

STUDI KOMPARATIF ALGORITMA MACHINE LEARNING PADA ANALISIS SENTIMEN MEDIA SOSIAL

Febriyanti Panjaitan¹, Win Ce², Hery Oktafiandy³, Ghanim Kanugrahan⁴,
Yudi Ramdhani⁵, Vito Hafizh Cahaya Putra⁶, Antika Permai⁷

^{1,4,6} Informatika, Satu University

^{3,5,7} Sistem Informasi, Satu University

² Sistem Informasi, Bina Nusantara University

Jl. BKR No.63, Ancol, Kec. Regol, Bandung, Indonesia

Pebrianti.panjaitan@univ.satu.ac.id

ABSTRAK

Analisis sentimen di Twitter telah menjadi salah satu topik utama dalam penelitian terkait opini publik di bidang ekonomi, politik, dan isu sosial. Penggunaan machine learning dalam analisis sentimen memungkinkan untuk memproses data teks secara efisien. Penelitian ini bertujuan untuk mengeksplorasi literatur terkait analisis sentimen menggunakan metode machine learning pada Twitter dalam konteks ekonomi, politik, dan isu sosial. Metode yang digunakan adalah Systematic Literature Review (SLR), dengan pengumpulan artikel dari tiga database utama: IEEE Xplore, Google Scholar, dan Scopus. Setelah menerapkan kriteria inklusi dan eksklusi, 45 artikel relevan terpilih untuk dianalisis. Hasil penelitian menunjukkan bahwa Support Vector Machine (SVM) memiliki performa terbaik dengan akurasi rata-rata 85.3%, diikuti oleh Random Forest (83.7%) dan Naïve Bayes (81.5%). KNN dan Decision Tree menunjukkan performa lebih rendah, kemungkinan karena sensitivitas terhadap data yang tidak seimbang. Tren penelitian mengindikasikan bahwa analisis sentimen di bidang ekonomi lebih banyak berkaitan dengan dampak kebijakan ekonomi, di bidang politik fokus pada opini publik terkait pemilu dan kebijakan pemerintah, sementara di bidang isu sosial berkaitan dengan gerakan sosial dan kebijakan kesehatan.

Kata kunci: Analisis Sentimen, Twitter, Machine Learning, Systematic Literature Review, Ekonomi, Politik, Isu Sosial

1. PENDAHULUAN

Media sosial, khususnya Twitter, telah menjadi salah satu platform utama bagi masyarakat dalam mengekspresikan opini mereka terhadap berbagai isu, termasuk politik, ekonomi, dan sosial terjadi di Indonesia [1].

Dengan jumlah pengguna yang terus meningkat dan produksi data yang sangat besar, pemahaman terhadap opini publik melalui analisis sentimen menjadi semakin mendesak. Informasi yang diperoleh dari analisis sentimen di Twitter dapat dimanfaatkan oleh berbagai pihak, seperti pemerintah, perusahaan, dan organisasi sosial, untuk merumuskan kebijakan, strategi pemasaran, serta memahami tren dan persepsi masyarakat secara real-time. Namun, tantangan utama dalam analisis ini adalah dinamika bahasa dan perubahan tren yang cepat, yang dapat memengaruhi akurasi model prediktif. Oleh karena itu, diperlukan pendekatan yang sistematis dalam mengevaluasi metode analisis sentimen guna mendapatkan hasil yang lebih akurat dan relevan dengan kondisi terkini [2], [3].

Namun, dalam penelitian ini, data yang digunakan dibatasi hanya dalam rentang tiga tahun terakhir. Batasan ini diterapkan untuk memastikan relevansi data dengan kondisi terkini, mengingat perubahan tren, kebijakan media sosial, serta dinamika bahasa yang terus berkembang di platform tersebut. Penggunaan data yang terlalu lama berisiko menghasilkan model yang kurang akurat karena

konteks sosial dan kebahasaan dapat berubah seiring waktu [4].

Metode yang digunakan dalam penelitian ini adalah studi literature review dengan pendekatan Systematic Literature Review (SLR). Pemilihan metode SLR dipertimbangkan karena SLR memungkinkan peneliti untuk secara sistematis mengidentifikasi, mengevaluasi, dan mengintegrasikan temuan-temuan dari berbagai studi yang relevan, guna memberikan dasar yang kuat bagi pemilihan algoritma dan analisis yang dilakukan [5] dalam penelitian ini.

Pendekatan ini juga membantu dalam mengidentifikasi tren dan kesenjangan penelitian yang ada, serta memberikan gambaran menyeluruh tentang penerapan machine learning dalam analisis sentimen di media sosial.

Dalam penelitian ini, lima algoritma machine learning—Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes, Decision Tree, dan Random Forest—dipilih karena karakteristik dan keunggulannya masing-masing dalam analisis teks, terutama dalam klasifikasi sentimen [6], [7].

