

PERBANDINGAN ALGORITMA NAIVE BAYES DAN KNN DALAM ANALISIS SENTIMEN TWITTER TENTANG DEPRESI DI INDONESIA

Micho Respati Putra¹⁾, Wina Ningsih²⁾, Juanny Cheristy Souisa³⁾, Muhammad Rifqi Asy'ari⁴⁾, Sumanto⁵⁾, Ade Surya Budiman⁶⁾

^{1,2,3,4,5,6}Ilmu Komputer, Fakultas Teknik Informatika, Universitas Bina Sarana Informatika
Jl.Kramat Raya No.98, Kwitang, Senen, Jakarta Pusat, 10450, Indonesia

Corresponding Author: ¹ 15220693@bsi.ac.id

Article Info

Article history:

Received: Juny 27, 2025

Revised: July, 20, 2025

Accepted: sept, 24, 2025

Published: Okt, 30, 2025

Keywords:

Analisis sentimen

Depresi

Twitter

Naïve Bayes

k-Nearest Neighbours

ABSTRACT

Depresi merupakan gangguan mental yang menjadi perhatian global, dengan prevalensi tinggi khususnya di kalangan usia muda. Media sosial seperti Twitter menjadi ruang ekspresi yang potensial dalam mencerminkan kondisi psikologis masyarakat. Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis sentimen publik terkait isu depresi di Indonesia melalui data Twitter menggunakan dua algoritma pembelajaran mesin, yakni Naïve Bayes dan k-Nearest Neighbours (kNN). Data dikumpulkan melalui proses web scraping terhadap 1.342 tweet, yang kemudian disaring menjadi 873 tweet relevan berbahasa Indonesia. Tahapan analisis meliputi praproses teks, ekstraksi fitur, dan klasifikasi sentimen menjadi tiga kategori: positif, negatif, dan netral. Evaluasi kinerja model dilakukan dengan metrik AUC, akurasi, presisi, recall, F1-score, dan Matthews Correlation Coefficient (MCC). Hasil menunjukkan bahwa algoritma kNN memiliki kinerja unggul dengan akurasi 96,4% dan MCC 0,645, sementara Naïve Bayes mencatatkan AUC tertinggi sebesar 0,985. Distribusi data memperlihatkan dominasi sentimen negatif, mengindikasikan bahwa Twitter kerap digunakan sebagai wadah pelampiasan emosi negatif terkait depresi, terutama di kalangan usia produktif. Temuan ini menegaskan bahwa media sosial berpotensi digunakan sebagai alat skrining awal untuk memantau kesehatan mental masyarakat secara dinamis dan real-time. Implikasi dari studi ini membuka peluang pengembangan sistem pemantauan berbasis kecerdasan buatan untuk mendukung intervensi dini dalam penanganan isu depresi di Indonesia.



This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. INTRODUCTION

Menurut WHO Depresi adalah gangguan mental yang ditandai rasa sedih berlarut, gangguan tidur, dan hilangnya nafsu makan. Kondisi ini berdampak besar pada kehidupan dan menjadi penyebab kecacatan pada lebih dari 300 juta orang tiap tahun [1]. Di berbagai negara, depresi telah menjadi isu kesehatan masyarakat yang sangat serius karena tergolong dalam penyakit penyebab disabilitas tertinggi [2]. Lebih jauh, depresi berkontribusi sebagai faktor risiko utama terhadap tindakan bunuh diri. Sekitar 40% individu dengan depresi memiliki keinginan atau pikiran untuk mengakhiri hidup, meskipun hanya sekitar 15% yang benar-benar melakukannya hingga tuntas [3]. Wanita mengalami depresi dua kali lebih sering dibandingkan pria, namun pria memiliki kecenderungan lebih tinggi untuk melakukan bunuh diri [4].

Gangguan ini dapat dialami oleh siapa saja, tanpa mengenal batas usia. Anak-anak, lansia, bahkan

kelompok usia produktif seperti Generasi Z, berisiko mengalami stres berkepanjangan yang berpotensi berkembang menjadi depresi apabila tidak ditangani secara tepat [5].

