

Perbandingan Algoritma Naive Bayes dan SVM dalam Analisis Sentimen Pengguna AI di Platform X

Noval Dinda Firdaus*, Ryan Randy Suryono

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}noval_dinda_firdaus@teknokrat.ac.id, ²ryan@teknokrat.ac.id

Email Penulis Korespondensi: noval_dinda_firdaus@teknokrat.ac.id

Submitted: 01/03/2025; Accepted: 17/03/2025; Published: 18/03/2025

Abstrak—Pertumbuhan pesat kecerdasan buatan (AI), hal ini juga telah menimbulkan pengaruh yang signifikan pada berbagai aspek kehidupan, termasuk interaksi di platform media sosial seperti Platform X. Di platform ini, pengguna secara aktif mendiskusikan berbagai topik terkait AI, mulai dari manfaat hingga tantangan yang ditimbulkannya. Memahami bagaimana publik merespons teknologi AI menjadi hal yang penting bagi pengembang, peneliti, dan pembuat kebijakan agar dapat merancang strategi yang lebih sesuai dengan kebutuhan serta ekspektasi masyarakat. Penelitian ini bertujuan untuk mengevaluasi dan melakukan perbandingan kinerja antara dua algoritma yang biasa dipakai dalam pengamatan sentimen, yakni *Naive Bayes* dan *Support Vector Machine* (SVM). Data dikumpulkan melalui teknik *crawling* dengan menggunakan *Google Colab*, yang menghasilkan 9.183 entri. Sebelum analisis dilakukan, data melewati serangkaian tahap pemrosesan awal, meliputi proses pembersihan teks, normalisasi huruf, *tokenisasi*, menghilangkan kata-kata yang sering digunakan. (*stopword removal*), serta *stemming* untuk menyederhanakan kata. Hasil analisis menunjukkan bahwa SVM memiliki keunggulan dalam hal akurasi serta kemampuannya yakni 96% akurasi dalam menangani data yang kompleks, sedangkan *Naive Bayes* lebih cepat dalam proses komputasi dan efisien untuk *dataset* berukuran besar yang menghasilkan akurasi yakni 84% lebih kecil dari akurasi SVM. Penilaian dilaksanakan dengan memakai metrik *akurasi*, *presisi*, *recall*, serta *F1-score* berdasarkan *confusion matrix*.

Kata Kunci: Analisis Sentimen; Naive Bayes; Support Vector Machine; Analisis Pengguna AI; Platform X

Abstract—The rapid development of artificial intelligence (AI) has also had a significant impact on various aspects of life, including interactions on social media platforms such as Platform X. On this platform, users actively discuss various topics related to AI, from the benefits to the challenges it poses. Understanding how the public responds to AI technology is important for developers, researchers, and policy makers in order to design strategies that are more in line with the needs and expectations of the community. This study aims to evaluate and compare the performance of two algorithms commonly used in sentiment analysis, namely Naive Bayes and Support Vector Machine (SVM). Data were collected through crawling techniques using Google Colab, which resulted in 9,183 entries. Before the analysis was carried out, the data went through a series of initial processing stages, including text cleaning, letter normalization, tokenization, removing frequently used words (stopword removal), and stemming to simplify words. The results of the analysis show that SVM has advantages in terms of accuracy and capability, namely 96% accuracy in handling complex data, while Naive Bayes is faster in the computational process and efficient for large datasets, resulting in an accuracy of 84% smaller than SVM accuracy. The assessment is carried out using accuracy, precision, recall, and F1-score metrics based on the confusion matrix.

Keywords: Sentiment Analysis; Naive Bayes; Support Vector Machine; AI User Analysis; Platform X

1. PENDAHULUAN

Di era digital yang berkembang dengan cepat, teknologi *Artificial Intelligence* (AI) menjadi salah satu terobosan utama dengan potensi yang signifikan untuk pengaruh terhadap berbagai aspek kehidupan manusia termasuk di platform media sosial dan forum diskusi daring. *Artificial Intelligence* (AI) merupakan salah satu cabang ilmu komputer yang secara khusus menitikberatkan perhatian pada pengembangan sistem dan program yang dirancang untuk mencontoh kemampuan berpikir dan memahami yang dimiliki oleh manusia [1]. AI kini digunakan secara luas, baik dalam bentuk chatbot, asisten virtual, hingga sistem rekomendasi yang secara otomatis menyesuaikan konten dengan preferensi pengguna [2]. Dengan kemampuan tersebut, AI dapat memproses dan menganalisis data secara cepat dan akurat, sehingga menghasilkan wawasan penting bagi pengembang dan pengambil keputusan. Platform X, sebagai salah satu media sosial terkemuka, tidak hanya menjadi tempat utama bagi pengguna untuk berdiskusi, berbagi pengalaman, dan mengekspresikan opini tentang teknologi AI, tetapi juga dirancang khusus agar pengguna dapat saling berbagi pendapat secara bebas [3]. Seiring meningkatnya adopsi AI, muncul beragam reaksi dari masyarakat, ada yang menyambut positif dengan harapan efisiensi dan inovasi, namun ada pula yang mengkhawatirkan etika, privasi, serta dampak AI terhadap lapangan kerja. Oleh karena itu, analisis sentimen terhadap pengguna AI di Platform X menjadi penting untuk memahami persepsi publik dan implikasinya bagi pengembang teknologi serta pembuat kebijakan. Dalam konteks tersebut, pemilihan algoritma yang tepat sangat krusial untuk memastikan keakuratan klasifikasi sentimen.

Dua algoritma yang sering digunakan dalam analisis ini adalah *Naive Bayes* dan *Support Vector Machine* (SVM) [4]. Penerapan kedua metode ini memungkinkan evaluasi langsung terhadap performa masing-masing, sehingga dapat mengidentifikasi kelebihan serta keterbatasan dari setiap pendekatan [5]. Analisis sentimen dengan membandingkan beberapa algoritma klasifikasi juga pernah dilakukan oleh beberapa penelitian sebelumnya. Seperti pada penelitian yang membahas produk *Apple Vision Pro*, dengan membandingkan tiga algoritma yakni *Random Forest*, *Support Vector Machine* serta *Naive Bayes*, yang memperoleh akurasi bentuk *Random Forest* sebesar 83%,

model SVM sebesar 80%, dan *Naïve Bayes* sebesar 75%, di antara ketiga metode tersebut, *Random Forest* mempunyai tingkat akurasi lebih unggul dari pada kedua metode lainnya [6]. Penelitian lainnya juga pada analisis sentimen terhadap penggunaan mobil listrik yang membandingkan dua algoritma yakni *Naïve Bayes* serta *Support Vector Machine*, yang menghasilkan akurasi model *Naïve Bayes* sebanyak 63,02%, sedangkan untuk model SVM sebesar 70,82%. Hasil tersebut menyatakan bahwa teknik *Support Vector Machine* menunjukkan tingkat akurasi yang lebih unggul daripada *Naïve Bayes* [7]. Selanjutnya, dalam penelitian pengamatan sentimen tentang *ChatGPT* dengan 1000 dataset menunjukkan bahwa akurasi SVM dengan label *RoBERT* sebesar 55%, lebih tinggi dibandingkan akurasi *Naïve Bayes* sebesar 47% [8].