SVM dipilih karena kemampuannya yang sangat kuat dalam menangani data yang tidak terstruktur, serta memiliki mekanisme untuk menemukan batas pemisah yang jelas antara kelas. Dalam konteks analisis sentimen, hal ini sangat berguna, mengingat karakteristik data teks yang sering kali bersifat ambigu dan memiliki batasan yang tidak selalu jelas. SVM, dengan pendekatan margin maksimalnya, dapat

memisahkan data dalam ruang vektor dengan cara yang efisien, sehingga meningkatkan kemampuan model dalam mengklasifikasikan data teks [8].

KNN dipilih karena kesederhanaannya dan kemampuannya untuk bekerja baik dengan data yang tidak memerlukan pelatihan yang rumit. KNN bekerja dengan pendekatan berbasis tetangga terdekat, yang memungkinkan algoritma ini untuk menangani masalah klasifikasi dengan cara yang intuitif dan efisien [9].

Dalam banyak kasus, KNN mampu memberikan hasil yang memadai meskipun tidak mengandalkan proses pelatihan yang intensif, menjadikannya pilihan yang tepat untuk penelitian ini. Naive Bayes, algoritma probabilistik yang sangat efisien, dipilih karena kemampuannya dalam menangani data teks dengan asumsi independensi antar fitur. Keunggulan utama Naive Bayes adalah kemampuannya untuk menangani skala data besar dengan cepat, karena proses pelatihannya relatif ringan dan tidak memerlukan banyak sumber daya komputasi. Pada analisis sentimen, Naive Bayes efektif dalam mengidentifikasi pola kata yang dapat menunjukkan sentimen positif atau negatif, dengan mengandalkan frekuensi kemunculan kata dalam teks untuk menghitung probabilitas kelas. Meskipun berbasis pada asumsi independensi yang sering kali tidak sepenuhnya akurat dalam dunia nyata, Naive Bayes tetap menunjukkan performa yang baik dalam banyak aplikasi analisis teks [10].

Decision Tree dipilih karena kemudahan dalam interpretasi hasil dan visualisasi proses pengambilan keputusan. Decision Tree membuat model yang mudah dipahami, di mana pengguna dapat dengan jelas melihat bagaimana setiap fitur memengaruhi keputusan klasifikasi. Keunggulan utama dari Decision Tree adalah transparansi dalam proses pembuatan keputusan, yang membuatnya sangat berguna dalam aplikasi yang memerlukan penjelasan dan interpretasi hasil. Selain itu, Decision Tree juga mampu menangani data yang beragam dan tidak terstruktur, yang sering ditemukan dalam analisis sentimen berbasis teks [11].

Random Forest, sebagai pengembangan dari Decision Tree, dipilih karena kemampuannya untuk mengurangi masalah overfitting yang sering kali terjadi pada Decision Tree dengan menggunakan teknik ensemble. Dengan menggabungkan banyak pohon keputusan untuk membuat prediksi akhir, Random Forest cenderung memberikan hasil yang lebih stabil dan robust, serta lebih akurat dalam klasifikasi sentimen. Keunggulan utama Random Forest adalah kemampuannya untuk menangani berbagai jenis data dengan baik, serta meningkatkan akurasi klasifikasi melalui pendekatan agregasi dari beberapa model [12].

Kelima algoritma ini sering digunakan oleh para peneliti dalam bidang analisis sentimen karena metode ini telah terbukti efektif dalam berbagai penelitian sebelumnya [13]–[15].

Masing-masing algoritma memberikan hasil yang baik dalam berbagai konteks dan dataset yang berbeda. Sering kali, para peneliti memilih algoritma-algoritma ini karena mereka dapat memberikan akurasi terbaik, baik secara individual maupun dalam komparasi. Dengan keunggulan masing-masing dalam hal interpretasi, efisiensi, dan stabilitas hasil, kelima algoritma ini menjadi pilihan populer dalam analisis teks, terutama untuk kasus-kasus seperti analisis sentimen di media sosial yang melibatkan data besar dan beragam.

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja lima algoritma machine learning dalam mengklasifikasikan sentimen dari tweet-tweet yang diperoleh selama tiga tahun terakhir. Dengan membandingkan metrik kinerja seperti akurasi, presisi, recall, dan F1-score, penelitian ini bertujuan untuk mengidentifikasi algoritma terbaik untuk analisis sentimen di platform Twitter. Penelitian ini juga bertujuan untuk memberikan wawasan yang lebih mendalam tentang dinamika opini publik terhadap isu-isu tertentu, serta membantu peneliti dan praktisi dalam memilih algoritma yang paling efektif dan sesuai dalam melakukan analisis sentimen pada media sosial.

