Arus Jurnal Sains dan Teknologi (AJST)
Website: http://jurnal.ardenjaya.com/index.php/ajst
Email: jurnal.ajst@ardenjaya.com

Studi Komparatif Evolusi Metode *Natural Language Processing* pada Analisis Sentimen Bahasa Indonesia: dari Algoritma Klasik Hingga Pre-trained Transformer

INFO PENPULIS

Wasti Harianja
Universitas Negeri Medan
wastiharianja36@gmail.com

Shinta Novaliana Aritonang
Universitas Negeri Medan
shintanovaliana99@gmail.com

Cetrin Yolanda Sinaga
Universitas Negeri Medan
cetrinyolandasinaga@gmail.com

Keysa Fransiska Siregar
Universitas Negeri Medan
keysasiregar751@gmail.com

Elsa Sabrina
Universitas Negeri Medan
elsasabrina@unimed.ac.id

INFO ARTIKEL

ISSN: 3026-3603
Vol. 4, No. 2 Oktober 2026
http://jurnal.ardenjaya.com/index.php/ajst

© 2026 Arden Jaya Publisher All rights reserved

Saran Penulisan Referensi:

Harianja, W., Aritonang, S. N., Sinaga, C. Y., Siregar, K. F., & Sabrina, E. (2026). Studi Komparatif Evolusi Metode Natural Language Processing pada Analisis Sentimen Bahasa Indonesia: dari Algoritma Klasik Hingga Pre-trained Transformer. *Arus Jurnal Sains dan Teknologi* , *4* (2), 326-339.

Abstrak

Pertumbuhan data teks berbahasa Indonesia di media sosial dan platform ulasan telah mendorong penerapan analisis sentimen yang memanfaatkan Natural Language Processing (NLP). Penelitian ini bertujuan untuk membandingkan evolusi metode NLP untuk analisis sentimen bahasa Indonesia, mulai dari pendekatan berbasis leksikon dan machine learning klasik hingga model Transformer yang telah dilatih sebelumnya (pre-trained). Studi literatur dilakukan terhadap 20 artikel yang diterbitkan antara tahun 2021 dan 2026 dengan menggunakan analisis deskriptif-komparatif. Hasil penelitian menunjukkan adanya pergeseran dari algoritma klasik—seperti Naïve Bayes, Support Vector Machine (SVM), dan Random Forest yang menggunakan TF-IDF—menuju model Transformer seperti IndoBERT dan IndoBERTweet. Tingkat akurasi untuk algoritma klasik berkisar antara 74% hingga 98%, sedangkan IndoBERT mencapai angka antara 79,5% hingga 98,27%. Rentang nilai tersebut mengindikasikan bahwa kinerja dipengaruhi oleh karakteristik dataset, domain spesifik, dan tahapan pra-pemrosesan (preprocessing), bukan semata-mata oleh jenis algoritmanya.

Kata kunci: Pemrosesan Bahasa Alami, Machine Learning, Analisis Sentimen, IndoBERT, Bahasa Indonesia

Abstract

The growth of Indonesian text data on social media and review platforms has driven the application of sentiment analysis utilizing Natural Language Processing (NLP). This study aims to compare the evolution of NLP methods for Indonesian sentiment analysis, ranging from lexicon-based approaches and classic machine learning to pre-trained Transformer models. A literature review was conducted on 20 articles published between 2021 and 2026 using a descriptive-comparative analysis. The results indicate a shift from classic algorithms—such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest using TF-IDF—toward Transformer models such as IndoBERT and IndoBERTweet. Accuracy rates for classic algorithms ranged from 74% to 98%, while IndoBERT achieved scores between 79.5% and 98.27%. This range indicates that performance is influenced by dataset characteristics, specific domains, and preprocessing steps, rather than solely by the choice of algorithm.

Keywords: Natural Language Processing, Machine Learning, Sentiment Analysis, IndoBERT, Indonesian Language

A. Pendahuluan

Perkembangan teknologi informasi dan media sosial yang masif telah memicu ledakan volume data teks berbahasa Indonesia secara signifikan (Nahdhudin et al., 2024) . Berbagai opini, ulasan pengguna, hingga pandangan masyarakat terhadap kebijakan publik kini tersimpan dalam bentuk data tidak terstruktur di internet (Rifki et al., 2024). Untuk mengolah dan memahami pola data teks dalam skala besar ini secara otomatis, teknik *Natural Language Processing* (NLP) khususnya analisis sentimen menjadi instrumen yang sangat vital (Zahara et al., 2025). Oleh karena itu, ekstraksi sentimen menjadi kebutuhan penting bagi akademisi maupun industri dalam mengukur persepsi publik secara *real-time* (Rifki et al., 2024).

Meskipun analisis sentimen sangat dibutuhkan, penerapannya pada teks berbahasa Indonesia memiliki kompleksitas dan tantangan tersendiri. Karakteristik bahasa Indonesia di media sosial sering kali tidak baku, menggunakan banyak kata gaul (*slang*), singkatan informal, serta memiliki struktur kalimat yang sangat bervariasi. Pendekatan awal untuk mengatasi hal ini banyak mengandalkan metode berbasis leksikon (*Lexicon-Based*) yang bekerja berdasarkan pencocokan kamus kata (Sapanji et al., 2023) . Namun, keterbatasan utama pendekatan leksikon terletak pada ketidakmampuannya memahami konteks ikatan kata dan perubahan makna kata yang terus berkembang (Manullang et al., 2023) .

Keterbatasan metode berbasis aturan tersebut mendorong pergeseran tren menuju pemanfaatan algoritma *Machine Learning* klasik. Algoritma populer seperti *Support Vector Machine* (SVM), Naïve Bayes, dan *Random Forest* mulai banyak diterapkan dengan bantuan ekstraksi fitur TF-IDF (Herjanto & Carudin, 2024). Penelitian pada era ini juga dipadukan dengan teknik optimasi parameter seperti GridSearch maupun penanganan ketidakseimbangan data menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) (Kumala Sari & Randy Suryono, 2024). Kendati mampu meningkatkan akurasi klasifikasi secara signifikan, algoritma tradisional ini masih bergantung pada pemrosesan awal (*preprocessing*) yang rumit serta belum optimal dalam menangkap ketergantungan konteks kalimat yang panjang.

Perkembangan metodologi NLP kemudian memasuki era baru dengan hadirnya arsitektur *Deep Learning* hingga *Pre-trained Transformer* modern (Abdurrohim & Rahman, 2024) . Model Transformer berbasis bahasa Indonesia seperti IndoBERT dan XLM-RoBERTa terbukti memberikan lonjakan performa yang pesat. Model ini mampu memahami konteks kata secara dua arah (*bidirectional*) serta sangat andal dalam menangani keragaman bahasa tidak formal (Asher & Rachmawati, 2025; Kurniawan et al., 2026).

Meskipun penelitian analisis sentimen di Indonesia terus berkembang, mayoritas studi terdahulu masih berfokus secara parsial. Sebagian besar peneliti hanya menguji satu jenis algoritma atau membandingkan metode dalam lingkup domain dataset yang sangat terbatas. Belum banyak studi literatur komprehensif yang memetakan secara sistematis bagaimana performa, kelebihan, dan keterbatasan dari tiap generasi metode ini berkembang dari masa ke masa. Ketiadaan pemetaan yang holistik ini menyebabkan minimnya panduan komparatif mengenai efektivitas transisi antar-metode NLP tersebut.

Berdasarkan celah penelitian tersebut, studi literatur ini bertujuan untuk mengkaji dan membandingkan secara komprehensif evolusi metode NLP pada analisis sentimen Bahasa Indonesia. Penelitian ini secara khusus memetakan perjalanan metodologi mulai dari pendekatan *Lexicon-Based*, *Machine Learning* klasik, hingga *Pre-trained Transformer*. Selain membandingkan tingkat akurasi dan efisiensi tiap model, studi ini juga mengidentifikasi tantangan teknis yang dihadapi di setiap era. Hasil kajian ini diharapkan dapat memberikan rekomendasi yang jelas serta menjadi acuan arah penelitian analisis sentimen Bahasa Indonesia di masa mendatang.

