

# ANALISIS SENTIMEN DALAM DATA BIG DATA: STUDI KASUS PADA MEDIA SOSIAL

## *Abstrak*

*Perkembangan media sosial menghasilkan big data yang memuat opini publik dalam volume besar dan beragam bentuk. Hal ini menuntut metode analisis sentimen yang mampu bekerja cepat dan akurat untuk mendukung pengambilan keputusan strategis. Penelitian ini bertujuan untuk mengidentifikasi tren metode, algoritma, dan tantangan utama dalam penerapan analisis sentimen pada big data media sosial. Penelitian dilakukan dengan pendekatan systematic literature review, menganalisis publikasi ilmiah terkini yang memanfaatkan algoritma machine learning dan deep learning, seperti SVM, Naive Bayes, LSTM, CNN, dan BERT. Hasil menunjukkan bahwa kombinasi metode pembelajaran mendalam dengan strategi pra-pemrosesan yang tepat dapat meningkatkan akurasi klasifikasi sentimen hingga di atas 90%. Metode transformer modern terbukti efektif dalam memahami konteks bahasa yang kompleks di media sosial. Penelitian ini menemukan bahwa tren penggunaan model hybrid dan model berbasis transformer membuka potensi baru untuk mengatasi keterbatasan model klasik dalam skenario big data yang terus bertambah. Analisis sentimen pada big data media sosial dapat berjalan optimal dengan pemilihan metode yang adaptif, pengelolaan data yang tepat, dan integrasi teknologi terkini. Penelitian ini memberikan arah pengembangan untuk studi lanjutan agar dapat mendukung pengambilan keputusan yang lebih responsif dan akurat.*

*Kata Kunci : Analisis Sentimen, BERT, Big Data, Deep Learning, Media Sosial, Naive Bayes, Neural Networks, SVM*

## *Abstract*

*The development of social media has generated big data containing public opinion in large volumes and various forms. This demands fast and accurate sentiment analysis methods to support strategic decision-making. This study aims to identify trends in methods, algorithms, and key challenges in applying sentiment analysis to social media big data. The study was conducted using a systematic literature review approach, analyzing recent scientific publications utilizing machine learning and deep learning algorithms, such as SVM, Naive Bayes, LSTM, CNN, and BERT. The results show that combining deep learning methods with appropriate pre-processing strategies can improve sentiment classification accuracy to above 90%. Modern transformer methods have proven effective in understanding complex language contexts in social media. This study found that the trend of using hybrid and transformer-based models opens up new potential to overcome the limitations of classical models in the ever-increasing big data scenario. Sentiment analysis in social media big data can be optimized with adaptive*

*method selection, appropriate data management, and the integration of the latest technologies. This research provides development directions for further studies to support more responsive and accurate decision-making.*

*Keywords: Sentiment Analysis, BERT, Big Data, Deep Learning, Social Media, Naive Bayes, Neural Networks, SVM*

## **PENDAHULUAN**

Perkembangan teknologi informasi yang pesat telah mendorong pertumbuhan data dalam skala besar atau big data yang bersumber dari berbagai platform digital, salah satunya media sosial. Media sosial menjadi wadah interaksi publik yang menghasilkan data teks dalam jumlah masif dan tidak terstruktur setiap harinya. Data ini dapat dimanfaatkan untuk menggali opini, emosi, dan persepsi publik melalui proses analisis sentimen. Analisis sentimen pada media sosial memiliki tantangan tersendiri, di antaranya ragam bahasa, ekspresi informal, serta volume data yang terus berkembang. (Trista & Asmoro, 2025). Berbagai metode telah dikembangkan untuk mendukung akurasi dan efektivitas analisis sentimen pada data media sosial. Pendekatan berbasis machine learning dan deep learning kini menjadi pilihan utama karena keunggulannya dalam memahami konteks dan pola bahasa yang kompleks. Penelitian sebelumnya menunjukkan bahwa metode Support Vector Machine (SVM) dan Naïve Bayes merupakan teknik klasik yang masih banyak digunakan. (Ryandi et al., 2025; Zulfa & Winarko, 2017) Namun demikian, model-model deep learning seperti Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT), dan Recurrent Neural Networks (RNN) terbukti memiliki performa lebih baik dalam menangani data teks yang besar dan bervariasi. (Hafidzah et al., 2024; Nurfauziyah et al., 2024; Sawitri, 2025)

Penggunaan algoritma mutakhir seperti BERT mampu mencapai tingkat akurasi yang signifikan. Misalnya, Putu, Saraswati, Yudistira, dan Adikara (2023) menunjukkan bahwa BERT berhasil mengklasifikasikan sentimen perundungan siber dengan akurasi hingga 81%. Selain itu, penelitian lain memadukan BERT dengan model CNN untuk meningkatkan akurasi dan F1-score secara optimal. (Hafidzah et al., 2024) Di sisi lain, model LSTM juga banyak digunakan untuk mendeteksi opini publik terkait isu sosial dan politik dengan hasil akurasi yang tinggi. (A. P. Prasetyo et al., 2024). Namun, penerapan analisis sentimen dalam skala big data masih menghadapi beberapa tantangan, seperti

kebutuhan komputasi yang besar, permasalahan data tidak seimbang, serta ambiguitas makna kata dalam konteks media sosial. (Sanusi et al., 2021; Trista & Asmoro, 2025) Oleh karena itu, penelitian ini berfokus pada bagaimana penerapan metode analisis sentimen yang efektif dapat digunakan untuk mengekstraksi opini publik pada data big data media sosial, serta mengevaluasi keunggulan model-model pembelajaran mesin dan pembelajaran mendalam yang digunakan pada studi-studi sebelumnya. Dengan demikian, diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan teknologi analisis sentimen pada era big data dan mendukung pengambilan keputusan yang berbasis data secara real-time.