2. TINJAUAN PUSTAKA

Analisis sentimen di media sosial, khususnya di platform seperti Twitter, telah menjadi topik yang banyak dibahas dalam berbagai penelitian terkait dengan pemrosesan bahasa alami (NLP) dan machine learning. Twitter sebagai platform sosial media memiliki karakteristik unik, yaitu penggunaan bahasa yang tidak baku, slang, emotikon, dan singkatan yang sering kali mempengaruhi akurasi dalam analisis sentimen. Oleh karena itu, pemilihan metode yang tepat sangat penting untuk memperoleh hasil analisis yang optimal.

2.1. Pemilihan Data Tiga Tahun Kebelakang

Penelitian ini mengambil data selama tiga tahun kebelakang untuk memberikan gambaran yang lebih terkini mengenai opini publik dan dinamika sosial yang terjadi. Dengan menggunakan data Twitter dari rentang waktu tersebut, peneliti dapat menangkap perubahan tren sentimen yang terkait dengan peristiwa besar dan isu-isu yang berkembang, seperti perubahan politik, ekonomi, dan sosial yang mempengaruhi pandangan masyarakat khususnya terjadi di Indonesia. Pemilihan periode tiga tahun memberikan keseimbangan antara relevansi data dengan kondisi saat ini serta memberikan analisis yang lebih mendalam terkait perubahan opini publik dalam jangka waktu yang cukup panjang. Hal ini juga membantu menghindari pengaruh fluktuasi data yang terlalu cepat, sehingga menghasilkan temuan yang lebih stabil dan representatif.

2.2. Pemilihan Algoritma Machine Learning

Dalam penelitian ini, lima algoritma machine learning yaitu SVM, KNN, Naive Bayes, Decision Tree, dan Random Forest, dipilih karena karakteristik dan keunggulannya yang telah terbukti dalam berbagai aplikasi analisis sentimen.

2.3. Support Vector Machine

SVM dipilih karena kemampuannya yang kuat dalam menangani data yang tidak terstruktur dan memiliki batas pemisah yang jelas antara kelas. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan kelas-kelas data, yang sangat berguna dalam menangani data teks yang kompleks dan beragam seperti yang ada di Twitter. Dalam banyak studi, SVM telah terbukti memberikan hasil yang akurat dalam berbagai tugas klasifikasi, termasuk analisis sentimen.

2.4. K-Nearest Neighbor

KNN dipilih karena kesederhanaannya dan kemampuannya untuk menangani data dengan baik tanpa memerlukan pelatihan model yang intensif. Dengan cara kerja berbasis kedekatan (nearest neighbor), KNN dapat secara efektif mengklasifikasikan data yang memiliki kemiripan fitur. Keunggulan lainnya adalah kemudahan dalam implementasi dan interpretasi, menjadikannya pilihan populer dalam penelitian-penelitian analisis sentimen.

2.5. Naive Bayes

Naive Bayes dipilih karena kemampuannya dalam menangani data teks dengan probabilitas, serta mengasumsikan independensi antar fitur. Metode ini telah terbukti efisien dan sangat baik dalam klasifikasi teks, terutama dalam analisis sentimen. Naive Bayes juga sangat cepat dalam proses pelatihan dan cocok untuk skala data besar, yang sering ditemui dalam analisis media sosial [16].

2.6. Decision Tree

Decision Tree dipilih karena kemudahan dalam interpretasi dan visualisasi proses pembuatan keputusan. Pengguna dapat dengan jelas melihat bagaimana setiap fitur berkontribusi pada klasifikasi, yang membuatnya sangat berguna untuk aplikasi yang membutuhkan penjelasan dan interpretasi dari hasil model. Selain itu, Decision Tree dapat menangani data yang tidak terstruktur dengan baik, menjadikannya algoritma yang sangat baik untuk analisis sentimen berbasis teks [17].

2.7. Random Forest

Sebagai pengembangan dari Decision Tree, Random Forest dipilih karena kemampuannya untuk mengurangi overfitting dengan menggunakan teknik ensemble learning. Dengan menggabungkan banyak pohon keputusan, Random Forest mampu menghasilkan prediksi yang lebih stabil dan akurat, sehingga sering digunakan dalam klasifikasi teks yang

lebih kompleks. Algoritma ini telah terbukti memberikan kinerja yang lebih baik dalam mengatasi masalah ketidakseimbangan data dan meningkatkan akurasi prediksi [18], [19].

Kelima algoritma ini dipilih karena keberhasilan mereka dalam berbagai penelitian sebelumnya dalam analisis sentimen, dan masing-masing memiliki keunggulan yang berbeda dalam mengatasi masalah-masalah khas dalam analisis teks dan data media sosial.

2.8. Metode Pendekatan Systematic Literature Review

Untuk memberikan gambaran yang lebih komprehensif mengenai pemilihan algoritma machine learning dan tren terbaru dalam analisis sentimen di Twitter, penelitian ini menggunakan pendekatan SLR. SLR merupakan metode yang terstruktur dan sistematis dalam meninjau dan menganalisis literatur yang relevan, yang memungkinkan peneliti untuk mengidentifikasi dan mengevaluasi temuan-temuan dari penelitian sebelumnya secara objektif [20].

