Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 9172 – 9184 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

METODE DAN ALGORITMA ANALISIS SENTIMEN: TINJAUAN LITERATUR SISTEMATIS 2023–2026

Purwadi[1], Berlilana[2], Bagus Adhi Kusuma[3], Dhanar Intan Surya Saputra[4], Hendra Marcos[5]
Universitas Amikom Purwokerto, Purwokerto
e-mail: [1]purwadi@amikompurwokerto.ac.id, [2]berli@amikompurwokerto.ac.id,
[3]bagus@amikompurwokerto.ac.id, [4]dhanarsaputra@amikompurwokerto.ac.id,
[5]hendra.marcos@amikompurwokerto.ac.id

Abstract: *Sentiment analysis has shifted from feature-based machine learning to deep learning, Transformer encoders, and large language models (LLMs), affecting model accuracy and approaches to data formulation, supervision, and evaluation. This systematic literature review maps developments from 2023 to 2026 using an auditable corpus of 30 primary studies classified into four methodological groups: machine learning, deep learning, Transformer, and LLM. The corpus composition changes over time: machine learning accounts for 77.8% of studies in 2023, falls to 40.0% in 2024, and is absent in 2025–2026. Transformer and LLM approaches rise from 0% in 2023 to 50.0% in 2024 and reach 100% in 2025 and 2026. Deep learning serves as a transitional stage, especially in 2023–2024. The evidence indicates a shift from feature engineering and class-balancing optimization toward contextual representations, domain adaptation, prompting, fine-tuning, and assessment of reliability, fairness, and language sensitivity. Greater model capacity, however, does not automatically improve robustness: recent studies highlight prompt sensitivity, demographic differences, and dependence on fine-tuning and domain-specific information. The review recommends controlled comparisons, multilingual evaluation, reproducibility, uncertainty assessment, and resource-aware execution.*

Keywords: *Sentiment Analysis; Systematic Literature Review; Machine Learning; Transformer; Large Language Models*

Abstrak: Perkembangan analisis sentimen bergeser dari pembelajaran mesin berbasis fitur menuju pembelajaran mendalam, enkoder Transformer, dan model bahasa besar (LLM). Perubahan tersebut memengaruhi akurasi model serta pendekatan perumusan data, pembelajaran terawasi, dan evaluasi. Tinjauan literatur sistematis ini memetakan perkembangan pada 2023–2026 berdasarkan 30 artikel penelitian primer yang dapat ditelusuri dan dikelompokkan ke dalam empat kategori metodologis: pembelajaran mesin, pembelajaran mendalam, Transformer, dan LLM. Komposisi kumpulan penelitian berubah dari tahun ke tahun: pembelajaran mesin mencakup 77,8% penelitian pada 2023, turun menjadi 40,0% pada 2024, dan tidak ditemukan pada 2025–2026. Pendekatan Transformer dan LLM meningkat dari 0% pada 2023 menjadi 50,0% pada 2024, kemudian mencapai 100% pada 2025 dan 2026. Pembelajaran mendalam berperan sebagai tahap peralihan, terutama pada 2023–2024. Hasil kajian menunjukkan pergeseran fokus metodologis dari rekayasa fitur dan optimasi keseimbangan kelas menuju representasi kontekstual, adaptasi domain, penyusunan prompt, penyetelan lanjutan (fine-tuning), serta penilaian keandalan, keadilan, dan sensitivitas bahasa. Namun, peningkatan kapasitas model tidak selalu diikuti peningkatan ketahanan. Penelitian terbaru menyoroti sensitivitas terhadap prompt, perbedaan antarkelompok demografis, serta ketergantungan pada penyetelan lanjutan dan informasi khusus domain. Kajian ini merekomendasikan perbandingan terkendali, evaluasi multibahasa, keterulangan penelitian, pengukuran ketidakpastian, dan penerapan yang memperhatikan kebutuhan sumber daya.

Kata kunci: Analisis Sentimen; Tinjauan Literatur Sistematis; Pembelajaran Mesin; Transformer; Model Bahasa Besar.

PENDAHULUAN

Analisis sentimen telah berkembang dari persoalan rekayasa fitur menjadi persoalan pembelajaran representasi dan pemenuhan instruksi. Dalam pendekatan tradisional, algoritma seperti support vector machine (SVM), Naïve Bayes, dan random forest sangat bergantung pada pembobotan token serta representasi yang dirancang secara manual atau bersifat sederhana. Penelitian di Indonesia menunjukkan bahwa teknik tersebut masih kompetitif pada domain yang terdefinisi dengan baik dan himpunan data berukuran sedang, terutama apabila prapemrosesan, penyeimbangan kelas, dan optimasi parameter dilakukan secara cermat (Ardana et al., 2023; Sumantiawan et al., 2023; Dewi dan Mailoa, 2023; Kusnawi et al., 2023; Ningsih et al., 2024).

Pembelajaran mendalam mengurangi ketergantungan pada rekayasa fitur secara manual dengan mempelajari representasi hierarkis melalui arsitektur berulang dan konvolusional. Temuan dari BiLSTM, LSTM–CNN, dan pendekatan hibrida pada 2023–2024 menunjukkan peningkatan kemampuan pemodelan kontekstual untuk teks sekuensial. Meskipun demikian, kinerjanya tetap dipengaruhi oleh himpunan data pelatihan, pemilihan embedding (representasi vektor), dan penyesuaian arsitektur (Barik et al., 2023; Al-Khazaleh et al., 2023; Lestari et al., 2024).