Berdasarkan Survei Kesehatan Indonesia (SKI) tahun 2023, kelompok usia 15 hingga 24 tahun tercatat sebagai kelompok dengan prevalensi depresi tertinggi di Indonesia, yakni sebesar 2 persen. Angka ini lebih tinggi dibandingkan kelompok usia lainnya [6]. Menurut Riskesdas 2018, terdapat 706.689 kasus depresi pada usia di atas 15 tahun, termasuk 157.659 kasus di usia 15–24 tahun. Namun, hanya 9% yang mendapat penanganan medis [7]. Selaras dengan data Riskesdas, Ikatan Dokter Anak Indonesia (IDAI) Yogyakarta mengungkapkan bahwa dalam satu dekade terakhir, depresi berat dialami oleh sekitar 3% anak usia sekolah dan 6% remaja di Indonesia [8]. Berdasarkan data Riskesdas, prevalensi depresi di Indonesia mencapai 6,1 per mil, namun hanya sekitar

9% penderitanya yang mendapatkan pengobatan medis [3].

Berbicara tentang depresi, tidak jauh dari social media. Dimana di dalam social media pengguna dapat mengekspresikan diri/emosi dan bisa mencari berbagai informasi yang berkaitan dengan penyakit mental depresi ini salah satunya twitter.

Media sosial berfungsi sebagai wadah bagi individu untuk menyampaikan opini dan pandangan terhadap berbagai permasalahan. Jumlah pengguna yang tinggi di Indonesia menghasilkan akumulasi data tekstual dalam jumlah besar [9]. Twitter merupakan platform mikroblogging yang cukup populer di kalangan pengguna internet Indonesia. Berdasarkan survei APJII, Twitter menempati posisi keempat sebagai media sosial yang paling sering diakses [9]. API adalah mekanisme yang memungkinkan program komputer saling berkomunikasi untuk bertukar data. Hal ini dilakukan melalui pemanggilan endpoint, yaitu alamat unik yang menyediakan jenis informasi tertentu [10]. API Twitter memberikan akses luas terhadap data publik yang dibagikan pengguna secara terbuka. Selain itu, Twitter juga menyediakan API untuk mengelola data non-publik dengan izin dari pengguna kepada pengembang tertentu [10].

Analisis sentimen adalah proses mengevaluasi pendapat seseorang terhadap suatu topik, produk, layanan, atau peristiwa tertentu. Proses ini biasanya menggunakan pendekatan pemrosesan bahasa alami (NLP) dan algoritma pembelajaran mesin [11]. Secara umum, analisis sentimen adalah proses mengidentifikasi sikap atau opini terhadap suatu objek atau individu. Dari jutaan cuitan di Twitter, dapat dianalisis seberapa besar proporsi sentimen positif dan negatif terhadap topik tertentu [9].

Naive Bayes adalah salah satu algoritma yang sering digunakan dalam proses text mining. Algoritma ini bekerja dengan memprediksi kemungkinan suatu kejadian di masa depan berdasarkan data atau pengalaman yang telah terjadi sebelumnya. Oleh karena itu, metode ini dikenal dengan sebutan Teorema Bayes [10].

Algoritma K-Nearest Neighbor (KNN) merupakan salah satu metode pembelajaran mesin yang sederhana dan banyak digunakan untuk menyelesaikan permasalahan klasifikasi maupun regresi. Dalam hal klasifikasi, KNN dikenal sebagai salah satu algoritma yang paling mudah dipahami dan diimplementasikan dalam pengolahan data [1]

Web scraping merupakan metode otomatisasi yang digunakan untuk mengambil data dari internet. Dalam pengumpulan data media sosial, teknik ini dimanfaatkan untuk memperoleh informasi yang relevan dari berbagai platform seperti Facebook, Twitter, Instagram, dan lainnya [12].