Pendekatan ini juga memungkinkan peneliti untuk mengintegrasikan berbagai temuan dan memberikan wawasan yang lebih holistik terkait penerapan algoritma machine learning dalam analisis sentimen.

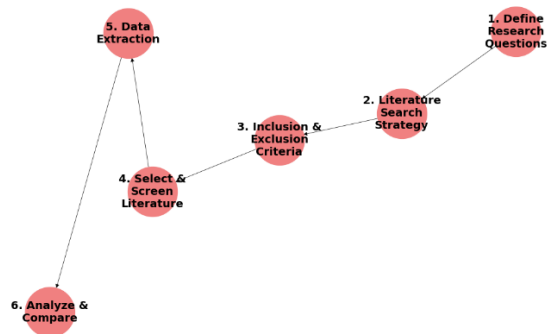
Metode SLR dipilih karena dapat membantu mengidentifikasi algoritma yang paling sering digunakan, serta memberikan gambaran tentang performa dan keterbatasan masing-masing algoritma dalam analisis sentimen di media sosial. Dengan melakukan tinjauan literatur yang mendalam, SLR memungkinkan peneliti untuk menggali tren-tren terkini dalam penelitian yang berkaitan dengan analisis sentimen dan memberikan dasar yang lebih kuat dalam pemilihan metode dan teknik yang digunakan dalam penelitian ini. Pendekatan ini juga memungkinkan untuk mengidentifikasi kesenjangan dalam penelitian yang ada, serta memberikan saran bagi penelitian lebih lanjut untuk mengatasi tantangan yang belum terselesaikan.

3. METODE PENELITIAN

Metode penelitian dalam studi ini menggunakan pendekatan SLR untuk menganalisis dan membandingkan algoritma machine learning dalam analisis sentimen di media sosial Twitter fokus pada bidang ekonomi, politik dan isu sosial di Indonesia. Proses penelitian ini mengikuti tahapan yang sistematis untuk memastikan bahwa hasil yang diperoleh berdasarkan literatur yang relevan dan valid.

3.1. Tahapan Systematic Literature Review

SLR terdiri dari beberapa tahapan utama yang mengikuti pendekatan yang ketat untuk memastikan validitas dan reliabilitas hasil yang diperoleh. Tahapan SLR telah dikembangkan dan mengacu pada Kitchenham et al. [21] untuk rekayasa perangkat lunak.



Gambar 1. Flowchart: Systematic Literature Review

3.1.1. Menentukan Pertanyaan Penelitian

SLR diawali dengan mendefinisikan pertanyaan penelitian yang jelas dan spesifik. pertanyaan penelitian yang berfokus pada analisis sentimen di Twitter dalam konteks ekonomi, politik, dan isu sosial. Pertanyaan Penelitian untuk SLR dalam Analisis Sentimen di Twitter:

- a. Ekonomi: Bagaimana tren sentimen publik terhadap kebijakan ekonomi pemerintah dalam tiga tahun terakhir?
- b. Politik: Algoritma machine learning apa yang paling efektif dalam mengklasifikasikan sentimen terkait pemilu, kebijakan pemerintah di Twitter?
- c. Isu Sosial: Bagaimana perubahan sentimen publik terhadap isu sosial tertentu, seperti korupsi dalam 3 tahun terakhir?
- d. Algoritma machine learning apa saja yang paling sering digunakan dalam analisis sentimen di Twitter?
- e. Bagaimana performa masing-masing algoritma dalam analisis sentimen berdasarkan metrik evaluasi seperti akurasi, precision, recall, dan F1-score?
- f. Tren terbaru apa yang muncul dalam penelitian analisis sentimen menggunakan machine learning dalam tiga tahun terakhir?
- g. Apa keunggulan dan kelemahan dari masing-masing algoritma yang digunakan dalam penelitian sebelumnya?

3.1.2. Strategi Pencarian Literatur (Literature Search Strategy)

Langkah ini melibatkan pencarian literatur menggunakan database akademik yang terpercaya untuk menemukan penelitian yang relevan.

Sumber Database yang Digunakan:

- a. IEEE Xplore
- b. Google Scholar
- c. Scopus

Kata Kunci Pencarian:

- a. Strategi pencarian menggunakan kombinasi Boolean (AND, OR) untuk mendapatkan hasil yang lebih relevan. Contoh pencarian:

b. Ekonomi:

- "Sentiment Analysis on Economic Policies using Machine Learning"
- "Public Opinion on Inflation and Stock Market using Twitter Data"

c. Politik:

- "Political Sentiment Analysis on Twitter using SVM and Naïve Bayes"
- "Comparative Study of Machine Learning in Election Sentiment Analysis"

d. Isu Sosial:

- "Social Issue Sentiment Analysis: A Machine Learning Approach"
- "Public Reaction to Corruption Cases using Twitter Sentiment Analysis"

3.1.3. Penentuan Kriteria Inklusi dan Eksklusi (Inclusion and Exclusion Criteria)

Setelah merumuskan pertanyaan penelitian, langkah selanjutnya adalah menentukan kriteria inklusi dan eksklusi agar hanya studi yang relevan yang dimasukkan dalam analisis.

