Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2026
Surabaya, 3 Juni 2026

ANALISIS KOMPARATIF EFEKTIVITAS VARIAN ARSITEKTUR TRANSFORMER PADA ANALISIS SENTIMEN BAHASA INFORMAL

COMPARATIVE ANALYSIS OF TRANSFORMER ARCHITECTURE VARIANT EFFECTIVENESS IN INFORMAL LANGUAGE SENTIMENT ANALYSIS

Nisfa Likaila Rahmatika Sari[1], Anggi Putri Lestari[2], Adinda Rahmayani Putri[3], Erna Daniati[4]*
***E-mail: ernadaniati@unpkediri.ac.id**

[1,2,3,4]Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

Abstrak

Bahasa informal yang kaya akan slang, sarkasme, dan fenomena code-switching di media sosial menjadi tantangan serius bagi sistem Natural Language Processing (NLP), khususnya dalam tugas analisis sentimen. Penelitian ini bertujuan menganalisis secara komparatif efektivitas berbagai varian arsitektur Transformer—meliputi BERT, RoBERTa, IndoBERT, XLM-R, FNet, dan FastFormer—dalam menangani teks bahasa informal berbahasa Indonesia. Evaluasi dilakukan pada tiga kategori dataset media sosial yang merepresentasikan teks slang, code-mixed, dan sarkasme, dengan menerapkan strategi fine-tuning berbasis domain serta augmentasi data melalui back-translation. Hasil menunjukkan bahwa IndoBERT-base mencapai performa tertinggi dengan akurasi 89,1% dan F1-score 88,7% pada teks slang, sementara XLM-R-base unggul pada teks code-mixed dengan F1-score 89,3%. Pendekatan ensemble keduanya menghasilkan F1-score terbaik sebesar 90,7%. Deteksi sarkasme menjadi tantangan tersulit dengan capaian terbaik hanya 74,8%. Kesimpulannya, tidak ada satu varian Transformer yang secara universal unggul; pemilihan model optimal sangat bergantung pada karakteristik linguistik data target..

Kata kunci**:** *analisis sentimen, arsitektur Transformer, bahasa informal, NLP, slang.*

Abstract

Informal language rich in slang, sarcasm, and code-switching phenomena on social media poses serious challenges for Natural Language Processing (NLP) systems, particularly in sentiment analysis tasks. This study aims to comparatively analyze the effectiveness of various Transformer architecture variants—including BERT, RoBERTa, IndoBERT, XLM-R, FNet, and FastFormer— in handling informal Indonesian text. Evaluation was conducted on three social media dataset categories representing slang, code-mixed, and sarcasm texts, applying domain-based fine-tuning strategies and back-translation data augmentation. Results show that IndoBERT-base achieved the highest performance with 89.1% accuracy and 88.7% F1-score on slang texts, while XLM-R-base excelled in code-mixed texts with an F1-score of 89.3%. An ensemble of both models yielded the best F1-score of 90.7%. Sarcasm detection proved the most difficult challenge, with a best F1-score of only 74.8%. In conclusion, no single Transformer variant is universally superior; optimal model selection is highly dependent on the linguistic characteristics of the target data.

Keywords*: informal language, NLP, sentiment analysis, slang, transformer architecture.*

1. PENDAHULUAN

Pesatnya pertumbuhan platform media sosial dalam satu dekade terakhir telah menghasilkan volume data teks yang belum pernah terjadi sebelumnya. Pengguna internet secara aktif mengekspresikan opini, emosi, dan sentimen mereka melalui berbagai kanal digital, mulai dari Twitter/X, TikTok, hingga forum komunitas daring lainnya [27][29]. Fenomena ini membuka peluang besar sekaligus tantangan serius dalam bidang Pemrosesan Bahasa Alami (*Natural Language Processing*/NLP), khususnya dalam tugas analisis sentimen yang bertujuan mengidentifikasi polaritas emosional suatu teks secara otomatis [26][33].

Tantangan utama yang dihadapi dalam analisis sentimen di media sosial bersumber dari karakteristik linguistik teks yang digunakan. Berbeda dengan teks formal, pengguna media sosial cenderung menggunakan bahasa informal yang kaya akan slang, singkatan, sarkasme, dan fenomena *code-switching* [14][20]. Penggunaan bahasa slang merupakan fenomena yang terus berkembang, khususnya di kalangan generasi muda yang aktif di media sosial [11][18]. Di Indonesia, percampuran Bahasa Indonesia, Bahasa Daerah, dan Bahasa Inggris dalam satu kalimat merupakan hal yang umum dan sulit dimodelkan oleh sistem NLP konvensional yang dilatih pada korpus formal [5]. Variasi ejaan dan neologisme yang terus berkembang secara dinamis semakin mempersulit generalisasi model ke ranah bahasa informal [9][21]. Sebagai contoh, kalimat informal seperti *"Bro anjir filmnya sadis banget, gue sampe gemeteran wkwkwk"* memuat slang (anjir, sadis, bro), onomatope tawa (wkwkwk), dan ekspresi emosional yang seluruhnya berkontribusi pada sentimen positif, namun sulit diidentifikasi secara tepat oleh model berbasis kamus konvensional [23][34].

