

Meta-Analisis Kinerja Metode *Natural Language Processing* dalam Analisis Sentimen Teks Berbahasa Indonesia

Dia Mia Zara¹, Elsa Sabrina², Fitri Muthmainnah³, Gilbert Sianipar⁴, Mery Christy Sinaga⁵,
Najwa Shafira Suha⁶, Yusuf Pratama Perangin-angin⁷.

Program Studi Pendidikan Teknologi Informatika dan Komputer
Universitas Negeri Medan

Email: diamiazara@gmail.com, elsasabrina@unimed.ac.id, fitrimuthmainnah13@gmail.com,
sianipargilbert4@gmail.com, merysng30@gmail.com, suhanajwa06@gmail.com,
yusufpratamaperanginangin56@gmail.com.

Abstrak

Penelitian ini mengkaji penerapan metode *Natural Language Processing* (NLP) dalam analisis sentimen teks berbahasa Indonesia melalui sintesis terhadap 25 artikel yang terbit pada rentang 2021–2026. Artikel dikelompokkan berdasarkan metode, sumber data, dan ketersediaan metrik evaluasi. Sebanyak 20 artikel melaporkan metrik kinerja yang dapat diekstraksi, sedangkan lima artikel hanya menyediakan hasil deskriptif seperti distribusi sentimen sehingga tidak dimasukkan dalam perbandingan metrik. Metode yang ditemukan mencakup algoritma machine learning, model deep learning, transformer, serta pendekatan berbasis leksikon/aturan. Nilai accuracy yang dilaporkan beragam, tetapi tidak dapat langsung dipakai untuk menentukan metode terbaik secara universal karena dataset, pelabelan, pembagian data, dan prosedur evaluasi berbeda antarpencapaian. Oleh sebab itu, temuan ditafsirkan secara deskriptif dan komparasi kuantitatif dibatasi pada metrik yang dilaporkan oleh studi sumber. Kajian ini menegaskan pentingnya pelaporan prosedur evaluasi yang transparan dan penggunaan dataset serta protokol pengujian yang jelas agar hasil antarpencapaian lebih mudah dibandingkan.

Kata kunci: *Natural Language Processing*, analisis sentimen, bahasa Indonesia, accuracy, evaluasi model, sintesis penelitian.

Pendahuluan

Perkembangan *Artificial Intelligence* (AI) telah mendorong pemanfaatan teknologi di berbagai bidang, salah satunya pengolahan bahasa manusia melalui *Natural Language Processing* (NLP) (Hawari, 2025). NLP memungkinkan komputer memproses dan menganalisis bahasa manusia dalam bentuk teks, termasuk teks berbahasa Indonesia (Tarumingkeng, 2024). Seiring perkembangannya, NLP telah menghasilkan berbagai metode yang dapat digunakan untuk

mengolah dan memahami informasi dalam teks (M. Amien, 2022). Dalam penerapannya pada Bahasa Indonesia, NLP menghadapi sejumlah tantangan yang berkaitan dengan karakteristik bahasa, variasi penggunaan bahasa, serta ketersediaan sumber daya bahasa (M. Amien & Gunawan, 2024). Meskipun demikian, berbagai tantangan tersebut tidak menghambat perkembangan penerapan NLP, salah satunya pada analisis sentimen yang terus mengalami perkembangan dalam pengolahan teks. Hal ini menunjukkan bahwa NLP memiliki peluang yang sangat luas untuk dimanfaatkan dalam menganalisis berbagai teks berbahasa Indonesia.

Analisis sentimen banyak digunakan untuk mengidentifikasi kecenderungan opini yang terdapat dalam suatu teks (Febriani & Sari, 2026). Analisis sentimen pada data berbahasa Indonesia telah dilakukan dengan menerapkan berbagai metode NLP, seperti *Support Vector Machine*, *Naive Bayes*, LSTM, dan IndoBERT (Irawanto, 2026; Nurwanda et al., 2024; Shobrina Fathoniah, 2022; Zahara et al., 2025). Sumber data yang digunakan mencakup media sosial, komentar, serta ulasan pengguna pada berbagai platform digital. Keragaman tersebut mencerminkan perkembangan yang pesat dalam penerapan metode NLP untuk analisis sentimen Bahasa Indonesia. Namun, banyaknya penelitian dengan pendekatan dan karakteristik data yang berbeda menyebabkan hasil penelitian mengenai kinerja metode NLP tersebar di berbagai studi dan sulit dipahami secara menyeluruh (Azzahra & Printo, 2025). Kondisi inilah yang mendorong perlunya suatu kajian yang mampu menelaah berbagai penelitian tersebut secara lebih terstruktur.

Kebutuhan untuk memahami penelitian terdahulu secara lebih komprehensif dapat dipenuhi melalui pendekatan meta-analisis. Meta-analisis memungkinkan hasil dari sejumlah penelitian yang relevan dikumpulkan dan disintesis berdasarkan kriteria yang telah ditetapkan (Awali, 2025). Pendekatan tersebut digunakan untuk mengkaji kinerja berbagai metode NLP yang telah diterapkan dalam analisis sentimen teks berbahasa Indonesia. Penelitian yang dikaji dapat mencakup berbagai metode dan sumber data selama memenuhi kriteria yang telah ditentukan. Melalui proses sintesis tersebut, diharapkan diperoleh gambaran yang lebih terstruktur mengenai kinerja metode NLP dalam analisis sentimen Bahasa Indonesia. Penelitian ini bertujuan melakukan meta-analisis kinerja metode NLP dalam analisis sentimen teks berbahasa Indonesia sebagai bahan referensi yang bermanfaat bagi penelitian selanjutnya.

Tinjauan Pustaka

2.1 Natural Language Processing

Natural Language Processing (NLP) adalah salah satu cabang dari kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia (Tarumingkeng, 2024). NLP digunakan untuk membantu komputer memahami, menelaah, dan mengelola informasi yang disampaikan dalam bentuk bahasa manusia (Tarumingkeng, 2024). Teknologi ini memungkinkan data teks diolah sehingga muatan informasinya dapat dianalisis secara komputasional (Nurwanda et al., 2024).

Dalam penerapannya, NLP berfungsi untuk memproses data teks sebelum masuk ke tahap analisis yang lebih mendalam (Widiantoro et al., 2024). Pemrosesan tersebut umumnya mencakup sejumlah langkah, yaitu data *cleaning*, *case folding*, *tokenisasi*, penghapusan *stopword*, dan *stemming*. Tahapan ini bertujuan menyiapkan data teks agar memiliki struktur yang lebih rapi sehingga siap digunakan dalam proses analisis (Nurwanda et al., 2024).

Selain itu, NLP berperan penting dalam analisis sentimen sebab teknologi ini memungkinkan teks berbahasa manusia diolah dan dianalisis secara otomatis. Dengan memanfaatkan NLP, informasi mengenai pandangan maupun tanggapan yang terkandung dalam sekumpulan teks dapat diperoleh dengan lebih mudah (Hadi et al., 2025).

2.2 Analisis Sentimen

Analisis sentimen adalah proses komputasi yang bertujuan mengidentifikasi sentimen atau emosi yang terkandung dalam sebuah teks (Franciska et al., 2025). Pendekatan ini dapat dimanfaatkan untuk menelaah opini maupun tanggapan yang diungkapkan melalui teks mengenai suatu objek, produk, layanan, ataupun topik tertentu (Hadi et al., 2025).