Enkoder Transformer mengubah pendekatan pengembangan model melalui prapelatihan kontekstual dan pembelajaran transfer (transfer learning). Penelitian yang menggunakan IndoBERT untuk analisis sentimen bahasa Indonesia semakin menekankan strategi penyetelan lanjutan (fine-tuning), laju pembelajaran, jumlah siklus pelatihan (epoch), adaptasi domain, dan konfigurasi khusus tugas, bukan sekadar pemilihan algoritma (Sayarizki et al., 2024; Zaidan et al., 2024; Ahmadian et al., 2024; Iskoko et al., 2025; Perwira et al., 2025).

Penelitian lain di Indonesia memberikan bukti pendukung mengenai klasifikasi sentimen berbasis SVM pada topik kesehatan mental, kendaraan listrik, opini publik, dan ulasan aplikasi, serta perbandingan model SVM dan random forest beserta konfigurasi optimasi parameternya (Darmawan et al., 2023; Ernawati et al., 2023; Khan et al., 2024; Sebastian et al., 2024; Simbolon et al., 2023; Winarto et al., 2024).

Kemunculan model bahasa besar (large language models/LLM) memperkenalkan paradigma yang berbeda secara mendasar. Analisis sentimen kini dapat dilakukan melalui pemberian prompt (perintah masukan) tanpa contoh (zero-shot) atau dengan sedikit contoh (few-shot), penyesuaian instruksi, pembuatan data sintetis, dan penerapan model dengan bobot terbuka secara lokal. Namun, bukti terbaru tidak selalu menunjukkan hasil yang lebih baik. Studi komparatif menunjukkan bahwa LLM berkinerja baik pada tugas sentimen sederhana, tetapi dapat berkinerja lebih rendah daripada model khusus pada analisis sentimen terstruktur yang kompleks. Penelitian terkini juga menyoroti variasi hasil akibat prompt, bias sosial, pengaruh bahasa, dan ketidakakuratan kalibrasi antarkelompok (Zhang et al., 2024; Radaideh et al., 2025; Kastrati et al., 2026; Girard et al., 2026).

Kesenjangan penelitian ini bukan terletak pada kurangnya tinjauan analisis sentimen, melainkan pada belum tersedianya sintesis temporal yang terperinci yang menempatkan pembelajaran mesin, pembelajaran mendalam, Transformer, dan LLM sebagai tahapan evolusi metodologis, sekaligus mengkaji perubahan fokus pertanyaan penelitian pada setiap tahap.

Evaluasi yang ada kerap merangkum teknik, penerapan, atau kendala tanpa membedakan pergeseran kerangka metodologis secara kronologis. Penelitian ini mengatasi kekurangan tersebut dengan menelaah kumpulan penelitian periode 2023–2026 melalui empat kategori serta membandingkan keluarga model, strategi pengolahan data, pendekatan pembelajaran terawasi, fokus evaluasi, dan keterbatasan penerapan.

Kompilasi bukti pendukung juga mencakup literatur Indonesia mengenai efektivitas algoritma pemrosesan bahasa alami (NLP), Naïve Bayes, penyetelan hiperparameter IndoBERT, analisis multibahasa IndoBERT–RoBERTa, dan penanganan ketidakseimbangan kelas (Triawan et al., 2025; Iskoko et al., 2025; Febrian et al., 2026; Ramadhan et al., 2026; Iskoko et al., 2025a).

METODE

Desain Penelitian

Penelitian ini menggunakan tinjauan literatur sistematis (systematic literature review/SLR) untuk menemukan, mengevaluasi, mengelompokkan, memeriksa, dan menyintesis penelitian analisis sentimen yang diterbitkan pada 1 Januari 2023 hingga 25 Agustus 2026. Fokus kajian adalah pergeseran dari pembelajaran mesin menuju pembelajaran mendalam, Transformer, dan LLM.

Cakupan Pencarian dan Sumber Informasi

Tinjauan ini mencakup periode 1 Januari 2023 hingga 25 Agustus 2026. Sebanyak 40 artikel kandidat diperoleh dari lembar kerja terdokumentasi dan diverifikasi melalui situs penerbit, catatan DOI, PubMed, Garuda, serta sumber tepercaya lainnya.

Tiga artikel tambahan mengenai pembelajaran mendalam juga dikonfirmasi. Data hasil ekspor resmi dari Scopus, Web of Science, dan Google Scholar tidak dapat diakses. Kumpulan akhir terdiri atas 30 penelitian primer.

Strategi Pencarian

Kata kunci pencarian menggabungkan istilah analisis sentimen dan nama model, meliputi sentiment analysis, SVM, Naïve Bayes, LSTM, CNN, BERT, IndoBERT, RoBERTa, LLM, ChatGPT, GPT, Llama, Mistral, dan Gemini untuk mencakup empat tahap metodologis.

Kriteria Kelayakan

Penelitian disertakan apabila diterbitkan pada 2023–2026, berfokus secara substansial pada analisis sentimen, telah melalui penelaahan sejawat, dapat ditelusuri secara bibliografis, dan dapat dimasukkan ke dalam salah satu kelompok metodologis.

Penyaringan dan Audit Bibliografi

Proses penyaringan mencakup pemeriksaan duplikasi, relevansi, konsistensi metodologis, dan keterlacakan bibliografi. Hanya artikel yang lolos pemeriksaan tersebut yang dimasukkan ke dalam analisis temporal.

Kerangka Klasifikasi Metodologis

Penelitian dikelompokkan ke dalam empat kategori: pembelajaran mesin (machine learning/ML: SVM, Naïve Bayes, dan random forest), pembelajaran mendalam (deep learning/DL: LSTM, BiLSTM, BiGRU, dan CNN), Transformer (BERT, IndoBERT, dan RoBERTa), serta model bahasa besar (LLM: pemberian prompt, penyesuaian instruksi, penyetelan lanjutan, dan pembuatan data sintetis).