Munculnya arsitektur Transformer yang diperkenalkan oleh Vaswani dkk. melalui makalah seminalnya yang berjudul "*Attention Is All You Need*" menandai babak baru dalam pengembangan model bahasa [6]. Mekanisme *self-attention* yang menjadi fondasi arsitektur ini memungkinkan model menangkap dependensi jarak jauh antar token secara lebih efektif dibandingkan arsitektur sebelumnya seperti *Recurrent Neural Network* (RNN) dan *Long Short-Term Memory* (LSTM) [37]. Perkembangan ini kemudian melahirkan serangkaian model pra-latih seperti BERT, RoBERTa, IndoBERT, XLM-R, dan berbagai varian lainnya yang telah menunjukkan kinerja unggul pada berbagai benchmark NLP standar [10][35].

Meskipun demikian, efektivitas berbagai varian arsitektur Transformer ini tidak seragam ketika dihadapkan pada karakteristik bahasa informal yang sangat beragam. Perbedaan mendasar pada desain arsitektur, mulai dari jumlah lapisan perhatian, mekanisme tokenisasi, strategi pra-pelatihan, hingga korpus yang digunakan menghasilkan kemampuan yang berbeda-beda dalam menangani slang, sarkasme, dan code-switching [3][12]. Sebagian besar model pra-latih yang tersedia dilatih pada korpus formal atau semi-formal, sehingga representasi semantiknya belum tentu optimal untuk menangani kekhasan bahasa informal yang terus berevolusi seiring tren budaya populer [10][30]. Studi komparatif yang secara sistematis mengevaluasi efektivitas varian-varian Transformer ini pada teks informal masih sangat terbatas dalam literatur.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk: (1) menganalisis secara komparatif efektivitas berbagai varian arsitektur Transformer meliputi BERT, RoBERTa, IndoBERT, XLM-R, dan model efisien seperti FNet dan FastFormer, dalam tugas analisis sentimen teks bahasa informal; (2) mengidentifikasi perbedaan struktural dan mekanisme di balik kinerja masing-masing varian dalam menangani fenomena slang, *code-switching*, dan sarkasme; (3) menganalisis pengaruh strategi *fine-tuning* dan augmentasi data terhadap performa model pada domain bahasa informal; serta (4) merumuskan rekomendasi pemilihan model yang optimal berdasarkan konteks bahasa dan karakteristik dataset yang digunakan [17][31]. Dengan demikian, artikel ini diharapkan dapat menjadi acuan yang komprehensif bagi peneliti dan praktisi yang bekerja di persimpangan antara NLP, analisis sentimen, dan komputasi media sosial.

2. METODOLOGI

Penelitian ini menggunakan pendekatan eksperimental komparatif untuk mengevaluasi efektivitas berbagai varian arsitektur Transformer dalam tugas analisis sentimen bahasa informal. Kerangka metodologi dirancang untuk memungkinkan perbandingan yang adil (fair comparison) antar varian model dengan mengontrol variabel-variabel eksternal seperti ukuran dataset, prosedur pelatihan, dan metrik evaluasi. Pendekatan ini dipilih karena mampu menghasilkan temuan yang dapat direplikasi dan memberikan wawasan praktis yang langsung dapat diterapkan dalam pengembangan sistem NLP berbasis bahasa informal [12][26].

2.1 Arsitektur Transformer dan Mekanisme Self-Attention

Fondasi dari seluruh varian yang dievaluasi adalah mekanisme self-attention yang diperkenalkan Vaswani et al. [6]. Dalam konteks teks informal, mekanisme ini memiliki peran kritis karena kemampuannya menangkap hubungan semantik antar token tanpa memandang jarak posisi dalam kalimat. Teks informal seperti "filmnya sadis banget wkwkwk" membutuhkan kemampuan model untuk mengasosiasikan "sadis" (yang dalam konteks informal bermakna positif) dengan token emosional lain secara bersamaan dengan sesuatu yang sulit dilakukan model sekuensial [3][37]. Secara matematis, bobot perhatian dihitung sebagai:

$$Attention\,(Q,\,K,\,V)\;=\;softmax\left(\frac{QK^{T}}{\sqrt{d_{k}}}\right)V \tag{1}$$

di mana Q (Query) adalah representasi dari token saat ini yang sedang diproses (mencari informasi apa), K (Key) adalah representasi dari semua token dalam input (informasi apa yang tersedia?), dan V (Value) adalah konten aktual dari token input (informasi apa yang akan diambil?). Mekanisme multi-head attention memungkinkan model menangkap berbagai aspek semantik secara paralel, yang sangat berguna untuk teks yang mengandung slang dan ekspresi ambigu [6][35].

2.2 Varian Arsitektur Transformer yang Dievaluasi

Lima varian arsitektur Transformer dipilih untuk evaluasi komparatif berdasarkan relevansinya dengan analisis sentimen bahasa informal:ditulis secara jelas dengan indeks seperti contoh berikut:

2.2.1 Bidirectional Encoder Representations from Transformers (BERT)

BERT menggunakan arsitektur encoder dua arah (bidirectional) dengan mekanisme Masked Language Modeling (MLM) sebagai tugas pra-pelatihan. Kemampuan pemahaman konteks dua arah ini memberikan keunggulan dalam menangkap makna kata yang bergantung pada konteks kiri dan kanan sekaligus relevan untuk slang yang makna literalnya bertentangan dengan sentimen yang dimaksud [35][38].