Kajian Pustaka

Natural Language Processing dan Analisis Sentimen

Natural Language Processing (NLP) merupakan pendekatan yang digunakan untuk mengolah dan menganalisis data berupa bahasa manusia agar dapat diproses oleh sistem komputer. Dalam konteks analisis sentimen, NLP digunakan untuk mengidentifikasi dan mengklasifikasikan kecenderungan sentimen yang terdapat dalam teks, seperti ulasan pengguna maupun komentar pada media sosial. Penerapan tersebut terlihat pada berbagai penelitian dalam kajian ini yang menggunakan data dari Google Play Store, Twitter, TikTok, dan platform perdagangan elektronik untuk memperoleh informasi mengenai opini atau respons pengguna (Julianti et al., 2024; Kurniawan et al., 2026; Larasati et al., 2022; Rizal Chandra Rivaldi & T.D. Wismarini, 2024)

Karakteristik data bahasa Indonesia menjadi salah satu aspek yang perlu diperhatikan dalam penerapan NLP. Data dari media sosial dan platform digital dapat mengandung kata tidak baku, slang, singkatan, simbol, emoji, serta variasi penulisan yang menyebabkan teks perlu melalui tahapan preprocessing sebelum digunakan dalam proses klasifikasi. Hal tersebut terlihat pada penelitian Larasati et al. (2022), Sari dan Suryono (2024), Julianti et al. (2024), serta Rahmat et al. (2026) yang menerapkan tahapan seperti cleaning, case folding, tokenisasi, normalisasi, stopword removal, dan stemming sebelum melakukan pemodelan (Julianti et al., 2024; Kumala Sari & Randy Suryono, 2024; Larasati et al., 2022; Seri Rahmat et al., 2026) .

Dengan demikian, analisis sentimen dalam NLP tidak hanya berkaitan dengan pemilihan algoritma klasifikasi, tetapi juga dengan bagaimana data teks dipersiapkan dan direpresentasikan. Pada pendekatan machine learning klasik, teks umumnya diubah menjadi representasi numerik seperti TF-IDF atau Bag-of-Words sebelum diproses oleh algoritma klasifikasi. Sementara itu, perkembangan model Transformer memperkenalkan representasi berbasis token dan embedding yang memungkinkan model memanfaatkan informasi kontekstual dari teks (Kurniawan et al., 2026; Tarwoto et al., 2025) .

Pendekatan Lexicon-Based dalam Analisis Sentimen

Pendekatan Lexicon-Based merupakan salah satu pendekatan yang digunakan dalam analisis sentimen dengan memanfaatkan daftar kata atau leksikon yang telah memiliki nilai atau kecenderungan sentimen tertentu. Dalam pendekatan ini, sentimen teks ditentukan berdasarkan keberadaan dan skor kata-kata yang terdapat dalam leksikon sehingga proses klasifikasi tidak sepenuhnya bergantung pada model pembelajaran mesin. Penelitian Sapanji et al. (2023) menggunakan pendekatan berbasis leksikon untuk menganalisis sentimen pada ulasan lima platform e-commerce di Indonesia, sedangkan Manullang et al. (2023) menggunakan Lexicon-Based bersama Random Forest dalam proses analisis sentimen(Manullang et al., 2023; Sapanji et al., 2023) .

Penggunaan Lexicon-Based juga dapat ditempatkan sebagai bagian dari tahapan pelabelan data sebelum klasifikasi menggunakan algoritma machine learning. Pada penelitian Manullang et al. (2023), pendekatan Lexicon-Based digunakan untuk memperoleh label atau skor sentimen sebelum data diproses menggunakan Random Forest(Manullang et al., 2023) . Pola tersebut menunjukkan bahwa pendekatan berbasis leksikon tidak selalu berdiri sendiri sebagai metode klasifikasi, tetapi dapat digunakan bersama metode pembelajaran mesin dalam suatu alur analisis sentimen.

Meskipun demikian, penggunaan leksikon memiliki keterbatasan karena hasil analisis bergantung pada kata dan nilai sentimen yang tersedia dalam leksikon. Variasi bahasa pada data digital, termasuk penggunaan istilah tidak baku dan perubahan penggunaan kata berdasarkan konteks, dapat menjadi tantangan dalam penerapan pendekatan tersebut. Oleh karena itu, perkembangan penelitian kemudian banyak memanfaatkan algoritma machine learning yang dapat mempelajari pola dari data berdasarkan fitur yang dibentuk dari teks.

Machine Learning Klasik dan Representasi TF-IDF

Machine learning klasik menjadi salah satu pendekatan yang banyak digunakan dalam penelitian analisis sentimen bahasa Indonesia. Algoritma seperti Naive Bayes, Support Vector Machine (SVM), Random Forest, Logistic Regression, dan K-Nearest Neighbors (KNN) diterapkan untuk mengklasifikasikan teks berdasarkan fitur yang telah diekstraksi dari data. Dalam 20 artikel yang dikaji, pendekatan tersebut digunakan pada berbagai objek, seperti ulasan aplikasi Dana, Shopee, SIREKAP, TikTok, Canva, SMILE Indonesia, Wattpad, serta data Twitter (Herjanto & Carudin, 2024; Idris et al., 2023; Larasati et al., 2022; Nur Adhan et al., 2024; Pratmanto et al., 2023) .

Salah satu representasi teks yang banyak digunakan pada pendekatan machine learning klasik adalah Term Frequency-Inverse Document Frequency (TF-IDF). Representasi ini mengubah teks menjadi bobot numerik berdasarkan kemunculan suatu kata dalam dokumen dan tingkat kekhasan kata tersebut dalam kumpulan dokumen. Penggunaan TF-IDF terlihat pada penelitian (Herjanto & Carudin, 2024; Julianti et al., 2024; Kumala Sari & Randy Suryono, 2024; Larasati et al., 2022; Manullang et al., 2023; Nahdhudin et al., 2024; Rhomaningtias et al., 2025) .

Beberapa penelitian juga menunjukkan bahwa pemilihan algoritma dapat menghasilkan perbedaan performa ketika model diuji pada dataset dan kondisi eksperimen yang sama. Sari dan Suryono (2024), misalnya, membandingkan Random Forest dan SVM pada data Metaverse, sedangkan Pratmanto dan Imaniawan (2023) membandingkan KNN dan Naive Bayes pada data ulasan Canva(Kumala Sari & Randy Suryono, 2024; Pratmanto et al., 2023) . Penelitian Rahmat et al. (2026) juga membandingkan SVM, Random Forest, dan Multi-Layer Perceptron (MLP) pada data ulasan Instagram, sehingga memperlihatkan bahwa perbandingan dalam dataset yang sama dapat memberikan gambaran yang lebih terkontrol mengenai perbedaan performa algoritma(Seri Rahmat et al., 2026) .

Selain pemilihan algoritma, beberapa penelitian menerapkan teknik untuk menangani karakteristik data tertentu, terutama ketidakseimbangan kelas. Sari dan Suryono (2024) menggunakan SMOTE dalam eksperimen perbandingan Random Forest dan SVM, sedangkan Larasati et al. (2022) melakukan pengujian pada data yang diseimbangkan dan memperoleh perubahan hasil evaluasi dibandingkan data awal(Kumala Sari & Randy Suryono, 2024; Larasati et al., 2022) . Sebaliknya, penelitian Adhan et al. (2024) menunjukkan bahwa penerapan SMOTE pada Random Forest tidak selalu meningkatkan seluruh metrik evaluasi, sehingga pengaruh teknik penyeimbangan data perlu dilihat berdasarkan kondisi dataset dan rancangan eksperimen yang digunakan(Nur Adhan et al., 2024) .