Ekstraksi Bukti

Untuk setiap penelitian, dicatat tahun publikasi, metodologi, model, bentuk representasi, mekanisme adaptasi, data dan konteks, fokus evaluasi, kontribusi, serta keterbatasannya.

Sintesis dan Analisis Temporal

Jumlah penelitian per tahun dihitung secara deskriptif untuk membandingkan pembelajaran mesin, pembelajaran mendalam, Transformer,

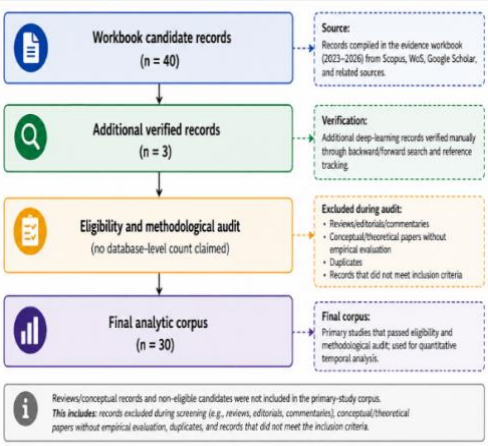
dan LLM. Skor prediksi tidak digabungkan karena adanya perbedaan himpunan data dan kondisi evaluasi. Sintesis berfokus pada optimasi, pembelajaran representasi, pembelajaran transfer, penyetelan lanjutan, pemberian prompt, keadilan, ketidakpastian, kemampuan multibahasa, dan penerapan.

Keterlacakan dan Keterulangan Penelitian

Penilaian ini mendokumentasikan 40 artikel kandidat, proses validasi terfokus, tiga artikel tambahan terkait pembelajaran mendalam, dan 30 penelitian primer dalam kumpulan akhir. Pedoman klasifikasi, kriteria kelayakan, batas waktu, dan aspek sintesis dijelaskan secara eksplisit. Persentase yang disajikan menggambarkan kumpulan penelitian yang ditelaah, bukan proporsi seluruh publikasi yang tersedia.

Evolusi Metodologis

Kumpulan penelitian ini telah diperiksa secara cermat dan mencakup beragam domain sehingga skor prediksi tidak dapat digabungkan secara langsung. Meskipun demikian, sintesis temporal tetap relevan karena menekankan perubahan metodologis, bukan hanya efektivitas secara keseluruhan. Gambar 1 merangkum prosedur evaluasi dan pemeriksaan yang digunakan untuk menentukan kumpulan akhir penelitian.



Gambar 1 Alur Seleksi dan Pemeriksaan Literatur untuk Korpus Analisis Akhir

Gambar 1 menyajikan kumpulan artikel kandidat, proses verifikasi, dan 30 penelitian yang akhirnya memenuhi kriteria kelayakan serta pemeriksaan metodologis. Gambar tersebut tidak menyajikan jumlah hasil pencarian resmi dari Scopus, Web of Science, atau Google Scholar karena data tersebut tidak tersedia.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan tinjauan literatur sistematis (systematic literature review/SLR) serta perkembangan metodologis dari pembelajaran mesin (machine learning/ML) menuju pembelajaran mendalam (deep learning/DL), Transformer, dan model bahasa besar (LLM) selama 2023–2026.

Perkembangan metodologis. Bagian ini membahas perkembangan metode analisis sentimen, mulai dari teknik pembelajaran mesin konvensional hingga pembelajaran mendalam, kerangka berbasis Transformer, dan model bahasa besar (LLM). Pembahasan menyoroti karakteristik setiap tahap metodologis serta perubahan prioritas penelitian seiring dengan perkembangan metode analisis sentimen.

Teknik pembelajaran mesin menganalisis sentimen melalui fitur eksplisit atau representasi ruang vektor; pembelajaran mendalam mempelajari representasi terdistribusi dan sekuensial; arsitektur Transformer menggunakan mekanisme perhatian-diri (self-attention) serta embedding kontekstual hasil prapelatihan; sedangkan LLM memperluas pemodelan bahasa pralatih untuk mendukung kepatuhan terhadap instruksi dan penggunaan beberapa contoh dalam proses inferensi (few-shot).

Pembedaan tersebut penting karena sasaran inovasi metodologis berbeda pada setiap tahap. Penelitian pembelajaran mesin menyempurnakan pengklasifikasi dan alur pengolahan fitur; penelitian pembelajaran mendalam

menyempurnakan arsitektur dan embedding; penelitian Transformer berfokus pada pembelajaran transfer dan penyetelan lanjutan; sedangkan penelitian LLM mengoptimalkan prompt, instruksi, modul penyesuai (adapter), atau protokol evaluasi.

Metode pembelajaran mesin seperti SVM, Naïve Bayes, dan random forest menitikberatkan representasi fitur, parameter pengklasifikasi, serta strategi penyeimbangan kelas. Teknik tersebut tetap relevan apabila himpunan data relatif kecil, terstruktur, atau terbatas pada topik tertentu. Namun, kinerjanya dapat sangat dipengaruhi oleh rekayasa fitur dan prapemrosesan.

Pembelajaran mendalam menggunakan arsitektur jaringan saraf seperti LSTM, BiLSTM, BiGRU, dan CNN untuk menghasilkan representasi yang dipelajari dari data. Arsitektur tersebut mengurangi ketergantungan pada fitur yang dirancang secara manual dan meningkatkan pemodelan sekuens, tetapi kinerjanya tetap dipengaruhi oleh data pelatihan, embedding, pilihan arsitektur, dan kapasitas komputasi.