Dalam analisis sentimen, teks pada umumnya dikelompokkan ke dalam beberapa kategori sentimen. Salah satu pembagian yang lazim digunakan dalam penelitian adalah sentimen positif, netral, dan negatif. Pengelompokan tersebut berguna untuk melihat kecenderungan sentimen yang muncul pada data teks yang dianalisis (Hadi et al., 2025).

Adapun dalam pelaksanaannya, data teks perlu melalui serangkaian tahap pemrosesan sebelum diklasifikasikan. Pada penelitian (Aryanti & Suria, 2025) data teks terlebih dahulu menjalani tahap *preprocessing*, lalu diberi label untuk menentukan apakah suatu teks memuat sentimen positif, negatif, atau netral. Dengan begitu, analisis sentimen dapat digunakan untuk mengenali kecenderungan sentimen dalam sekumpulan data teks (Aryanti & Suria, 2025).

2.3 Analisis Sentimen pada Teks Bahasa Indonesia

Analisis sentimen pada teks berbahasa Indonesia memiliki ciri karakteristik tersendiri karena data teksnya kerap memuat kata tidak baku dan bentuk penulisan yang bervariasi. (Amien et al., 2021) meneliti klasifikasi aspek pada ulasan pelanggan berbahasa Indonesia, menjelaskan bahwa ulasan pelanggan dapat mengandung kata-kata yang tidak formal. Keadaan ini menuntut

teks untuk melalui tahap prapemrosesan terlebih dahulu agar dapat dianalisis dengan lebih terstruktur.

Dalam pengolahan teks bahasa Indonesia, tahap prapemrosesan yang lazim digunakan antara lain *case folding*, *stopword removal*, dan *stemming*. *Case folding* mengubah seluruh huruf menjadi huruf kecil sekaligus membuang tanda baca. *Stopword removal* berfungsi menyingkirkan kata-kata umum yang dinilai kurang berperan dalam pengolahan teks, sementara *stemming* mengembalikan kata berimbuhan ke bentuk dasarnya. (Amien et al., 2021) menerapkan ketiga tahap tersebut pada dataset ulasan pelanggan berbahasa Indonesia dari tiga domain, yakni restoran, hotel, dan *e-commerce*, dalam tugas klasifikasi aspek.

Penerapan prapemrosesan ini dapat dilihat pada contoh teks dalam penelitian tersebut. Kata "ENAK" menjadi "enak" melalui *case folding*, sedangkan kata "suasananya", "menenangkan", dan "disajikan" berubah menjadi "suasana", "tenang", dan "saji" setelah melalui *stemming*. Contoh ini memperlihatkan bahwa pengolahan teks bahasa Indonesia perlu memperhatikan bentuk kata dan penggunaan imbuhan, sehingga teks dapat direpresentasikan secara lebih seragam pada tahap analisis berikutnya (Amien et al., 2021).

Hal serupa juga ditemukan pada penelitian (Aryanti & Suria, 2025) yang menganalisis tweet berbahasa Indonesia mengenai pemutusan hubungan kerja. Data tweet tersebut terlebih dahulu melalui tahapan *case folding*, *cleaning*, tokenisasi, normalisasi, penghapusan *stopword*, dan *stemming* sebelum diklasifikasikan. Pada tahap normalisasi, penelitian tersebut mengganti kata tidak baku dengan padanan bakunya, misalnya "yg" menjadi "yang", sedangkan *stemming* menyederhanakan setiap kata menjadi kata dasarnya. Dengan demikian, karakteristik penggunaan bahasa Indonesia dalam data teks menjadi aspek yang perlu diperhatikan dalam penerapan analisis sentimen.

2.4 Metode NLP Dalam Analisis Sentimen

NLP menyediakan berbagai pendekatan untuk mengolah dan mengklasifikasikan data teks (Tarumingkeng, 2024). Dalam penerapan analisis sentimen, pendekatan yang ditemukan dalam artikel-artikel kajian ini mencakup tujuh metode atau alat utama: *Support Vector Machine* (SVM), *Naive Bayes*, *Long Short-Term Memory* (LSTM), IndoBERT, BERT, VADER, dan *TextBlob*. Ketujuhnya tidak berada pada tingkat kategori yang sepenuhnya sama: SVM dan *Naive Bayes* merupakan algoritma *machine learning*, LSTM merupakan model *deep learning*, BERT dan IndoBERT merupakan model bahasa berbasis transformer, sedangkan VADER dan *TextBlob* pada penggunaan standarnya menyediakan analisis sentimen berbasis leksikon/aturan. Pemilihan ini mengikuti artikel-artikel yang tersedia dan tetap memperhatikan

apakah hasil evaluasi kinerja dapat dibandingkan secara kuantitatif (Ningsih et al., 2024; Rahmawati & Aini, 2025; Setiadi & Sulisty, 2025).

Support Vector Machine (SVM) merupakan salah satu metode *machine learning* yang dapat digunakan untuk melakukan klasifikasi data (Asana & Yanti, 2023). Pada pengolahan teks, data yang berbentuk kalimat atau dokumen perlu diubah terlebih dahulu menjadi bentuk numerik agar dapat diproses oleh algoritma (Prasetya et al., 2024). Representasi teks tersebut dapat dilakukan menggunakan teknik pembobotan kata seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) (Emarapenta et al., 2024). Setelah data direpresentasikan dalam bentuk fitur, SVM melakukan proses klasifikasi dengan menentukan batas pemisah antar kelompok data. Dalam analisis sentimen, batas tersebut digunakan untuk membedakan teks berdasarkan kategori sentimen yang telah ditentukan. Penerapan SVM dalam analisis sentimen berbahasa Indonesia ditemukan pada penelitian yang menggunakan data media sosial maupun data ulasan pengguna (Aryanti & Suria, 2025; Irawanto, 2026).

Naive Bayes merupakan metode klasifikasi yang menggunakan pendekatan probabilistik dalam menentukan kategori suatu data (Pebdika et al., 2023). Pada klasifikasi teks, metode ini memanfaatkan informasi yang terdapat pada fitur atau kata dalam suatu dokumen untuk memperkirakan kemungkinan dokumen tersebut termasuk ke dalam kelas tertentu. Penerapan *Naive Bayes* dalam analisis sentimen telah dilakukan pada berbagai jenis data berbahasa Indonesia, termasuk ulasan pengguna pada platform digital dan data dari media sosial (Siregar et al., 2025).

Long Short Term Memory (LSTM) merupakan salah satu jenis model deep learning yang dikembangkan dari *Recurrent Neural Network* (RNN) (Adha et al., 2025). Model ini dirancang untuk menangani data yang memiliki hubungan secara berurutan. Karakteristik tersebut sesuai dengan data teks karena susunan kata dalam sebuah kalimat dapat memberikan informasi terhadap makna yang ingin disampaikan. LSTM memiliki mekanisme penyimpanan informasi yang memungkinkan model mempertahankan informasi tertentu dari bagian sebelumnya selama proses pengolahan data (Saputra et al., 2024). Dalam analisis sentimen, LSTM digunakan untuk mempelajari pola yang terdapat pada rangkaian teks dan selanjutnya menentukan kategori sentimen dari data tersebut. Penggunaan LSTM pada analisis sentimen berbahasa Indonesia dapat ditemukan pada penelitian (Fathoniah & Rozikin, 2022) serta penelitian (Rahmawati & Aini, 2025).