Selain itu, pada penelitian pengamatan sentimen tentang adanya *ransomware* pada data KIP-K yang membandingkan teknik *Naïve Bayes* serta SVM. Hasil menyatakan bahwa SVM lebih tinggi pada nilai akurasi sebesar 80% dibandingkan *Naïve Bayes* 70% [9]. Dan analisis sentimen tentang ulasan pengguna aplikasi *LinkedIn* dengan 2000 dataset, menunjukkan skor akurasi sebanyak 88%, *precision* 88%, *recall* 85%, serta *F1-score* 86% pada model *Naïve Bayes*. Sementara itu, SVM mencapai akurasi 90%, *precision* 89%, *recall* 88%, serta *F1-score* 88%. yang membuktikan SVM memiliki kinerja lebih unggul pada penelitian ini [10]. *Naïve Bayes* merupakan algoritma berdasarkan probabilitas yang memiliki konsep sederhana, efisien, dan bekerja dengan berdasarkan prinsip *teorema Bayes* dengan dugaan bahwa tiap bagian pada data dianggap saling independen. [11]. Keunggulan *Naïve Bayes* terletak pada kecepatannya dalam memproses data, menjadikannya ideal untuk data yang besar dan tidak terlalu kompleks. Namun, asumsi independensi fitur yang menjadi dasar algoritma ini sering kali tidak memenuhi didalam data dunia nyata, sehingga dapat menurunkan akurasi hasil dari klasifikasi. Sementara, *Support Vector Machine* (SVM) ialah algoritma yang berfokus pada pencarian *hyperplane* optimal yang mampu membagi data masuk kelas-kelas yang berbeda pada margin maksimal [12].

Selain itu, SVM (*Support Vector Machine*) ialah satu dari tipe algoritma yang biasa dipakai pada teknik *machine learning* [13]. SVM sangat efektif untuk data dengan dimensi tinggi dan pola yang kompleks, terutama jika data memiliki distribusi non-linear. Dengan menggunakan *kernel trick*, SVM bisa memproyeksikan dataset ke dimensi yang lebih baik dan tinggi, akibatnya mungkin akan dapat memisahkan dataset yang unggul baik di antara kelas. Namun, SVM memiliki keterbatasan dalam hal kecepatan pemrosesan, terutama ketika digunakan pada dataset yang besar. Selain itu, pengaturan parameter SVM, seperti *kernel*, regularisasi, dan margin, membutuhkan perhatian khusus untuk mencapai performa optimal. Studi ini dilaksanakan untuk menganalisis rasio kinerja diantara algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) pada menganalisis sentimen pengguna AI di Platform X, dengan evaluasi kinerja yang dilakukan melalui metrik-metrik yang dihitung memakai *confusion matrix*, yakni akurasi, *precision*, *recall*, serta *F1-score*. Selain itu, Penelitian ini diharapkan mampu mengenali model dengan efektivitas tertinggi serta menyajikan wawasan mendalam mengenai efektivitas metode analisis sentimen.

Dengan demikian, studi ini diharapkan mampu memberi wawasan yang lebih lanjut dan luas bagi pengembang AI, peneliti, serta pemangku kebijakan dalam merancang metode interaksi yang tepat serta mengembangkan teknologi yang selaras pada keperluan dan dinamika penduduk. Di lain sisi, hasil studi ini pula bisa sebagai rujukan untuk akademisi pada studi lebih lanjut terkait penerapan kecerdasan buatan dalam analisis sentimen dan pengolahan data. Dengan wawasan yang lebih baik terhadap kinerja algoritma *Naïve Bayes* serta *Support Vector Machine* (SVM), implementasi teknologi ini diharapkan bisa semakin optimal pada berbagai bidang, termasuk bisnis, pendidikan, kesehatan, serta pemerintahan. Dalam dunia bisnis, hasil penelitian ini dapat membantu perusahaan dalam memahami opini pelanggan, menambah pengetahuan pengguna, serta merancang pendekatan pemasaran yang lebih efisien dan berhasil.

Sementara itu, di sektor pendidikan, teknologi berbasis AI dapat digunakan untuk menganalisis sentimen siswa dan memberikan wawasan bagi tenaga pendidik dalam meningkatkan metode pembelajaran yang lebih adaptif dan personal. Di bidang kesehatan, penerapan AI dalam analisis sentimen dapat mendukung pemantauan persepsi masyarakat terhadap layanan kesehatan serta membantu tenaga medis dalam memahami kebutuhan pasien secara lebih mendalam. Sementara itu, dalam sektor pemerintahan, teknologi ini dapat digunakan untuk menganalisis opini publik terhadap kebijakan tertentu, sehingga para pengambil keputusan dapat merancang Peraturan yang lebih dapat diadaptasi serta selaras dengan keperluan masyarakat. Dengan hasil studi ini, diharapkan perkembangan teknologi kecerdasan buatan dapat semakin disesuaikan dengan kebutuhan nyata di berbagai sektor, sehingga tidak hanya meningkatkan efisiensi dan efektivitas, tetapi juga menyediakan dampak positif yang lebih besar bagi masyarakat secara keseluruhan.

2. METODOLOGI PENELITIAN

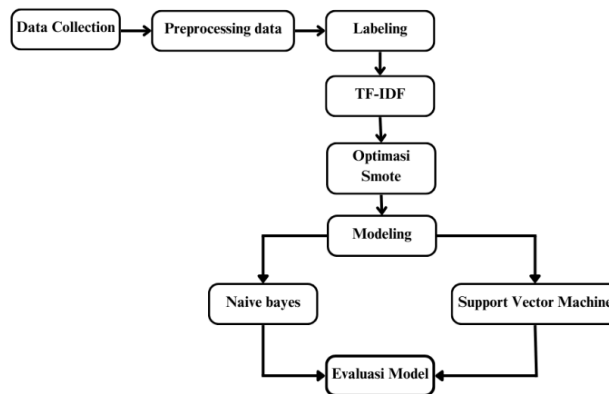
2.1 Tahapan penelitian

Dalam penelitian ini disusun agar memastikan analisis sentimen pengguna AI di Platform X dilakukan secara Dalam penelitian ini, analisis sentimen pengguna AI di Platform X disusun secara sistematis untuk memastikan keakuratan hasil. Penelitian ini melibatkan beberapa tahapan utama, sebagaimana ditampilkan dalam Gambar 1. Proses diawali dengan pengambilan data, yang kemudian dilanjutkan dengan pra-pemrosesan teks untuk membersihkan dan menyiapkan data agar dapat dianalisis secara optimal. Setelah itu, dilakukan penamaan data sebagai bagian dari proses persiapan klasifikasi. Selanjutnya, ekstraksi fitur dilakukan untuk mendapatkan representasi data yang lebih bermakna

bagi model klasifikasi. Pada tahap ini, algoritma Naive Bayes dan SVM diterapkan untuk mengklasifikasikan sentimen berdasarkan pola yang ditemukan dalam data.