Kriteria Inklusi:

- a. Studi yang membahas analisis sentimen menggunakan algoritma machine learning (SVM, KNN, Naive Bayes, Decision Tree, Random Forest).
- b. Studi yang membahas analisis sentimen di Twitter dalam konteks ekonomi, politik, atau isu sosial di Indonesia.
- c. Studi yang menggunakan dataset dari Twitter.
- d. Studi yang dipublikasikan dalam tiga tahun terakhir untuk menjaga keterkinian penelitian.
- e. Studi yang ditulis dalam bahasa Inggris atau bahasa Indonesia.
- f. Studi yang menyediakan evaluasi performa algoritma menggunakan metrik seperti akurasi, precision, recall, dan F1-score.

Kriteria Eksklusi:

- a. Studi yang tidak berfokus pada analisis sentimen.
- b. Studi yang tidak menggunakan metode machine learning yang ditentukan dalam penelitian ini.
- c. Studi yang hanya membahas teori tanpa implementasi atau eksperimen.
- d. Studi yang menggunakan dataset selain dari Twitter.

3.1.4. Seleksi dan Penyaringan Literatur (Study Selection and Screening)

Setelah literatur dikumpulkan dari berbagai database, tahap seleksi dilakukan berdasarkan:

- a. Penyaringan judul dan abstrak: Artikel yang tidak relevan dieliminasi berdasarkan judul dan abstrak.
- b. Penyaringan teks lengkap: Studi yang lolos dari penyaringan abstrak diperiksa secara menyeluruh untuk memastikan relevansinya dengan pertanyaan penelitian.

- c. Tahapan ini sering menggunakan alat bantu seperti Mendeley atau Zotero untuk mengelola referensi dan menyaring artikel yang sesuai dengan kriteria yang telah ditetapkan.

3.1.5. Ekstraksi Data (Data Extraction)

Tahap ini bertujuan untuk mengekstrak informasi yang relevan dari studi yang telah dipilih. Data yang diekstrak mencakup:

- Referensi artikel (judul, tahun, penulis, sumber jurnal/konferensi).
- Jenis algoritma machine learning yang digunakan (SVM, KNN, Naive Bayes, Decision Tree, Random Forest).
- Dataset yang digunakan (Twitter dataset, jumlah tweet, preprocessing data).
- Metode evaluasi yang digunakan (akurasi, precision, recall, F1-score).

3.1.6. Menganalisis & Membandingkan Algoritma

Pada tahap ini, semua hasil dari studi yang dikumpulkan dianalisis untuk membandingkan performa SVM, KNN, Naive Bayes, Decision Tree, dan Random Forest. Analisis dilakukan berdasarkan:

- Akurasi tertinggi dan stabilitas model.
- Kelebihan dan kekurangan masing-masing algoritma.
- Efektivitas algoritma terhadap berbagai jenis dataset.
- Tren perkembangan metode dalam tiga tahun terakhir.

4. HASIL DAN PEMBAHASAN

Pada bagian ini, akan dibahas hasil dari tahapan SLR yang telah dilakukan untuk menganalisis penelitian terkait analisis sentimen di Twitter dengan fokus pada isu-isu ekonomi, politik, dan isu sosial menggunakan algoritma machine learning. Proses SLR yang melibatkan pencarian literatur dari berbagai sumber database akademik telah berhasil menghasilkan artikel-artikel relevan yang kemudian dianalisis lebih lanjut.

4.1. Identifikasi Sumber Dataset

Dalam penelitian ini, tiga sumber database utama digunakan untuk mencari literatur terkait analisis sentimen di Twitter dalam bidang ekonomi, politik, dan isu sosial di Indonesia menggunakan machine learning. Setelah dilakukan penyaringan berdasarkan kriteria inklusi dan eksklusi, jumlah artikel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Identifikasi Sumber Database

Sumber Database	Jumlah Artikel Awal	Jumlah Artikel Terpilih
IEEE Xplore	142	13
Google Scholar	223	6
Scopus	83	26
Total	448	45

4.2. Penerapan Kriteria Inklusi dan Eksklusi

Agar hanya artikel yang relevan digunakan, dilakukan penyaringan dengan kriteria inklusi dan eksklusi.