IndoBERT merupakan model pemrosesan bahasa berbasis transformer yang dikembangkan untuk menangani teks berbahasa Indonesia (Apriansyah et al., 2025). Model ini merupakan pengembangan dari pendekatan *Bidirectional Encoder Representations from Transformers*

(BERT) yang menggunakan representasi bahasa berdasarkan konteks. Dengan pendekatan tersebut, informasi suatu kata dapat dipahami dengan mempertimbangkan kata-kata lain yang terdapat di dalam kalimat. Dalam analisis sentimen, IndoBERT dapat digunakan untuk melakukan klasifikasi terhadap teks berdasarkan konteks bahasa yang terdapat dalam data. Penggunaan IndoBERT dalam analisis sentimen berbahasa Indonesia dapat ditemukan pada penelitian (Azzahra & Printo, 2025) dan (Aryanti & Suria, 2025).

VADER (*Valence Aware Dictionary and Sentiment Reasoner*) merupakan alat analisis sentimen berbasis leksikon dan aturan (Leonardi et al., 2025). VADER menghitung skor sentimen berdasarkan muatan emosional kata serta aturan yang berkaitan dengan penekanan bahasa, tanda baca, dan huruf capital (Wahyuni & Februariyanti, 2026). Dalam artikel yang dikaji (Meliala et al., 2025), VADER digunakan untuk menganalisis cuitan X bertagat #KaburAjaDulu. Karena VADER pada penggunaan standar terutama dikembangkan untuk bahasa Inggris, penerapannya pada teks Indonesia perlu memperhatikan proses penerjemahan atau adaptasi leksikon dan keterbatasan konteks Bahasa.

TextBlob merupakan pustaka Python yang menyediakan fungsi pemrosesan teks, termasuk analisis sentimen berbasis leksikon pada penggunaan standarnya (Wijaya et al., 2025). Hasil analisis sentimen *TextBlob* umumnya berupa polarity dan subjectivity. Dalam penelitian (Meliala et al., 2025), *TextBlob* digunakan untuk menganalisis komentar YouTube tentang pelantikan menteri dan cuitan X mengenai #KaburAjaDulu. Keluaran berupa distribusi sentimen tidak sama dengan metrik evaluasi seperti *accuracy*; karena itu, hasil *TextBlob* yang hanya melaporkan proporsi kelas digunakan sebagai temuan deskriptif, bukan bukti langsung tentang akurasi model.

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model bahasa berbasis transformer yang membangun representasi kata berdasarkan konteks dua arah. BERT dapat disesuaikan untuk klasifikasi sentimen melalui proses fine-tuning (Kamal & Astri, 2025). Berbeda dari IndoBERT yang dikembangkan khusus untuk bahasa Indonesia, BERT pada tiap studi perlu diperiksa berdasarkan model prelatih, bahasa data, serta prosedur adaptasi yang digunakan. Artikel (Meliala et al., 2025) #KaburAjaDulu membandingkan BERT dengan VADER dan *TextBlob*, tetapi melaporkan distribusi kelas sentimen dan bukan metrik evaluasi akurasi.

Ketujuh pendekatan atau model tersebut digunakan sebagai ruang lingkup kajian yang diperluas. SVM dan *Naive Bayes* mewakili *machine learning*; LSTM mewakili *deep learning*; BERT dan IndoBERT mewakili model berbasis transformer; sedangkan VADER dan *TextBlob* mewakili pendekatan analisis sentimen berbasis leksikon/aturan. Pengelompokan ini

membantu proses identifikasi dan pengodean penelitian terdahulu. Karena cara kerja dan jenis keluaran tiap pendekatan berbeda, tidak semua hasil dapat langsung digabungkan ke dalam satu perbandingan kuantitatif (Ningsih et al., 2024; Wahyudi et al., 2025).

2.5 Kinerja Metode Analisis Sentimen

Evaluasi kinerja diperlukan untuk mengetahui kemampuan suatu model dalam melakukan klasifikasi terhadap data sentimen. Penelitian analisis sentimen yang termasuk dalam kajian menggunakan sejumlah ukuran evaluasi untuk menggambarkan ketepatan model dalam menghasilkan klasifikasi. Beberapa ukuran yang ditemukan dalam jurnal yang dianalisis adalah *accuracy*, *precision*, *recall*, dan *F1-score* (Ningsih et al., 2024; Rahmawati & Aini, 2025).

Accuracy menggambarkan tingkat ketepatan model berdasarkan perbandingan antara jumlah prediksi yang benar dan keseluruhan data yang diuji. Ukuran tersebut digunakan dalam berbagai penelitian analisis sentimen, salah satunya penelitian (Agustina et al., 2022) pada ulasan Shopee di *Google Play Store*. *Accuracy* juga digunakan oleh (Ningsih et al., 2024) ketika mengevaluasi model SYM dan *Naive Bayes* pada data Twitter mengenai penggunaan mobil listrik.

Precision digunakan untuk melihat tingkat ketepatan prediksi model ketika menetapkan suatu data ke dalam kelas tertentu. Dengan ukuran ini, evaluasi tidak hanya melihat jumlah prediksi yang benar secara keseluruhan, tetapi juga memperhatikan ketepatan hasil prediksi pada kelas yang dituju. (Ainunnisa & Sulastri, 2023) menggunakan *precision* sebagai salah satu ukuran dalam mengevaluasi SVM dan *Naive Bayes* pada analisis sentimen ulasan TikTok.

Sementara itu, *recall* digunakan untuk menggambarkan kemampuan model dalam mengenali data yang memang termasuk dalam suatu kategori. Ukuran tersebut memberikan informasi mengenai seberapa banyak data dari suatu kelas yang berhasil ditemukan oleh model. Penggunaan *recall* bersama metrik evaluasi lainnya dapat ditemukan dalam penelitian (Rahmawati & Aini, 2025) pada klasifikasi sentimen komentar YouTube berbasis LSTM.

F1-score digunakan untuk memperoleh ukuran yang mempertimbangkan *precision* dan *recall* secara bersamaan. Metrik ini menjadi salah satu ukuran yang dilaporkan dalam penelitian analisis sentimen karena dapat memberikan gambaran mengenai keseimbangan antara ketepatan prediksi dan kemampuan model dalam menemukan data dari suatu kelas. (Rahmawati & Aini, 2025), misalnya, menggunakan *F1-score* bersama *accuracy*, *precision*, dan *recall* dalam mengevaluasi model LSTM.

Metrik evaluasi yang digunakan pada penelitian terdahulu tidak selalu sama pada setiap studi. Beberapa penelitian lebih menekankan *accuracy* sebagai indikator utama, sedangkan penelitian

lainnya melaporkan beberapa metrik secara bersamaan. Perbedaan tersebut terlihat pada penelitian (Agustina et al., 2022; Ningsih et al., 2024; Rahmawati & Aini, 2025) yang memiliki rancangan evaluasi dengan ukuran kinerja yang berbeda.

Dalam meta-analisis, perbedaan pelaporan metrik tersebut perlu diperhatikan pada tahap ekstraksi data. Setiap ukuran kinerja dicatat berdasarkan informasi yang benar-benar dilaporkan oleh penelitian sumber sehingga tidak terdapat penambahan nilai yang tidak tersedia dalam studi asli. Dengan pendekatan tersebut, data kinerja yang diperoleh tetap sesuai dengan karakteristik penelitian yang menjadi sumber meta-analisis.