## **METODE PENELITIAN**

Penelitian ini menggunakan metode Systematic Literature Review (SLR), yaitu pendekatan terstruktur dan sistematis untuk mengidentifikasi, mengevaluasi, dan mensintesis berbagai hasil penelitian terdahulu yang relevan dengan topik analisis sentimen pada data Big Data di media sosial. Tujuan penggunaan metode ini adalah untuk memberikan pemahaman yang komprehensif dan objektif mengenai penerapan teknik Machine Learning dan Deep Learning dalam analisis sentimen, serta untuk menilai efektivitas berbagai algoritma seperti Support Vector Machine (SVM), Naive Bayes, Neural Networks, Bidirectional Encoder Representations from Transformers (BERT), dan Recurrent Neural Networks (RNN). Proses seleksi artikel dalam penelitian ini mengikuti kerangka kerja PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), yang mencakup empat tahapan utama, yaitu identifikasi (identification), penyaringan (screening), penilaian kelayakan (eligibility), dan penyertaan akhir (inclusion). Penelusuran literatur dilakukan melalui beberapa basis data dan mesin pencari ilmiah seperti Google Scholar, Garuda Ristekbrin, IEEE Xplore, dan DOAJ, dengan menggunakan kombinasi kata kunci dalam Bahasa Indonesia dan Bahasa Inggris, antara lain “analisis sentimen” (sentiment analysis), “big data”, “media sosial” (social media), “machine learning”, “deep learning”, “SVM”, “Naive Bayes”, “Neural Networks”, “BERT”, dan “Recurrent Neural Networks”. Dalam proses pencarian, digunakan operator Boolean AND dan OR untuk memperluas atau mempersempit hasil pencarian, dengan fokus pada kemunculan kata kunci pada judul, abstrak, atau kata kunci

artikel. Penggunaan kombinasi bahasa Indonesia dan Inggris bertujuan untuk menjangkau literatur nasional maupun internasional yang relevan dengan topik penelitian ini.

Kriteria inklusi dalam penelitian ini meliputi artikel yang membahas penerapan analisis sentimen pada data Big Data di platform media sosial; penelitian yang mengevaluasi atau membandingkan metode Machine Learning dan Deep Learning untuk analisis sentimen; studi yang secara eksplisit mengulas algoritma seperti SVM, Naive Bayes, Neural Networks, BERT, atau RNN dalam konteks analisis sentimen; serta artikel ilmiah yang diterbitkan antara tahun 2020–2025 dalam jurnal terakreditasi. Sementara itu, kriteria eksklusi mencakup artikel yang tidak membahas penerapan praktis analisis sentimen pada data Big Data atau media sosial, penelitian yang hanya membahas teori dasar tanpa eksperimen atau studi kasus nyata, serta artikel yang tidak memiliki ISSN, tidak diterbitkan dalam jurnal terakreditasi, atau tidak tersedia dalam versi lengkap (full text). Dari hasil penelusuran awal, ditemukan 173 artikel ilmiah, namun setelah melalui proses penyaringan berdasarkan kriteria inklusi dan eksklusi, diperoleh 15 artikel yang dianggap memenuhi syarat dan relevan untuk dianalisis lebih lanjut. Artikel-artikel tersebut kemudian direview secara mendalam untuk mengidentifikasi tren penelitian terkini, algoritma yang paling populer digunakan, tingkat akurasi setiap model, kelebihan dan kelemahan tiap pendekatan, serta potensi pengembangan di masa depan. Proses analisis dilakukan secara kualitatif-deskriptif, dengan penyajian hasil dalam bentuk sintesis tematik untuk menggambarkan kondisi terkini, tantangan yang dihadapi, dan peluang pengembangan penerapan analisis sentimen berbasis Machine Learning dan Deep Learning pada data Big Data di media sosial.