Kriteria Inklusi seperti

- Artikel dipublikasikan dalam tiga tahun terakhir (2021–2024).
- Studi membahas analisis sentimen di Twitter dalam ekonomi, politik, atau isu sosial.
- Artikel menggunakan machine learning sebagai metode klasifikasi sentimen.
- Artikel menyajikan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

Selanjutnya untuk kriteria Eksklusi seperti

- Artikel yang berfokus pada media sosial selain Twitter
- Penelitian yang tidak menggunakan machine learning dalam analisis sentimen.
- Studi yang hanya membahas analisis kata-kata atau leksikon tanpa klasifikasi sentimen. Hasil dari tahap ini mengeliminasi 403 artikel, menyisakan 45 artikel yang relevan untuk analisis lebih lanjut.

4.3. Analisis Algoritma Machine Learning

Dari 45 artikel yang dianalisis, Jumlah artikel yang menggunakan machine learning ada di Tabel 2., rata-rata jumlah data dan distribusi sentimen dalam analisis sentimen di Twitter ditampilkan pada Tabel 3. Sementara itu, hasil perbandingan performa algoritma berdasarkan rata-rata metrik evaluasi terlihat pada Tabel 4.

Tabel 2. Jumlah Artikel yang Menggunakan Algoritma dalam Analisis Sentimen di Twitter (2021–2024).

Metode Machine Learning	2021	2022	2023	Jumlah Artikel
SVM	4	5	3	12
KNN	2	3	3	8
Naive Bayes	3	4	2	9
Decision Tree	2	3	2	7
Random Forest	3	4	2	9
Jumlah	14	19	12	45

Tabel 2. Tabel ini menunjukkan distribusi jumlah artikel yang menggunakan masing-masing algoritma dalam analisis sentimen di Twitter dari tahun 2021 hingga 2023. Dari data ini, dapat dilihat bahwa SVM menjadi algoritma yang paling banyak digunakan setiap tahunnya, sementara KNN dan Decision Tree memiliki jumlah penggunaan yang lebih rendah dibanding algoritma lainnya.

Tabel 3. Rata-rata Jumlah Data dan Distribusi Sentimen dalam Analisis Sentimen di Twitter

Metode Machine Learning	Jumlah Data	Negatif	Positif	Netral
SVM	± 12.500	4.200	5.100	3.200
KNN	± 10.800	3.700	4.500	2.600
Naïve Bayes	± 11.200	3.900	4.700	2.600
Decision Tree	± 9.700	3.300	4.000	2.400
Random Forest	± 11.500	4.000	4.900	2.600

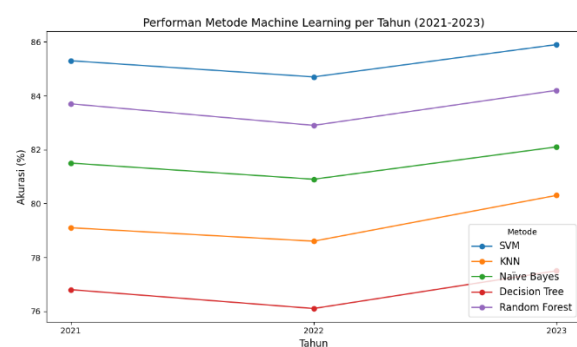
Tabel 3 ini menggambarkan rata-rata jumlah data yang digunakan dalam penelitian analisis sentimen berbasis Twitter berdasarkan 45 artikel dalam tiga tahun terakhir. Data ini mencerminkan distribusi sentimen negatif, positif, dan netral untuk setiap algoritma machine learning yang dianalisis.

Tabel 4. Perbandingan Kinerja Algoritma Machine Learning dalam Analisis Sentimen di Twitter

Metode Machine Learning	Akurasi Rata-rata	Presisi	Recall	F1-Score
SVM	85.3%	84.7%	85.9%	85.2%
KNN	79.1%	78.6%	80.3%	79.4%
Naïve Bayes	81.5%	80.9%	82.1%	81.5%
Decision Tree	76.8%	76.1%	77.5%	76.8%
Random Forest	83.7%	82.9%	84.2%	83.5%

Temuan Utama penelitian ini adalah:

- SVM memiliki performa terbaik dalam analisis sentimen di Twitter, dengan akurasi rata-rata 85.3%, menjadikannya metode yang paling stabil untuk klasifikasi sentimen.
- Random Forest (83.7%) dan Naïve Bayes (81.5%) juga menunjukkan performa yang baik, terutama dalam menangani dataset yang lebih besar.
- KNN dan Decision Tree memiliki performa lebih rendah, kemungkinan karena sensitivitas terhadap data yang tidak seimbang atau jumlah fitur yang besar.



Gambar 2. Performa Metode Machine Learning per Tahun berdasarkan Akurasi

Pada Gambar 2 memvisualisasikan perbandingan akurasi lima metode machine learning (SVM, KNN, Naïve Bayes, Decision Tree, dan Random Forest) selama tiga tahun (2021, 2022, dan 2023). Dengan demikian, kita dapat melihat tren atau perubahan performa masing-masing metode dari tahun ke tahun.