Deep Learning dalam Analisis Sentimen

Perkembangan analisis sentimen selanjutnya melibatkan penggunaan pendekatan deep learning yang memungkinkan model mempelajari representasi dan pola dari data secara lebih kompleks. Salah satu pendekatan yang ditemukan dalam kajian ini adalah Recurrent Neural Network (RNN) dengan arsitektur Long Short-Term Memory (LSTM). Abdurrohim dan Rahman menggunakan RNN-LSTM bersama pendekatan Aspect-Based Sentiment Analysis (ABSA) dan Named Entity Recognition (NER), sedangkan Fauzianto dan Supatman (2023) membandingkan Neural Network dengan beberapa algoritma machine learning klasik pada data Twitter(Abdurrohim & Rahman, 2024; Fauzianto & Supatman, 2023).

Penggunaan deep learning dalam penelitian yang dikaji tidak hanya diarahkan pada klasifikasi sentimen secara umum. Pada penelitian Abdurrohim dan Rahman, model digunakan bersama ABSA dan NER untuk mengolah komentar berdasarkan aspek tertentu, sedangkan penelitian Fauzianto dan Supatman membandingkan beberapa model untuk melihat performa klasifikasi pada data mengenai isu teknologi dan ketenagakerjaan. Hal tersebut menunjukkan bahwa deep learning mulai digunakan untuk menangani permasalahan yang membutuhkan pemodelan pola teks yang lebih kompleks dibandingkan pendekatan berbasis fitur konvensional.

Namun, keberadaan deep learning dalam kajian ini belum menggantikan seluruh penggunaan machine learning klasik. Sebagian penelitian masih menggunakan algoritma klasik karena pendekatan tersebut dapat diterapkan dengan representasi fitur seperti TF-IDF dan memiliki alur pemrosesan yang relatif terstruktur. Perkembangan berikutnya dalam kajian ini lebih terlihat pada penggunaan model pre-trained Transformer yang memanfaatkan representasi bahasa berbasis konteks.

Pre-trained Transformer untuk Analisis Sentimen Bahasa Indonesia

Pre-trained Transformer merupakan perkembangan penting dalam pemrosesan bahasa karena model dapat memanfaatkan representasi kontekstual yang diperoleh melalui proses pre-training. Dalam kajian analisis sentimen bahasa Indonesia, model seperti IndoBERT dan IndoBERTweet mulai digunakan pada data ulasan aplikasi maupun media sosial. Tarwoto et al. (2025) menerapkan fine-tuning IndoBERT pada ulasan Mobile JKN, Ningrum et al. (2025) menggunakan IndoBERT pada komentar TikTok, sedangkan Asher dan Rachmawati (2025) menerapkan IndoBERT pada ulasan Instagram(Asher & Rachmawati, 2025; Ningrum et al., 2025; Tarwoto et al., 2025) .

Penggunaan Transformer juga ditemukan pada data dengan karakteristik bahasa yang lebih kompleks. Tribuana et al. (2025) membandingkan representasi TF-IDF dengan pretrained IndoBERTweet pada data layanan publik di Twitter, sedangkan Kurniawan et al. (2026) menggunakan XLM-RoBERTa dan IndoBERTweet pada komentar TikTok yang menggunakan bahasa Indonesia dan Inggris(Kurniawan et al., 2026; Tribuana et al., 2025) . Penggunaan model tersebut menunjukkan bahwa penelitian NLP mulai diarahkan pada pengolahan data yang memiliki variasi bahasa, termasuk penggunaan bahasa tidak baku dan code-switching.

Penelitian berbasis Transformer dalam kajian ini menunjukkan variasi hasil evaluasi pada dataset yang berbeda. Tarwoto et al. (2025) dan Ningrum et al. (2025) melaporkan performa tinggi pada dataset masing-masing, sedangkan Asher dan Rachmawati (2025) memperoleh hasil yang lebih rendah pada data ulasan bintang lima Instagram(Asher & Rachmawati, 2025; Ningrum et al., 2025; Tarwoto et al., 2025) . Perbedaan tersebut menunjukkan bahwa penggunaan arsitektur Transformer tidak secara otomatis menghasilkan nilai evaluasi yang sama karena performa model tetap berkaitan dengan karakteristik dataset, distribusi kelas, proses preprocessing, dan rancangan eksperimen.

Perkembangan Transformer juga memperluas penggunaan NLP dari sekadar klasifikasi sentimen menuju analisis yang lebih kompleks. Kurniawan et al. (2026), misalnya, menggabungkan model Transformer dengan analisis deret waktu dan pemodelan topik untuk mengamati perubahan isu pada data TikTok multibahasa(Kurniawan et al., 2026) . Ningrum (2025) juga menerapkan model IndoBERT dalam pengembangan bot komentar otomatis sehingga hasil analisis sentimen dapat digunakan dalam suatu sistem yang memberikan respons terhadap komentar(Ningrum et al., 2025).

Perkembangan Metode dan Research Gap

Berdasarkan 20 artikel yang dikaji, perkembangan metode analisis sentimen bahasa Indonesia dapat dilihat melalui perubahan pendekatan dari Lexicon-Based, machine learning klasik, deep learning, hingga pre-trained Transformer. Perubahan tersebut tidak hanya terjadi pada algoritma klasifikasi, tetapi juga pada cara teks direpresentasikan, mulai dari representasi berbasis fitur seperti TF-IDF dan Bag-of-Words hingga representasi kontekstual pada model Transformer. Perkembangan ini menunjukkan bahwa penelitian NLP semakin memperhatikan kemampuan sistem dalam menangani variasi bahasa dan konteks data yang digunakan.

Meskipun demikian, hasil penelitian terdahulu belum dapat digunakan untuk menyatakan bahwa satu kelompok metode selalu menghasilkan performa yang lebih tinggi daripada kelompok lainnya. Nilai evaluasi yang dilaporkan berasal dari dataset, jumlah data, distribusi kelas, preprocessing, pembagian data, serta konfigurasi eksperimen yang berbeda, sehingga perbandingan langsung antarpenelitian memiliki keterbatasan. Oleh sebab itu, nilai akurasi dan metrik lainnya perlu dipahami sebagai hasil pada kondisi eksperimen masing-masing, bukan sebagai dasar untuk membuat pemeringkatan universal terhadap algoritma.

Selain perbedaan performa, terdapat variasi yang cukup jelas pada tujuan dan karakteristik data penelitian. Sebagian penelitian berfokus pada ulasan aplikasi dan produk, sementara penelitian lainnya menggunakan data media sosial dengan bahasa yang lebih informal, data multibahasa, maupun data yang membutuhkan analisis aspek dan pemodelan topik. Keragaman tersebut memperlihatkan bahwa perkembangan NLP dalam analisis sentimen tidak hanya didorong oleh pencarian algoritma baru, tetapi juga oleh kebutuhan untuk menyesuaikan metode dengan karakteristik permasalahan dan data.

Berdasarkan kondisi tersebut, masih diperlukan kajian yang memetakan perkembangan metode analisis sentimen bahasa Indonesia secara lebih terintegrasi. Dua puluh penelitian yang digunakan dalam studi ini memberikan bahan untuk melihat hubungan antara jenis metode, representasi teks, karakteristik data, dan hasil evaluasi pada berbagai penelitian. Kajian komparatif ini kemudian digunakan untuk mengidentifikasi pola perkembangan metode NLP

dari pendekatan berbasis leksikon dan machine learning klasik hingga penggunaan deep learning dan pre-trained Transformer.