## **HASIL DAN PEMBAHASAN**

### **Hasil Penelitian**

Dalam kajian literatur ini, hasil data penelitian yang telah dikumpulkan berisi analisis dan ringkasan dari artikel-artikel yang didokumentasikan terkait dengan analisis sentimen dalam data big data: studi kasus pada media sosial yaitu sebanyak 15 artikel yang tersaji dalam tabel berikut ini

Tabel 1 Implementasi Sistem Informasi Penjualan Untuk Meningkatkan Efisiensi Operasional

No.	Penulis (Tahun)	Hasil Penelitian
1	(Tania Ananda Paramitha et al., 2019)	Pada penelitian ini membantu untuk mengklasifikasikan review kedalam rating. Pengujian metode ini menggunakan prediksi dengan metode Extreme Learning Machine (ELM). Proses prediksi ini menggunakan tahapan pre-processing, pembobotan kata dengan TF-IDF, dan perhitungan metode Extreme Learning Machine (ELM). Terdapat tahapan-tahapan pada metode ELM antara lain normalisasi, proses training, dan proses testing. Metode ELM ini menghasilkan akurasi sebesar 80,01% Dengan jumlah k yaitu 10 menggunakan hidden neuron sebanyak 25 dengan Interval bobot -0,5 hingga 0,5 fungsi aktivasi Sigmoid biner. Dapat disimpulkan bahwa metode ELM dapat menyelesaikan masalah prediksi dengan cukup baik.
2	(Kawatu et al., 2019)	Dengan melakukan analisis ini, perusahaan dapat mengidentifikasi area yang memerlukan perbaikan, memahami tren kepuasan pelanggan, dan merancang strategi yang lebih baik untuk meningkatkan kualitas layanan. Selain itu, hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang manajemen layanan pelanggan dan pengembangan bisnis berbasis data.
3	(Muhammad Arief Rahman et al., 2019)	Hasil dari penelitian menunjukkan dari 4 aspek dan 3 sentimen maka didapatkan nilai rata-rata precision, recall, dan f1- score adalah precision – 95.80%, recall- 88.12%, f1_score- 91.62%. Bisa disimpulkan score representation ini dapat digunakan untuk klasifikasi sentimen.

4	(Khomsah & Agus Sasmito Aribowo, 2020)	Validasi menunjukkan penggunaan fitur unigram lebih baik akurasi daripada kombinasi fitur unigram dan bigram. Sedangkan konversi kosakata slang mampu menaikkan akurasi. Penghapusan stop words juga menyumbang kenaikan akurasi. Kesimpulannya, kombinasi preprocessing yang terdiri preprocessing standar, penghapusan stop words, konversi kosakata slang menjadi kosakata standar KKBI menaikkan akurasi sekitar 3,5 % pada fitur term tunggal (unigram).
5	(Faturrohman & Rosmala, 2022)	Dalam penelitian ini diusulkan metode Bidirectional GRU (BiGRU) untuk ekstraksi fitur teks dengan parameter yang diukur adalah perubahan nilai dari Learning Rate, Batch Size, dan Epoch. Dalam penerapan metodenya dibantu dengan word embedding GloVe dalam menghasilkan model terbaik, yang akhirnya digunakan untuk melakukan klasifikasi polaritas pada data sentimen dari sosial media Twitter. Hasil yang didapatkan pada model terbaik menghasilkan akurasi sebesar 97.7% dengan nilai precision 97.8%, recall 97.5% dan nilai loss 1.3%
6	(Pakpahan & Jasman Pardede, 2023)	Hal ini bertujuan untuk meringankan proses pelatihan dengan menghilangkan karakter atau kata yang tidak dibutuhkan. Selanjutnya kata divektorkan menggunakan FastText, tujuannya untuk mengubah tipe data string menjadi vektor array, sehingga kata dapat diproses di dalam LSTM. Performa akhir model diukur berdasarkan nilai precision, recall, accuracy, dan fMeasure. Berdasarkan pengujian parameter dropout layer pada hidden layer terhadap 10 fold cross validation, rata-rata akurasi pengujian model yang dihasilkan dari keseluruhan k fold adalah sebesar 72.4%. dengan perfoma model maksimum yang dicapai terdapat pada k fold = 9, saat menggunakan

		dropout layer sebesar 0.4, dengan nilai yang dicapai adalah precision, recall, accuracy, dan fmeasure secara berurutan: 76.74%, 80.49%, 78.31%, 78.57%.
7	(Putu et al., 2023)	Selama beberapa tahun terakhir, Bidirectional Encoder Representation from Transformers (BERT) telah menjadi model representasi yang digunakan secara luas dan efisien yang mencapai kinerja mutakhir pada tugas-tugas tingkat kalimat dan token-level, mengungguli banyak arsitektur tugas khusus. Tingkat akurasi pada klasifikasi perundungan siber dengan menggunakan data uji dan data validasi dalam Bahasa Indonesia yang dihasilkan dengan menggunakan algoritma BERT sebesar 0,81 yang dimana dalam persentase akurasi adalah 81%. Algoritma BERT dapat membantu mengidentifikasi masukan yang mengarah pada perundungan siber dengan memberikan keluaran berupa kelas dengan kategori yang telah ditentukan, yakni netral, bahasa kasar maupun mengandung ujaran kebencian. Sistem juga memberikan persentase terhadap kategori yang didapatkan melalui program tersebut
8	(Lestari, 2023)	Setelah sentimen dalam teks diidentifikasi, hasilnya dievaluasi untuk menilai reputasi merek produk Perusahaan X. Hal ini dapat dilakukan dengan menghitung persentase sentimen positif dan negatif, serta menganalisis komentar dan umpan balik yang terkait. Berdasarkan hasil analisis sentimen, Perusahaan X dapat mengambil tindakan perbaikan jika reputasi merek produknya dinilai negatif. Ini bisa berupa meningkatkan kualitas produk, meningkatkan layanan pelanggan, atau berkomunikasi dengan konsumen secara aktif di media sosial. Dengan melakukan analisis