Data mining adalah proses menemukan pola tersembunyi dari kumpulan data berukuran besar. Proses ini melibatkan kombinasi metode dari kecerdasan buatan, machine learning, statistika, dan sistem basis data [9].

Orange Data Mining merupakan perangkat lunak open source yang berfungsi untuk melakukan pengolahan data, analisis data, maupun data mining. Di dalam Orange, tersedia berbagai widget yang dapat diletakkan pada drawing board dan saling dihubungkan satu sama lain sesuai kebutuhan. Keunggulan utama dari software ini adalah kemampuannya dalam visualisasi data melalui pemrograman visual. Selain itu, Orange juga mendukung proses klasifikasi data berbasis gambar, sehingga sangat membantu dalam penelitian ini untuk menganalisis dan mengklasifikasikan gambar [13]

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis sentimen publik terhadap isu depresi di Indonesia yang diekspresikan melalui media sosial, khususnya Twitter. Seiring meningkatnya prevalensi depresi di kalangan usia muda dan masih rendahnya tingkat penanganan medis yang diterima, penelitian ini mengangkat permasalahan mengenai bagaimana media sosial, yang sering menjadi ruang pelampiasan emosi, dapat dimanfaatkan sebagai alat bantu untuk mendeteksi dini kondisi kesehatan mental masyarakat. Salah satu tantangan utama dalam pemantauan kondisi psikologis secara real-time adalah terbatasnya sistem yang mampu mengolah dan menginterpretasi data dalam jumlah besar secara efektif dan efisien. Oleh karena itu, penelitian ini memanfaatkan pendekatan machine learning dengan menggunakan dua algoritma klasifikasi teks, yaitu Naïve Bayes dan k-Nearest Neighbours (kNN), untuk melakukan analisis sentimen terhadap data tweet berbahasa Indonesia yang mengandung kata kunci “depresi”. Data dikumpulkan melalui metode web scraping menggunakan bahasa pemrograman Python, kemudian diproses melalui tahapan pembersihan data, ekstraksi fitur dengan metode Bag of Words, dan klasifikasi sentimen menjadi tiga kategori: positif, negatif, dan netral. Kinerja kedua model dievaluasi menggunakan metrik AUC, akurasi, presisi, recall, F1-score, dan Matthews Correlation Coefficient (MCC). Dengan membandingkan performa kedua algoritma, penelitian ini juga bertujuan untuk menentukan metode klasifikasi yang paling optimal dalam mendeteksi sentimen terkait isu depresi, sehingga dapat membuka peluang pengembangan sistem pemantauan berbasis kecerdasan buatan untuk intervensi dini di bidang kesehatan mental.

2. MATERIALS AND METHODS

Untuk mengidentifikasi sentimen terkait depresi di Indonesia, penelitian ini akan menggunakan data yang diperoleh melalui Web Scraping dengan Python, dimulai pada 1 Januari 2023. Selanjutnya, analisis

sentimen akan dilakukan melalui klasifikasi teks pada Orange Data Mining menggunakan dan membandingkan kinerja algoritma Naïve Bayes dan k-Nearest Neighbours (kNN). Hasil dari webscraping Twitter didapatkan 1.342 tweet berbahasa Indonesia yang kemudian dilakukan cleansing data dan menghasilkan 873 data tweet.

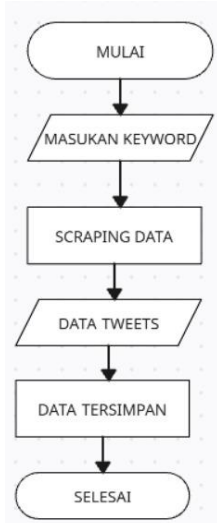


Figure 1. Alur Scraping Data

2.1. Dataset

Dataset yang digunakan didalam penelitian ini berasal dari hasil scraping data pada twitter yang menggunakan Bahasa pemrograman Python. Dengan cara memasukan auth_token pada akun twitter dan memasukan keywords yang berupa “Depresi”. Proses penyaringan data ini dibatasi dari mulai tanggal 9 januari 2023 sampai sekarang. Dataset ini berisi beberapa atribut yang dapat membantu pemrosesan data seperti *full_text* yang berupa isi dari tweet yang ditulis oleh akun tersebut dan sentiment yang bernilai positif, negative, dan juga netral. Atribut tersebut dapat dilihat dari tabel berikut ini.