Selain metrik evaluasi, karakteristik proses pengolahan data juga menjadi informasi yang relevan dalam membaca hasil penelitian. Tahapan seperti *case folding*, *cleaning*, tokenisasi, normalisasi, *stopword removal*, dan *stemming* ditemukan pada beberapa penelitian yang menggunakan data teks berbahasa Indonesia (Ainunnisa & Sulastri, 2023; Aryanti & Suria, 2025). Perbedaan tahapan prapemrosesan tersebut perlu dicatat karena setiap penelitian dapat memiliki prosedur pengolahan data yang tidak sepenuhnya sama.

Berdasarkan hal tersebut, *accuracy*, *precision*, *recall*, dan *F1-score* digunakan sebagai indikator kinerja yang diekstraksi apabila dilaporkan oleh penelitian sumber. Data tersebut belum digunakan untuk menentukan metode yang lebih unggul pada bagian ini, tetapi menjadi dasar sintesis. Kajian diperluas untuk mencakup SVM, Naive Bayes, LSTM, IndoBERT, BERT, VADER, dan TextBlob; namun, perbandingan kuantitatif hanya dilakukan pada studi yang menyediakan metrik evaluasi yang memadai dan dapat ditafsirkan secara bertanggung jawab.

2.6 Penelitian Terdahulu

Penelitian mengenai analisis sentimen teks berbahasa Indonesia telah dilakukan pada berbagai sumber data dan topik. Data yang digunakan dalam penelitian terdahulu antara lain berasal dari Twitter, TikTok, YouTube, Google Play Store, dan Shopee. Keragaman sumber tersebut memperlihatkan bahwa penelitian NLP tidak hanya diterapkan pada satu bentuk data, tetapi juga pada berbagai teks yang dihasilkan oleh pengguna platform digital (Wahyudi et al., 2025). (Agustina et al., 2022) melakukan analisis sentimen terhadap ulasan aplikasi Shopee yang diperoleh dari Google Play Store dengan menggunakan *Naive Bayes*. Penelitian tersebut menunjukkan penerapan metode klasifikasi berbasis *machine learning* pada data ulasan pengguna. Sumber data berupa ulasan aplikasi memungkinkan opini pengguna diolah menjadi data tekstual yang selanjutnya diklasifikasikan berdasarkan kategori sentimen.

Penelitian (Fathoniah & Rozikin, 2022) menggunakan data Twitter untuk menganalisis sentimen masyarakat terhadap isu terorisme. Penelitian tersebut menerapkan NLP pada tahap pengolahan data dan menggunakan LSTM sebagai model klasifikasi. Penggunaan Twitter

sebagai sumber data menunjukkan bahwa media sosial menjadi salah satu sumber teks yang dapat digunakan untuk mengamati kecenderungan opini masyarakat

Pada tahun berikutnya, (Ainunnisa & Sulastri, 2023) melakukan analisis sentimen terhadap ulasan aplikasi TikTok dengan menggunakan SVM, *Naive Bayes*, dan *Logistic Regression*. Data yang digunakan berupa ulasan pengguna yang kemudian melalui tahapan *case folding*, tokenisasi, penyaringan, dan *stemming*. Penelitian tersebut menjadi salah satu sumber yang memperlihatkan penggunaan dua metode yang termasuk dalam ruang lingkup meta-analisis ini, yaitu SVM dan *Naive Bayes*.

Penelitian (Sayogo et al., 2023) menggunakan ulasan Instagram yang tersedia pada Google Play Store sebagai data penelitian. Klasifikasi dilakukan menggunakan *Naive Bayes* untuk mengelompokkan ulasan berdasarkan sentimen yang terkandung di dalamnya. Penelitian ini memperluas konteks penerapan *Naive Bayes* karena data yang dianalisis berasal dari ulasan pengguna aplikasi media sosial.

Penelitian (Ningsih et al., 2024) membandingkan SVM dan *Naive Bayes* dalam analisis sentimen Twitter mengenai penggunaan mobil listrik di Indonesia. Data penelitian diperoleh berdasarkan pencarian topik terkait pada Twitter dan kemudian melalui proses pelabelan, prapemrosesan, serta pemodelan. Studi tersebut menjadi salah satu sumber yang relevan untuk meta-analisis karena secara langsung menggunakan dua metode yang menjadi fokus penelitian ini.

(Nurwanda et al., 2024) menerapkan NLP untuk menganalisis sentimen pengguna Telegram berdasarkan ulasan pada Google Play Store. Penelitian tersebut mengelompokkan opini pengguna ke dalam sentimen positif, negatif, dan netral serta menggunakan beberapa ukuran evaluasi untuk menilai hasil klasifikasi. Penggunaan data ulasan aplikasi dalam penelitian tersebut menunjukkan kembali bahwa platform *app store* dapat menjadi sumber data sentimen berbahasa Indonesia.

Penelitian lain pada tahun yang sama dilakukan oleh (Prastyo et al., 2024) terhadap ulasan Spotify. Penelitian tersebut menerapkan tahapan prapemrosesan NLP sebelum melakukan klasifikasi terhadap tiga kategori sentimen, yaitu positif, netral, dan negatif. Studi tersebut juga menunjukkan bahwa proses prapemrosesan menjadi bagian penting sebelum data digunakan dalam proses klasifikasi.

Penelitian (Rivaldi & Wismarini, 2024) menggunakan ulasan produk Zalika Store 88 pada Shopee sebagai objek analisis sentimen. Penelitian tersebut menerapkan NLP melalui beberapa proses pengolahan teks, termasuk tokenisasi dan *stemming*, sebelum melakukan pengelompokan sentimen. Penggunaan data ulasan produk memperlihatkan bahwa analisis

sentimen juga dapat diterapkan untuk memahami tanggapan pengguna terhadap produk yang ditawarkan pada platform perdagangan elektronik.

Pada penelitian yang menggunakan model LSTM, (Rahmawati & Aini, 2025) menganalisis komentar YouTube pada program *Clash of Champions* Ruangguru. Data berupa komentar pengguna diproses sebelum digunakan dalam model LSTM, kemudian kinerja model dievaluasi menggunakan beberapa metrik. Penelitian tersebut menjadi sumber yang relevan dalam kajian karena secara langsung merepresentasikan penerapan LSTM pada teks berbahasa Indonesia.

Pendekatan berbasis IndoBERT ditemukan dalam penelitian (Zahara et al., 2025) yang menganalisis komentar TikTok mengenai program Makan Siang Gratis. Model IndoBERT digunakan untuk melakukan klasifikasi sentimen terhadap data yang diperoleh dari media sosial tersebut. Studi ini menunjukkan penggunaan model berbasis *transformer* pada data teks berbahasa Indonesia yang berasal dari platform media sosial.

(Setiadi & Sulisty, 2025) juga menggunakan IndoBERT dalam penelitian mengenai sentimen publik terhadap kebijakan Makan Bergizi Gratis. Penelitian tersebut memanfaatkan data berbahasa Indonesia dari platform X dan menggabungkan representasi IndoBERT dengan fitur lainnya dalam proses analisis. Sementara itu, (Aryanti & Suria, 2025) menggunakan IndoBERT dan SVM untuk menganalisis sentimen terhadap isu pemutusan hubungan kerja di Indonesia.

Penelitian (Hadi et al., 2025) menggunakan *Naive Bayes* untuk menganalisis komentar pengguna pada video pembelajaran matematika di YouTube. Data yang digunakan berasal dari komentar pada dua video pembelajaran dengan topik matematika yang berbeda. Studi tersebut memperlihatkan bahwa analisis sentimen dapat diterapkan pada komentar yang berkaitan dengan aktivitas pembelajaran dan bukan hanya pada isu sosial atau ulasan produk.

