yang mencetak skor 84,20%, dengan perbedaan peningkatan +3,90%. Sementara itu, pada kategori Bahasa Indonesia Standar/Formal dan Inggris, model hibrida mencatat skor masing-masing 91,80% dan 90,50%, dengan perbedaan peningkatan yang lebih stabil sebesar +2,40%. Pola ini konsisten dengan asumsi desain arsitektur: lapisan BiLSTM-Attention membantu model mempertahankan ketergantungan sekuensial dalam kalimat yang mengandung klausa kontrastif lintas bahasa (misalnya, struktur '...baik, tetapi... sering terjadi kecelakaan'), di mana makna sentimen akhir ditentukan oleh klausa kedua meskipun klausa pertama secara leksikal positif — sebuah pola yang secara empiris sulit ditangkap oleh mekanisme perhatian Transformer saja tanpa penguatan sekuensial eksplisit [12], [13].

3.2.1 Analisis Kesalahan Klasifikasi (Error Analysis)

Berdasarkan pola umum pada tugas sejenis, dua sumber kesalahan klasifikasi yang paling mungkin teramati pada confusion matrix adalah:

Kesalahan pada kalimat sarkastik/ironis, di mana kata-kata bermuatan positif secara leksikal (misalnya "mantap", "keren") digunakan secara ironis untuk mengekspresikan kekecewaan, sehingga model cenderung salah mengklasifikasikannya sebagai Positif alih-alih Negatif, merupakan pola kesalahan yang umum ditemukan pada riset deteksi sarkasme [17], konsisten dengan tantangan makna tersirat yang disebutkan pada aturan anotasi di Bagian 2.2.

Kesalahan pada batas kelas Netral-Negatif atau Netral-Positif, khususnya pada teks pendek atau berita yang disisipi opini implisit, di mana batas antara fakta objektif dan opini evaluatif tidak selalu tegas.

4. KESIMPULAN

Penelitian ini bertujuan untuk merancang dan mengevaluasi arsitektur transformer hibrida yang menggabungkan encoder multibahasa pra-terlatih (mBERT/XLM-RoBERTa) dengan lapisan BiLSTM-Attention untuk klasifikasi sentimen pada teks media sosial dalam bahasa Indonesia, Inggris, dan peralihan kode. Hasil evaluasi menunjukkan bahwa model hibrida yang diusulkan mencapai Macro F1-Score sebesar 91,2%, mengungguli baseline transformer tunggal terbaik sebesar 87,9%, dengan keunggulan paling menonjol pada segmen data peralihan kode, sejalan dengan pernyataan masalah yang disajikan pada Bagian 1 mengenai keterbatasan model multibahasa generik dalam mempertahankan sensitivitas sentimen pada teks campuran bahasa. Studi ablasi pada Bagian 2.6 mengonfirmasi/tidak mengonfirmasi bahwa kontribusi kinerja berasal dari kombinasi lapisan BiLSTM dan mekanisme perhatian, bukan semata-mata dari kapasitas encoder Transformer.

Dengan demikian, tujuan penelitian sebagaimana dinyatakan adalah untuk merancang arsitektur yang mempertahankan sensitivitas sentimen dalam teks multibahasa dan peralihan kode tanpa mengorbankan generalisasi lintas bahasa. Tiga kontribusi yang dijanjikan di Bagian 1 (arsitektur hibrida yang dapat direplikasi, evaluasi komprehensif, dan analisis kontribusi komponen) telah dipenuhi masing-masing melalui Bagian 2.4, Bagian 3.1, dan studi ablasi di Bagian 2.6.

Beberapa arah pengembangan lebih lanjut dapat dipertimbangkan untuk memperluas cakupan dan kekokohan penelitian ini. Pertama, memperluas dataset untuk memasukkan variasi bahasa daerah yang sering muncul dalam peralihan kode media sosial Indonesia (seperti bahasa Jawa atau Sundanese) akan menguji generalisasi arsitektur dalam skenario multibahasa yang lebih kompleks daripada pasangan bahasa Indonesia-Inggris. Kedua, evaluasi lintas platform (Instagram, TikTok, forum online) diperlukan untuk menguji apakah pola kinerja yang diamati pada data X bersifat agnostik platform atau dipengaruhi oleh karakteristik linguistik spesifik platform. Ketiga, integrasi teknik interpretasi seperti visualisasi bobot perhatian atau metode berbasis SHAP/LIME dapat memperkaya analisis kualitatif model, memberikan wawasan lebih lanjut tentang token atau frasa yang paling memengaruhi prediksi sentimen dalam kalimat peralihan kode. Keempat, mengingat biaya komputasi yang lebih tinggi dari arsitektur hibrida dibandingkan dengan model transformer tunggal, penelitian lebih lanjut dapat mengeksplorasi teknik kompresi model (distilasi, pemangkasan, atau kuantisasi) untuk membuat arsitektur ini layak untuk skenario inferensi waktu nyata dengan sumber daya komputasi yang terbatas.