		sentimen pada data media sosial, Perusahaan X dapat memperoleh wawasan yang berharga tentang bagaimana merek produk mereka dipersepsikan oleh konsumen dan mengambil langkah-langkah yang tepat untuk meningkatkan reputasi merek tersebut.
9	(A. P. Prasetyo et al., 2024)	Hasil penelitian menunjukkan bahwa sebanyak 58,0% dari sampel menunjukkan sentimen positif, sementara 42,0% menunjukkan sentimen negatif. Model LSTM yang diterapkan berhasil mencapai tingkat akurasi sebesar 88% dalam mengklasifikasikan sentimen masyarakat di Twitter terkait isu penundaan pemilu. Kesimpulan penelitian ini adalah mayoritas masyarakat yang terlibat dalam percakapan di Twitter terkait isu penundaan pemilu 2024 menunjukkan sentimen positif. Namun, masih ada sebagian yang menunjukkan sentimen negatif. Tingkat akurasi model LSTM yang tinggi menunjukkan efektivitas metode ini dalam menganalisis data teks dari media sosial. Oleh karena itu, penelitian ini memberikan wawasan yang berharga tentang pandangan publik terhadap isu penting ini di era digital.
19	(Bouchra et al., 2024)	Melalui analisis sentimen terhadap opini masyarakat, penelitian ini bertujuan untuk memahami persepsi terhadap kualitas konten yang disajikan dengan penggunaan data media sosial X (Twitter) menggunakan metode Bi-LSTM dan Word2Vec. Penelitian diawali dari pengumpulan data, preprocessing, vektorisasi Word2Vec, pembagian data, pelatihan dan pengujian model, evaluasi penentuan model terbaik, klasifikasi data sentimen, serta analisis data sentimen. Digunakan data sentimen sejumlah 515.492 yang menghasilkan akurasi sebesar 96,4%, precision sebesar 72,1%, recall sebesar 72,0%, dan f1-score sebesar 72,8%.

		Hasil analisis sentimen pengguna X(Twitter) cenderung netral dan positif. Hasil tren kepuasan masyarakat menunjukkan perubahan pola kepuasan masyarakat terhadap kualitas konten stasiun televisi
11	(A. W. Prasetyo & Putra, 2024)	Hasil yang diperoleh setelah dilakukan analisis sentimen, review pengguna mayoritas menyatakan netral dengan persentase, sentimen positif 23,9% ; sentimen netral 40,3% dan sentimen negatif 35,8%. Dengan ini dapat ditarik kesimpulan bahwa metode BERT dan RCNN kurang efektif untuk implementasi analisis sentiment.
12	(Mizan, 2024)	Data dikumpulkan dari platform media sosial Twitter, dengan total 1336 tweet yang diambil dari dataset Kaggle. Data tersebut terdiri dari 988 tweet dengan sentimen positif, 187 tweet dengan sentimen netral, dan 161 tweet dengan sentimen negatif. Data kemudian diolah menggunakan teknik prapemrosesan teks seperti tokenisasi, penghapusan stopwords, dan stemming. Selanjutnya, metode NLP seperti menganalisis sentimen dan mengklasifikasikan teks untuk identifikasi pola sentimen dalam dataset. Model yang dikembangkan memiliki hasil akhir akurasi sebesar 88.06%, dengan nilai precision, recall, dan f1-score yang beragam untuk setiap kategori sentimen.
13	(Hafidzah et al., 2024)	Hasil penelitian menunjukkan bahwa model-model pembelajaran mendalam seperti BERT, CNN, LSTM, dan GRU unggul dalam hal akurasi, ingatan, dan skor F1 dibandingkan dengan metode tradisional seperti SVM dalam analisis sentimen dalam media sosial. Kombinasi model BERT+CNN terbukti memiliki kinerja terbaik dalam hal akurasi dan skor F1. Arsitektur pembelajaran mendalam yang sering digunakan adalah model Transformer,