Arsitektur Transformer memanfaatkan embedding kontekstual hasil prapelatihan dan mekanisme perhatian-diri (self-attention). Dalam analisis sentimen bahasa Indonesia, penelitian yang menggunakan IndoBERT semakin berfokus pada pembelajaran transfer, penyetelan lanjutan, penentuan laju pembelajaran, pengoptimalan jumlah epoch, adaptasi domain, dan pemodelan khusus tugas.

LLM memperluas perkembangan tersebut melalui model bahasa pralatih berskala besar yang dapat digunakan untuk pemberian prompt tanpa contoh (zero-shot) atau dengan sedikit contoh (few-shot), penyempurnaan instruksi, pembuatan data sintetis, serta penyetelan lanjutan. Fokus penelitian pun meluas dari pemilihan model menuju stabilitas, keadilan, ketidakpastian, kemampuan multibahasa, dan penerapan.

Studi terdahulu yang mewakili. Untuk memperlihatkan pergeseran metodologis pada tingkat penelitian, Tabel 1 merangkum penelitian utama dari kumpulan bukti yang dikaji beserta model, konteks, dan kontribusi metodologisnya.

Tabel 1 Ringkasan Studi Terdahulu yang Mewakili

No.	Studi Penelitian	Kerangka Penelitian	Informasi dan Latar Belakang	Kesimpulan Utama
1	Ardana et al. (2023)	SVM, Naïve Bayes	Analisis sentimen Twitter dan pembelajaran daring.	Pembelajaran mesin tetap relevan sebagai tolok ukur awal.
2	Sumantiawan et al. (2023)	SVM + SMOTE-Tomek	Klasifikasi sentimen pada ulasan pelanggan.	Akurasi, presisi, dan skor F1 meningkat.
3	Dewi & Mailoa (2023)	SVM + SMOTE	Opini masyarakat mengenai Mixue di Twitter.	Akurasi dan presisi meningkat, sedangkan sensitivitas (recall) dan skor F1 menurun.
4	Barik et al. (2023)	LSTM dengan optimasi	Analisis sentimen ulasan pelanggan daring.	LSTM yang dioptimalkan meningkatkan klasifikasi.
5	Sayarizki et al. (2024)	IndoBERT	Analisis sentimen mengenai calon presiden di Twitter.	Transformer efektif untuk teks bahasa Indonesia.

No.	Studi Penelitian	Kerangka Penelitian	Informasi dan Latar Belakang	Kesimpulan Utama
6	Zaidan et al. (2024)	IndoBERT dengan penyetelan lanjutan	Analisis sentimen ulasan MyTelkomsel.	Laju pembelajaran dan jumlah epoch memengaruhi kinerja.
7	Zhang et al. (2024)	LLM dengan jumlah contoh terbatas	Evaluasi LLM pada 26 himpunan data dan 13 tantangan.	LLM menunjukkan kinerja yang baik ketika data berlabel terbatas.
8	Perwira et al. (2025)	IndoBERT khusus domain	Ulasan pariwisata untuk analisis sentimen berbasis aspek (ABSA) dalam bahasa Indonesia.	Penyetelan lanjutan khusus domain meningkatkan kesesuaian model.
9	Tarwoto et al. (2025)	IndoBERT	Analisis sentimen tiga kelas aplikasi Mobile JKN.	IndoBERT menunjukkan kinerja yang tangguh.
10	Triawan et al. (2025)	BERT, RoBERTa, ELECTRA	Tinjauan metode sentimen berbasis Transformer.	Model Transformer tangguh, tetapi membutuhkan sumber daya komputasi yang besar.
11	Girard et al. (2026)	24 model LLM dengan bobot terbuka	Evaluasi transkrip percakapan alami.	Keadilan dan kalibrasi masih menjadi tantangan.
12	Febrian et al. (2026)	IndoBERT + Latent-SMOTE + fungsi kerugian berbobot kelas	3.767 ulasan aplikasi Ruangguru.	Fungsi kerugian berbobot kelas menghasilkan nilai Macro-F1 tertinggi.

Tabel 1 menunjukkan pergeseran dari optimasi pengklasifikasi dan penyeimbangan kelas menuju pemodelan sekuens, pembelajaran transfer, penyetelan lanjutan, adaptasi domain, pemberian prompt, pembuatan data sintetis, serta perhatian terhadap keadilan dan ketidakpastian pada LLM.

Komposisi Kumpulan Penelitian dan Perubahan Tahunan

Untuk memberikan dasar kuantitatif bagi analisis temporal,

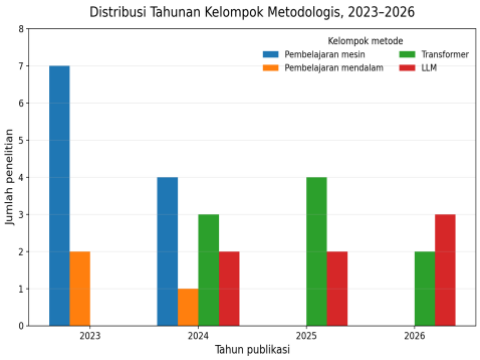
Tabel 2 Distribusi Tahunan Kumpulan Penelitian Berdasarkan Kelompok Metode

Tahun	Pembelajaran Mesin	Pembelajaran Mendalam	Transformer	LLM	Jumlah
2023	7	2	0	0	9
2024	4	1	3	2	10
2025	0	0	4	2	6
2026	0	0	2	3	5
Jumlah	11	3	9	7	30

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 9172 – 9184 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Tabel 2 menunjukkan pergeseran dari pembelajaran mesin dan pembelajaran mendalam menuju metode Transformer dan LLM. Pada 2026, LLM menjadi kelompok dengan jumlah penelitian terbanyak.