2.2.2 Robustly Optimized BERT Pretraining Approach (RoBERTa)

RoBERTa merupakan penyempurnaan dari BERT dengan pelatihan pada korpus yang jauh lebih besar, tanpa tugas Next Sentence Prediction (NSP), dan menggunakan dynamic masking. Modifikasi ini menghasilkan representasi yang lebih robust pada teks yang mengandung variasi leksikal tinggi seperti slang dan neologisme [10][22].

2.2.3 IndoBERT
IndoBERT adalah model BERT yang dilatih secara khusus pada korpus Bahasa Indonesia skala besar dari berbagai sumber termasuk teks media sosial. Spesialisasi domain ini menjadikannya kandidat utama untuk teks informal Bahasa Indonesia, karena tokenizer-nya dioptimalkan untuk mengenali morfologi dan leksikon Bahasa Indonesia, termasuk slang yang lazim digunakan di media sosial lokal [12][17].

2.2.4 XLM-R (XLM-RoBERTa)
XLM-R dilatih pada korpus multilingua yang mencakup lebih dari 100 bahasa menggunakan pendekatan RoBERTa. Keunggulan utamanya terletak pada kemampuan menangani teks code-mixed (campuran dua bahasa atau lebih dalam satu ujaran), yang merupakan fenomena dominan di media sosial Asia Tenggara [14][24][25].

2.2.5 FNet dan FastFormer
FNet mengganti mekanisme self-attention dengan transformasi Fourier yang lebih efisien secara komputasional, sementara FastFormer menggunakan additive attention untuk mereduksi kompleksitas dari O(n^2) menjadi O(n). Kedua model ini dievaluasi sebagai alternatif efisien untuk skenario dengan sumber daya komputasi terbatas [2].

2.3 Dataset dan Persiapan Data
Evaluasi dilakukan pada tiga kategori dataset yang mewakili spektrum bahasa informal media sosial: (1) Dataset Twitter/X Bahasa Indonesia yang mengandung campuran slang, singkatan, dan code-mixing; (2) Dataset komentar YouTube dalam Bahasa Indonesia yang kaya akan sarkasme dan ironi; serta (3) Dataset multilingual yang mengandung teks code-mixed Bahasa Indonesia-Inggris. Pemilihan ketiga kategori ini didasarkan pada representativitas fenomena linguistik informal yang paling sering ditemui di media sosial [20][31].

2.4 Strategi Preprocessing Teks Informal
Preprocessing teks informal memerlukan pipeline yang berbeda dari teks formal. Tahapan yang diterapkan meliputi: (1) Normalisasi slang dan singkatan menggunakan kamus slang yang dikembangkan secara khusus berdasarkan tipologi slang Generasi Z dan Alpha di platform media sosial [11][13]; (2) Penanganan emoji dengan konversi ke deskripsi tekstual yang kontekstual; (3) Tokenisasi berbasis subword menggunakan Word Piece (untuk BERT/IndoBERT) dan Byte-Pair Encoding (untuk RoBERTa/XLM-R); (4) Normalisasi karakter berulang (contoh: "sadisssss" menjadi "sadis"); serta (5) Penghapusan elemen non-informatif seperti URL dan mention pengguna [3][30].
Khusus untuk teks code-mixed, dilakukan deteksi bahasa per token menggunakan pendekatan berbasis karakter sebelum dinormalisasi. Hal ini penting karena XLM-R memproses code-mixed secara berbeda dengan cara melakukan pemisahan representasi per bahasa, sedangkan IndoBERT lebih rentan terhadap token bahasa asing yang tidak dikenali dalam kosakatanya [5][14].

2.5 Prosedur Fine-Tuning
Setiap model pra-latih dimodifikasi dengan menambahkan lapisan klasifikasi (classification head) di atas representasi token [CLS] untuk tugas klasifikasi sentimen tiga kelas (positif, negatif, netral). Fine-tuning dilakukan dengan pendekatan domain-adaptive pre-training (DAPT) dua tahap: tahap pertama melanjutkan pra-pelatihan pada korpus teks informal yang tidak berlabel untuk mengadaptasi representasi leksikal model; tahap kedua melakukan supervised fine-tuning pada data berlabel dengan label sentimen [30][36].
Konfigurasi hiperparameter yang digunakan: learning rate awal 2e-5 dengan scheduler linear warm-up selama 10% langkah pertama, batch size 32, epoch maksimal 10 dengan early

stopping berdasarkan F1-score pada validation set, dan dropout rate 0.1 untuk regularisasi. Augmentasi data diterapkan pada kelas minoritas menggunakan teknik back-translation (Indonesia-Inggris-Indonesia) dan synonym replacement berbasis WordNet Bahasa Indonesia untuk mengatasi ketidakseimbangan distribusi kelas [8][28].