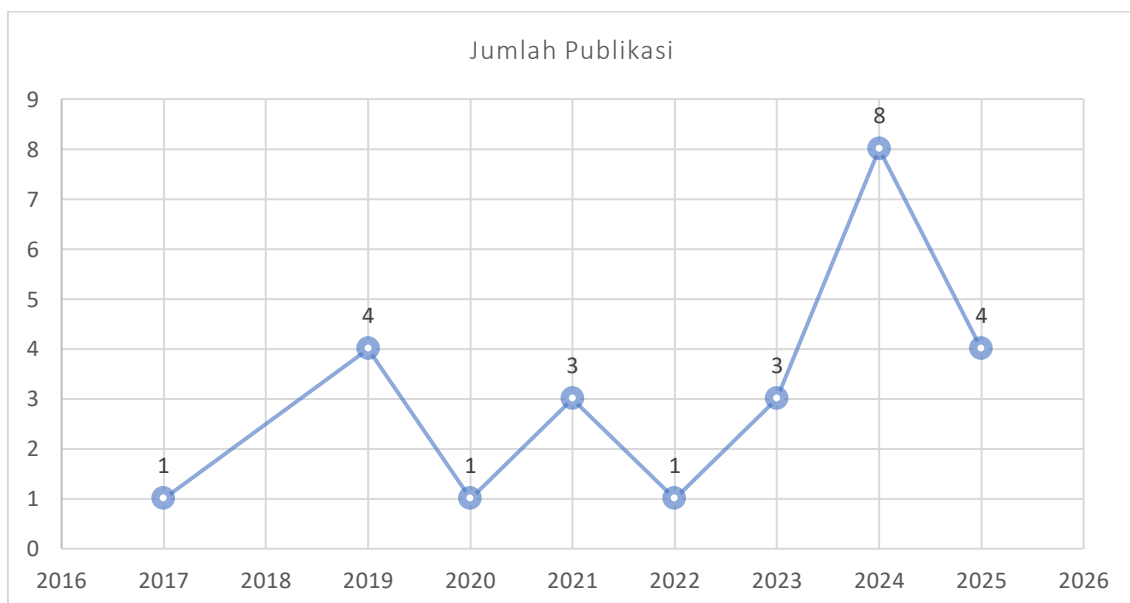
		<p>khususnya BERT, karena kemampuannya mencapai akurasi tinggi. Dataset Hinglish umum digunakan untuk melatih model deep learning dalam analisis sentimen media sosial.</p>
14	(Razzak & Setiawan, 2025)	<p>Hasil ini menunjukkan bahwa integrasi RNN dan CNN, bersama dengan teknik ekspansi fitur FastText, meningkatkan kemampuan analisis sentimen, memberikan metode yang lebih efisien dan akurat dalam mengklasifikasikan sentimen publik.</p>
15	(Sawitri, 2025)	<p>Peran big data dan deep learning dapat digunakan secara bersama untuk meningkatkan efisiensi dan akurasi analisis data serta memfasilitasi kemampuan analisis agar dapat mendeteksi masalah keuangan untuk kepentingan dunia usaha. Pada penelitian ini merupakan tinjauan literatur, membahas bagaimana deep learning dan big data berperan dalam mendeteksi masalah keuangan, termasuk deteksi penipuan, prediksi kebangkrutan, dan analisis risiko kredit. Tinjauan literatur bertujuan untuk mengeksplorasi bagaimana teknologi diterapkan dalam sektor keuangan menggunakan arsitektur CNN dan RNN dimana model LSTM juga GRU efektif dalam menangani data untuk mendeteksi aktivitas resiko keuangan. Deep learning dapat mengenali anomali secara real-time dengan tingkat keandalan tinggi, mengoptimal keamanan data dan memaksimalkan pengelolaan risiko</p>

**Perkembangan penelitian tentang Analisis Sentimen Dalam Data Big Data: Studi Kasus Pada Media Sosial**

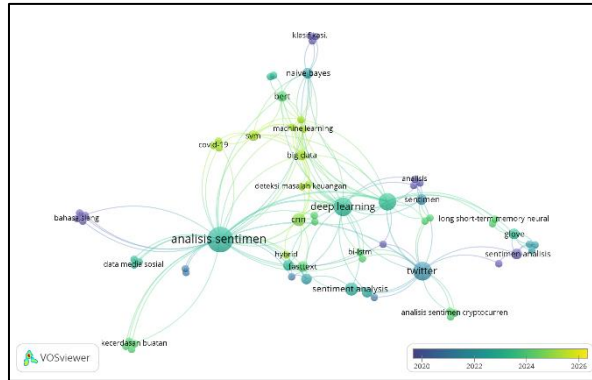
Berdasarkan analisis data penelitian mengenai Sistem Informasi Penjualan Untuk Meningkatkan Efisiensi Operasional, diperoleh data bahwa perkembangan penelitian pada topik ini adalah sebagai berikut:

Tabel 4. 2 Pengembangan penelitian tentang Sistem Informasi Penjualan Untuk Meningkatkan Efisiensi Operasional

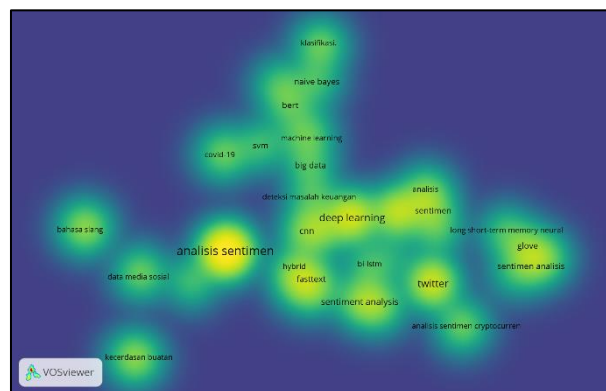
Tahun Publikasi	Jumlah Publikasi
2017	1
2019	4
2020	1
2021	3
2022	1
2023	3
2024	8
2025	4