Pelabelan data dalam penelitian ini dilakukan secara manual, di mana setiap sampel dianalisis dan dikategorikan berdasarkan sentimen yang terkandung di dalamnya. Pendekatan manual ini memastikan bahwa label yang diberikan sesuai dengan konteks yang sebenarnya, sehingga meningkatkan akurasi model dalam memahami pola sentimen. Meskipun pelabelan manual memerlukan waktu dan usaha lebih, metode ini memberikan hasil yang lebih dapat diandalkan dibandingkan pelabelan otomatis yang berisiko mengalami kesalahan interpretasi.

Tahap akhir dari penelitian ini adalah evaluasi performa algoritma yang digunakan. Evaluasi dilakukan dengan pendekatan berbasis data guna memastikan bahwa model yang dikembangkan mampu mengklasifikasikan sentimen dengan akurat dan konsisten. Dengan tahapan yang terstruktur ini, penelitian diharapkan dapat memberikan wawasan yang lebih akurat mengenai sentimen pengguna AI di Platform X.



Gambar 1. Proses Studi

2.2 Data Collection

Data *Collection* merupakan tahap pengumpulan data yang digunakan agar memperoleh informasi yang sesuai dan mendukung studi [14]. Studi ini juga mengumpulkan data oleh unggahan pengguna di Platform X menggunakan teknik *crawling* dengan menggunakan token autentikasi aplikasi X untuk menjangkau API (*Application Programming Interface*). Proses pengambilan data dilakukan dengan memanfaatkan *library tweet harvest* dengan menggunakan *keyword* yang relevan terkait AI, dengan fokus pada tweet yang berbahasa Indonesia. Teknik ini akan mengumpulkan data secara otomatis, sehingga memperlancar proses pengumpulan data [15]. Data yang berhasil dikumpulkan sebanyak 9.152 data, kemudian disimpan pada bentuk format CSV untuk proses lanjut dalam pengamatan.

2.3 Preprocessing data

Preprocessing data menjadi tahap penting pada proses pengamatan sentimen guna memastikan bahwa data yang dipakai bersih, terstruktur, serta siap diproses oleh model klasifikasi [16].

2.3.1 Cleaning data

Pada ini, data teks diusahakan tetap relevan dengan menghilangkan komponen-komponen yang tidak relevan. atau dapat mengganggu analisis, seperti emoji, simbol, *hashtag*, *mention* (@username), tanda baca, angka, dan URL [17]. Proses pembersihan ini bertujuan untuk menghilangkan semua komponen yang tidak memberikan kontribusi terhadap analisis sentimen, sehingga hanya informasi penting yang tersisa untuk dianalisis lebih lanjut.

2.3.2 Case Folding

Tahapan ini merupakan proses normalisasi teks dilakukan dengan mengonversi seluruh karakter didalam teks menjadi huruf kecil guna memastikan konsistensi dalam analisis [18]. Misalnya, kata "AI" dan "ai" akan dianggap sama, sehingga menghindari duplikasi atau perbedaan yang tidak diperlukan.

2.3.3 Tokenizing

Tahap memisahkan teks ke dalam bagian-bagian yang lebih kecil, dikenal sebagai token [19]. Token ini bisa meliputi kata, frasa, ataupun karakter, sesuai dalam metode yang diterapkan. Tujuan utama tokenisasi adalah menyederhanakan teks agar dapat dianalisis dengan lebih efektif, sehingga elemen-elemen individual dapat diproses dan diinterpretasikan oleh algoritma.

2.3.4 Stopword Removal

Tahapan yang digunakan untuk menghilangkan kata-kata yang sering digunakan tetapi tidak memiliki makna signifikan dalam analisis teks [20]. Kata-kata ini umumnya mencakup kata penghubung, kata ganti, atau kata yang

sering digunakan namun tidak berperan besar dalam pemahaman makna. dalam suatu teks, seperti “yang”, “di”, “ke” dan sebagainya.

2.3.5 Stemming

Tahapan ini merupakan proses terakhir dari Preprocessing data. proses merubah kata menjadi model dasarnya dengan menghilangkan imbuhan, seperti awalan, sisipan, dan akhiran [21]. Adapun tujuannya untuk menyederhanakan kata sehingga kata-kata dengan makna yang sama dapat diidentifikasi sebagai satu entitas, meskipun memiliki variasi bentuk.

2.3.6 Labeling data

Proses memberikan kategori atau kelas tertentu pada data untuk memudahkan analisis dan pemodelan. Dalam analisis sentimen, data umumnya diklasifikasikan sebagai positif, negatif, ataupun netral. Pelabelan dapat dilaksanakan dengan manual maupun otomatis pada memanfaatkan kamus sentimen ataupun model pengelompokan berbasis machine learning [22].

2.3.7 Feature Extraction

Feature extraction merupakan proses mengonversi data teks mentah untuk menjadikan representasi numerik dengan cara memberikan bobot pada setiap kata atau fitur, sehingga dapat lebih mudah dipahami oleh algoritma pembelajaran mesin [23]. Term Frequency (TF) mengukur jumlah dan munculnya sebuah kata dalam bentuk dokumen, sementara Inverse Document Frequency (IDF) mengukur kelangkaan kata tersebut pada semua kumpulan dokumen. Percampuran kedua konsep ini memungkinkan penyesuaian bobot kata sesuai dengan frekuensinya pada dokumen serta distribusinya di seluruh dataset [24].

2.4 Optimasi Smote

Optimasi SMOTE (Synthetic Minority Over-sampling Technique) ialah teknik yang dipakai untuk menyelesaikan ketidakseimbangan data dengan cara menambahkan sampel sintetis pada kelas yang jumlahnya lebih sedikit [25]. Dalam analisis sentimen ini setelah dilakukan pelabelan terjadi perbedaan yang signifikan antara jumlah kedua kelas, untuk kelas yang bersentimen positif berjumlah 6192, sedangkan kelas sentimen negatif hanya 1775 data. Oleh karena itu SMOTE bekerja dengan membuat sampel baru berdasarkan data yang sudah ada, bukan hanya dengan menduplikasi data, tetapi dengan menciptakan data sintetis yang mirip dengan sampel asli, sehingga didapat hasil kelas jumlah kelas minoritas telah bertambah hingga menyamai jumlah kelas mayoritas, yaitu 6.192 data. Metode ini berkontribusi dalam menyeimbangkan dataset, untuk meningkatkan kualitas model klasifikasi, serta dapat menghasilkan perkiraan yang lebih tepat tanpa tanpa terpengaruh oleh ketidakseimbangan kelas.