2.6 Metrik Evaluasi

Kinerja model dievaluasi menggunakan metrik standar NLP: Accuracy, Precision, Recall, dan F1-score (macro-averaged untuk menangani ketidakseimbangan kelas). Selain metrik kinerja, evaluasi juga mencakup efisiensi komputasional (waktu inferensi per sampel dan ukuran model dalam jumlah parameter) untuk mempertimbangkan kelayakan deployment pada sistem produksi dengan sumber daya terbatas. Uji statistik Wilcoxon signed-rank digunakan untuk mengkonfirmasi signifikansi perbedaan kinerja antar varian model [12][19].

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil evaluasi komparatif kinerja varian arsitektur Transformer pada tiga skenario uji utama: (1) klasifikasi sentimen umum teks informal, (2) kinerja pada sub kategori bahasa informal spesifik, dan (3) efektivitas strategi fine-tuning. Analisis dilanjutkan dengan pembahasan mendalam mengenai faktor-faktor arsitektur yang memengaruhi kinerja masing-masing model.

3.1 Perbandingan Performa Umum Antarvarian Transformer

Tabel 1 merangkum hasil evaluasi kinerja keseluruhan enam varian Transformer pada dataset analisis sentimen teks informal Bahasa Indonesia. IndoBERT-base menunjukkan performa tertinggi dengan akurasi 89,1% dan F1-score 88,7%, diikuti XLM-R-base (87,5% akurasi) dan RoBERTa-base (86,3% akurasi). Model efisien FNet dan FastFormer berada di posisi terbawah dari sisi akurasi, meskipun unggul dalam hal efisiensi parameter.

Table 1. Perbandingan Performa Umum Varian Transformer pada Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Params (M)
BERT-base	84.7	83.9	84.2	84.0	**110**
RoBERTa-base	86.3	85.8	86.1	85.9	**125**
IndoBERT-base	89.1	88.6	88.9	88.7	**124**
XLM-R-base	87.5	87.0	87.3	87.1	**278**
FNet	79.2	78.5	79.0	78.7	**82**
FastFormer	80.8	80.1	80.5	80.3	**98**

Dominasi IndoBERT-base konsisten dengan temuan Karim dan Wibowo [17] yang melaporkan keunggulan IndoBERT dalam klasifikasi emosi multi-kelas pada teks informal Bahasa Indonesia, serta Daniati et al. [12] yang membuktikan superioritas fine-tuned IndoBERT dibandingkan SVM dan LSTM dalam analisis sentimen opini masyarakat Indonesia. Keunggulan ini secara fundamental bersumber dari korpus pra-pelatihan yang mencakup teks media sosial Bahasa Indonesia, sehingga tokenizer dan representasi semantik model lebih selaras dengan fenomena linguistik lokal.

XLM-R-base menempati posisi kedua dengan keunggulan signifikan pada teks yang mengandung elemen multibahasa. Meskipun jumlah parameternya jauh lebih besar (278 juta) dibandingkan IndoBERT (124 juta), performa keseluruhannya sedikit lebih rendah pada teks monolingual Bahasa Indonesia. Hal ini mengindikasikan adanya trade-off antara kemampuan

multibahasa dan spesialisasi domain monolingual [19][24]. RoBERTa-base mengungguli BERT-base sebesar 1,6 poin persentase F1-score, mengkonfirmasi bahwa optimisasi prosedur pra-pelatihan dengan korpus yang lebih besar dan penghapusan tugas NSP memberikan representasi yang lebih robust untuk teks informal [10].

3.2 Kinerja pada Subkategori Bahasa Informal Spesifik

Evaluasi pada tiga subkategori bahasa informal slang, code-mixed, dan sarkasme mengungkap pola kinerja yang lebih nuansif. Tabel 2 menyajikan F1-score masing-masing model pada setiap subkategori.

Table 2. Perbandingan F1-Score per Subkategori

Model	F1 Slang (%)	F1 Code-Mixed	F1 Sarkasme	Rata-rata F1
BERT-base	81.3	76.4	68.2	75.3
RoBERTa-base	83.7	79.1	71.5	78.1
IndoBERT-base	91.2	83.6	74.8	83.2
XLM-R-base	86.4	89.3	73.2	83.0
FNet	75.1	72.8	61.9	69.9
FastFormer	76.3	74.2	63.1	71.2

3.2.1 Penanganan Slang dan Neologisme

IndoBERT-base meraih F1-score tertinggi sebesar 91,2% pada kategori slang, unggul 4,8 poin dari RoBERTa-base (83,7%) dan 7,5 poin dari BERT-base (81,3%). Keunggulan ini mencerminkan efektivitas pra-pelatihan berbasis korpus lokal dalam menginternalisasi representasi vektor untuk leksikon slang Indonesia. Kanapala et al. [1] mencatat bahwa model yang mampu memperbarui representasi slang secara adaptif akan secara konsisten melampaui performa model statis pada data Twitter dan Reddit. Temuan ini selaras dengan observasi Rachmijati dan Cahyati [11] mengenai cepatnya evolusi slang Generasi Alpha di platform X, yang menuntut model dengan representasi leksikal yang lebih dinamis.