UCAPAN TERIMA KASIH[Times New Roman 10 tebal]

Penulis mengucapkan terima kasih kepada ... . Dalam kebanyakan kasus, ucapan terima kasih kepada sponsor dan dukungan finansial disertakan.

REFERENSI[Times New Roman 10 tebal]

[1] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, dan J. Dean, "Representasi terdistribusi dari kata dan frasa dan komposisionalitasnya," dalam Prosiding Konferensi ke-27 tentang Sistem Pemrosesan Informasi Neural (NeurIPS), 2013, hlm. 3111–3119.

[2] J. Pennington, R. Socher, dan C. D. Manning, "GloVe: Vektor global untuk representasi kata," dalam Prosiding Konferensi Metode Empiris dalam Pemrosesan Bahasa Alami (EMNLP), 2014, hlm. 1532–1543.

[3] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, dan L. Zettlemoyer, "Representasi kata kontekstual mendalam," dalam Prosiding Konferensi Asosiasi Bab Amerika Utara Linguistik Komputasi: Teknologi Bahasa Manusia (NAACL-HLT), 2018, hlm. 2227–2237.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, dan I. Polosukhin, "Perhatian . . . . . " Konferensi ke-31. Informasi Saraf. Proses. sistem. (NeurIPS), 2017, hlm.1-11. 5998–6008.

[5] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, "BERT: Pra-pelatihan transformer bidirectional mendalam untuk pemahaman bahasa," dalam Prosiding Konferensi Asosiasi Bab Amerika Utara Linguistik Komputasi: Teknologi Bahasa Manusia (NAACL-HLT), 2019, hlm. 4171–4186.

[6] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, dan V. Stoyanov, "Pembelajaran representasi lintas bahasa tanpa pengawasan dalam skala besar," dalam Prosiding Pertemuan Tahunan ke-58 Asosiasi Linguistik Komputasi (ACL), 2020, hlm. 8440–8451.

[7] E. M. Mercha dan H. Benbrahim, "Pembelajaran mesin dan pembelajaran mendalam untuk analisis sentimen lintas bahasa: Sebuah survei," Neurocomputing, vol. 531, hlm. 195–216, 2023.

[8] M. Wankhade, A. C. S. Rao, dan C. Kulkarni, "Sebuah survei tentang metode, aplikasi, dan tantangan analisis sentimen," Artificial Intelligence Review, vol. 55, no. 7, hlm. 5731–5780, 2022.

[9] M. K. Nazir, C. N. Faisal, M. A. Habib, dan H. Ahmad, "Memanfaatkan transformer multibahasa untuk analisis sentimen multikelas dalam data campuran kode bahasa-bahasa dengan sumber daya terbatas," IEEE Access, vol. 13, hlm. 7538–7554, 2025.

[10] S. Sitaram, K. R. Chandu, S. K. Rallabandi, dan A. W. Black, "Sebuah survei tentang pemrosesan ucapan dan bahasa peralihan kode," arXiv preprint arXiv:1904.00784, 2019.

[11] E. W. Pamungkas, A. Fatmawati, Y. S. Nugroho, D. Gunawan, dan E. Sudarmilah, “Deteksi ujaran kebencian di media sosial campuran kode Indonesia: Memanfaatkan sumber daya bahasa multibahasa,” dalam Proc. ke-7 Int. Konf. Informatika dan Komputasi (ICIC), 2022, hlm.1–5.

[12] M. M. Rahman, A. I. Shiplu, Y. Watanobe, dan M. A. Alam, "RoBERTa-BiLSTM: Model hibrida sadar konteks untuk analisis sentimen," IEEE Trans. Emerging Topics in Computational Intelligence, 2025, doi: 10.1109/TETCI.2025.3572150.

[13] F. Barbieri, L. Espinosa Anke, dan J. Camacho-Collados, "XLM-T: Model bahasa multibahasa di Twitter untuk analisis sentimen dan seterusnya," dalam Prosiding Konferensi Sumber Daya dan Evaluasi Bahasa ke-13 (LREC), 2022, hlm. 258–266.

[14] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, "IndoLEM dan IndoBERT: Kumpulan data benchmark dan model bahasa pra-terlatih untuk NLP Indonesia," dalam Prosiding