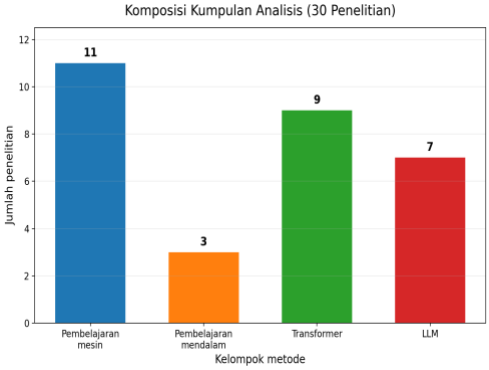
Gambar 2 memperlihatkan penurunan penggunaan metode pembelajaran mesin dan pembelajaran mendalam serta peningkatan penggunaan Transformer dan LLM selama 2023–2026.



Gambar 2 Tren Tahunan Metode Pembelajaran Mesin, Pembelajaran Mendalam, Transformer, dan LLM pada 2023–2026

Gambar 2 menyoroti pergeseran dari pembelajaran mesin dan pembelajaran mendalam menuju Transformer dan LLM, dengan perubahan penting yang terlihat pada 2024 dan 2025.

Gambar 3 merangkum komposisi keseluruhan 30 penelitian dalam kumpulan akhir berdasarkan empat kelompok metodologis.



Gambar 3 Komposisi Kumpulan Analisis yang Terdiri atas 30 Penelitian

Distribusi pada Gambar 3 mencakup 11 penelitian pembelajaran mesin, 3 penelitian pembelajaran mendalam, 9 penelitian Transformer, dan 7 penelitian LLM, sehingga jumlah keseluruhannya mencapai 30 penelitian. Gambar ini melengkapi Gambar 2 dengan memperlihatkan komposisi keseluruhan, bukan perubahan tahunan.

Karakteristik Metodologis pada Empat Tahap

Untuk memperjelas perbedaan utama di antara generasi metodologi, Tabel 3 membandingkan sasaran pengoptimalan, metode yang lazim digunakan, dan keterbatasan utama setiap kategori.

3 Perbandingan Karakteristik Metodologis Antarkelompok

Kelompok Metode	Sasaran Pengoptimalan	Contoh Metode yang Umum Digunakan	Keterbatasan Utama
Pembelajaran Mesin	Pembobotan fitur, parameter pengklasifikasi, dan penyeimbangan kelas	SVM, Naïve Bayes, TF-IDF, SMOTE, dan PSO	Ketergantungan pada fitur dan keterbatasan representasi kontekstual
Pembelajaran Mendalam	Arsitektur, representasi, dan pemodelan sekuens	LSTM, BiLSTM, CNN, model hibrida berulang	Biaya pelatihan tinggi, ketergantungan pada data, dan sensitivitas arsitektur

Kelompok Metode	Sasaran Pengoptimalan	Contoh Metode yang Umum Digunakan	Keterbatasan Utama
Transformer	Penyetelan lanjutan dan adaptasi domain	IndoBERT, pengoptimalan laju pembelajaran dan jumlah epoch, serta ABSA khusus domain	Kebutuhan komputasi tinggi, sensitivitas penyetelan lanjutan, dan pergeseran domain
LLM	Pemberian prompt, penyesuaian instruksi, penyetelan lanjutan, dan generalisasi	Contoh few-shot, data sintetis, dan model dengan bobot terbuka	Variasi prompt, keadilan, ketakterdugaan hasil, dan sensitivitas bahasa

Tabel 3 menunjukkan bahwa pergeseran metodologis berkaitan dengan sasaran pengoptimalan, bukan sekadar nama model. Pembelajaran mesin dan pembelajaran mendalam terutama menyempurnakan proses yang disesuaikan dengan tugas; penelitian Transformer menekankan representasi pralatih dan pembelajaran transfer; sedangkan penelitian LLM memperluas pengoptimalan hingga mencakup pemberian prompt, desain interaksi, dan strategi evaluasi.

Pembelajaran Mendalam sebagai Tahap Transisi

Walaupun pembelajaran mendalam bukan tahap akhir yang dominan dalam kumpulan penelitian, metode ini penting secara analitis karena menggambarkan peralihan dari fitur yang dirancang secara manual menuju representasi kontekstual yang dipelajari. Dalam penelitian tahun 2023, Barik et al. dan Al-Khazaleh et al. menggunakan variasi LSTM/CNN untuk mengenali pola sekuens, sedangkan Lestari et al. (2024) menunjukkan bahwa BiLSTM dengan Word2Vec dapat mengungguli model LSTM satu arah. Model hibrida menggambarkan pergeseran menuju arsitektur kontekstual pralatih, seperti penggabungan IndoBERT dan struktur berulang oleh Ahmadian et al. (2024). Oleh karena itu, pembelajaran mendalam lebih tepat dipandang sebagai tahap penghubung daripada sebagai tujuan akhir evolusi metodologis.

Pemantapan Transformer dalam Analisis Sentimen Bahasa Indonesia

Periode Transformer ditandai oleh pergeseran fokus penelitian dari pemilihan pengklasifikasi menuju penentuan strategi penyetelan lanjutan. Zaidan et al. (2024) menunjukkan bahwa pemilihan laju pembelajaran dan jumlah epoch sangat memengaruhi kinerja IndoBERT; Sayarizki et al. (2024) menerapkan IndoBERT untuk analisis sentimen politik di Indonesia.