Gambar 3.b Visualisasi Hamparan



Gambar 4.c Visualisasi Kepadatan

Visualisasi bibliometrik dalam penelitian ini dianalisis menggunakan perangkat lunak VOSviewer dengan menampilkan tiga jenis tampilan utama: visualisasi jaringan (*network visualization*), hamparan (*overlay visualization*), dan kepadatan (*density visualization*) untuk memetakan hubungan antar kata kunci pada tema *Analisis Sentimen* menggunakan teknologi *Machine Learning* dan *Deep Learning*. Hasil visualisasi jaringan menunjukkan bahwa kata kunci “analisis sentimen”, “deep learning”, dan “machine learning” menjadi pusat utama yang saling terhubung dengan kata kunci lain seperti *Twitter*, *big data*, *CNN*, *LSTM*, *BERT*, *Naive Bayes*, *SVM*, dan *kecerdasan buatan*. Jaringan ini membentuk beberapa klaster, seperti klaster *deep learning* dan *machine learning* (Trista & Asmoro, 2025; Hafidzah et al., 2024; Zulfa & Winarko, 2017), klaster *Twitter* dan media sosial (Pakpahan & Jasman Pardede, 2023; Prasetyo, Sri Kusuma, & Rizki, 2024), serta klaster pengembangan metode hibrida (*hybrid*) dan fitur *Word Embedding* seperti *FastText* dan *GloVe* (Razzak & Setiawan, 2025; Faturrohman & Rosmala, 2022).

Pada visualisasi hamparan, terlihat gradasi waktu publikasi mulai dari 2020 hingga 2025. Kata kunci seperti *deep learning*, *big data*, *LSTM*, *BERT*, dan *Twitter* mendominasi tren publikasi terbaru pada rentang 2024–2025 (Hafidzah et al., 2024; Sawitri, 2025; Ryandi, Pratiwi, & Sari, 2025). Sementara kata kunci seperti *Naive Bayes* dan *SVM* lebih banyak muncul pada penelitian sebelumnya di periode 2017–2021 (Zulfa & Winarko, 2017; Sulthan, Wahyudi, & Suhartini, 2021). Hal ini menunjukkan adanya pergeseran fokus dari metode klasifikasi tradisional ke pendekatan *deep learning* yang lebih mutakhir. Pada visualisasi kepadatan, kata kunci “analisis sentimen” dan “*deep learning*” memperlihatkan kepadatan tertinggi, yang berarti istilah ini paling sering muncul dalam literatur yang dianalisis. Temuan ini sejalan dengan ringkasan penelitian yang menunjukkan dominasi topik *deep learning* (CNN, LSTM, BERT) dalam mendukung klasifikasi sentimen media sosial (Paputungan & Agustinus Jacobus, 2021; Prasetyo, Sri Kusuma, & Rizki, 2024; Mizan, 2024). Kepadatan kata kunci juga memperlihatkan integrasi metode *Word Embedding* seperti *FastText* dan *GloVe* untuk meningkatkan akurasi model (Pakpahan & Jasman Pardede, 2023; Dwijo Utomo Rahino Putroa et al., 2024).

Berdasarkan publikasi terbaru, penelitian Hafidzah et al. (2024) dan Razzak & Setiawan (2025) menekankan pentingnya penerapan arsitektur *deep learning* berbasis *Transformer* dan *Recurrent Neural Network* untuk memahami konteks dan nuansa teks yang kompleks. Penelitian Sawitri (2025) juga menyoroti integrasi *big data* dengan *deep learning* untuk mendeteksi pola sentimen dalam data keuangan, sementara Trista & Asmoro (2025) menyoroti potensi *machine learning* dan *deep learning* dalam menangani tantangan data media sosial yang besar dan bervariasi. Dengan demikian, hasil visualisasi ini menunjukkan bahwa topik *analisis sentimen* berbasis *deep learning* dan *machine learning* semakin relevan dan diminati, dengan potensi pengembangan riset lanjutan seperti penerapan model *Transformer*, integrasi *big data*, dan optimalisasi metode *Word Embedding* untuk mendukung klasifikasi teks secara lebih akurat di era digital.

## **Pembahasan**

Perkembangan teknologi informasi yang masif mendorong munculnya data berukuran besar (*big data*) dari berbagai platform digital, terutama media sosial yang

menghasilkan teks tak terstruktur secara real-time. Tantangan terbesar dalam analisis sentimen pada media sosial terletak pada ragam bahasa, konteks lokal, slang, hingga kecepatan pertumbuhan data yang sangat tinggi (Trista & Asmoro, 2025). Karena itulah, kebutuhan akan teknik analisis yang akurat dan adaptif menjadi semakin penting. Beberapa studi menunjukkan bahwa pendekatan machine learning konvensional seperti Naive Bayes dan SVM masih sering digunakan, meski performanya cenderung lebih rendah dibandingkan metode deep learning pada data big data yang lebih kompleks (Zulfa & Winarko, 2017; Ryandi, Pratiwi, & Sari, 2025). Penelitian terdahulu membuktikan bahwa penerapan model deep learning membawa peningkatan akurasi signifikan. Misalnya, metode Long Short-Term Memory (LSTM) terbukti ampuh dalam menangani data sekuensial dari media sosial, seperti yang terlihat pada penelitian Prasetyo, Sri Kusuma, & Rizki (2024) dengan akurasi mencapai 88% dalam mendeteksi sentimen publik terkait isu sosial-politik. Hal serupa juga dicapai oleh Paputungan & Agustinus Jacobus (2021) yang memanfaatkan parameter LSTM untuk klasifikasi opini publik di media sosial dengan akurasi optimal. Hasil ini menegaskan bahwa model RNN dan LSTM cocok diterapkan untuk data teks panjang yang kontekstual dan tidak terstruktur.