(Irawanto, 2026) mengkaji penerapan NLP pada analisis sentimen media sosial dengan menggunakan SVM dan *Naive Bayes*. Data penelitian berasal dari Twitter dan diklasifikasikan ke dalam kategori positif, negatif, dan netral. Penelitian tersebut relevan dengan meta-analisis karena melibatkan dua metode yang menjadi fokus utama penelitian ini dalam satu konteks data.

(Prastyo et al., 2024) menganalisis 1.729 komentar YouTube pada debat Pilkada Banten 2024 menggunakan SVM dan *Naive Bayes*. Distribusi sentimen yang dilaporkan adalah 49% positif, 35% negatif, dan 16% netral. Dalam evaluasi model, SVM memperoleh *accuracy* 73%, *precision* 73%, *recall* 73%, dan *F1-score* 71%; sedangkan *Naive Bayes* memperoleh *accuracy*

71%, precision 76%, recall 71%, dan F1-score 65%. Studi ini menambah bukti evaluasi untuk dua metode *machine learning*, meskipun konteks politik dan datasetnya berbeda dari studi lain. (Pratama et al., 2025) menerapkan NLP pada 465 komentar YouTube mengenai pelantikan menteri di kanal Sekretariat Presiden. Penelitian menggunakan pembersihan teks, TF-IDF, analisis *n-gram*, analisis sentimen berbasis leksikon melalui *TextBlob*, serta pemodelan topik LDA. Hasilnya melaporkan 94,2% komentar bersentimen netral. Karena artikel ini menyajikan distribusi sentimen tetapi tidak melaporkan *accuracy*, *precision*, *recall*, atau *F1-score* untuk memvalidasi klasifikasi, temuannya digunakan dalam sintesis deskriptif, bukan perbandingan kuantitatif kinerja.

Penelitian (Merdiansah et al., 2024) menganalisis sentimen pengguna X di Indonesia mengenai kendaraan listrik menggunakan IndoBERT dengan membandingkan model yang dilatih tanpa dan dengan data IndoNLU. Pada epoch kelima, *accuracy* validasi yang dilaporkan adalah 82% untuk model tanpa data IndoNLU dan 98% untuk model dengan data IndoNLU. Artikel juga melaporkan *precision*, *recall*, dan *F1-score* pada tiap epoch. Nilai ini dicatat sebagai hasil validasi pada konfigurasi tertentu dan tidak boleh dianggap langsung sebanding dengan *accuracy* dari dataset atau prosedur evaluasi penelitian lain.

Penelitian (Meliala et al., 2025) membandingkan VADER, *TextBlob*, dan BERT pada 967 cuitan yang relevan dengan tagar #KaburAjaDulu di platform X. VADER menghasilkan distribusi 38,5% positif, 17,6% negatif, dan 43,9% netral; *TextBlob* menghasilkan 41,9% positif, 13,3% negatif, dan 44,8% netral; sedangkan BERT menghasilkan 47,7% positif dan 52,3% negatif tanpa kelas netral. Artikel menjelaskan bahwa cuitan diterjemahkan ke bahasa Inggris untuk analisis. Karena artikel ini melaporkan distribusi hasil klasifikasi, bukan metrik evaluasi terhadap label acuan, temuannya dipakai secara deskriptif dan perlu dibaca dengan mempertimbangkan dampak penerjemahan terhadap sentimen.

Jika keseluruhan penelitian terdahulu diperhatikan, terlihat adanya variasi dalam metode, sumber data, topik, tahapan pengolahan, serta indikator evaluasi yang digunakan. Variasi tersebut menyebabkan hasil penelitian individual memiliki karakteristik yang berbeda sehingga diperlukan proses pengumpulan dan sintesis untuk memperoleh gambaran yang lebih terstruktur. Atas dasar tersebut, penelitian ini menggunakan pendekatan meta-analisis dengan memusatkan perhatian pada empat metode, yaitu SVM, *Naive Bayes*, LSTM, dan IndoBERT. Penelitian terdahulu dalam bagian ini berfungsi sebagai dasar untuk mengidentifikasi studi yang sesuai dengan ruang lingkup kajian. Informasi mengenai metode, sumber data, karakteristik pengolahan, dan ukuran kinerja dari masing-masing penelitian selanjutnya digunakan dalam proses ekstraksi data. Hasil ekstraksi tersebut menjadi dasar untuk tahap

analisis berikutnya, sedangkan perbandingan kinerja antarmetode akan dibahas secara khusus pada bagian hasil dan pembahasan.

Metode Penelitian

Penelitian ini menggunakan metode meta-analisis untuk mengidentifikasi dan menggabungkan hasil penelitian terdahulu mengenai kinerja metode NLP dalam analisis sentimen teks berbahasa Indonesia. Meta-analisis merupakan pendekatan yang menggabungkan hasil dari beberapa penelitian secara sistematis untuk memperoleh sintesis kuantitatif dari temuan yang tersedia (Hansen et al., 2022). Penerapan metode meta-analisis dilakukan melalui serangkaian tahapan yang meliputi identifikasi penelitian yang relevan, pengumpulan dan pengodean data, serta pengolahan hasil penelitian secara statistik (Suharyat et al., 2022).

Berdasarkan tahapan tersebut, penelitian diawali dengan penelusuran artikel ilmiah yang membahas analisis sentimen teks berbahasa Indonesia menggunakan metode NLP sesuai dengan rentang waktu 2021 – 2026. Penelusuran dan pemilihan artikel dilakukan dengan menetapkan pertanyaan penelitian, strategi pencarian, serta kriteria inklusi dan eksklusi agar studi yang digunakan sesuai dengan tujuan meta-analisis (Hansen et al., 2022). Penggunaan kriteria dalam pemilihan sumber penelitian juga diterapkan dalam penelitian meta-analisis sebelumnya untuk memastikan data yang digunakan memiliki hubungan dengan variabel penelitian (Suharyat et al., 2022).

Setelah artikel yang memenuhi kriteria ditetapkan, data dari setiap penelitian dikumpulkan dan dikodekan berdasarkan informasi yang diperlukan untuk analisis, seperti karakteristik penelitian, ukuran sampel, serta hasil atau ukuran efek yang dilaporkan (Hansen et al., 2022; Hu et al., 2026). Data hasil penelitian yang memiliki bentuk atau ukuran statistik berbeda dapat disesuaikan ke dalam bentuk yang dapat dibandingkan sebelum dilakukan sintesis, karena meta-analisis memerlukan hasil penelitian yang dapat diolah dalam format yang sesuai (Abbas et al., 2024). Selanjutnya, data yang telah diekstraksi dianalisis secara statistik untuk memperoleh ukuran efek gabungan dan mengidentifikasi variasi hasil antarpelitian sesuai dengan karakteristik data dan tujuan penelitian (Deeks et al., 2024).

Hasil & Pembahasan

4.1 Hasil Seleksi Artikel

Melalui tahap penelusuran dan ekstraksi artikel, terkumpul 25 artikel yang membahas penggunaan NLP untuk menganalisis sentimen pada teks berbahasa Indonesia. Artikel-artikel

ini berbeda dalam metode, sumber data, karakteristik dataset, dan jenis keluaran evaluasi. Sebanyak 20 artikel melaporkan metrik kinerja yang dapat diekstraksi, sedangkan lima artikel tidak menyediakan metrik evaluasi kinerja yang dapat digunakan untuk perbandingan kuantitatif. Artikel yang hanya melaporkan distribusi kelas sentimen tetap dapat dimanfaatkan dalam sintesis deskriptif.