Penelitian berikutnya mengarah pada adaptasi domain khusus (Perwira et al., 2025), klasifikasi sentimen tiga kelas (Tarwoto et al., 2025), dan pemodelan yang lebih fleksibel (Iskoko et al., 2025b). Studi yang ditelaah memperkuat kecenderungan ini: Iskoko et al. (2025a) menemukan konfigurasi IndoBERT yang efektif melalui Grid Search, sedangkan Febrian et al. (2026) menunjukkan bahwa optimasi dengan bobot kelas dapat melampaui Latent-SMOTE dalam pengujian IndoBERT yang terkendali. Ramadhan et al. (2026) memperluas evaluasi multibahasa dengan membandingkan IndoBERT dan RoBERTa menggunakan terjemahan NLLB-200.

Secara keseluruhan, tantangan utama bergeser dari pemilihan pengklasifikasi menuju pengelolaan pembelajaran transfer, ketidakseimbangan kelas, bahasa, dan domain.

LLM: Dari Pemilihan Model menuju Keandalan Evaluasi

LLM menawarkan kemampuan yang lebih luas, tetapi juga menghadirkan risiko. LLM menunjukkan kinerja yang baik pada evaluasi sentimen sederhana dan terbatas, tetapi dapat berkinerja lebih rendah daripada model khusus pada skenario yang kompleks.

Penelitian terkini mengkaji pembuatan data sintetis, keadilan, ketidakpastian, kemampuan multibahasa, penerapan model dengan bobot terbuka, serta analisis sentimen berbasis aspek (aspect-based sentiment analysis/ABSA) pada lingkungan dengan sumber daya terbatas. Dengan demikian, fokus penelitian berkembang dari akurasi semata menuju keterkendalian, keterulangan hasil, keadilan, ketangguhan, dan kemampuan beradaptasi terhadap tugas.

Data, Ketidakseimbangan Kelas, dan Evaluasi: Perubahan Setelah 2024

Penelitian awal dalam pembelajaran mesin menangani ketidakseimbangan kelas melalui SMOTE dan SMOTE-Tomek, tetapi hasilnya tidak konsisten. Sumantiawan et al. (2023) melaporkan peningkatan, sedangkan Dewi dan Mailoa (2023) menunjukkan bahwa kenaikan akurasi dan presisi dapat disertai penurunan sensitivitas (recall) dan skor F1. Analisis ulasan Ruangguru pada 2026 menunjukkan bahwa pembobotan kelas menghasilkan nilai Macro-F1 tertinggi dibandingkan dengan Latent-SMOTE.

Oleh karena itu, penyeimbangan kelas tidak selalu meningkatkan kinerja dan perlu dievaluasi menggunakan metrik untuk setiap kelas, bukan hanya akurasi. Pada era LLM, evaluasi juga mencakup variasi prompt, keadilan, kalibrasi, dan ketangguhan multibahasa (Radaideh et al., 2025; Kastrati et al., 2026; Girard et al., 2026).

Sintesis Kesenjangan Penelitian dan Agenda Penelitian

Kesenjangan utama lintas tahap adalah belum tersedianya protokol longitudinal terstandar yang memungkinkan perbandingan adil antartahap metodologis. Nilai akurasi dari himpunan data yang berbeda tidak dapat digabungkan secara seragam. Sebagai gantinya, prioritas penelitian dapat dibandingkan: pembelajaran mesin berfokus pada pengklasifikasi dan fitur, pembelajaran mendalam pada representasi yang dipelajari, Transformer pada pembelajaran transfer dan penyempurnaan model, serta LLM pada interaksi dan keandalan evaluasi.

Penelitian selanjutnya perlu menggunakan tolok ukur terstandar, pembagian data yang konsisten, metrik khusus kelas, pengulangan dengan benih acak (random seed) yang berbeda, validasi statistik, dokumentasi biaya dan latensi, serta evaluasi ketangguhan. Penelitian analisis sentimen bahasa Indonesia perlu menggabungkan evaluasi multibahasa dan teks campuran bahasa (code-mixed) dengan adaptasi domain, pengukuran ketidakpastian, serta inferensi lokal yang memperhatikan privasi.

Implikasi Praktis

Untuk himpunan data yang kecil dan terstruktur dengan baik, pembelajaran mesin tradisional tetap menjadi tolok ukur yang sahih dan hemat biaya. Jika tersedia data pelatihan dalam jumlah besar, pembelajaran mendalam dapat menjadi pilihan perantara yang bermanfaat untuk konteks sekuensial. Saat ini, enkoder Transformer merupakan pilihan yang andal untuk penyetelan lanjutan pada domain khusus bahasa Indonesia, selama tersedia data berlabel dan kapasitas komputasi yang memadai. LLM menarik untuk digunakan ketika label data terbatas, tugas beragam, kebutuhan multibahasa tinggi, atau prototipe perlu dikembangkan dengan cepat. Namun, LLM tidak seharusnya dipilih semata-mata karena merupakan teknologi terbaru. Keputusan penerapan perlu mempertimbangkan sensitivitas data, biaya inferensi, keandalan, ketersediaan model, dan keadilan selain akurasi

prediksi.