Hasil akurasi SVM pada tahun 2021 adalah 85.3%, sedikit menurun menjadi 84.7% pada 2022, namun kemudian meningkat kembali pada 2023 menjadi 85.9%. Hal ini menunjukkan bahwa SVM memberikan performa yang stabil dan terus meningkatkan hasil akurasi seiring dengan pengumpulan data dan penerapan teknik-teknik baru dalam pemodelannya.

Akurasi KNN pada 2021 adalah 79.1%, sedikit menurun menjadi 78.6% pada 2022, tetapi kemudian meningkat menjadi 80.3% pada 2023. Perubahan ini mungkin mencerminkan peningkatan pemahaman dan penggunaan KNN yang lebih baik dalam menangani data Twitter dengan berbagai pola dan variasi sentimen.

Pada tahun 2021, Naïve Bayes mencapai akurasi sebesar 81.5%, yang sedikit menurun menjadi 80.9% pada 2022, namun kembali meningkat menjadi 82.1% pada 2023. Peningkatan akurasi ini mungkin disebabkan oleh penerapan teknik pra-pemrosesan teks yang lebih baik, seperti pembersihan data dan penghapusan noise dalam teks Twitter.

Pada tahun 2021, akurasi Decision Tree adalah 76.8%, yang sedikit menurun menjadi 76.1% pada 2022, dan meningkat sedikit menjadi 77.5% pada 2023. Meskipun Decision Tree mungkin tidak sekuat SVM dalam hal akurasi, ia tetap merupakan pilihan populer karena kemampuannya untuk interpretasi yang mudah dan pengambilan keputusan berbasis aturan.

Performa Random Forest pada tahun 2021 adalah 83.7%, yang menunjukkan performa yang cukup baik dibandingkan dengan metode lainnya. Angka ini sedikit menurun menjadi 82.9% pada tahun 2022, namun kembali meningkat pada 2023 menjadi 84.2%. Peningkatan ini menunjukkan bahwa Random Forest terus memberikan hasil yang solid dalam analisis sentimen, dengan kemampuannya untuk menangani data yang besar dan kompleks, serta mengurangi potensi bias atau overfitting yang dapat terjadi jika hanya menggunakan satu pohon keputusan.

4.4. Tren Penelitian Berdasarkan Topik

Setelah menganalisis literatur berdasarkan kategori ekonomi, politik, dan isu sosial, ditemukan beberapa tren menarik seperti:

- Analisis Sentimen dalam Ekonomi

Mayoritas penelitian dalam ekonomi berfokus pada dampak kebijakan pemerintah terhadap pasar dan sentimen masyarakat, SVM dan Random Forest sering digunakan dalam menganalisis reaksi publik terhadap kebijakan ekonomi seperti kenaikan harga BBM, inflasi, dan investasi.

- b. Analisis Sentimen dalam Politik
Twitter menjadi sumber utama dalam menganalisis opini publik terhadap pemilu, kebijakan pemerintah, dan partai politik. SVM dan Naïve Bayes terbukti efektif dalam mengklasifikasikan sentimen positif dan negatif terkait kampanye politik, kinerja presiden, dan kasus korupsi.
- c. Analisis Sentimen dalam Isu Sosial
Penelitian dalam isu sosial banyak membahas opini publik terhadap gerakan sosial, perubahan lingkungan, dan kebijakan kesehatan. Random Forest dan Decision Tree sering digunakan dalam topik ini karena dapat menangani variasi opini yang luas di media sosial.

4.5. Implikasi dan Kesenjangan Penelitian

Berdasarkan hasil analisis, ditemukan beberapa kesenjangan dalam penelitian yang telah dilakukan seperti Kurangnya studi tentang kombinasi machine learning dan deep learning dalam analisis sentimen di Twitter, Minimnya dataset yang seimbang, yang dapat mempengaruhi performa algoritma tertentu seperti KNN dan Decision Tree. Serta Fokus yang lebih banyak pada politik, sementara ekonomi dan isu sosial masih kurang mendapat perhatian dalam penelitian terbaru

5. KESIMPULAN DAN SARAN

Penelitian ini mengumpulkan literatur terkait analisis sentimen di Twitter menggunakan machine learning dari tiga database utama: IEEE Xplore, Google Scholar, dan Scopus, dengan 45 artikel terpilih untuk analisis. Hasil perbandingan algoritma menunjukkan SVM memiliki akurasi tertinggi (85.3%), diikuti oleh Random Forest (83.7%) dan Naïve Bayes (81.5%), sementara KNN dan Decision Tree memiliki performa lebih rendah. Tren penelitian menunjukkan fokus analisis sentimen di ekonomi pada dampak kebijakan, di politik pada opini publik terkait pemilu dan kebijakan, serta di isu sosial pada gerakan sosial dan kebijakan kesehatan. Penelitian ini juga menemukan beberapa kesenjangan, termasuk minimnya studi yang menggabungkan machine learning dengan deep learning, serta kebutuhan akan dataset yang lebih seimbang.