2.5 Evaluasi Model

Evaluasi model ini dilakukan guna mengukur performa algoritma *Naïve Bayes* serta *Support Vector Machine (SVM)* dalam proses pengelompokan opini publik terkait perbandingan kedua metode dalam analisis pengguna AI di Platform X. Proses ini memanfaatkan *confusion matrix* Untuk menguji berbagai evaluasi, yakni akurasi, presisi, *recall*, serta *F1-score*, guna memastikan efektivitas model dalam melakukan prediksi [26].

$$Akurasi = \frac{TP+TN}{FP+FN+TP+TN} \times 100 \quad (1)$$

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \cdot \frac{Presisi \cdot Recall}{Presisi + Recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Di studi ini, data didapat dari platform x dengan teknik Crawling, yaitu teknik otomatis untuk mengumpulkan data dari internet. Teknik ini dilakukan dengan menggunakan Library Tweet Harvest, yang memungkinkan pengambilan data berdasarkan Istilah tertentu. kata kunci yang diterapkan dalam penelitian ini berfokus pada Artificial Intelligence (AI) agar data yang diperoleh relevan dengan topik yang diteliti. Proses ini dijalankan menggunakan Google Colab, sebuah platform berbasis cloud yang mendukung pemrograman Python. Google Colab dipilih karena menyediakan lingkungan yang mudah digunakan serta memiliki akses ke berbagai pustaka dan sumber daya komputasi yang memadai. Untuk mengakses data dari X, penelitian ini menggunakan API resmi yang memerlukan kode autentikasi agar dapat mengumpulkan tweet sesuai dengan aturan yang ditetapkan oleh platform. Hasil dari proses crawling ini terkumpul sebanyak 9.183 tweet mentah yang mengandung berbagai informasi terkait Artificial Intelligence. Data



yang diperoleh masih dalam bentuk mentah, sehingga perlu dilakukan proses pembersihan sebelum diamati lebih dalam. Seluruh data yang sudah diambil berikutnya data dikumpulkan dalam format CSV. (*Comma-Separated Values*), yang mempermudah proses pengolahan dan analisis pada tahap selanjutnya. Hasil pengumpulan dataset bisa diamati dalam Tabel 1.

Tabel 1. Dataset

No	Pelabelan Data
1	Program Studi Pendidikan Matematika Universitas Muhammadiyah Malang telah mengadakan kegiatan pengabdian kepada masyarakat berupa pelatihan pemanfaatan kecerdasan buatan (Artificial Intelligence/AI) dalam penyusunan instrumen pembelajaran~.
2	Sudah mulai merasakan kan dampak ai, makanya penggunaannya harus diatur biar tidak timbul. seperti ini sayang dulu ada paslon yang melegalkan penggunaan ai jadinya pada merasa boleh pakai ai!~

3.2 Preprocessing Data

Dalam analisis ini *dataset* yang digunakan awalnya terdiri dari sekumpulan teks yang tidak memiliki struktur sehingga memerlukan serangkaian langkah pra-pemrosesan guna menyiapkan dataset secara optimal. Tahap ini mencakup berbagai proses terdiri dari *Cleaning Data*, *Case Folding*, *Tokenizing Data*, *StopWord* serta *Stemming*. Didapat hasil data bersih berjumlah 7.967 tweet setelah melalui serangkaian tahapan *preprocessing*. Proses *cleaning data* dilakukan dengan menghapus karakter tertentu berupa tanda baca, numerik, URL, serta simbol yang tidak sesuai dalam pengamatan. Selanjutnya, *case folding* diterapkan dengan merubah semua teks jadi huruf kecil agar memastikan konsistensi pada pemrosesan kata. Setelah itu, dilakukan *tokenizing*, yaitu memisah teks jadi kata-kata sendiri supaya lebih gampang diamati. Pada tahap *stopword removal*, kata-kata yang dianggap tidak memiliki makna signifikan dalam pengamatan, seperti “kepada”, serta “telah”, dihapus untuk meningkatkan relevansi data. Terakhir, tahap *stemming* dilaksanakan dengan mengonversi kata-kata ke dalam model dasarnya, seperti kata “berjalan” diubah menjadi “jalan”. Setelah melalui seluruh tahapan tersebut, dataset yang telah dibersihkan siap digunakan untuk analisis lebih lanjut. Dapat dilihat pada Tabel 1 hasil dari *preprocessing data*.

Tabel 2. Hasil Preprocessing data

Tahapan	Hasil Preprocessing data
Cleaning Data	Program studi pendidikan matematika universitas muhammadiyah malang telah mengadakan kegiatan pengabdian kepada masyarakat berupa pelatihan pemanfaatan kecerdasan buatan artificial intelligence ai dalam penyusunan instrumen pembelajaran
Case Folding	program studi pendidikan matematika universitas muhammadiyah malang telah mengadakan kegiatan pengabdian kepada masyarakat berupa pelatihan pemanfaatan kecerdasan buatan artificial intelligence ai dalam penyusunan instrumen pembelajaran
Tokenizing Data	['program', 'studi', 'pendidikan', 'matematika', 'universitas', 'muhammadiyah', 'malang', 'telah', 'mengadakan', 'kegiatan', 'pengabdian', 'kepada', 'masyarakat', 'berupa', 'pelatihan', 'pemanfaatan', 'kecerdasan', 'buatan', 'artificial', 'intelligence', 'ai', 'dalam', 'penyusunan', 'instrumen', 'pembelajaran']
StopWord	['program', 'studi', 'pendidikan', 'matematika', 'universitas', 'muhammadiyah', 'malang', 'mengadakan', 'kegiatan', 'pengabdian', 'masyarakat', 'berupa', 'pelatihan', 'pemanfaatan', 'kecerdasan', 'buatan', 'artificial', 'intelligence', 'ai', 'penyusunan', 'instrumen', 'pembelajaran']
Stemming	program studi didik matematika universitas muhammadiyah malang ada giat abdi masyarakat upa latih manfaat cerdas buat artificial intelligence ai susun instrumen ajar

3.3 Pelabelan Data

Selanjutnya, dilakukan tahapan pelabelan data untuk mengkategorikan tweet sesuai dengan konteks yang diinginkan dalam penelitian ini. Pelabelan dilakukan secara manual untuk memastikan setiap label mencerminkan isi tweet dengan akurat. Pendekatan ini dipilih karena metode otomatis sering kali kurang tepat dalam memahami konteks tertentu, terutama jika terdapat ambiguitas dalam bahasa atau istilah teknis. Dalam proses ini, setiap tweet dianalisis dengan cermat berdasarkan kata kunci, struktur kalimat, dan maknanya. Pelabelan dilakukan mengikuti pedoman yang telah ditetapkan agar konsistensi tetap terjaga. Setelah semua data diberi label, hasilnya disusun Dalam tabel yang dapat ditemukan pada Tabel 2. Tahap ini penting karena kualitas pelabelan akan berpengaruh pada hasil analisis selanjutnya. Dengan pelabelan yang akurat, penelitian dapat memperoleh wawasan yang lebih valid mengenai tren dan pola terkait *Artificial Intelligence* di media sosial.

Tabel 3. Hasil pelabelan

Tweet	Pelabelan Data
Program Studi Pendidikan Matematika Universitas Muhammadiyah Malang telah menyelenggarakan kegiatan pengabdian kepada masyarakat berupa pelatihan pemanfaatan kecerdasan buatan (AI) dalam pengembangan instrumen pembelajaran.	Positif

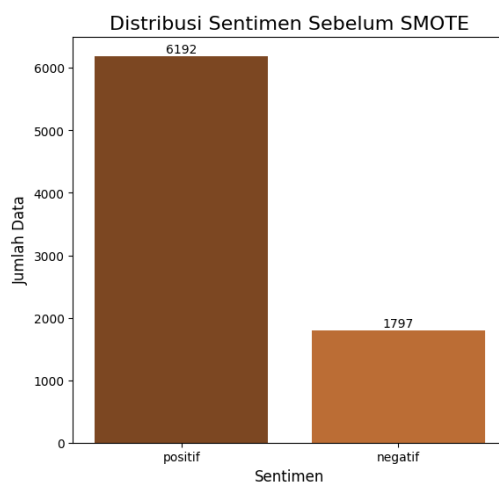
sudah mulai merasakan kan dampak ai makanya penggunaannya harus diatur biar tidak timbul seperti ini sayang dulu ada paslon yang melegalkan penggunaan ai jadinya pada merasa boleh pakai ai.