SIMPULAN

Tinjauan sistematis ini menemukan perkembangan metodologis yang jelas pada kumpulan penelitian periode 2023–2026, yaitu dari pembelajaran mesin menuju pembelajaran mendalam, kemudian Transformer, dan akhirnya LLM. Pada 2023, penelitian didominasi pembelajaran mesin dengan pembelajaran mendalam sebagai tahap penghubung. Metode Transformer mendominasi pada 2025, sedangkan model bahasa besar menjadi kelompok terbesar pada 2026. Perkembangan tersebut menunjukkan pergeseran fokus dari pengoptimalan pengklasifikasi dan fitur menuju pembelajaran representasi, pembelajaran transfer dan penyetelan lanjutan, hingga pemberian prompt, generalisasi, keadilan, dan keandalan.

Temuan utama ini menegaskan bahwa kemajuan metodologis tidak seharusnya dipandang sebagai persaingan sederhana untuk memperoleh akurasi tertinggi. Setiap tahap mengatasi keterbatasan tertentu sekaligus memunculkan tantangan baru. Oleh karena itu, arah penelitian setelah 2026 perlu berfokus pada perbandingan lintas generasi yang terkendali dengan mempertimbangkan keragaman bahasa, ketidakseimbangan kelas, validasi statistik, efisiensi, ketidakpastian, dan keadilan.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Universitas AMIKOM Purwokerto atas dukungan akademik, fasilitas, dan dukungan kelembagaan yang diberikan selama penyusunan serta pelaksanaan penelitian ini.

DAFTAR PUSTAKA

Ahmadian, M. S., et al. (2024). Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian. *Applied Computational Intelligence and Soft Computing*. https://doi.org/10.1155/2024/2826773.

Al-Khazaleh, M., Alian, M., Biltawi, M., & Al-Hazaimeh, B. (2023). Sentiment Analysis for People's Opinions about COVID-19 Using LSTM and CNN Models. *International Journal of Online and Biomedical Engineering*, 19(1), 135–154. https://doi.org/10.3991/ijoe.v19i01.35645.

Ardana, Y., Kambau, R. A., & Mustikasari. (2023). Comparison of Sentiment Analysis of Support Vector Machine and Naïve Bayes Algorithms on Public Response About Online Learning During the COVID-19 Pandemic. *AGENTS: Journal of Artificial Intelligence and Data Science*, 3(1). https://doi.org/10.24252/jagti.v3i1.46.

Barik, K., Misra, S., Ray, A. K., & Bokolo, A. (2023). LSTM-DGWO-Based Sentiment Analysis Framework for Analyzing Online Customer Reviews. *Journal of Environmental and Public Health*, 2023, 6348831. https://doi.org/10.1155/2023/6348831.

Darmawan, F., Joe, M., Kurniawan, Y. I., & Afuan, L. (2023). Analisis Sentimen Kemungkinan Depresi dan Kecemasan pada Twitter Menggunakan Support Vector Machine. *Jurnal Eksplora Informatika*, 13(1), 24–36. https://doi.org/10.30864/eksplora.v13i1.854.

Dewi, T. A., & Mailoa, E. (2023). Perbandingan Implementasi Metode SMOTE pada Algoritma Support Vector Machine (SVM) dalam Analisis Sentimen Opini Masyarakat tentang Mixue. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 4(3), 849–855. https://doi.org/10.35870/jimik.v4i3.289

Ernawati, N. W., Satya Kumara, I. N., & Setiawan, W. (2023). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes pada Analisis Sentimen Kendaraan Listrik. *Jurnal SPEKTRUM*, 10(3), 106–114. https://doi.org/10.24843/SPEKTRUM.2023.v10.i03.p12.

Febrian, A. T., Purwadi, P., & Kuncoro, A. P. (2026). Comparative Analysis of Data-Level and Cost-Sensitive Learning in IndoBERT-Based Sentiment Analysis of Ruangguru App Reviews. *Jambura Journal of Electrical and Electronics Engineering*, 8(2), 250–259. https://doi.org/10.37905/jjeee.v8i2.39885

Fehle, J., Kruschwitz, U., Hellwig, N. C., & Wolff, C. (2026). Leveraging Fine-Tuning of Large Language Models for Aspect-Based Sentiment Analysis in Resource-Scarce Environments. *Knowledge-Based Systems*, 336, 115277. https://doi.org/10.1016/j.knosys.2026.115277

Girard, J. M., Jun, D., Ong, D. C., Liebenthal, E., & Baker, J. T. (2026). Sentiment Analysis of Naturalistic Speech Using Open-Weight Large Language Models. *Affective Science*, 7(2), 166–180. https://doi.org/10.1007/s42761-025-00352-7.

Hellwig, N. C., Fehle, J., & Wolff, C. (2024). Exploring Large Language Models for the Generation of Synthetic Training Samples. *Expert Systems with Applications*, 125514. https://doi.org/10.1016/j.eswa.2024.125514.

Herrera-Poyatos, D., Peláez-González, C., Zuheros, C., Herrera-Poyatos, A., Tejedor, V., Herrera, F., & Montes, R. (2025). An Overview of Model Uncertainty and Variability in LLM-Based Sentiment Analysis: Challenges, Mitigation Strategies and the Role of Explainability. *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2025.1609097.

Iskoko, A., Tahyudin, I., & Purwadi, P. (2025a). Hyperparameter Optimization of IndoBERT Using Grid Search, Random Search, and Bayesian Optimization in Sentiment Analysis of E-Government Application Reviews. *Jurnal Teknik Informatika (JUTIF)*, 6(5), 3430–3444. https://doi.org/10.52436/1.jutif.2025.6.5.4897.