Konferensi Linguistik Komputasional Internasional ke-28 (COLING), 2020, hlm. 757–770.
[15] F. Koto, J. H. Lau, dan T. Baldwin, "IndoBERTweet: Model bahasa pra-terlatih untuk Twitter Indonesia dengan inisialisasi kosakata spesifik domain yang efektif," dalam Prosiding Konferensi Metode Empiris dalam Pemrosesan Bahasa Alami (EMNLP), 2021, hlm. 10660–10668.
[16] V. Barriere dan A. Balahur, "Meningkatkan analisis sentimen pada tweet non-Inggris menggunakan transformer multibahasa dan terjemahan otomatis untuk augmentasi data," arXiv preprint arXiv:2010.03486, 2020.
[17] A. Joshi, P. Bhattacharyya, dan M. J. Carman, "Deteksi sarkasme otomatis: Sebuah survei," ACM Computing Surveys, vol. 50, no. 5, Art. 73, 2017.
[18] J. L. Fleiss, "Mengukur kesepakatan skala nominal di antara banyak penilai," Buletin Psikologi, vol. 76, no. 5, hlm. 378–382, 1971.
[19] S. Hochreiter dan J. Schmidhuber, "Memori jangka pendek panjang," Komputasi Neural, vol. 9, no. 8, hlm. 1735–1780, 1997.
[20] M. Schuster dan K. K. Paliwal, "Jaringan saraf berulang dua arah," IEEE Trans. Signal Processing, vol. 45, no. 11, hlm. 2673–2681, 1997.
[21] D. Bahdanau, K. Cho, dan Y. Bengio, "Terjemahan mesin saraf dengan pembelajaran bersama untuk menyelaraskan dan menerjemahkan," dalam Prosiding Konferensi Internasional ke-3 tentang Pembelajaran Representasi (ICLR), 2015.
[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, dan R. Salakhutdinov, "Dropout: Cara sederhana untuk mencegah jaringan saraf dari overfitting," Journal of Machine Learning Research, vol. 15, no. 1, hlm. 1929–1958, 2014.
[23] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, dan A. M. Rush, "Transformers: Pemrosesan bahasa alami terkini," dalam Prosiding Konferensi Metode Empiris dalam Pemrosesan Bahasa Alami: Demonstrasi Sistem (EMNLP), 2020, hlm. 38–45.
[24] I. Loshchilov dan F. Hutter, "Regularisasi peluruhan bobot yang terpisah," dalam Prosiding Konferensi Internasional ke-7 tentang Pembelajaran Representasi (ICLR), 2019.
[25] B. Gambäck dan A. Das, "Membandingkan tingkat peralihan kode dalam korpus," dalam Prosiding Konferensi Internasional ke-10 tentang Sumber Daya Bahasa dan Evaluasi (LREC), 2016, hlm. 1850–1855.
[26] F. Wilcoxon, "Perbandingan individu dengan metode pemeringkatan," Buletin Biometrik, vol. 1, no. 6, hlm. 80–83, 1945.
[27] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[28] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Auli, A. Joulin, and T. Grave, "Unsupervised cross-lingual representation learning at scale," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), Online, 2020, pp. 8440–8451.

[29] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014, pp. 1532–1543.

[30] Y. Kim, "Convolutional neural networks for sentence classification," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014, pp. 1746–1751.

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[32] A. Vaswani, N. Shazeer, P. N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008.

[33] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), New Orleans, LA, USA, 2018, pp. 2227–2237.

[34] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of the Association for Computational Linguistics (TACL), vol. 5, pp. 135–146, 2017.

[35] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in Advances in Neural Information Processing Systems (NeurIPS), Vancouver, Canada, 2019, pp. 5753–5763.

[36] A. Z. Lim, A. Pesaranghader, and G. W. S. Hsu, "Multilingual and multi-aspect sentiment analysis with context-aware self-attention," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020, pp. 6814–6823.

[37] C. Xing, Y. Wang, J. Liu, Y. Huang, and W. Y. Wang, "Hierarchical attention networks for document classification," in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), San Diego, CA, USA, 2016, pp. 1480–1489.

[38] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 7370–7377, Jul. 2019.

[39] Y. Li and T. Yang, "Word embedding and text classification research in big data era," Journal of Electrical and Computer Engineering, vol. 2016, p. 7285079, 2016.

[40] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, nos. 1–2, pp. 1–135, 2008.

[41] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 task 4: Sentiment analysis in Twitter," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Vancouver, Canada, 2017, pp. 502–518.

[42] T. Chen, R. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in Proceedings of the 37th International Conference on Machine Learning (ICML), Online, 2020, pp. 1597–1607.

[43] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in Advances in Neural Information Processing Systems (NeurIPS), Montreal, Canada, 2015, pp. 649–657.

[44] F. Liu, J. G. Eisner, and D. Ruths, "Fast and accurate deep network learning by exponential linear units (ELUs)," in Proceedings of the 2016 International Conference on Learning Representations (ICLR), San Juan, Puerto Rico,

2016, pp. 1–13.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proceedings of the 2015 International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015, pp. 1–15.

[46] L. Yao, G. Huang, M. Liu, L. Deng, and X. Maybank, "Deep learning from crowds with belief propagation," in Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021, pp. 4701–4710.

[47] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[48] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), Seattle, WA, USA, 2013, pp. 1631–1642.

[49] J. Turian, L. Ratinov, and Y. Bengio, "Word representations: A simple and general method for semi-supervised learning," in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), Uppsala, Sweden, 2010, pp. 384–394.

[50] S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL), Jeju Island, Korea, 2012, pp. 90–94.

[51] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Baltimore, MD, USA, 2014, pp. 655–665.