Negatif

3.4 Optimasi SMOTE

Setelah Setelah pelabelan, yang dapat dilihat pada Gambar 2, diperoleh 1.797 tweet negatif dan 6.192 tweet positif, yang menunjukkan adanya ketidakseimbangan data. Ketidakseimbangan ini dapat menyebabkan model machine learning lebih cenderung mengenali sentimen positif dibandingkan negatif, sehingga mengurangi akurasi klasifikasi. Masalah ini juga bisa memengaruhi evaluasi model, karena akurasi mungkin tampak tinggi tetapi tidak mencerminkan performa sebenarnya. Oleh karena itu, diperlukan solusi untuk menyeimbangkan data agar model dapat bekerja lebih adil dan akurat. Salah satu solusinya adalah menggunakan SMOTE (*Synthetic Minority Over-Sampling Technique*), yaitu metode yang membuat sampel baru agar kelompok yang kurang terwakili. (negatif) berdasarkan data yang telah ada. Dengan teknik ini, jumlah data minoritas bisa ditingkatkan hingga lebih seimbang dengan data mayoritas, Sehingga model dapat belajar lebih optimal tanpa terpengaruh bias terhadap kelas tertentu Selain SMOTE, metode lain yang dapat digunakan untuk menangani ketidakseimbangan data adalah undersampling, di mana jumlah sampel dari kelas mayoritas dikurangi agar lebih seimbang dengan kelas minoritas. Namun, pendekatan ini berisiko menghilangkan informasi yang mungkin berharga dalam proses pelatihan model. Alternatif lain yang bisa diterapkan adalah penyesuaian bobot kelas (*class weighting*) saat melatih model, sehingga model lebih mempertimbangkan kelas minoritas meskipun jumlahnya lebih sedikit. Dengan menerapkan strategi penyeimbangan data seperti SMOTE, undersampling, atau class weighting, diharapkan Model dapat mengidentifikasi kedua sentimen dengan lebih tepat dan menghasilkan prediksi yang lebih baik serta seimbang. Langkah ini sangat penting dalam memastikan bahwa analisis sentimen yang dilakukan tidak bias terhadap kelas mayoritas, sehingga hasilnya lebih representatif dan dapat diandalkan.

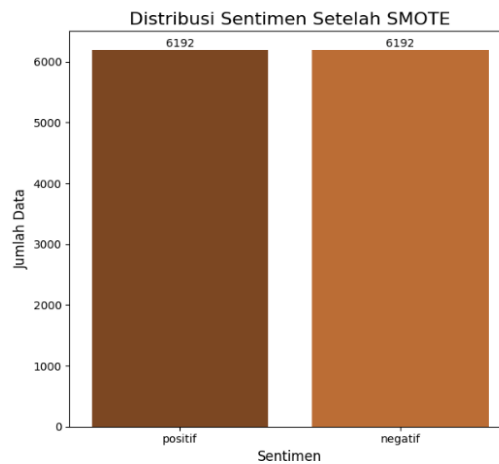
Dalam menangani ketidakseimbangan kelas pada data, berbagai metode dapat digunakan, seperti SMOTE (*Synthetic Minority Over-sampling Technique*), class weighting, dan undersampling. Namun, SMOTE sering kali menjadi pilihan yang lebih unggul karena beberapa alasan utama. Pertama, SMOTE tidak hanya meningkatkan jumlah sampel dalam kelas minoritas tetapi juga menciptakan data sintesis baru berdasarkan distribusi yang ada. Hal ini membantu model untuk belajar pola yang lebih representatif dibanding sekadar menduplikasi data yang ada, seperti yang terjadi pada oversampling tradisional. Kedua, dibandingkan dengan class weighting yang hanya memberikan bobot lebih besar pada kelas minoritas selama pelatihan, SMOTE secara langsung menyeimbangkan distribusi data sehingga model tidak hanya mengandalkan penyesuaian bobot dalam proses pembelajaran. Dengan demikian, SMOTE lebih efektif dalam mengatasi masalah bias model terhadap kelas mayoritas. Ketiga, metode undersampling mengurangi jumlah data dari kelas mayoritas untuk mencapai keseimbangan, tetapi hal ini dapat menyebabkan hilangnya informasi berharga dan mengurangi kemampuan model dalam mengenali pola secara menyeluruh. Sebaliknya, SMOTE mempertahankan seluruh data dari kelas mayoritas sambil memperkaya kelas minoritas, sehingga model dapat bekerja dengan lebih baik dalam kondisi data yang lebih seimbang. Dengan keunggulan-keunggulan tersebut, SMOTE menjadi metode yang sering dipilih dalam penanganan ketidakseimbangan kelas, karena dapat meningkatkan performa model tanpa mengorbankan informasi yang tersedia.



Gambar 2. Jumlah klasifikasi sentimen sebelum Smote

Gambar 3 ini menunjukkan distribusi sentimen setelah diterapkan metode SMOTE (*Synthetic Minority Over-Sampling Technique*). Sebelumnya, terdapat ketidakseimbangan data, di mana jumlah tweet positif (6.192) jauh lebih banyak dibandingkan tweet negatif (1.797), yang dapat menyebabkan model lebih cenderung mengklasifikasikan data ke kelas mayoritas dan menghasilkan bias dalam prediksi. Setelah SMOTE diterapkan, jumlah tweet negatif ditingkatkan menjadi 6.192, sehingga kedua kelas memiliki jumlah data yang seimbang. Teknik SMOTE bekerja

dengan membuat sampel sintesis berdasarkan pola dalam data asli, bukan sekadar menduplikasi data yang ada. Sampel sintesis ini dihasilkan dengan menganalisis hubungan antar data dalam kelas minoritas, sehingga data baru tetap memiliki karakteristik yang menyerupai data asli. Dengan data yang lebih seimbang, model dapat belajar secara lebih adil tanpa terlalu condong ke salah satu kelas. Hal ini penting karena model yang dilatih menggunakan data dengan distribusi yang tidak seimbang cenderung memiliki performa buruk dalam mengenali kelas dengan jumlah data yang lebih sedikit. Dengan penerapan *SMOTE*, diharapkan akurasi klasifikasi sentimen meningkat model yang mampu menghasilkan prediksi yang lebih tepat serta dapat dipercaya.