Tabel 1 Ringkasan Hasil Seleksi Artikel

Kategori	Jumlah
Artikel yang diekstraksi	25
Artikel dengan data kinerja	20
Artikel tanpa data kinerja	5

Lima artikel yang tidak mencantumkan metrik evaluasi kinerja tetap dipertahankan dalam kajian literatur untuk sintesis deskriptif, tetapi tidak diikutsertakan dalam perbandingan kuantitatif *accuracy*, *precision*, *recall*, atau *F1-score*.

Gambar Proses Seleksi:



4.2 Karakteristik Metode NLP yang Digunakan

Hasil ekstraksi memperlihatkan bahwa penelitian-penelitian sebelumnya memakai beragam pendekatan NLP dalam mengklasifikasikan sentimen. Metode dan alat yang muncul meliputi SVM, *Naive Bayes*, *Random Forest*, *Decision Tree*, *Logistic Regression*, LSTM, *Neural Network*, IndoBERT, BERT, VADER, dan *TextBlob*. Sebagian penelitian membandingkan lebih dari satu pendekatan dalam satu kajian. Keragaman ini mencakup algoritma *machine learning*, model *deep learning*, model bahasa berbasis transformer, serta analisis sentimen berbasis leksikon/aturan. Karena keluaran yang dilaporkan tidak selalu berupa metrik evaluasi yang sama, distribusi sentimen tidak diperlakukan sebagai pengganti *accuracy* atau ukuran kinerja lainnya.

Tabel 2. Kelompok Metode yang Ditemukan

Kelompok	Metode
Machine Learning	SVM, Naive Bayes, Random Forest, Decision Tree, Logistic Regression
Deep Learning	LSTM, Neural Network
Transformer	BERT, IndoBERT
Lexicon/rule-based	VADER, TextBlob

Merujuk pada tabel tersebut, penelitian-penelitian yang ditinjau tidak terpaku pada satu jenis algoritma saja. Kajian ini mencakup metode yang menghasilkan metrik evaluasi model maupun pendekatan leksikon yang sering kali melaporkan distribusi sentimen. Perbedaan bentuk keluaran tersebut perlu dipertahankan saat sintesis: hanya metrik evaluasi yang sebanding digunakan untuk perbandingan kuantitatif, sedangkan distribusi kelas digunakan sebagai temuan deskriptif.

4.3 Hasil Kinerja Berdasarkan Accuracy

Accuracy adalah metrik yang paling sering tercantum dalam tabel ekstraksi, sehingga digunakan sebagai salah satu indikator kinerja model. Nilai *accuracy* bervariasi antarpelitian. Artikel (Prastyo et al., 2024) melaporkan *accuracy* 73% untuk SVM dan 71% untuk Naive Bayes. (Merdiansah et al., 2024) melaporkan *accuracy* validasi pada epoch kelima sebesar 82% untuk IndoBERT tanpa data IndoNLU dan 98% untuk IndoBERT dengan data IndoNLU. Nilai-nilai ini berasal dari desain eksperimen yang berbeda, sehingga tidak boleh langsung ditafsirkan sebagai peringkat universal antarmetode.

Tabel 3. Accuracy yang Dilaporkan dalam Artikel

No.	Penelitian	Metode/Model	Accuracy
1	Irawanto (2026)	SVM	87%
1	Irawanto (2026)	Naïve Bayes	82%
2	Nurwanda et al. (2024)	Naïve Bayes	85,31%
3	Fathoniah & Rozikin (2022)	LSTM	59,34%
4	Zahara et al. (2025)	IndoBERT	84,41%
5	Rahmawati & Aini (2025)	LSTM	90%
6	Ningsih et al. (2024)	SVM/Naïve Bayes	70,83%
7	Setiadi & Sulisty (2025)	IndoBERT	93,63%

8	Fauzan et al. (2024)	SVM	83%
8	Fauzan et al. (2024)	Random Forest	82%
9	Aryanti & Suria (2025)	IndoBERT	89,60%
9	Aryanti & Suria (2025)	SVM	88%
9	Aryanti & Suria (2025)	Random Forest	78,04%
9	Aryanti & Suria (2025)	Decision Tree	70,40%
10	Ainunnisa & Sulastri (2023)	SVM	82%
10	Ainunnisa & Sulastri (2023)	Logistic Regression	84%
10	Ainunnisa & Sulastri (2023)	Naïve Bayes	79%
11	Agustina et al. (2022)	Naïve Bayes	83%
11	Agustina et al. (2022)	Naïve Bayes	82%
12	Sayogo et al. (2023)	Naïve Bayes	89%
13	Rahmat et al. (2026)	SVM	90,7%
13	Rahmat et al. (2026)	Random Forest	81,5%
13	Rahmat et al. (2026)	Neural Network	89,2%
16	Fauzianto & Supatman (2023)	Model yang dilaporkan	95,56%
17	Rivaldi & Wismarini (2024)	Logistic Regression	87%
18	Agustina & Herliana (2025)	Naïve Bayes	93,01%
19	Tribuana et al. (2025)	Model yang dilaporkan	86,5%
21	Utama et al. (2024)	SVC	69,81%
22	Prastyo et al. (2024)	SVM	73%
22	Prastyo et al. (2024)	Naive Bayes	71%
24	Merdiansah et al. (2024)	IndoBERT tanpa IndoNLU (validasi, epoch 5)	82%
24	Merdiansah et al. (2024)	IndoBERT dengan IndoNLU (validasi, epoch 5)	98%

Catatan: Apabila satu penelitian menguji beberapa model, hasil tiap model tetap disajikan secara terpisah. Nilai 82% dan 98% dari (Merdiansah et al., 2024) merupakan *accuracy* validasi pada epoch kelima untuk dua konfigurasi IndoBERT, bukan hasil eksperimen yang identik dengan penelitian lain. Artikel (Meliala et al., 2025; Pratama et al., 2025) hanya melaporkan distribusi sentimen dan tidak dimasukkan ke tabel *accuracy* karena tidak menyajikan metrik evaluasi kinerja.

4.4 Pembahasan

Hasil ekstraksi menunjukkan bahwa kinerja yang dilaporkan dalam artikel sangat bervariasi. Dalam tabel yang tersedia, *accuracy* terendah adalah 59,34% pada model LSTM dalam penelitian (Fathoniah & Rozikin, 2022), sedangkan nilai tertinggi yang tercantum adalah 95,56% pada penelitian (Fauzianto & Supatman, 2023). Kedua nilai tersebut merupakan hasil dari penelitian berbeda dan tidak cukup untuk menyimpulkan bahwa satu metode selalu lebih baik daripada metode lain. Perbandingan langsung perlu mempertimbangkan ukuran serta komposisi dataset, jumlah kelas sentimen, proses pelabelan, metode *preprocessing*, pembagian data latih-uji, strategi validasi, dan apakah nilai yang dilaporkan berasal dari data uji atau validasi.