3.2.2 Penanganan Code-Switching dan Code-Mixed

Pola kinerja yang berbeda muncul pada subkategori code-mixed, di mana XLM-R-base mengungguli IndoBERT-base dengan selisih 5,7 poin F1-score (89,3% vs 83,6%). Keunggulan XLM-R pada skenario ini secara konsisten dilaporkan dalam literatur: Nazir et al. [24] mendemonstrasikan keunggulan XLM-R pada analisis sentimen multi-kelas data code-mixed bahasa rendah sumber daya, sementara Yadav et al. [7] mengkonfirmasi superioritas XLM-R pada prediksi sentimen dan ujaran kebencian teks code-mixed. Hashmi et al. [14] secara khusus menyoroti bahwa model multilingual dengan pra-pelatihan berbasis korpus multibahasa memiliki kemampuan lebih baik dalam menangkap transisi semantik antarbahasa dalam satu ujaran. Sebaliknya, keterbatasan IndoBERT pada skenario ini diduga bersumber dari kosakata yang terbatas pada token Bahasa Indonesia, sehingga frasa asing yang tidak dikenali akan dipecah menjadi subword yang kehilangan makna kontekstualnya [5].

3.2.3 Deteksi Sarkasme dan Ironi

Deteksi sarkasme merupakan subkategori tersulit bagi semua model yang dievaluasi, dengan F1-score terbaik hanya mencapai 74,8% (IndoBERT-base). Fenomena ini konsisten dengan temuan Perdana dan Sugandi [15] yang mengidentifikasi keterbatasan serius generalisasi model ketika diuji lintas domain pada tugas deteksi sarkasme Bahasa Indonesia. Memon et al.

[32] mengungkap bahwa akar permasalahan terletak pada kebutuhan akan pemahaman konteks pragmatik yang melampaui makna literal teks sebuah kemampuan yang belum sepenuhnya dimiliki oleh model Transformer berbasis teks semata. Fanani dan Wahyudin [4] menunjukkan bahwa penanganan ketidakseimbangan kelas pada data sarkasme komentar YouTube Indonesia dengan teknik augmentasi khusus dapat meningkatkan F1-score secara signifikan, namun pencapaian di atas 75% masih tetap menantang.

3.3 Kinerja pada Subkategori Bahasa Informal Spesifik

Tabel 3 membandingkan kinerja model sebelum dan sesudah penerapan Domain-Adaptive Pre-Training (DAPT), yaitu kelanjutan pra-pelatihan pada korpus teks informal tidak berlabel sebelum supervised fine-tuning.

Table 3. Tabel Pengaruh Domain-Adaptive Pre-Training (DAPT) terhadap Akurasi Model

Model	Tanpa DAPT (%)	Dengan DAPT (%)	Peningkatan (%)
BERT-base	80.2	84.7	+4.5
RoBERTa-base	82.6	86.3	+3.7
IndoBERT-base	85.3	89.1	+3.8
XLM-R-base	83.9	87.5	+3.6
FNet	76.1	79.2	+3.1
FastFormer	77.4	80.8	+3.4

Strategi DAPT secara konsisten meningkatkan performa seluruh varian model, dengan peningkatan berkisar antara 3,1% (FNet) hingga 4,5% (BERT-base). BERT-base memperoleh manfaat terbesar karena korpus pra-pelatihannya yang lebih generik memberikan ruang adaptasi yang lebih luas dibandingkan IndoBERT yang sudah lebih terspesialisasi. Temuan ini selaras dengan Calizaya-Milla et al. [30] yang mengkonfirmasi efektivitas fine-tuning model BERT monolingual untuk konteks slang lokal pada data Peru. Adhim dan Cahyono [28] secara khusus menunjukkan bahwa kombinasi fine-tuning dengan hybrid labeling mampu mengoptimalkan representasi IndoBERT untuk analisis sentimen FOMO di media sosial Bahasa Indonesia.

Strategi augmentasi data melalui back-translation menghasilkan peningkatan konsisten pada semua model, dengan rata-rata peningkatan F1-score sebesar 2,3 poin pada kelas minoritas. Temuan ini mendukung Pratama et al. [8] yang menunjukkan bahwa augmentasi kelas minoritas secara signifikan mengurangi bias model pada dataset tidak seimbang. Pendekatan ensemble yang menggabungkan IndoBERT dan XLM-R menghasilkan F1-score tertinggi sebesar 90,7% melampaui performa model individual mana pun mengkonfirmasi potensi strategi ensemble seperti yang dilaporkan Bilehsavar et al. [22].

3.4 Analisis Efisiensi Komputasional

Dari perspektif efisiensi komputasional, terdapat trade-off yang signifikan antara akurasi dan sumber daya yang diperlukan. XLM-R-base dengan 278 juta parameter membutuhkan waktu inferensi rata-rata 47 ms per sampel, dibandingkan IndoBERT-base (124 juta parameter, 23 ms) dan FNet (82 juta parameter, 11 ms). Untuk skenario deployment produksi dengan latensi rendah, model FNet dan FastFormer menjadi alternatif yang layak dipertimbangkan, khususnya apabila teks yang dianalisis tidak mengandung banyak fenomena code-mixed atau sarkasme tingkat tinggi [2].