Gambar 3. Total klasifikasi sentimen sesudah Smote

3.5 Evaluasi Model

Di proses ini, pengujian dapat dilakukan dengan proses dua bentuk pengelompokan, yaitu *Naïve Bayes* dan *Support Vector Machine (SVM)*, yang kemudian dibandingkan agar mengevaluasi performa model yang paling cocok dalam analisis pada *dataset* ini. Pengujian dilakukan menggunakan berbagai metrik penilaian, berupa *confusion matrix*, akurasi, *presisi*, *recall*, dan *F1-score*. yang dapat ditunjukkan dalam Tabel 3, dari hasil pengukuran bisa membuktikan bahwa SVM mempunyai tingkat akurasi yang lebih baik serta lebih unggul, yaitu 96%, sedangkan *Naïve Bayes* hanya mencapai 84%. Hal tersebut menyatakan bahwa bentuk SVM membuktikan kinerja yang lebih unggul daripada algoritma *Naïve Bayes* pada pengamatan sentimen pada *dataset* yang digunakan. Kelebihan ini mencerminkan kemampuan SVM dalam mendeteksi pola data secara lebih efektif, terutama ketika *dataset* telah dioptimalkan melalui teknik *Synthetic Minority Over-Sampling Technique (SMOTE)*. Teknik SMOTE membantu menambah jumlah *dataset* di kelas minoritas dengan menciptakan sampel sintesis berdasarkan pola yang ada, sehingga mengurangi ketidakseimbangan data dan memungkinkan model untuk mengenali kedua kelas sentimen dengan lebih baik. Selain itu, SVM bekerja lebih optimal dalam memisahkan data berdasarkan *hyperplane* yang maksimal, sehingga mampu menangani data dengan dimensi tinggi dan distribusi yang kompleks. Hal ini membuat SVM lebih baik dalam mengklasifikasikan sentimen. secara akurat daripada dengan *Naïve Bayes*, yang cenderung mempunyai tanggapan keterpisahan antar karakteristik yang tidak selalu relevan dalam analisis teks. Dengan demikian, pemilihan model klasifikasi yang tepat sangat bergantung pada karakteristik *dataset* yang dipakai. Pada studi ini, hasil evaluasi menyatakan SVM lebih handal dalam melakukan klasifikasi sentimen dibandingkan dengan *Naïve Bayes*, terutama setelah proses *preprocessing* dan penyeimbangan data dilakukan.

Tabel 4. Hasil dari akurasi

Model	Akurasi
Naïve Bayes	84%
Support Vector Machine	96%

Dari Tabel 4 hasil klasifikasi model sentimen, *Naïve Bayes* dan *Support Vector Machine (SVM)* dibandingkan berdasarkan *Precision*, *Recall*, dan *F1-Score* untuk sentimen positif dan negatif. Pada *Naïve Bayes*, untuk sentimen positif, *precision* 90%, *recall* 75%, dan *F1-Score* 82%, membuktikan bahwa model cukup unggul dalam memahami data positif. Untuk sentimen negatif, *precision* 78%, *recall* 91%, dan *F1-Score* 86%, yang berarti model lebih akurat dalam mendeteksi data negatif dibandingkan positif. Sementara itu, SVM memiliki kinerja lebih baik dalam kedua kategori sentimen. Untuk sentimen positif, *precision* 97%, *recall* 94%, serta *F1-Score* 96%, sedangkan untuk sentimen negatif, *precision* 94%, *recall* 97%, serta *F1-Score* 96%. Secara keseluruhan, SVM lebih tinggi daripada *Naïve Bayes*, karena memiliki nilai *precision*, *recall*, serta *F1-Score* yang lebih unggul, sehingga lebih akurat pada mengklasifikasikan sentimen positif maupun negatif.

Tabel 5. Hasil klasifikasi

Model	Sentimen	Precision	Recall	F1-Score
-------	----------	-----------	--------	----------

Naive bayes	Positif	90%	75%	82%
	Negatif	78%	91%	84%
SVM	Positif	97%	94%	96%
	Negatif	94%	97%	96%

Pada hasil perbandingan evaluasi klasifikasi *Naïve Bayes* serta *Support Vector Machine (SVM)* bisa dibuktikan pada Gambar 4, yang menunjukkan *confusion matrix* dari masing-masing model.

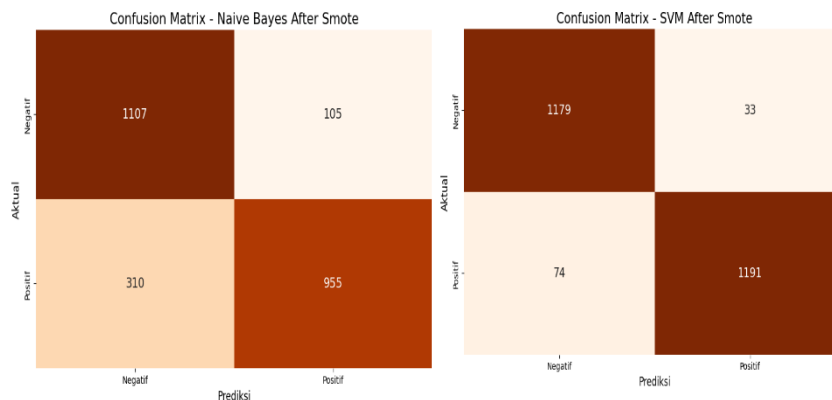
a. Pada model *Naïve Bayes*, nilai matriks yang diperoleh adalah:

1. *True Negative*, (TN): 1107
2. *False Positive*, (FP): 105
3. *False Negative*, (FN): 310
4. *True Positive*, (TP): 995

b. Sementara itu, model *SVM* menunjukkan hasil sebagai berikut:

1. *True Negative*, (TN): 1179
2. *False Positive*, (FP): 33
3. *False Negative*, (FN): 74
4. *True Positive*, (TP): 1191

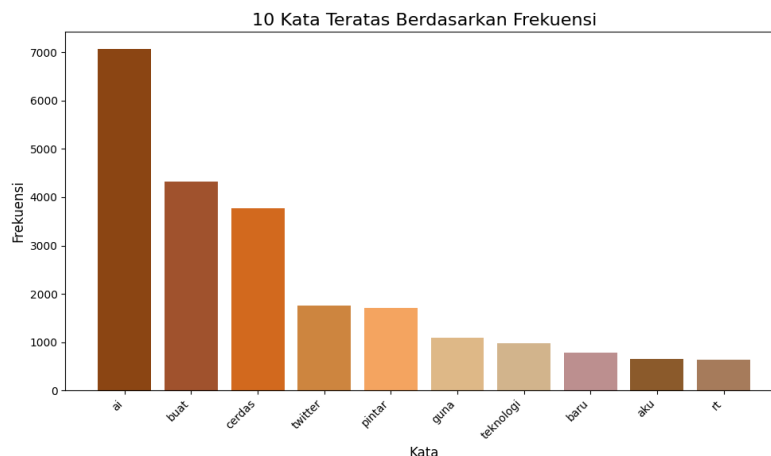
Jika dibandingkan berdasarkan nilai FP dan FN yang dimiliki oleh kedua model klasifikasi, terlihat bahwa *SVM* memiliki tingkat kesalahan yang lebih rendah pada membaca kedua kelas, baik positif maupun negatif. Nilai FP yang lebih kecil pada *SVM* menunjukkan bahwa bentuk ini lebih sedikit membuat kesalahan pada mengklasifikasikan data negatif menjadi positif, sedangkan nilai FN yang lebih rendah menandakan bahwa *SVM* lebih mampu mendeteksi data positif dengan lebih akurat dibandingkan *Naïve Bayes*. Dengan kata lain, *SVM* memiliki tingkat prediksi yang lebih akurat karena menghasilkan lebih sedikit kesalahan klasifikasi dibandingkan dengan *Naïve Bayes*. Keunggulan ini dapat dikaitkan dengan cara kerja *SVM* yang membangun *hyperplane* optimal untuk memisahkan kelas data secara lebih baik, terutama dalam kasus data yang memiliki pola kompleks. Dari hasil studi ini, bisa ditarik kesimpulan bahwa *Support Vector Machine (SVM)* lebih unggul untuk memprediksi hasil klasifikasi dibandingkan dengan *Naïve Bayes*, menjadikannya model ini lebih efektif untuk analisis sentimen pada dataset yang digunakan dalam penelitian ini.