B. Metodologi

Penelitian ini menggunakan metode studi literatur dengan pendekatan deskriptif-komparatif untuk mengkaji perkembangan metode *Natural Language Processing* (NLP) dalam analisis sentimen bahasa Indonesia. Studi literatur digunakan untuk mengumpulkan dan mensintesis hasil penelitian terdahulu sehingga dapat memberikan gambaran mengenai perkembangan suatu bidang penelitian dari berbagai hasil studi yang telah dilakukan (Snyder, 2021) . Pendekatan deskriptif digunakan untuk mengidentifikasi karakteristik metode yang digunakan pada setiap penelitian, sedangkan pendekatan komparatif digunakan untuk membandingkan perkembangan metode, representasi data, serta hasil evaluasi yang dilaporkan dalam penelitian-penelitian yang dikaji.

Kata Kunci Pencarian

Pencarian literatur dilakukan dengan menggunakan beberapa kata kunci yang berkaitan dengan topik penelitian, yaitu "analisis sentimen bahasa Indonesia", "*Natural Language Processing", "sentiment analysis", "machine learning", "deep learning", "Transformer", "IndoBERT", "Naive Bayes", "Support Vector Machine",* dan *"Random Forest*". Kata kunci tersebut digunakan untuk memperoleh artikel yang membahas penerapan metode NLP dan algoritma pembelajaran mesin dalam analisis sentimen terhadap data berbahasa Indonesia.

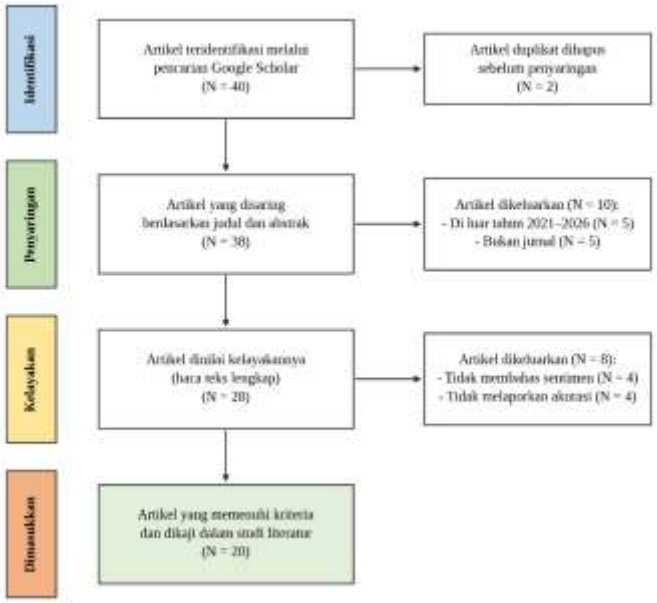
Sumber Data

Sumber data penelitian berupa 20 artikel ilmiah yang telah ditetapkan sebagai bahan kajian. Artikel-artikel tersebut membahas analisis sentimen pada berbagai objek dan platform, seperti Google Play Store, Twitter, TikTok, aplikasi digital, serta media sosial lainnya. Kedua puluh artikel tersebut digunakan sebagai sumber utama untuk mengidentifikasi perkembangan metode analisis sentimen, mulai dari pendekatan berbasis *lexicon*, algoritma *machine learning* klasik, hingga pendekatan berbasis *Transformer*. Daftar artikel yang digunakan sebagai sumber kajian ditetapkan berdasarkan dokumen sumber yang telah disusun sebelumnya dan seluruh artikel tetap dipertahankan dalam penelitian ini.

Kriteria Pemilihan Artikel

Proses pemilihan literatur dalam penelitian ini dilakukan secara sistematis mengikuti pedoman diagram alur PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*). Proses ini terdiri dari empat tahapan utama, yaitu identifikasi, penyaringan, uji kelayakan, dan penentuan artikel yang dimasukkan. Pada tahap pertama (identifikasi), pencarian awal dilakukan melalui pangkalan data Google Scholar dan berhasil mengidentifikasi sebanyak 40 artikel. Dari jumlah tersebut, 2 artikel dihapus karena merupakan publikasi ganda (duplikat), sehingga menyisakan 38 artikel. Pada tahap kedua (penyaringan), evaluasi awal dilakukan dengan meninjau kesesuaian judul dan abstrak. Sebanyak 10 artikel dikeluarkan dari seleksi karena berada di luar rentang tahun publikasi 2021–2026 (5 artikel) dan bukan merupakan artikel jurnal/prosiding (5 artikel).

Selanjutnya, pada tahap ketiga (kelayakan), 28 artikel yang tersisa dievaluasi secara menyeluruh dengan membaca teks lengkap (*full text*). Hasilnya, sebanyak 8 artikel dieksklusi karena tidak berfokus pada metode NLP untuk analisis sentimen (4 artikel) dan tidak melaporkan metrik evaluasi seperti akurasi (4 artikel). Pada tahap akhir, didapatkan sebanyak 20 artikel yang sepenuhnya memenuhi kriteria dan secara resmi dikaji dalam studi literatur ini. Alur lengkap proses seleksi ini dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur PRISMA

Prosedur Analisis Data

Analisis dilakukan melalui beberapa tahap. Tahap pertama adalah identifikasi artikel, yaitu mengumpulkan informasi utama dari 20 artikel yang telah ditetapkan. Informasi yang dicatat meliputi nama penulis, tahun publikasi, objek atau platform data, metode atau algoritma yang digunakan, teknik pengolahan atau representasi teks, serta hasil evaluasi yang dilaporkan.

Tahap kedua adalah pengelompokan artikel berdasarkan pendekatan yang digunakan. Artikel dikelompokkan ke dalam beberapa perkembangan metode, yaitu pendekatan berbasis *lexicon*, *machine learning* klasik seperti Naive Bayes, Support Vector Machine (SVM), dan Random Forest, serta pendekatan *deep learning* dan *Transformer* seperti BERT dan IndoBERT. Pengelompokan ini digunakan untuk melihat perubahan pendekatan dalam analisis sentimen dari metode yang menggunakan representasi teks berbasis fitur menuju metode yang mampu memanfaatkan representasi kontekstual.

Tahap ketiga adalah analisis deskriptif dan komparatif. Analisis deskriptif digunakan untuk menjelaskan karakteristik dan penerapan masing-masing metode berdasarkan artikel yang dikaji. Selanjutnya, analisis komparatif digunakan untuk melihat perbedaan metode, objek data, teknik pengolahan teks, serta hasil evaluasi yang dilaporkan. Perbandingan nilai akurasi atau metrik lainnya dilakukan secara deskriptif dan tidak dimaksudkan sebagai perbandingan langsung antarpenelitian apabila dataset, jumlah data, pembagian data, maupun kondisi eksperimen yang digunakan berbeda. Hal ini penting karena perbedaan rancangan penelitian dapat memengaruhi hasil evaluasi masing-masing model.

Tahap terakhir adalah sintesis hasil kajian. Hasil dari setiap kelompok metode kemudian disintesiskan untuk menggambarkan perkembangan penggunaan NLP dalam analisis sentimen bahasa Indonesia. Sintesis tidak hanya mempertimbangkan nilai evaluasi, tetapi juga memperhatikan karakteristik data, representasi teks, metode yang digunakan, serta perkembangan kemampuan model dalam menangani konteks bahasa. Pendekatan sintesis seperti ini sesuai dengan fungsi studi literatur untuk mengintegrasikan temuan dari berbagai penelitian dan melihat pola perkembangan pengetahuan dalam suatu bidang (Snyder, 2021).

C. Hasil dan Pembahasan

Hasil Kajian Literatur

Hasil kajian diperoleh dari 20 artikel ilmiah yang membahas penerapan *Natural Language Processing* (NLP) dalam analisis sentimen pada berbagai sumber data berbahasa Indonesia. Artikel-artikel tersebut menggunakan objek yang beragam, mulai dari ulasan aplikasi pada Google Play Store hingga data Twitter, TikTok, dan platform perdagangan elektronik, dengan metode yang mencakup pendekatan berbasis leksikon, *machine learning* klasik, *deep learning*, dan model *Transformer*. Keragaman objek, jumlah data, metode, serta teknik pengolahan teks

menunjukkan bahwa penerapan NLP dalam analisis sentimen berkembang sesuai dengan karakteristik data dan tujuan masing-masing penelitian.