3.5 Sintesis Temuan dan Implikasi Praktis

Secara keseluruhan, evaluasi ini mengkonfirmasi bahwa tidak ada satu varian Transformer yang secara universal unggul di semua subkategori bahasa informal. IndoBERT-base

merupakan pilihan optimal untuk teks monolingual Bahasa Indonesia yang kaya slang, sedangkan XLM-R-base lebih sesuai untuk teks code-mixed multibahasa. Kombinasi ensemble keduanya menghasilkan performa terbaik secara keseluruhan dengan mengorbankan efisiensi komputasional. FNet dan FastFormer menawarkan alternatif efisien dengan penurunan akurasi yang dapat diterima untuk use case dengan batasan komputasi [2][31]. Bahasa informal yang terus berevolusi, fenomena code-switching, dan kompleksitas pragmatik sarkasme masih menjadi tantangan terbuka yang membutuhkan inovasi metodologis berkelanjutan terutama untuk bahasa dengan sumber daya terbatas seperti Bahasa Indonesia dalam konteks media sosial [16][19].

4. KESIMPULAN DAN SARAN

Penelitian ini bertujuan menganalisis secara komparatif efektivitas varian arsitektur Transformer dalam analisis sentimen bahasa informal, dengan fokus pada karakteristik teks media sosial berbahasa Indonesia yang mencakup slang, code-mixing, dan sarkasme. Berdasarkan seluruh rangkaian evaluasi eksperimental yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut. Kesimpulan utama penelitian ini menegaskan bahwa tidak ada satu varian arsitektur Transformer yang secara universal superior di semua subkategori bahasa informal. IndoBERT-base terbukti paling efektif untuk analisis sentimen teks monolingual Bahasa Indonesia yang kaya akan slang dan neologisme, mencapai F1-score 88,7% secara keseluruhan dan 91,2% pada subkategori slang. Keunggulan ini secara langsung bersumber dari spesialisasi domain korpus pra-pelatihan yang mencakup teks media sosial Bahasa Indonesia. Sebaliknya, XLM-R-base mengungguli varian lain pada subkategori teks code-mixed dengan F1-score 89,3%, mengkonfirmasi hipotesis bahwa kemampuan representasi multibahasa merupakan prasyarat fundamental untuk menangani fenomena code-switching yang dominan di media sosial Asia Tenggara. Temuan ini secara langsung menjawab permasalahan yang dirumuskan: efektivitas model sangat bergantung pada kesesuaian antara karakteristik linguistik data target dengan sifat korpus pra-pelatihan model yang digunakan.

Strategi Domain-Adaptive Pre-Training (DAPT) terbukti secara konsisten meningkatkan performa seluruh varian model dengan rata-rata peningkatan akurasi 3,5 poin persentase, dengan BERT-base memperoleh manfaat terbesar (+4,5%). Pendekatan ensemble yang menggabungkan IndoBERT dan XLM-R menghasilkan F1-score tertinggi sebesar 90,7%, melampaui performa model individual mana pun. Augmentasi data melalui back-translation juga terbukti efektif meningkatkan robustitas model pada kelas minoritas dan distribusi data yang tidak seimbang.

Sebagai kesimpulan tambahan, penelitian ini mengidentifikasi bahwa deteksi sarkasme dan ironi tetap menjadi tantangan paling sulit bagi semua varian yang dievaluasi, dengan F1-score terbaik hanya 74,8% (IndoBERT-base). Hal ini mencerminkan fakta bahwa sarkasme membutuhkan pemahaman konteks pragmatik yang melampaui kapabilitas representasi semantik berbasis teks semata sebuah gap yang belum terjembatani oleh mekanisme self-attention yang ada. Selain itu, model efisien FNet dan FastFormer, meskipun menunjukkan akurasi lebih rendah (79,2% dan 80,8%), tetap merupakan alternatif yang layak untuk skenario deployment dengan batasan komputasi ketat, mengingat efisiensi inferensinya yang jauh lebih tinggi. Pengembangan lebih lanjut hendaknya difokuskan pada pembuatan korpus pra-pelatihan teks informal Bahasa Indonesia yang lebih komprehensif, pengembangan mekanisme adaptasi leksikal dinamis untuk menangani evolusi slang, serta eksplorasi arsitektur multimodal yang mengintegrasikan sinyal teks dan emoji untuk meningkatkan pemahaman sentimen kontekstual..

5. DAFTAR RUJUKAN

[1] A. Kanapala, M. Sandeep, J. Teja, A. Manikonda, B. Mekala, and S. Mahipal, "ADAPTIVE SLANG AND LANGUAGE MODELING FOR MINING INFORMAL WEB CONTENT," *J. Theor. Appl. Inf. Technol.*, vol. 31, no. 24, 2025, [Online]. Available: www.jatit.org

[2] A. Kumar, A. Pandey, S. Ahlawat, and Y. Prasad, "On Enhancing Code-Mixed Sentiment and Emotion Classification Using FNet and FastFormer," in *International Conference on Agents and Artificial Intelligence*, Science and Technology Publications, Lda, 2025, pp. 670–678. doi: 10.5220/0013173600003890.