Selain LSTM, pendekatan Transformer seperti BERT juga makin banyak diadopsi untuk mendongkrak pemahaman konteks kalimat secara mendalam. Putu et al. (2023) membuktikan bahwa BERT dapat mengklasifikasikan sentimen terkait perundangan siber dengan akurasi 81%, sedangkan Hafidzah et al. (2024) menemukan bahwa kombinasi BERT dan CNN mampu menghasilkan akurasi dan F1-score yang lebih tinggi dibandingkan model tradisional. Hal ini membuktikan bahwa model Transformer sangat relevan untuk diterapkan pada analisis sentimen big data yang memiliki variasi bahasa tinggi di media sosial. Kombinasi berbagai pendekatan juga menjadi tren pengembangan ke depan. Razzak & Setiawan (2025) memperlihatkan bahwa penggabungan CNN, RNN, dan ekspansi fitur menggunakan FastText mampu meningkatkan akurasi klasifikasi. Integrasi ini memperkuat hasil model dalam memahami pola bahasa slang, kosakata lokal, dan kata-kata baru yang sering muncul di media sosial (Khomsah & Aribowo, 2020). Penelitian ini menunjukkan bahwa strategi hibrida dapat menjadi solusi untuk mengatasi keterbatasan metode tunggal, terutama ketika menghadapi data dalam skala besar dan dinamis.

Dari sisi praktik, penelitian analisis sentimen terbukti bermanfaat untuk berbagai tujuan, seperti pengukuran kepuasan pelanggan (Kawatu, Jacobus, & Lantang, 2019), deteksi reputasi merek (Lestari, 2023), hingga prediksi isu politik (Nurfauziyah et al., 2024). Sanusi, Astuti, & Buryadi (2021) juga menyoroti tantangan overfitting akibat data tidak seimbang di media sosial, yang menuntut evaluasi model secara berkelanjutan. Oleh karena itu, penelitian mendatang perlu menekankan pentingnya preprocessing, balancing data, dan optimasi parameter model untuk menjaga kinerja klasifikasi tetap stabil. Dengan semakin tingginya minat penelitian pada tema analisis sentimen berbasis big data, terbuka peluang untuk mengeksplorasi integrasi metode deep learning mutakhir dengan teknologi big data analytics dan cloud computing agar proses analisis dapat dilakukan secara real-time (Sawitri, 2025). Penelitian ini diharapkan dapat berkontribusi pada pengembangan strategi manajemen data teks yang lebih adaptif dan akurat, mendukung pengambilan keputusan berbasis data, serta menjawab tantangan kecepatan, volume, dan variasi data pada media sosial di era digital (Trista & Asmoro, 2025).

## **KESIMPULAN**

Berdasarkan hasil penelitian dapat disimpulkan bahwa analisis sentimen pada data big data media sosial memerlukan pendekatan yang adaptif dengan menggabungkan metode machine learning dan deep learning. Hasil telaah menunjukkan bahwa metode seperti SVM, Naive Bayes, LSTM, CNN, hingga BERT memiliki keunggulan masing-masing, namun akurasi tertinggi dicapai saat algoritma dikombinasikan dengan teknik pra-pemrosesan dan optimasi fitur. Dengan demikian, permasalahan terkait besarnya volume data, keberagaman bahasa, serta kecepatan arus informasi dapat diatasi melalui pemilihan algoritma yang tepat dan pengembangan pipeline analitik yang terstruktur. Penelitian ini menjawab rumusan masalah terkait efektivitas, tantangan, dan arah pengembangan metode analisis sentimen pada konteks big data di media sosial. Saran untuk peneliti selanjutnya adalah agar mengembangkan model hybrid dengan memanfaatkan pendekatan berbasis transformer terbaru, mengeksplorasi dataset multibahasa, serta memanfaatkan komputasi paralel dan cloud computing agar proses analitik dapat berjalan real-time dengan akurasi optimal. Penelitian lanjutan juga diharapkan dapat fokus pada integrasi hasil analisis sentimen dengan pengambilan

keputusan praktis, baik di bidang pemasaran digital, kebijakan publik, maupun pemantauan opini publik.