Gambar 4. *Confusion matrix .Naïve Bayes & SVM.*

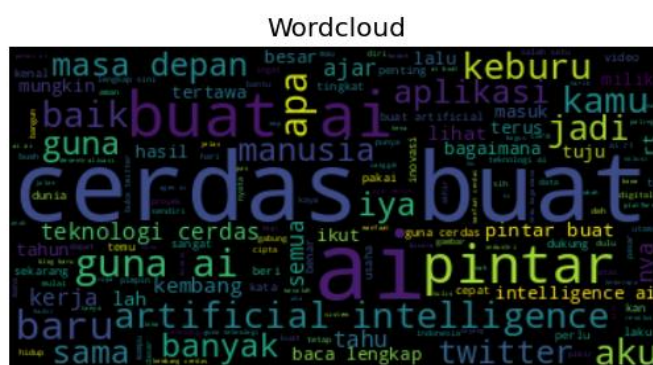
3.6 Visualisasi

Dalam penelitian ini visualisasi Memanfaatkan diagram batang untuk memunculkan sepuluh kata yang paling umum muncul didalam tweet. Sebagaimana ditampilkan pada Gambar 5. Kata "ai" memiliki frekuensi tertinggi, menunjukkan bahwa topik yang dibahas dalam tweet, yang berkaitan dengan Kecerdasan Buatan (AI). Sedangkan "buat" menggambarkan proses atau tindakan menciptakan sesuatu yang berkaitan dengan inovasi. "Cerdas" dan "pintar" mengindikasikan kualitas yang diharapkan dari sistem AI, menunjukkan bahwa teknologi tersebut dirancang untuk memiliki kemampuan tinggi. "Twitter" mengacu pada platform media sosial yang menjadi sumber pengumpulan data, sedangkan "guna" menekankan fungsi atau manfaat yang ditawarkan oleh teknologi. Kata "teknologi" merupakan inti dari pembahasan, menyoroti alat dan sistem yang mendukung AI, sementara "baru" menunjukkan adanya fitur atau inovasi terkini. Kata "aku" mencerminkan perspektif personal pengguna, dan "it" merupakan istilah bahasa Inggris yang sering digunakan dalam konteks teknologi informasi. Diagram batang ini membantu mengidentifikasi tema utama dan fokus diskusi masyarakat terkait AI, sehingga memberikan gambaran yang lebih jelas tentang topik yang sering muncul. mencerminkan



Gambar 5. 10 frekuensi kata tertinggi

Selain itu, penelitian ini menggunakan word cloud Guna memperlihatkan kata yang paling umum muncul terlihat dalam dataset. Teknik ini membantu dalam mempermudah analisis teks dengan memberikan visualisasi yang jelas mengenai kata-kata dominan. *Word cloud* bekerja dengan menampilkan Kata-kata yang paling sering digunakan atau yang sering muncul ditunjukkan ukurannya lebih banyak dibandingkan kata-kata yang jarang muncul, sehingga memudahkan peneliti dalam mengidentifikasi istilah-istilah yang paling relevan dengan topik yang diteliti. Seperti pada Gambar 6, kata "Cerdas Buat" muncul dengan ukuran yang paling besar, menunjukkan bahwa kata tersebut memiliki frekuensi yang tinggi dan keterkaitan erat dengan tema penelitian. Selain itu, beberapa kata lain yang muncul dengan ukuran cukup besar mencerminkan konsep utama dalam diskusi terkait kecerdasan buatan. Dengan visualisasi ini, dapat dilihat bahwa fokus utama pembicaraan dalam dataset berkisar pada aspek kecerdasan (*intelligence*), teknologi (*technology*), serta inovasi (*innovation*). Penggunaan word cloud memberikan wawasan tambahan tentang pola penggunaan kata dalam analisis sentimen, serta dapat membantu mengidentifikasi tren kata yang sering digunakan dalam berbagai kategori sentimen. Teknik ini juga berguna dalam memahami bagaimana pengguna media sosial berinteraksi dengan topik tertentu dan sejauh mana suatu istilah menjadi perhatian utama dalam diskusi. Dengan teknik ini, analisis menjadi lebih mudah dipahami dan dapat dikembangkan lebih lanjut dengan metode visualisasi lainnya, seperti analisis jaringan kata (*word network analysis*) atau pemetaan hubungan semantik (*semantic mapping*), untuk memperoleh Pemahaman yang lebih komprehensif dalam opini pengguna terhadap kecerdasan buatan.



Gambar 6. Wordcloud

4. KESIMPULAN

Berdasarkan hasil studi ini, diperoleh rasio akurasi antara teknik *Naive Bayes* serta *Support Vector Machine (SVM)* pada pengamatan sentimen pemakai Kecerdasan Buatan pada Platform X. Hasil pengujian membuktikan bahwa *Naive Bayes* mempunyai tingkat akurasi sebanyak 85%, sedangkan SVM mencapai akurasi 93%. Performa SVM yang lebih tinggi membuktikan untuk kemampuan dalam mengelola pola dataset yang memiliki tingkat kompleksitas lebih tinggi, terutama dengan penggunaan teknik pemisahan data yang lebih optimal. Meskipun *Naive Bayes* lebih efisien secara perhitungan dan cepat dalam menangani data besar, asumsi bahwa setiap fitur bersifat mandiri membatasi keakuratannya dalam analisis sentimen yang lebih kompleks. Oleh karena itu, SVM lebih direkomendasikan untuk analisis sentimen di Platform X karena tingkat ketepatannya yang lebih tinggi. Namun, pemilihan metode harus mempertimbangkan beberapa faktor lain seperti kecepatan pemrosesan, ketersediaan sumber daya komputasi, serta kompleksitas data yang dianalisis. *Naive Bayes* tetap menjadi pilihan yang baik jika kecepatan analisis lebih diutamakan, terutama dalam skenario yang membutuhkan pemrosesan data dalam jumlah besar secara cepat.