[3] A. Kurniasih and L. Parningotan Manik, "On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts." [Online]. Available: www.ijacsa.thesai.org

[4] A. M. Fanani and Moh. I. Wahyuddin, "Sarcasm Detection in Indonesian YouTube Comments using Fine-Tuned IndoBERT with Class Imbalance Handling," *sinkron*, vol. 10, no. 1, Jan. 2026, doi: 10.33395/sinkron.v10i1.15607.

[5] A. N. Sabrina, "INTERNET SLANG CONTAINING CODE-MIXING OF ENGLISH AND INDONESIAN USED BY MILLENNIALS ON TWITTER (Slang Internet Mengandung Campur-Kode Bahasa Inggris dan Indonesia yang Digunakan oleh Milenial
di Twitter)," *Kandai*, vol. 17, no. 2, p. 153, Nov. 2021, doi: 10.26499/jk.v17i2.3422.

[6] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 2023, [Online]. Available: http://arxiv.org/abs/1706.03762

[7] A. Yadav, T. Garg, M. Klemen, M. Ulcar, B. Agarwal, and M. R. Sikonja, "Code-mixed Sentiment and Hate-speech Prediction," May 2024, doi: 10.1109/TAFFC.2025.3553399.

[8] A. Yoga Pratama, G. Ananda Sanjaya, N. Khairunisa Lubis, and M. Rangga Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," *METIK Jurnal*, vol. 9, p. 2025, 2025, doi: 10.47002/metik.v9i1.1055.

[9] Ari Putra WIbowo and Nurul Hidayat, "Eksplorasi Linguistik Komputasional dalam Analisis Bahasa Alami untuk Mengungkap Evolusi Dialek Digital di Era Media Sosial Global," *Journal of New Trends in Sciences*, vol. 1, no. 3, pp. 45–52, Oct. 2023, doi: 10.59031/jnts.v1i3.778.

[10] B. V. P. Kumar and M. Sadanandam, "A Fusion Architecture of BERT and RoBERTa for Enhanced Performance of Sentiment Analysis of Social Media Platforms," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 51–66, 2024, doi: 10.12785/ijcds/150105.

[11] C. Rachmijati and S. S. Cahyati, "Know Your Skibidi: Navigating Gen Alpha's Slang Types and Trend on Social Media X," vol. 3, no. 2, 2024, doi: 10.47841/icorad.v3i2.219.

[12] E. Daniati *et al.*, "Perbandingan Kinerja Algoritma SVM, LSTM, dan Fine-tuned IndoBERT dalam Analisis Sentimen Opini Masyarakat Indonesia terhadap Mobil Listrik," 2026. [Online]. Available: https://subset.id/index.php/IJCSR

[13] E. F. B. T. M. H. A. E. N. Alysia Cynthia, "Bahasa slang pada media sosial 'x' di era gen z," *Journal Of Social Science Research*, 2024.

[14] E. Hashmi, S. Y. Yayilgan, and S. Shaikh, "Augmenting sentiment prediction capabilities for code-mixed tweets with multilingual transformers," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, Dec. 2024, doi: 10.1007/s13278-024-01245-6.

[15] F. S. Andreas Perdana, "Evaluasi Lintas Domain Deteksi Sarkasme Bahasa Indonesia:

Pendekatan Hybrid IndoBERT dan Fitur Pragmatis," 2026.
[16] G. BANDARUPALLI, "Enhancing Sentiment Analysis in Multilingual Social Media Data Using Transformer-Based NLP Models A Synthetic Computational Study," Apr. 11, 2025. doi: 10.36227/techrxiv.174440282.23013172/v1.
[17] H. F. Karim and A. P. Wibowo, "BULLETIN OF COMPUTER SCIENCE RESEARCH Kinerja Metode Fine-Tuning IndoBERT untuk Klasifikasi Emosi Multi-Kelas pada Teks Informal Bahasa Indonesia," *Media Online)*, vol. 6, no. 1, pp. 63–74, 2025, doi: 10.47065/bulletincsr.v6i1.850.
[18] I. Gede Budiasa, W. Savitri, A. A. S. Shanti, and S. Dewi, "HUMANIS Journal of Arts and Humanities Penggunaan Bahasa Slang di Media Sosial," *Journal of Arts and Humanities*, vol. 25, pp. 192–200, 2021, doi: 10.24843/JH.20.
[19] K. K. R Anitha, "SENTIMENT ANALYSIS IN LOW RESOURCE LANGUAGE: EXPLORING BERT, MBERT, XLM-R, AND RNN ARCHITECTURES TO UNDERPIN THE DEEP LANGUAGE UNDERSTANDING," *Journal of Nonlinear Analysis and Optimization*, vol. 15, no. 3, p. 2024, 2024.
[20] L. Widya Astuti and Y. Sari, "Code-Mixed Sentiment Analysis using Transformer for Twitter Social Media Data," 2023. [Online]. Available: www.ijacsa.thesai.org
[21] M. Asim, M. Waqar, and I. Alam, "Spectrum of Engineering Sciences A COMPARATIVE ANALYSIS OF DEEP LEARNING METHODS FOR SLANG DETECTION IN TWITTER DATA Iftikhar Alam", doi: 10.5281/zenodo.17906419.
[22] M. S. Bilehsavar, N. Mahmoudi, M. J. Torkamani, and K. Kiashemshaki, "Ensembling Multilingual Transformers for Robust Sentiment Analysis of Tweets," Sep. 2025, [Online]. Available: http://arxiv.org/abs/2509.24080
[23] M. Khazeni, M. Heydari, and A. Albadvi, "Persian Slang Text Conversion to Formal and Deep Learning of Persian Short Texts on Social Media for Sentiment Classification," Mar. 2024, doi: 10.22061/jecei.2024.10745.731.
[24] M. K. Nazir, C. N. Faisal, M. A. Habib, and H. Ahmad, "Leveraging Multilingual Transformer for Multiclass Sentiment Analysis in Code-Mixed Data of Low-Resource Languages," *IEEE Access*, vol. 13, pp. 7538–7554, 2025, doi: 10.1109/ACCESS.2025.3527710.
[25] M. Krasitskii, O. Kolesnikova, L. Chanona Hernandez, G. Sidorov, and A. Gelbukh, "Advancing Sentiment Analysis in Tamil-English Code-Mixed Texts: Challenges and Transformer-Based Solutions."
[26] M. M. N. M. Reddy, "CONTEXT-AWARE SENTIMENT CLASSIFICATION OF SOCIAL MEDIA TEXT USING ATTENTION-BASED TRANSFORMER MODELS,"
2026. [Online]. Available: www.ajaccm.com
[27] M. Rivaro Farrelino Gozali and D. Wisnu Brata, "Analisis Sentimen Pengguna Twitter / X Terhadap Fenomena #KaburAjaDulu Menggunakan Metode Support Vector Machine dan IndoBERT Embedding," 2026. [Online]. Available: http://j-ptiik.ub.ac.id
[28] N. Fauzil Adhim and N. Cahyono, "Optimization of IndoBERT for Sentiment Analysis of
FOMO on Social Media Through Fine-Tuning and Hybrid Labeling," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC
[29] R. Mandal, J. Chen, S. Becken, and B. Stantic, "Tweets Topic Classification and Sentiment Analysis Based on Transformer-Based Language Models," in *Vietnam Journal of Computer Science*, World Scientific, May 2023, pp. 117–134. doi: 10.1142/S2196888822500269.

[30] S. E. Calizaya-Milla, J. Santos, and F. A. Huanca Torres, "Fine-Tuning Monolingual Pre-Trained BERT Models for Sentiment Analysis in Peruvian Slang Contexts," *Applied Artificial Intelligence*, vol. 40, no. 1, 2026, doi: 10.1080/08839514.2026.2641381.

[31] S. Informasi, "Perbandingan Model BERT dan RNN-LSTM pada Analisis Sentimen Aplikasi BRI Mobile Dea Yuliana Ayu Ningrum 1a) , Erna Daniati* 2a) , M. Najibulloh Muzaki 3a) a)," 2025. [Online]. Available: https://subset.id/index.php/IJCSR

[32] S. K. B. M. N. A. F. N. & F. Mehreen. Muhammad Qasim Memon, "DETECTING SARCASM IN SOCIAL MEDIA POSTS USING TRANSFORMER-BASED LANGUAGE MODELS WITH CONTEXTUAL AND SENTIMENT-AWARE FEATURES," *Spectrum of Engineering Sciences*, 2024, doi: 10.5281/zenodo.15387702.

[33] S. Padmalal, I. Edwin Dayanand, G. S. Rao, and S. Gore, "Enhancing Sentiment Analysis in Social Media Texts Using Transformer-Based NLP Models," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 11, no. 8, pp. 208–216, Aug. 2024, doi: 10.14445/23488379/IJEEE-V11I8P118.

[34] S. Priccilia, "Sentiment Analysis of Slang Language Trends in Generation Alpha on Social Media Using BERT," vol. 7, no. 2, pp. 223–228, 2025, doi: 10.21512/emacsjournal.v6.

[35] S. Tabinda Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, Jul. 2022, doi: 10.1016/j.array.2022.100157.

[36] U. Pranitha, "Advancing Sentiment Prediction for Code- mixed Tweets with Transformer models," *International Journal of Computer Engineering in Research Trends*, vol. 12, no. 7, pp. 1–12, Jul. 2025, doi: 10.22362/ijcert/2025/v12/i7/v12i701.

[37] Y. Lashyn, O. Trofymchuk, S. Zabolotnyi, O. Voitko, and E. Seabra, "SENTIMENT ANALYSIS OF TEXTS USING RECURRENT NEURAL NETWORKS OF THE TRANSFORMER ARCHITECTURE," *Advanced Information Systems*, vol. 9, no. 3, pp. 91–101, Aug. 2025, doi: 10.20998/2522-9052.2025.3.11.

[38] Y. Tang and N. L. Ali, "Cat-related slang and sentiment analysis: the CatSlang-SA dataset and model evaluation using social media posts," *Lang. Resour. Eval.*, vol. 60, no. 1, Mar. 2026, doi: 10.1007/s10579-025-09901-9.