## DAFTAR PUSTAKA

- Bouchra, F., Suarjaya, I. M. A. D., & Rusjyanthi, N. K. D. (2024). Analisis Sentimen Masyarakat terhadap Tayangan Televisi Nasional menggunakan Metode Deep Learning. *Jurnal Buana Informatika*, 15(02), 89–99. <https://ojs.uajy.ac.id/index.php/jbi/article/view/9118>
- Faturrohman, F., & Rosmala, D. (2022). Analisis Sentimen Sosial Media dengan Metode Bidirectional Gated Recurrent Unit. *Prosiding Diseminasi FTI*, X(X), 1–10. <https://eproceeding.itenas.ac.id/index.php/fti/article/view/962>
- Hafidzah, P., Maryani, S., Ihsani, B. Y., Erwin, E., & Niswariyana, A. K. (2024). Penerapan Deep Learning dalam Menganalisis Sentimen di Media Sosial. *Seminar Nasional Paedagoria*, 04, 328–339.
- Kawatu, C. I., Jacobus, A. E., & Lantang, O. A. (2019). Penerapan Metode Gated Recurrent Unit Pada Analisis Sentimen Media Sosial Clinton Imanuel Kawatu. *UNSRAT Repository*, 1–8.
- Khomsah, S., & Agus Sasmito Aribowo. (2020). Text-Preprocessing Model Youtube Comments in Indonesian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 04(04), 648–654. <https://doi.org/10.29207/resti.v4i4.2035>
- Lestari, A. (2023). Analisis Sentimen pada Data Media Sosial untuk Menilai Reputasi Merek Produk Perusahaan X. *Cyberarea.Id*, 03(03), 1–19.
- Mizan, M. (2024). Analisis Sentimen Masyarakat Mengenai Pemilihan Umum 2024 Menggunakan Natural Language Processing (NLP) (Issue Table 10).
- Muhammad Arief Rahman, Herman Budianto, & Esther Irawati Setiawan. (2019). Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network. *Journal of Intelligent System and Computation*, 01(02), 50–57. <https://doi.org/10.52985/insyst.v1i2.83>

- Nurfauziyah, N., DwiYansaputra, R., Ika Murpratiwi, S., & Aranta, A. (2024). Analisis Sentimen Pada Pengguna Aplikasi X Terhadap Pemilihan Umum Presiden 2024 Menggunakan Metode Convolutional Neural Network (CNN). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 09(01), 635–642. <https://doi.org/10.36040/jati.v9i1.12437>
- Pakpahan, I., & Jasman Pardede. (2023). Analisis Sentimen Penanganan Covid-19 Menggunakan Metode Long Short-Term Memory Pada Media Sosial Twitter. *Jurnal Publikasi Teknik Informatika*, 2(1), 12–25. <https://doi.org/10.55606/jupti.v1i1.767>
- Prasetyo, A. P., Sri Kusuma, C., & Rizki, D. (2024). Analisis Sentimen Twitter terhadap Isu Penundaan Pemilu 2024 Menggunakan Metode Long Short-Term Memory (LSTM). *Jurnal Repositor*, 06(03), 293–304. <https://doi.org/10.22219/repositor.v6i3.32643>
- Prasetyo, A. W., & Putra, R. E. (2024). Analisis Sentimen Twitter Pada Cryptocurrency Menggunakan Bidirectional Encoder Representations From Transformers Dan Region-Based Convolutional Neural Network. *Journal of Informatics and Computer Science (JINACS)*, 06(01), 11–21. <https://doi.org/10.26740/jinacs.v6n01.p11-21>
- Putu, N., Saraswati, V. D., Yudistira, N., & Adikara, P. P. (2023). Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 07(02), 909–916. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12345>
- Razzak, R. N., & Setiawan, E. B. (2025). Ekspansi Fitur dengan FastText untuk Analisis Sentimen di Media Sosial X Menggunakan Recurrent Neural Network dan Convolutional Neural Network. *E-Proceeding of Engineering*, 12(01), 2025.
- Ryandi, F. A., Pratiwi, D., & Sari, S. (2025). Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM. *Jurnal Sains Dan Teknologi*, 07(01), 1–6. <https://doi.org/10.55338/saintek.v7i1.4615>
- Sanusi, R., Astuti, F. D., & Buryadi, I. Y. (2021). Sentiment Analysis On Twitter Towards Pre-Employment Card Program With Recurrent Neural Network. *JIKO (Jurnal Informatika Dan Komputer)*, 05(02), 89–99.

- Sawitri, D. (2025). Peran Deep Learning Dan Big Data Dalam Mendekteksi Masalah Keuangan. *Djtechno: Jurnal Teknologi Informasi*, 06(01), 193–207. <https://doi.org/10.46576/djtechno.v6i1.6037>
- Tania Ananda Paramitha, D., Cholissodin, I., & Dewi, C. (2019). Prediksi Rating Otomatis Berdasarkan Review Restoran pada Aplikasi Zomato dengan menggunakan Extreme Learning Machine (ELM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 03(05), 4687–4693. <http://j-ptiik.ub.ac.id>
- Trista, R. T., & Asmoro, E. T. (2025). Analisis Sentimen dalam Data Big Data (Studi Kasus pada Media Sosial). *Journal of Multidisciplinary Inquiry in Science Technology and Educational Research*, 02(01), 700–706. <https://doi.org/10.32672/mister.v2i1.2518>
- Zulfa, I., & Winarko, E. (2017). Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(02), 187–198. <https://doi.org/10.22146/ijccs.24716>