Studi (Prastyo et al., 2024) melaporkan *accuracy* 73% untuk SVM dan 71% untuk *Naive Bayes* pada komentar YouTube tentang debat Pilkada Banten. Sementara itu, (Merdiansah et al., 2024) melaporkan *accuracy* validasi 82% untuk konfigurasi IndoBERT tanpa data IndoNLU dan 98% untuk konfigurasi dengan data IndoNLU pada epoch kelima. Perbedaan ini memberi gambaran bahwa konfigurasi pelatihan dan data tambahan dapat berkaitan dengan hasil yang diperoleh dalam studi tertentu, tetapi tidak membuktikan keunggulan universal IndoBERT atas SVM atau *Naive Bayes* karena kedua penelitian menggunakan data dan prosedur eksperimen yang berbeda.

Lima artikel yang tidak melaporkan metrik evaluasi tetap memberikan informasi mengenai kecenderungan sentimen atau penerapan alat berbasis leksikon. Namun, persentase sentimen positif, negatif, atau netral bukanlah *accuracy*. Tanpa label acuan atau evaluasi terhadap data berlabel, distribusi tersebut tidak dapat digunakan untuk menilai ketepatan klasifikasi. Hal ini khususnya penting pada pendekatan VADER dan *TextBlob* serta artikel yang menerjemahkan teks sebelum analisis, karena adaptasi bahasa dapat memengaruhi interpretasi sentimen.

Secara keseluruhan, bukti dalam korpus mendukung kesimpulan bahwa beragam pendekatan NLP telah diterapkan pada berbagai sumber teks berbahasa Indonesia, tetapi heterogenitas desain penelitian membatasi perbandingan lintas-studi. Penelitian selanjutnya perlu melaporkan dataset dan jumlah sampel, skema pelabelan, pembagian data, konfigurasi model, metrik per kelas, *macro/weighted average*, serta hasil pada data uji yang terpisah. Untuk meta-analisis statistik, diperlukan data yang cukup sebanding misalnya jumlah prediksi benar dan jumlah sampel uji atau informasi confusion matrix serta metode penggabungan yang sesuai.

Kesimpulan

Penelitian ini mengkaji 25 artikel tentang analisis sentimen teks berbahasa Indonesia menggunakan pendekatan NLP. Sebanyak 20 artikel menyediakan metrik kinerja yang dapat diekstraksi, sedangkan lima artikel hanya digunakan dalam sintesis deskriptif karena melaporkan distribusi sentimen tanpa metrik evaluasi. Metode yang ditemukan mencakup *machine learning*, *deep learning*, transformer, serta pendekatan berbasis leksikon/aturan. Nilai *accuracy* yang dilaporkan menunjukkan variasi antarpelitian, tetapi perbedaan dataset dan prosedur evaluasi membuat nilai tersebut tidak dapat digunakan untuk menentukan satu metode yang paling unggul secara universal. Oleh karena itu, hasil penelitian ini perlu dipahami sebagai sintesis deskriptif terhadap bukti yang tersedia, bukan sebagai estimasi statistik gabungan. Penelitian berikutnya disarankan menggunakan strategi pencarian yang terdokumentasi, kriteria seleksi yang dapat direplikasi, pelaporan metrik secara lengkap, dan protokol pengujian yang lebih beragam. Jika data statistik yang memadai tersedia, meta-analisis kuantitatif dapat dilakukan untuk menghitung estimasi gabungan dan mengevaluasi heterogenitas antarpelitian.

Daftar Pustaka

- Abbas, A., Hefnawy, M. T., & Negida, A. (2024). Meta-analysis accelerator: a comprehensive tool for statistical data conversion in systematic reviews with meta-analysis. *BMC Medical Research Methodology*, 24(1). <https://doi.org/10.1186/s12874-024-02356-6>
- Adha, D. A., Ramadhan, A., Maulana, H., Harahap, P. P. H., & Ismanto, E. (2025). Peramalan Harga Emas Berbasis Time Series Menggunakan Arsitektur LSTM Deep Learning. *Jurnal Computer Science and Information Technology*, 6(2), 329–336.
- Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi algoritma Naive Bayes untuk analisis sentimen ulasan Shopee pada Google Play Store. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54.
- Ainunnisa, I. R., & Sulastri. (2023). Analisis sentimen aplikasi TikTok dengan metode Support Vector Machine (SVM), Logistic Regression dan Naïve Bayes. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 6(3), 423–430. <https://doi.org/10.32493/jtsi.v6i3.31076>
- Amien, M. (2022). *Sejarah dan perkembangan teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam Bahasa Indonesia*. arXiv.
- Amien, M., & Gunawan, G. F. (2024). BERT dan Bahasa Indonesia: Studi tentang efektivitas

- model NLP berbasis Transformer. *ELANG: Journal of Interdisciplinary Research*.
- Amien, S., Perdana, P., Bharata Aji, T., & Ferdiana, R. (2021). Aspect Category Classification with Machine Learning Approach Using Indonesian Language Dataset. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(3), 229–235.
- Apriansyah, F. M., Ramadhan, T. I., Hidayat, C. R., & Karta, A. (2025). Perbandingan IndoBERT dan IndoRoBERTa Untuk Analisis Sentimen Pada Film Dokumenter Dirty Vote. *Jurnal Informatika: Jurnal Pengembangan IT*, 10(3), 593–605. <https://doi.org/10.30591/jpit.v10i3.8607>
- Aryanti, N. N. A., & Suria, O. (2025). Analisis sentimen terhadap keputusan hubungan kerja di Indonesia: Komparasi IndoBERT dengan SVM, Random Forest, dan Decision Tree dengan optimasi TF-IDF. 10(2), 1158–1176.
- Asana, D. P., & Yanti, N. P. D. T. (2023). Sistem Klasifikasi Pengajuan Kredit Dengan Metode Support Vector Machine (SVM). *Jurnal Sistem Cerdas*, 06(02), 123–133.
- Awali, R. M. (2025). Meta-Analisis Efektivitas Media Quizizz dalam Meningkatkan Hasil Belajar Matematika Siswa. *SESIOMADIKA*, 210–223.
- Azzahra, F. S. P., & Printo, C. (2025). Systematic Literature Review: Analysis of AI Implementation for Document Verification. *Jurnal Ilmiah Sistem Informasi (JUSI)*, 4(3), 1–19.
- Deeks, J. J., Higgins, J. P., Altman, D. G., McKenzie, J. E., & Veroniki, A. A. (2024). Chapter 10: Analysing data and undertaking meta-analyses | Cochrane. In *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.). Cochrane.
- Emarapenta, J., Sinulingga, B., Cesar, H., & Sitorus, K. (2024). Analisis Sentimen Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA)*, 14(April), 42–53.
- Fathoniah, S., & Rozikin, C. (2022). Analisis sentimen masyarakat terhadap teroris dalam media sosial Twitter menggunakan NLP. *Jurnal Ilmiah Wahana Pendidikan*, 8(13), 412–419.
- Fauzianto, R. A., & Supatman. (2023). ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP TECH WINTER PADA TWITTER. *Jurnal Syntax DMIRATION*, 4(9), 1577–1585. <https://doi.org/10.46799/jsa.v3i9.909>