Tabel 1. Hasil Kajian 20 Artikel Analisis Sentimen Bahasa Indonesia

No.	Penulis & Tahun	Objek/ Data	Metode Utama	Preprocessing/Representasi	Hasil Utama
1	Larasati et al. (2022)	1.354 ulasan Dana di Google Play Store	Random Forest, 3 kelas	Cleaning, tokenisasi, stopword removal, normalisasi, stemming, TF-IDF	Data seimbang menghasilkan akurasi 84%, precision 84%, recall 84%, F1-score 84%.
2	Sari & Suryono (2024)	6.348 tweet Metaverse; 5.857 setelah cleaning	SVM dan Random Forest	Cleaning, case folding, tokenisasi, stopword removal, stemming, TF-IDF, SMOTE	Random Forest 91%; SVM 90%.
3	Nahdhudin et al. (2024/2025)	100.000 ulasan TikTok di Google Play Store	Multinomial Naive Bayes	Tokenisasi, stopword removal, stemming, lowercasing, TF-IDF	Akurasi 88,2%, precision 89,1%, recall 87,4%, F1-score 88,2%.
4	Herjanto & Carudin (2024)	5.000 ulasan SIREKAP; eksperimen 1.000 data seimbang	Random Forest	Cleaning, case folding, tokenisasi, stopword removal, stemming, TF-IDF	Akurasi 74%, precision 75%, recall 74%, F1-score 74%.
5	Idris et al. (2023)	3.000 ulasan Shopee	SVM	Slang normalization, case folding, stopword removal, cleansing	Akurasi terbaik 98%, dengan precision, recall, dan F1-score sekitar 98%.
6	Rivaldi & Wismarini (2024)	410 ulasan produk Zalika Store 88 di Shopee	Logistic Regression, 3 kelas	Tokenisasi, stopword removal, stemming & lemmatization, TF-IDF	Akurasi klasifikasi 87%.
7	Julianti et al. (2024)	5.500 data Twitter; 952 data bersih tentang judi online	Naive Bayes dengan pelabelan Lexicon	Cleaning, lowercase, tokenisasi, stopword removal, stemming, TF-IDF	Akurasi 76%, precision 58%, recall 76%.
8	Tribuana et al. (2025)	4.357 tweet layanan BPJS Ketenagakerjaan; 4.102 bersih	NB, SVM, RF, Logistic Regression; TF-IDF vs IndoBERTweet	Cleaning, normalisasi, tokenisasi, stopword removal, stemming	IndoBERTweet + Logistic Regression: akurasi 86,5%, F1-score 85,6%.
9	Fauzianto & Supatman (2023)	161 tweet tentang Tech Winter dan layoff	SVM, NB, NN, Logistic Regression, RF	Cleaning, normalisasi, stopword removal, stemming, TF-IDF + Chi-Square	Pada testing SVM dan NB 94,44%; pada cross-validation NN mencapai mean accuracy 95,56%.

No.	Penulis & Tahun	Objek/ Data	Metode Utama	Preprocessing/Representasi	Hasil Utama
10	Abdurrohim & Rahman (2024)	75.000 komentar tentang kebijakan Omnibus Law Cipta Kerja	RNN-LSTM, ABSA, NER	Stopword removal, tokenisasi, stemming, anotasi manual	Distribusi sentimen: 51% negatif, 29% netral, 20% positif; analisis juga dilakukan berdasarkan aspek.
11	Tarwoto et al. (2025)	5.404 ulasan Mobile JKN	Fine-tuning IndoBERT	Cleaning, lowercasing, tokenisasi BERT, stopword removal, stemming, koreksi ejaan	Akurasi rata-rata 97,28%; akurasi terbaik 98,27%, F1-score 98,27%.
12	Pratmanto & Imaniawan (2023)	2.000 ulasan Canva	Naive Bayes dan KNN	Case folding, tokenisasi, stopword removal, stemming	KNN 83,92%; Naive Bayes 77,41%.
13	Rhomaningtias et al. (2025)	383/382 ulasan SMILE Indonesia	Multinomial Naive Bayes dan SVM	Cleaning, case folding, tokenisasi, stopword removal, stemming, normalisasi, TF-IDF	Naive Bayes 79%; SVM 75%; perbedaan tidak signifikan berdasarkan uji McNemar.
14	Sapanji et al. (2023)	10.000 ulasan lima platform e-commerce	Lexicon-based sentiment analysis	Cleaning, lowercasing, stopword removal, Document Term Matrix	Positif 36,64%, netral 22,82%, negatif 40,54%.
15	Rahmat et al. (2026)	7.236 ulasan Instagram dari 10.000 data awal	SVM, Random Forest, MLP	Cleaning, case folding, slang normalization, tokenisasi, stopword removal, stemming, BoW	SVM 90,7%; MLP 89,2%; Random Forest 81,5%.
16	Ningrum et al. (2025)	241 komentar untuk klasifikasi sentimen TikTok UMKM	IndoBERT dan LSTM	Cleaning, lowercasing, normalisasi, augmentasi, tokenisasi BERT	IndoBERT 97,51% untuk sentimen; LSTM 73,47%.
17	Asher & Rachmawati (2025)	1.000 ulasan bintang lima Instagram	IndoBERT	Cleaning, lowercasing, stopword removal, normalisasi, stemming, tokenisasi BERT	Akurasi keseluruhan 79,5%; F1 positif 0,877, negatif 0,768, netral 0,505.
18	Kurniawan et al. (2026)	Komentar TikTok multibahasa Indonesia–Inggris	XLM-RoBERTa dan IndoBERTweet; pembanding mBERT,	Noise cleaning, emoji-to-text, slang normalization, tokenisasi	Model utama mencapai akurasi 93,4%; mBERT 89,1%, LSTM

No.	Penulis & Tahun	Objek/ Data	Metode Utama	Preprocessing/Representasi	Hasil Utama
			LSTM, SVM		82,3%, SVM 78,8%.
19	Manullang et al. (2023)	3.862 tweet terkait calon Presiden 2024	Lexicon-Based + Random Forest	Cleaning, case folding, tokenisasi, normalisasi, stopword removal, stemming, TF-IDF	Random Forest 94,56%; pendekatan Lexicon-Based 88%.
20	Adhan et al. (2024)	9.975 ulasan Wattpad; 8.743 data bersih	Random Forest dan Random Forest + SMOTE	Cleaning, case folding, tokenisasi, normalisasi, stopword removal, stemming, TF-IDF	Random Forest tanpa SMOTE: akurasi 84,05%; dengan SMOTE: 83,65%.

Sumber: diolah dari 20 artikel yang menjadi bahan kajian.

Berdasarkan Tabel 1, sebagian besar penelitian menerapkan tahapan preprocessing yang relatif serupa, seperti *case folding*, *cleaning*, tokenisasi, *stopword removal*, normalisasi, dan *stemming*. Representasi teks berbasis TF-IDF juga banyak digunakan pada penelitian yang menerapkan algoritma *machine learning* klasik, sedangkan penelitian berbasis BERT menggunakan tokenisasi dan representasi yang disesuaikan dengan arsitektur *Transformer*. Pola tersebut menunjukkan bahwa perkembangan metode tidak hanya terlihat dari perubahan algoritma klasifikasi, tetapi juga dari perubahan cara teks direpresentasikan sebelum diproses oleh model.

Perkembangan Metode Analisis Sentimen

Hasil kajian menunjukkan bahwa pendekatan analisis sentimen dalam 20 artikel dapat dipetakan ke dalam beberapa kelompok berdasarkan karakteristik metode yang digunakan. Kelompok pertama menggunakan pendekatan berbasis leksikon, kelompok kedua menggunakan algoritma *machine learning* klasik, sedangkan kelompok berikutnya menggunakan *deep learning* dan model *Transformer*. Pembagian tersebut digunakan untuk memudahkan analisis perkembangan pendekatan, bukan untuk menyatakan bahwa setiap penelitian mengikuti urutan perkembangan yang sama.