Berdasarkan hasil yang diperoleh, penelitian selanjutnya dapat mengeksplorasi metode hybrid yang menggabungkan machine learning dan deep learning untuk meningkatkan akurasi analisis sentimen. Selain itu, penelitian masa depan disarankan untuk menggunakan dataset yang lebih beragam dan seimbang, guna menghindari bias dalam klasifikasi sentimen.

DAFTAR PUSTAKA

- [1] R. N. Muhammad and B. Tanggahma, "Pengaruh Media Sosial Pada Persepsi Publik Terhadap Sistem Peradilan: Analisis Sentimen di Twitter," *UNES Law Rev.*, vol. 7, no. 1, pp. 507–516, 2024.
- [2] R. Merdiansah, S. Siska, and A. A. Ridha, "Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 221–228, 2024.
- [3] G. Kanugrahan, V. H. C. Putra, and Y. Ramdhani, "Analisis Sentimen Aplikasi Gojek Menggunakan SVM, Random Forest dan Decision Tree," *J. Infortech*, vol. 6, no. 2, pp. 171–178, 2024.
- [4] M. R. Pugu, S. Riyanto, and R. N. Haryadi, *Metodologi Penelitian; Konsep, Strategi, dan Aplikasi*. PT. Sonpedia Publishing Indonesia, 2024.
- [5] S. H. Wahid and A. Kususiyanah, "Systematic Review Menggunakan Artificial Intelligence (AI) dalam Ilmu Sosial," 2024.
- [6] Y. Hagiwara *et al.*, "Computer-aided diagnosis of atrial fibrillation based on ECG Signals: A review," *Inf. Sci. (Nyl.)*, vol. 467, pp. 99–114, 2018, doi: 10.1016/j.ins.2018.07.063. <https://doi.org/10.1016/j.ins.2018.07.063>
- [7] S. Rahmaddeni *et al.*, *Machine Learning*. Serasi Media Teknologi, 2024.
- [8] M. V. P. A. Devani, M. Chulkamdi, and M. F. Rahmat, "Analisis Sentimen Non-Fungible Token Solana Di Media Sosial X Dengan Random Forest Dan Support Vector Machine." Universitas Islam Balitar, 2024.
- [9] W. Andriyani *et al.*, *Matematika Pada Kecerdasan Buatan*. TOHAR MEDIA, 2024.
- [10] M. Saiful, A. M. Nur, A. E. Sutriandi, E. Puspita, and B. N. Nuzululnisa, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penyaluran Dana Zakat," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 239–249, 2025.
- [11] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, and D. Wulandari, "Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix: Implementation of Decision Tree Algorithm for Movie Recommendation and Rating Classification on the Netflix Platform," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 570–579, 2024.
- [12] J. PRANATA, "Penggunaan Model Bahasa indoBERT pada Metode Random Forest Untuk Klasifikasi Sentimen Dengan Dataset Terbatas," *Build. Informatics, Technol. Sci.*, vol. 6, no. 3, pp. 1668–1676, 2025.
- [13] R. W. Purba, A. A. Waskita, and M. Makshun, "Analisis Sentimen Opini Debat Calon Presiden Dengan Menggunakan Classifier Machine Learning (Studi Kasus: Pada Data Twitter 2024)," *Infotech J. Technol. Inf.*, vol. 10, no. 2, pp. 221–232, 2024.
- [14] H. Leidiyana, T. Misriati, and R. Aryanti, "Klasifikasi Sentimen Terhadap Kebijakan Tapera Menggunakan Komparasi Machine

- Learning dan SMOTE,” *J. Komtika (Komputasi dan Inform.*, vol. 8, no. 2, pp. 125–135, 2024.
- [15] I. Syahrohim, S. D. Saputra, R. W. Saputra, V. H. Pranatawijaya, and R. Priskila, “Perbandingan analisis sentimen setelah pilpres 2024 di Twitter menggunakan algoritma machine learning,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024.
- [16] E. Setiani and W. Ce, “Text Classification Services Using Naïve Bayes for Bahasa Indonesia,” in *2018 International Conference on Information Management and Technology (ICIMTech)*, 2018, pp. 361–366.
- [17] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiarsari, “Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE,” *Aiti*, vol. 18, no. 2, pp. 173–184, 2021.
- [18] M. A. Fauzi, “Random forest approach for sentiment analysis in Indonesian,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, pp. 46–50, 2018.
- [19] M. Melia, B. Irawan, and O. Nurdian, “ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK DAN GRAB PADA MEDIA SOSIAL TWITTER MENGGUNAKAN RANDOM FOREST,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 5, pp. 3614–3618, 2023.
- [20] A. Hadi et al., *Buku Ajar Metodologi Penelitian Ilmu Komputer: Dilengkapi dengan Studi Kasus*. PT. Sonpedia Publishing Indonesia, 2024.
- [21] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering—a systematic literature review,” *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.