Iskoko, A., Tahyudin, I., & Purwadi, P. (2025b). Analisis Sentimen Ulasan Pengguna Aplikasi Sinaga Mobile pada Google Play Store Menggunakan Algoritma Naive Bayes. *Jurnal Pendidikan dan Teknologi Indonesia*, 5(6), 1635–1645. https://doi.org/10.52436/1.jpti.843.

Kastrati, M., et al. (2026). Do Language Families Matter? Evaluating LLMs for Sentiment Analysis through a Language-Family Lens. *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2026.1854873.

Khan, T. A., Sadiq, R., Shahid, Z., Alam, M. M., & Su'ud, M. B. M. (2024).

Sentiment Analysis Using Support Vector Machine and Random Forest. *Journal of Informatics and Web Engineering*, 3(1). https://doi.org/10.33093/jiwe.2024.3.1.5.

Kusnawi, K., Rahardi, M., & Pandiangan, V. D. (2023). Sentiment Analysis of Neobank Digital Banking Using Support Vector Machine Algorithm in Indonesia. *JOIV: International Journal on Informatics Visualization*, 7(2), 377–383. https://doi.org/10.30630/joiv.7.2.1652.

Lestari, V. B., Utami, E., & Hanafi. (2024). Combining Bi-LSTM and Word2vec Embedding for Sentiment Analysis Models of Application User Reviews. *The Indonesian Journal of Computer Science*, 13(1). https://doi.org/10.33022/ijcs.v13i1.3647.

Ningsih, M. R., Unjung, J., Pertiwi, D. A. A., Prasetiyo, B., & Muslim, M. A. (2024). Optimized Support Vector Machine with Particle Swarm Optimization to Improve the Accuracy Amazon Sentiment Analysis Classification. *Kinetik*, 9(1), 101–108. https://doi.org/10.22219/kinetik.v9i1.1888.

Perwira, R. I., Permadi, V. A., Purnamasari, D. I., & Agusdin, R. P. (2025). Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content. *Journal of Information Systems Engineering and Business Intelligence*, 11(1), 30–40. https://doi.org/10.20473/jisebi.11.1.30-40.

Radaideh, M. I., et al. (2025). Fairness and Social Bias Quantification in Large Language Models for Sentiment Analysis. *Knowledge-Based Systems*, 113569. https://doi.org/10.1016/j.knosys.2025.113569.

Ramadhan, F. E., Karyono, G., & Purwadi, P. (2026). Multilingual Sentiment Detection: IndoBERT Vs. RoBERTa Using NLLB-200 Translations. *RABIT: Jurnal Teknologi dan Sistem Informasi Univrab*, 11(2), 4855–4865.

Sayarizki, P., Hasmawati, & Nurrahmi, H. (2024). Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates. *Indonesian Journal on Computing*, 9(2), 61–72. https://doi.org/10.34818/indojc.2024.9.2.934.

Sebastian, D. F., Sulistiani, H., & Isnain, A. R. (2024). Sentiment Analysis of Public Opinion on the Right of Inquiry in Indonesia in 2024 Using the Support Vector Machine (SVM) Method. *Jurnal Teknik Informatika (JUTIF)*, 5(4), 1025–1034. https://doi.org/10.52436/1.jutif.2024.5.4.1968.

Simbolon, O. S., Manullang, M. E., Alvarez, S., Brutu, L. F. M., & Indra, E. (2023). Sentiment Analysis of MyPertamina Application Using Support Vector Machine and Naïve Bayes Algorithms. *JUSIKOM: Jurnal Sistem Informasi dan Ilmu Komputer*, 7(1), 177–189. https://doi.org/10.34012/jurnalsistem informasidanilmukomputer.v7i1.4078.

Sumantiawan, D. I., Suseno, J. E., & Syafei, W. A. (2023). Sentiment Analysis of Customer Reviews Using Support Vector Machine and SMOTE-Tomek Links for Identify Customer Satisfaction. *Jurnal Sistem Informasi Bisnis*, 13(1), 1–9. https://doi.org/10.21456/vol13iss1pp1-9.

Tarwoto, Nugroho, Azka, & Graha, W. S. R. (2025). Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT. *Jurnal JTIK*, 9(2), 495–505. https://doi.org/10.35870/jtik.v9i2.3340.

Triawan, P., Tahyudin, I., & Purwadi, P. (2025). Impact of NLP Algorithms on Sentiment Analysis Efficiency and Accuracy. *Journal of Information Systems and Informatics*, 7(3), 2684–2709. https://doi.org/10.51519/journalisi.v.

Winarto, W., Musdar, I. A., & Hasniati. (2024). Sentiment Analysis of 2024 Presidential Candidate Using the Support Vector Machine Algorithm on Twitter. *KHARISMA Tech*, 19(1), 86–98. https://doi.org/10.55645/kharismatech.v19i1.464.

Wojdecka, A., Gromadziński, J., & Walczak, K. (2025). Effectiveness of Large Language Models and Software Libraries in Sentiment Analysis. *Applied Computer Science*, 21(2), 128–138. https://doi.org/10.35784/acs_6936.

Zaidan, M. N., Sibaroni, Y., & Prasetyowiyati, S. S. (2024). Learning Rate and Epoch Optimization in the Fine-Tuning Process for IndoBERT's Performance on Sentiment Analysis of MyTelkomsel App Reviews. *Jurnal Teknik Informatika (JUTIF)*, 5(5), 1443–1450. https://doi.org/10.52436/1.jutif.2024.5.5.2396.

Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2024). Sentiment Analysis in the Era of Large Language Models: A Reality Check. *Findings of NAACL 2024*. https://doi.org/10.18653/v1/2024.findings-naacl.246.