Pendekatan berbasis leksikon terlihat pada penelitian Sapanji (2023) dan Manullang (2023), meskipun penelitian Manullang et al. juga menggabungkannya dengan Random Forest untuk tahap klasifikasi(Manullang et al., 2023; Sapanji et al., 2023) . Pada penelitian Sapanji (2023), klasifikasi dilakukan melalui pemberian skor berdasarkan leksikon kata positif dan negatif terhadap 10.000 ulasan dari lima platform e-commerce(Sapanji et al., 2023) . Sementara itu, Manullang (2023) menggunakan pendekatan Lexicon-Based untuk otomatisasi pelabelan atau pemberian skor sentimen sebelum data diproses menggunakan Random Forest, sehingga memperlihatkan penggunaan pendekatan leksikon bersama algoritma pembelajaran mesin(Manullang et al., 2023) .

Pendekatan *machine learning* klasik merupakan kelompok yang paling banyak ditemukan dalam sumber kajian. Larasati (2022), Herjanto dan Carudin (2024), serta Adhan (2024) menggunakan Random Forest, sedangkan Sari dan Suryono (2024), Idris (2023), dan Rahmat (2026) menggunakan atau membandingkan SVM dengan algoritma lain (Herjanto & Carudin, 2024; Idris et al., 2023; Kumala Sari & Randy Suryono, 2024; Larasati et al., 2022; Nur Adhan et al., 2024; Seri Rahmat et al., 2026) . Selain itu, Naive Bayes, KNN, Logistic Regression, dan MLP juga digunakan pada beberapa penelitian, sehingga menunjukkan bahwa pemilihan algoritma dalam analisis sentimen masih sangat beragam dan bergantung pada karakteristik dataset serta rancangan eksperimen masing-masing penelitian.

Penggunaan TF-IDF menjadi salah satu pola yang paling konsisten pada kelompok *machine learning* klasik. Larasati (2022) menggunakan TF-IDF sebelum proses klasifikasi Random Forest, sedangkan Sari dan Suryono (2024) menggunakan TF-IDF pada perbandingan SVM dan Random Forest (Kumala Sari & Randy Suryono, 2024; Larasati et al., 2022) . Pola serupa juga ditemukan pada penelitian Nahdhudin, Herjanto dan Carudin, Julianti, serta Rhomaningtias, sehingga TF-IDF dapat dilihat sebagai salah satu bentuk representasi teks yang banyak

digunakan pada pendekatan klasik dalam kumpulan penelitian ini(Herjanto & Carudin, 2024; Julianti et al., 2024; Nahdhudin et al., 2024; Rhomaningtias et al., 2025) .

Selain pemilihan algoritma, beberapa penelitian menunjukkan bahwa karakteristik data dan teknik penyeimbangan kelas dapat memengaruhi hasil evaluasi. Larasati et al. (2022), misalnya, memperoleh akurasi 76% pada data awal yang tidak seimbang dan 84% setelah menggunakan data yang diseimbangkan dengan konfigurasi eksperimen tertentu. Sebaliknya, pada penelitian Adhan (2024), penerapan SMOTE tidak meningkatkan akurasi Random Forest karena akurasi turun dari 84,05% menjadi 83,65%, meskipun precision untuk kelas minoritas meningkat(Nur Adhan et al., 2024) . Temuan tersebut menunjukkan bahwa teknik penanganan ketidakseimbangan data tidak selalu menghasilkan peningkatan pada seluruh metrik evaluasi.

Beberapa penelitian mulai menggunakan *deep learning* untuk menangani data teks yang lebih kompleks. Abdurrohim dan Rahman menggunakan arsitektur RNN-LSTM serta menggabungkannya dengan pendekatan seperti *Aspect-Based Sentiment Analysis* (ABSA) dan *Named Entity Recognition* (NER) untuk menganalisis komentar mengenai kebijakan publik. Fauzianto dan Supatman juga membandingkan Neural Network dengan beberapa algoritma klasik, dan memperoleh *mean accuracy* 95,56% pada *cross-validation* untuk Neural Network. Hasil tersebut memperlihatkan bahwa penggunaan *deep learning* dalam penelitian yang dikaji tidak hanya diarahkan pada klasifikasi sentimen, tetapi juga pada pengolahan konteks dan analisis informasi yang lebih kompleks.

Perkembangan berikutnya terlihat pada penggunaan model *pre-trained Transformer*, khususnya IndoBERT dan IndoBERTweet. Tarwoto (2025) menggunakan *fine-tuning* IndoBERT pada ulasan Mobile JKN dan memperoleh akurasi terbaik 98,27% dengan F1-score 98,27%, sedangkan Ningrum (2025) memperoleh akurasi 97,51% untuk klasifikasi sentimen menggunakan IndoBERT(Ningrum et al., 2025; Tarwoto et al., 2025). Pada penelitian Tribuana (2025), representasi pretrained IndoBERTweet yang dipadukan dengan Logistic Regression menghasilkan akurasi 86,5%, sementara pendekatan TF-IDF dengan Random Forest mencapai akurasi tertinggi 78,5% dalam penelitian tersebut(Tribuana et al., 2025) .

Penggunaan model Transformer juga terlihat pada data yang memiliki karakteristik lebih kompleks, seperti data media sosial dan data multibahasa. Kurniawan (2026) menggunakan XLM-RoBERTa dan IndoBERTweet pada komentar TikTok berbahasa Indonesia dan Inggris serta menggabungkannya dengan analisis deret waktu dan pemodelan topik(Kurniawan et al., 2026) . Model utama yang digunakan dalam penelitian tersebut memperoleh akurasi 93,4%, sedangkan mBERT, LSTM, dan SVM masing-masing memperoleh 89,1%, 82,3%, dan 78,8%. Hasil ini memperlihatkan bahwa model Transformer dalam kumpulan kajian tidak hanya digunakan untuk klasifikasi sentimen, tetapi juga mulai diintegrasikan dengan analisis isu dan pemodelan topik.

Meskipun beberapa penelitian Transformer melaporkan nilai evaluasi yang tinggi, angka tersebut tidak dapat langsung digunakan untuk menyatakan bahwa Transformer selalu lebih baik daripada metode klasik. Perbedaan jumlah data, objek penelitian, distribusi kelas, teknik preprocessing, pembagian dataset, dan metrik evaluasi menyebabkan nilai akurasi dari penelitian yang berbeda tidak sepenuhnya berada pada kondisi eksperimen yang sama. Oleh karena itu, hasil kajian ini lebih tepat digunakan untuk menggambarkan perkembangan dan variasi penggunaan metode NLP daripada membuat peringkat performa antarpenelitian.

Perbandingan Karakteristik Pendekatan

Perbandingan 20 penelitian menunjukkan adanya perbedaan mendasar pada cara masing-masing pendekatan merepresentasikan teks. Pada metode klasik, teks umumnya diubah menjadi fitur numerik melalui TF-IDF atau Bag-of-Words sebelum diberikan kepada algoritma klasifikasi seperti SVM, Random Forest, Naive Bayes, atau Logistic Regression. Pada model Transformer, representasi teks diperoleh melalui mekanisme embedding dan *self-attention*, dengan tokenisasi khusus yang disesuaikan dengan arsitektur model.

Dari sisi preprocessing, sebagian besar penelitian tetap melakukan tahapan dasar seperti pembersihan teks, *case folding*, tokenisasi, penghapusan *stopword*, normalisasi, dan stemming. Tahapan tersebut terlihat pada penelitian Larasati (2022), Sari dan Suryono (2024), Julianti (2024), serta beberapa penelitian lainnya yang menggunakan TF-IDF sebagai representasi teks;(Julianti et al., 2024; Kumala Sari & Randy Suryono, 2024; Larasati et al., 2022) . Pada penelitian berbasis Transformer, preprocessing juga tetap dilakukan, tetapi ditambah dengan tokenisasi khusus BERT serta proses encoding seperti *input IDs* dan *attention mask*, sebagaimana terlihat pada penelitian Tarwoto (2025) dan Ningrum (2025)(Ningrum et al., 2025; Tarwoto et al., 2025) .