Sementara itu, SVM lebih cocok digunakan ketika keakuratan menjadi prioritas utama, meskipun memerlukan waktu pemrosesan yang lebih lama dan daya komputasi yang lebih besar. Selain itu, dalam pengembangan analisis sentimen ke depan, dapat dipertimbangkan penggunaan metode lain seperti pembelajaran mendalam yang menggunakan jaringan saraf tiruan untuk meningkatkan akurasi lebih lanjut. Kombinasi metode, seperti penggabungan *Naive Bayes* dan SVM atau penggunaan teknik ensemble learning, juga dapat menjadi strategi yang efektif. Dalam merancang model analisis sentimen yang lebih efektif, presisi dan efisien. Dengan terus berkembangnya teknologi dan peningkatan jumlah data yang tersedia, penelitian lebih lanjut sangat diperlukan untuk menemukan metode yang paling optimal dalam analisis sentimen pengguna kecerdasan buatan di Platform X. Batasan penelitian ini terletak pada sumber data yang digunakan, di mana data hanya diambil dari satu platform media sosial. Hal ini dapat menyebabkan hasil analisis kurang mencerminkan persepsi keseluruhan sentimen masyarakat secara luas. Setiap platform memiliki karakteristik pengguna dan pola interaksi yang berbeda, sehingga sentimen yang terdeteksi mungkin tidak sepenuhnya mewakili opini masyarakat secara keseluruhan. Untuk penelitian selanjutnya, disarankan agar data dikumpulkan dari berbagai platform guna memperoleh gambaran yang lebih komprehensif mengenai sentimen publik.

REFERENCES

- [1] Arnadi Arnadi, Aslan Aslan, & Arnes Yuli Vandika. “Penggunaan Kecerdasan Buatan Untuk Personalisasi Pengalaman Belajar,” *Jurnal Ilmu Pendidikan dan Kearifan Lokal*, vol. 15, no. 1, pp. 37–48, 2024, <https://www.jipkl.com/index.php/JIPKL/article/view/101>
- [2] Siti Masrichah, “Ancaman Dan Peluang Artificial Intelligence (AI),” *Khatulistiwa J. Pendidik. dan Sos. Hum.*, vol. 3, no. 3, pp. 83–101, 2023, doi: 10.55606/khatulistiwa.v3i3.1860.
- [3] A. F. Adhani and A. Aripudin, “Perspektif Generasi Z di Platform X Terhadap Penurunan Angka Pernikahan di Indonesia,” *J-KIs J. Komun. Islam*, vol. 5, no. 1, pp. 185–198, 2024, doi: <https://doi.org/10.53429/j-kis.v5i1.1001>
- [4] D. Ananda and R. R. Suryono, “Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 748, 2024, doi: 10.30865/mib.v8i2.7517.
- [5] C. F. Hasri and D. Alita, “Penerapan Metode Naive Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.
- [6] R. R. Pratama, R. R. Suryono, “Performance Comparison Of Naive Bayes , Support Vector Machine And Random Forest Algorithms For Apple Vision Pro Sentiment Analysis,” *JUTIF*, vol. 6, no. 1, pp. 31–39, 2025, <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/download/4035/716/>
- [7] W. Ningsih, B. Alfaanda, R. Rahmadden, and D. Wulandari, “Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [8] D. Atmajaya, A. Febrianti, and H. Darwis, “Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter,” *Indones. J. Comput. Sci.*, vol. 12, no. 4, pp. 2173–2181, 2023, doi: 10.33022/ijcs.v12i4.3341.
- [9] N. S. Ramadan and D. Darwis, “Perbandingan Metode Naïve Bayes Dan Svm Untuk Sentimen Analisis Masyarakat Terhadap Serangan Ransomware Pada Data Kip-K,” *Simika*, vol. 8, no. 1, pp. 12–23, 2025, doi: <https://doi.org/10.47080/simika.v8i1.3621>
- [10] Gishella Septania Al-Husna, Dian Asmarajati, Iman Ahmad Ihsannuddin, and Rina Mahmudati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi LinkedIn,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 3, no. 2, pp. 139–144, 2024, doi: 10.55123/storage.v3i2.3602.
- [11] Rayuwati, Husna Gemasih, and Irma Nizar, “Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid,” *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [12] Z. Rani and B. K. Khotimah, “Sapi Di Twitter Menggunakan Kombinasi Metode K-Means Dan Support Vector Machine,” *Jitet*, vol. 13, no. 1, 2025, doi: <http://dx.doi.org/10.23960/jitet.v13i1.5685>
- [13] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, “Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM),” *Informatics Digit. Expert*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.
- [14] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, and R. Rahmadden, “Perbandingan Kinerja Naïve Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara,” *J. Ilm. Inform.*, vol. 12, no. 01, pp. 15–21, 2024, doi: 10.33884/jif.v12i01.8721.
- [15] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [16] D. R. Bakti, R. R. Suryono, “Sentiment Analysis Of Post-Covid Online Education Among Gen Z With Various Classification Methods,” *Jutif*, vol. 6, no. 1, pp. 301–310, 2025, <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/4003>
- [17] D. B. Reynaldi and R. R. Suryono, “Comparison Of Accuracy Of Various Text Classification Methods In Sentiment Analysis Of E-Stamps At X Komparasi Akurasi Berbagai Metode Klasifikasi Teks Dalam Analisis Sentimen E-Materai Di X,” *Jutif*, vol. 6, no. 1, pp. 281–290, 2025, <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/download/3999/739/>
- [18] A. Setiawan and R. R. Suryono, “Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.
- [19] S. Syafrizal, M. Afdal, and R. Novita, “Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.
- [20] M. Taufik Sugandi, Martanto, and U. Hayati, “Analisis Sentimen Komentar Pengguna Youtube terhadap Kebijakan Baru



- Badan Penyelenggara Jaminan Kesehatan Sosial Menggunakan Naïve Bayes,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 218–227, 2024, doi: <https://publikasiilmiah.unwahas.ac.id/JINRPL/article/view/10301>
- [21] Aditya, eka, Astawa, I. G. K., Limbong, K. G., Indrawan, G., Indrawan, G., & Gunawan, M. A. O. , “Analisis Sentimen Pengguna Sistem E-Kinerja Desa Kabupaten Jembrana Menggunakan Metode Naive Bayes,“. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, Vol 7, No 1, pp 8-14. <https://doi.org/10.47233/jteksis.v7i1.1693>
- [22] K. Shin, T. Yong, “Analisis Sentimen Kepuasan Publik Terhadap Masa,” *JTIK*, vol. 9, no. March, pp. 149-158, 2025, <https://journal.lembagakita.org/index.php/jtik/article/download/3020/2425/10976>
- [23] C. F. Alifa and D. Alita, “Analisis Opini Publik Tentang Boikot Produk Pro-Israel di Twitter Berbahasa Indonesia Menggunakan Metode SVM,” *J. Inform. J. Pengemb. IT*, vol. 9, no. 2, pp. 112–120, 2024, doi: 10.30591/jpit.v9i2.6559.
- [24] Puspitasari, R., & Dwi Indriyanti, A, ”Analisis Sentimen Opini Publik Terhadap Kebijakan Baru Skripsi Pada Media Sosial Twitter Menggunakan Metode Naive Bayes, ” *Journal of Emerging Information System and Business Intelligence (JEISBI)*, Vol 5, No 3, pp 37–42, <https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/61273>
- [25] R. Aryanti, T. Misriati, and A. Sagiyanto, “Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data,” *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 218–227, 2023, doi: 10.47065/josyc.v5i1.4562.
- [26] R. R. S. Natasha, “Sentiment Analysis Of The Influence Of The Korean Wave In Indonesia Using The Naive Bayes Method And Support Vector Machine,” *Jurnal Inovtek Polbeng*, vol. 10, no. 1, pp. 308–319, 2025, doi: <https://doi.org/10.35314/85x4wd90>