- Febriani, S. A., & Sari, V. (2026). Analisis komparasi metode SVM dan ModernBERT dalam klasifikasi sentimen opini publik di platform TikTok mengenai banjir bandang di Sumatra. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 10(3), 4268–4275.
- Franciska, D., Dina, M., Haryanti, T., & Haq, M. A. (2025). ANALISIS SENTIMEN TERHADAP KOMENTAR PADA MEDIA SOSIAL TIKTOK YANG BERPOTENSI MENYEBABKAN DEPRESI. *Jurnal Ilmiah Computing Insight*, 7(1), 1–9.
- Hadi, A. La, Safilda, R., & Riansyah, F. (2025). Pemanfaatan NLP untuk analisis sentimen: Perbandingan matematika konseptual vs. prosedural dalam video pembelajaran YouTube. *Insearch: Information System Research Journal*, 5(1), 17–29.
- Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: a practical guide. *Management Review Quarterly*, 72(1), 1–19. <https://doi.org/10.1007/s11301-021-00247-4>
- Hawari, F. D. (2025). Implementasi chatbot dalam berbagai bidang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4650–4658.
- Hu, M., Koh, P. S., Soh, X. C., Hartanto, A., & Majeed, N. M. (2026). From review to synthesis: A step-by-step methodological guide to systematic reviews and multilevel meta-analyses. In *Behavior Research Methods* (Vol. 58, Issue 7). Springer US. <https://doi.org/10.3758/s13428-026-02961-x>
- Irawanto, I. (2026). Penerapan Natural Language Processing (NLP) dalam analisis sentimen pada media sosial. *Journal of Informatics and Computer Research (JICR)*, 2019–2022.
- Jumeilah, F. S. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *JURNAL RESTI*, 1(1), 19–25.
- Kamal, A., & Astri, R. (2025). Eksplorasi Sentimen Pengguna pada Aplikasi E-Commerce dengan Deep Learning. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 7(3), 435–441.
- Leonardi, V. H., Ibrahim, A., Kurnia, R. D., & Afrina, M. (2025). Analisis Sentimen Ulasan Aplikasi Pembelajaran Bahasa Menggunakan Metode VADER. *Jurnal Algoritma*, 22. <https://doi.org/10.33364/algoritma/v.22-1.2285>
- Meliala, R. J., Chasanah, N. I., Manik, J. S. R., Pasya, T. M., & Lestari, H. R. (2025). Analisis sentimen tagar #KaburAjaDulu pilihan migrasi ke Jepang pada platform X dengan NLP. *Journal of Digital Business and Technology Innovation (DBESTI)*, 2(1), 1–7.

<https://journal.nurulfikri.ac.id/index.php/DBESTI>

- Merdiansah, R., Siska, & Ridha, A. A. (2024). Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 7(1), 221–228. <https://ejournal.sisfokomtek.org/index.php/jikom>
- Ningsih, W., Alfiana, B., Rahmadden, R., & Wulandari, D. (2024). Perbandingan algoritma SVM dan Naïve Bayes dalam analisis sentimen Twitter pada penggunaan mobil listrik di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 556–562. <https://doi.org/10.57152/malcom.v4i2.1253>
- Nurwanda, N., Suarna, N., & Prihartono, W. (2024). Penerapan NLP (Natural Language Processing) dalam analisis sentimen pengguna Telegram di Playstore. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1841–1846. <https://doi.org/10.36040/jati.v8i2.8469>
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 452–458.
- Prasetya, M. A., Wulandari, M., & Nikmah, S. A. (2024). Implementasi NLP (Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify. *PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS*, 3, 145–153.
- Prastyo, D., Mursyidin, I. H., & Irawan, D. (2024). Klasifikasi sentimen komentar YouTube dengan NLP pada debat Pilkada Banten 2024. *Bit-Tech (Binary Digital - Technology)*, 7(2), 414–422. <https://doi.org/10.32877/bt.v7i2.1833>
- Pratama, S., Yuniansyah, & Fransiska, W. (2025). Penerapan NLP untuk menganalisis komentar publik pada pelantikan menteri di kanal Sekretariat Presiden. *Teknomatika*, 15(2).
- Rahmawati, I., & Aini, N. (2025). Analisis sentimen komentar YouTube pada program Clash of Champions Ruangguru menggunakan deep learning berbasis LSTM. *Intellect: Indonesian Journal of Learning and Technological Innovation*, 4(1), 96–106. <https://doi.org/10.57255/intellect.v4i1.1376>
- Rivaldi, R. C., & Wismarini, T. D. (2024). Analisis sentimen pada ulasan produk dengan metode Natural Language Processing (NLP): Studi kasus Zalika Store 88 Shopee.

- ELKOM: Jurnal Elektronika Dan Komputer*, 17(1), 120–128.
<https://doi.org/10.51903/elkom.v17i1.1680>
- Saputra, R., Alamsyah, A. G., Tjoanda, M., Cornelius, A., & Herdiatmoko, H. F. (2024). ANALISIS PREDIKSI SAHAM TESLA MENGGUNAKAN ALGORITMA LONG SHORT TERM MEMORY (LSTM). *Journal of Computer Science and Information Technology (JCSIT)*, 2(1), 81–90.
- Sayogo, D. S., Irawan, B., & Bahtiar, A. (2023). Analisis sentimen ulasan Instagram di Google Play Store menggunakan algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3314–3319. <https://doi.org/10.36040/jati.v7i6.8178>
- Setiadi, E., & Sulistyo, D. A. (2025). Analisis sentimen kebijakan Makan Bergizi Gratis menggunakan IndoBERT dan machine learning. *Jurnal FASILKOM*, 15(3), 507–513. <https://doi.org/10.37859/jf.v15i3.10546>
- Shobrina Fathoniah, C. R. (2022). Analisis sentimen masyarakat terhadap teroris dalam media sosial Twitter menggunakan NLP. *Jurnal Ilmiah Wahana Pendidikan*, 8(13), 412–419.
- Siregar, M. Z., Elhanafi, A. M., & Irwan, D. (2025). ANALISIS SENTIMEN TERHADAP ULASAN APLIKASI MEDIA SOSIAL DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 3373–3381.
- Suharyat, Y., Ichsan, I., Santosa, T. A., Aprilisia, S., & Yulianti, S. (2022). Meta-Analysis Study: The Effectiveness of Problem Solving Learning in Science Learning in Indonesia. *International Journal of Education and Literature*, 1(3), 6–13. <https://doi.org/10.55606/ijel.v1i3.36>
- Tarumingkeng, R. C. (2024a). *Natural Language Processing (NLP)* (Issue November).
- Tarumingkeng, R. C. (2024b). *Natural Language Processing (NLP)*.
- Wahyudi, T., Purwati, N., Hasan, N., & Sulistyo, G. B. (2025). Tren NLP dalam analisis sentimen media sosial: Tinjauan sistematis dan bibliometrik. *Jurnal ICT: Information Communication & Technology*, 25(2), 206–212. <https://doi.org/10.36054/jict-ikmi.v25i2.341>
- Wahyuni, S. D., & Februariyanti, H. (2026). Analisis Sentimen dan Rating Pengguna Aplikasi Traveloka Menggunakan Vader. *Semantik*, 11(1), 79–88.

Widiantoro, A. D., Mustafid, & Sanjaya, R. (2024). *Pengantar NLP dan Topik Model LDA*. ASOSIASI DOKTOR SISTEM INFORMASI INDONESIA.

Wijaya, F., Ramadhani, R. A., & Setiawan, A. B. (2025). Analisis Sentimen Komentar Video Youtube Dengan Leksikon Textblob Menggunakan Algoritma Svm Berdasarkan Rasio Data Uji. *INOTEK*, 9, 1894–1902.

Zahara, M., Rizka, M., & Mahyar, H. (2025). Analisis sentimen program makan siang gratis pada TikTok dengan pendekatan NLP berbasis IndoBERT. *Jurnal Infomedia: Teknik Informatika, Multimedia, Dan Jaringan*, 10(2).