Perbedaan lain terdapat pada kemampuan metode dalam menangani karakteristik data tertentu. Pendekatan klasik relatif banyak digunakan untuk data ulasan aplikasi dengan preprocessing dan representasi fitur yang terstruktur, sedangkan beberapa model *deep learning* dan Transformer digunakan pada data media sosial yang lebih beragam dan memiliki konteks bahasa yang lebih kompleks. Contohnya, Kurniawan (2026) menangani data multibahasa dan *code-switching*, sedangkan Tribuana (2025) membandingkan representasi TF-IDF dengan pretrained embedding IndoBERTweet (Kurniawan et al., 2026; Tribuana et al., 2025) . Hal ini menunjukkan bahwa perkembangan metode juga berkaitan dengan kebutuhan untuk menangani variasi bahasa, konteks, dan karakteristik data yang semakin kompleks.

Perbandingan Hasil Evaluasi

Hasil evaluasi dalam 20 artikel menunjukkan rentang performa yang cukup beragam. Pada penelitian yang menggunakan algoritma klasik, akurasi yang dilaporkan antara lain 74% pada Random Forest untuk data SIREKAP, 76% pada Naive Bayes untuk data judi online, 84% pada Random Forest untuk data Dana yang telah diseimbangkan, dan 98% pada SVM untuk data ulasan Shopee. Perbedaan tersebut memperlihatkan bahwa performa suatu algoritma tidak hanya ditentukan oleh nama algoritmanya, tetapi juga berkaitan dengan dataset, preprocessing, distribusi kelas, serta konfigurasi eksperimen yang digunakan.

Perbandingan dalam satu penelitian memberikan gambaran yang lebih langsung karena model diuji pada dataset dan kondisi eksperimen yang sama. Sari dan Suryono (2024), misalnya, memperoleh akurasi 91% untuk Random Forest dan 90% untuk SVM pada data Metaverse yang sama, sedangkan Pratmanto dan Imaniawan (2023) memperoleh akurasi 83,92% untuk KNN dan 77,41% untuk Naive Bayes pada data Canva(Kumala Sari & Randy Suryono, 2024; Pratmanto et al., 2023) . Pada penelitian Rahmat (2026), SVM memperoleh akurasi 90,7%, MLP 89,2%, dan Random Forest 81,5% pada dataset Instagram yang sama, sehingga perbandingan internal seperti ini lebih informatif untuk melihat perbedaan performa model dalam kondisi eksperimen yang sama(Seri Rahmat et al., 2026) .

Pada kelompok Transformer, beberapa penelitian melaporkan performa yang relatif tinggi, tetapi hasilnya tetap perlu dibaca berdasarkan konteks masing-masing dataset. Tarwoto (2025) memperoleh akurasi terbaik 98,27% menggunakan IndoBERT, sedangkan Ningrum (2025) memperoleh akurasi 97,51% untuk klasifikasi sentimen dengan IndoBERT (Ningrum et al., 2025; Tarwoto et al., 2025) . Namun, Asher dan Rachmawati (2025) memperoleh akurasi 79,5% pada data ulasan bintang lima Instagram, yang menunjukkan bahwa penggunaan model yang sama atau serupa tidak secara otomatis menghasilkan nilai evaluasi yang sama(Asher & Rachmawati, 2025) .

Dengan demikian, nilai akurasi dalam kajian ini lebih tepat digunakan sebagai indikator hasil masing-masing penelitian daripada sebagai dasar pemeringkatan seluruh model. Perbandingan lintas penelitian harus mempertimbangkan perbedaan dataset, jumlah data, distribusi kelas, preprocessing, pembagian data, serta tujuan klasifikasi yang digunakan. Berdasarkan pertimbangan tersebut, perkembangan metode NLP dalam analisis sentimen bahasa Indonesia dapat dipahami sebagai perkembangan dalam representasi teks dan kemampuan pemodelan konteks, bukan sekadar peningkatan angka akurasi dari satu generasi metode ke generasi berikutnya.

Sintesis Hasil Kajian

Secara keseluruhan, 20 penelitian menunjukkan bahwa analisis sentimen bahasa Indonesia telah menggunakan pendekatan yang semakin beragam, mulai dari Lexicon-Based, *machine learning* klasik, *deep learning*, hingga model *pre-trained Transformer*. Pendekatan klasik masih banyak digunakan karena dapat diterapkan dengan representasi fitur seperti TF-IDF dan berbagai algoritma klasifikasi, sedangkan Transformer mulai digunakan untuk memperoleh representasi teks yang lebih kontekstual. Perubahan tersebut menunjukkan adanya perluasan pendekatan dalam penelitian NLP, terutama ketika penelitian berhadapan dengan data media sosial, bahasa tidak baku, data multibahasa, dan konteks yang lebih kompleks.

Meskipun demikian, hasil kajian tidak menunjukkan bahwa satu metode dapat dianggap paling unggul untuk seluruh kondisi penelitian. Beberapa algoritma klasik menghasilkan akurasi tinggi pada dataset tertentu, sementara beberapa model Transformer juga menghasilkan performa tinggi pada penelitian lainnya, tetapi kondisi pengujiannya berbeda. Oleh karena itu, pemilihan metode analisis sentimen perlu mempertimbangkan karakteristik data, tujuan penelitian, jumlah data, distribusi kelas, kebutuhan interpretasi, serta rancangan evaluasi yang digunakan.

Temuan lainnya menunjukkan bahwa preprocessing dan representasi data tetap memiliki peran penting dalam keseluruhan proses analisis sentimen. Pada penelitian berbasis *machine learning*, TF-IDF masih menjadi representasi yang dominan, sedangkan model Transformer memanfaatkan tokenisasi dan embedding yang disesuaikan dengan arsitektur model. Dengan demikian, perkembangan NLP dalam analisis sentimen bahasa Indonesia tidak hanya ditandai oleh perubahan algoritma klasifikasi, tetapi juga oleh perkembangan cara sistem merepresentasikan dan memahami informasi yang terdapat dalam teks.

D. Kesimpulan

Berdasarkan hasil studi literatur terhadap 20 artikel, perkembangan metode Natural Language Processing (NLP) dalam analisis sentimen bahasa Indonesia menunjukkan adanya perubahan pendekatan dari metode berbasis Lexicon-Based dan machine learning klasik menuju penggunaan deep learning dan pre-trained Transformer. Pendekatan machine learning klasik seperti Naive Bayes, Support Vector Machine (SVM), Random Forest, Logistic Regression, dan K-Nearest Neighbors (KNN) banyak digunakan dengan representasi teks seperti TF-IDF dan Bag-of-Words. Perkembangan selanjutnya ditunjukkan melalui penggunaan RNN-LSTM, IndoBERT, IndoBERTweet, dan XLM-RoBERTa pada berbagai jenis data ulasan aplikasi maupun media sosial.

Hasil kajian menunjukkan bahwa setiap pendekatan memiliki karakteristik dan hasil evaluasi yang berbeda. Metode *machine learning* klasik dapat menghasilkan performa yang tinggi pada beberapa penelitian, tetapi hasilnya tetap berkaitan dengan *preprocessing*, representasi fitur, karakteristik dataset, dan konfigurasi eksperimen yang digunakan. Di sisi lain, model *pre-trained* Transformer juga menunjukkan hasil yang bervariasi pada penelitian yang dikaji, sehingga penggunaan Transformer tidak dapat dipandang sebagai jaminan bahwa model tersebut selalu menghasilkan performa tertinggi pada semua jenis data. Perbandingan terhadap 20 artikel mengindikasikan bahwa perbedaan dataset dan rancangan eksperimen menjadi faktor krusial dalam interpretasi hasil. Nilai akurasi dari penelitian yang berbeda tidak dapat dibandingkan secara langsung tanpa mempertimbangkan kondisi pengujiannya. Oleh karena itu, perbandingan metode akan lebih objektif apabila dilakukan pada dataset dan kondisi eksperimen yang setara. Untuk penelitian di masa mendatang, disarankan adanya standarisasi dataset *benchmark* berbahasa Indonesia agar evaluasi model NLP, baik klasik maupun *deep learning*, dapat diukur dan dibandingkan secara lebih akurat.

Secara keseluruhan, evolusi metode NLP dalam analisis sentimen bahasa Indonesia tidak hanya ditandai oleh perubahan algoritma, tetapi juga oleh perkembangan cara teks diproses dan direpresentasikan. Pergeseran dari fitur berbasis TF-IDF menuju representasi kontekstual pada model Transformer menunjukkan adanya perkembangan dalam pendekatan pengolahan bahasa, sementara variasi hasil penelitian menunjukkan bahwa pemilihan metode tetap perlu disesuaikan dengan karakteristik data dan tujuan analisis. Hasil studi ini memberikan pemetaan mengenai perkembangan metode NLP dari pendekatan berbasis leksikon, machine learning klasik, deep learning, hingga pre-trained Transformer berdasarkan 20 artikel yang dikaji.

E. Referensi

Abdurrohim, I., & Rahman, A. (2024). *Penerapan Natural Language Processing Untuk Analisis Sentimen Terhadap Kebijakan Pemerintah*.

Asher, J., & Rachmawati, E. P. (2025). *Analisis Sentimen Ulasan Bintang Lima Aplikasi Instagram Di Google Play Store Menggunakan Indobert*. *30*(2), 2025.

Fauzianto, R. A., & Supatman. (2023). Analisis Sentimen Opini Masyarakat Terhadap Tech Winter Pada Twitter Menggunakan Natural Language Processing. *Jurnal Syntax Admiration*, *3*(9), 1577–1585. Https://Doi.Org/10.46799/Jsa.V3i9.909

Herjanto, M. F. Y., & Carudin, C. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifer. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). Https://Doi.Org/10.23960/Jitet.V12i2.4192

Idris, K., Mustofa, A., & Salihi, A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (Svm). *Journal Of Electrical And Electronics Engineering*, *5*(6), 823–848. Https://Doi.Org/10.1177/0165551510388123

Julianti, O. N., Suarna, N., & Prihartono, W. (2024). Penerapan Natural Language Processing Pada Analisis Sentimen Judi Online Di Media Sosial Twitter. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Number 3).

Kumala Sari, P., & Randy Suryono, R. (2024). *Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse* (Vol. 7, Number 1).

Kurniawan, D., Triyanto, D., Wahyudi, M., Pujiastuti, L., Christian, A., Hidayat Lubis, F., Bina Sarana Informatika, U., Antar Bangsa, S., Kramat Raya No, J., Senen, K., Jakarta Pusat, K., Hos Cokroaminoto, J., Tengah, K., & Tangerang, K. (2026). Analisis Sentimen Multibahasa Pada Tiktok Menggunakan Model Nlp Berbasis Transformer Untuk Deteksi Dini Isu Publik. *Jurnal Insan (Journal Of Information Systems Management Innovation*, *6*(1). Http://Jurnal.Bsi.Ac.Id/Index.Php/Jinsan

Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). *Analisis Sentimen Ulasan Aplikasi Dana Dengan Metode Random Forest* (Vol. 6, Number 9). Http://J-Ptiik.Ub.Ac.Id

Manullang, O., Prianto, C., & Harani, N. H. (2023). *Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest*.

Nahdhudin, M., Cahyo Hendro Wibowo, N., Rini Handayani, M., Umam, K., Sains Dan Teknologi, F., Walisongo, U., & Semarang, K. (2024). *Jutekom Jurnal Teknologi Dan Ilmu Komputer Klasifikasi Sentimen Masyarakat Terhadap Aplikasi Tiktok Menggunakan Algoritma Naive Bayes*. Https://Jurnal.Nolsatu.Co.Id/Jutekom

Ningrum, A. C., Rizal Tjut Adek, & Nunsina. (2025). Pengembangan Bot Komentar Otomatis Dengan Analisis Sentimen Berbasis Bert Pada Tiktok Untuk Umkm Di Lhokseumawe. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, *10*(2), 431–444. Https://Doi.Org/10.36341/Rabit.V10i2.6152

Nur Adhan, S., Ngurah Adhi Wibawa, G., Christien Arisona, D., Yahya, I., Studi Statistika, P., & Matematika Dan Ilmu Pengetahuan Alam, F. (2024). *Analisis Sentimen Ulasan Aplikasi Wattpad Di Google Play Store Dengan Metode Random Forest* (Vol. 2, Number 1).

Pratmanto, D., Fandi, F., Imaniawan, D., & Author, C. (2023). Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors. In *Universitas Bina Sarana Informatika Jl. Sipelem No* (Vol. 3, Number 2). Http://Jurnal.Bsi.Ac.Id/Index.Php/Co-Science

Rhomaningtias, L., Khairunisa, A., Shella, S., Wara, M., & Hindrayani, K. M. (2025). *Analisis Sentimen Ulasan Aplikasi Smile Indonesia Menggunakan Metode Naive Bayes Dan Support Vector Machine (Svm)*. Https://Doi.Org/10.52972/Hoaq.Vol16no1

Rifki, M., Utami, Y., & Harsadi. Paulus. (2024). Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma Naïve Bayes. *Jurnal Sistem & Teknologi Informasi Komunikasi*.

Rizal Chandra Rivaldi, & T.D. Wismarini. (2024). Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (Nlp). *Elkom: Jurnal Elektronika Dan Komputer*, *17*(1), 120–128. Https://Doi.Org/10.51903/Elkom.V17i1.1680

Sapanji, R., Hamdani, D., & Harahap, P. (2023). Sentiment Analysis Of The Top 5 E-Commerce Platforms In Indonesia Using Text Mining And Natural Language Processing (Nlp). In *Journal Of Applied Informatics And Computing (Jaic)* (Vol. 7, Number 2). Http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic

Seri Rahmat, Y., Saskia Renata, Albertus Belo, Weiskhy Steven Dharmawan, Fitriana, L. A., & Riski Annisa. (2026). Perbandingan Algoritma Machine Learning Untuk Analisis Sentimen Aplikasi Instagram Berbasis Nlp. *Jurnal Komputer Teknologi Informasi Sistem Komputer (Juktisi)*, *5*(1), 532–543. Https://Doi.Org/10.62712/Juktisi.V5i1.1057

Snyder, H. (2021). Literature Review As A Research Methodology: An Overview And Guidelines. *Journal Of Business Research*, *104*, 333–339. Https://Doi.Org/10.1016/J.Jbusres.2019.07.039

Tarwoto, Nugroho, R., Azka, N., & Sayudha Rendra Graha, W. (2025). *Analisis Sentimen Ulasan Aplikasi Mobile Jkn Di Google Playstore Menggunakan Indobert* (Vol. 3, Number 2).

Tribuana, D., Usman, U., & Dayanti, D. (2025). Penerapan Natural Language Processing Untuk Analisis Sentimen Terhadap Layanan Publik Di Media Sosial Twitter. *Jurnal Teknologi Dan Bisnis Cerdas*, *1*(1), 28–37. Https://Doi.Org/10.64476/Jtbc.V1i1.3

Zahara, M., Rizka, M., Abdi, Inul, & Mahyar, H. (2025). *Analisis Sentimen Program Makan Siang Gratis Pada Tiktok Dengan Pendekatan Nlp Berbasis Indobert*. Https://Www.Tiktok.Com/@Danicathayaa/Video