Table 1. Daftar Fitur Pada Dataset

<i>Nama</i>	<i>Type</i>	<i>Role</i>	<i>Deskripsi</i>
<i>Sentimen</i>	Categorical	Target	Jenis sentimen dari tweet apakah negative, positif, atau netral
<i>full_text</i>	text	meta	Isi tweet tersebut
<i>lang</i>	text	meta	Bahasa yang digunakan
<i>username</i>	text	meta	Nama dari akun
<i>user_id_str</i>	text	meta	Id dari user twitter

2.2. Alur Analisis

Penelitian ini menggunakan aplikasi Orange Data Mining. Proses ini mengandalkan visualisasi data dan bantuan dari berbagai widget yang digunakan. Alur kerja widget dari analisis ini sebagai berikut:

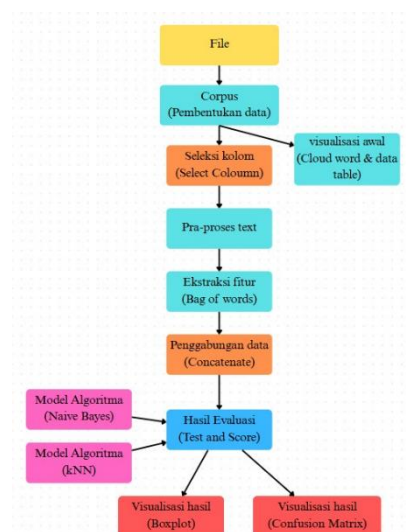


Figure 2. Alur Penelitian

2.3. Pra-Processing Data

Tahapan awal dari penelitian ini menggunakan Corpus yang berguna untuk mengelola dan mempresentasikan data teks yang akan dianalisis. Lalu setelahnya menggunakan Select Coloumn dan Preprocess Text untuk membersihkan dan menormalisasi data sebelum diubah menjadi numerik yang diproses melalui widget Bag Of Words lalu dimasukan ke dalam Test and Score dengan menggunakan 2 algoritma yaitu kNN dan Naïve bayes untuk membuat dua perbandingan.

3. RESULTS AND DISCUSSION

3.1. Hasil

Penelitian ini memanfaatkan 873 data tweet berbahasa Indonesia mengenai isu depresi, yang dikumpulkan melalui proses web scraping Twitter dan telah melalui tahap pembersihan data. Sentimen diklasifikasikan ke dalam tiga kelas, yakni negatif, netral, dan positif, menggunakan dua algoritma pembelajaran mesin: *Naïve Bayes* dan *k-Nearest Neighbours* (kNN). Evaluasi kinerja model dilakukan menggunakan metrik *AUC*, *Classification Accuracy* (CA), *F1-score*, *Precision*, *Recall*, dan *Matthews Correlation Coefficient* (MCC).

Tabel 2. Hasil Test And Score

MODEL	AUC	CA	F1	PREC	RECALL	MCC
NAIVE BAYES	0.985	0.955	0.953	0.958	0.955	0.487
KNN	0.936	0.964	0.963	0.966	0.964	0.645

Dari tabel yang disajikan, terlihat bahwa algoritma kNN menunjukkan performa superior pada sebagian besar metrik evaluasi. Secara spesifik, kNN mencapai akurasi tertinggi sebesar 96,4%, diikuti oleh F1-score dan recall yang juga kuat masing-masing 96,3% dan 96,4%. Selain itu, nilai Matthews Correlation Coefficient (MCC) kNN yang tinggi, yaitu 0.645, lebih lanjut mengindikasikan kinerja yang sangat seimbang dalam klasifikasi multi-kelas. Sementara itu, Naïve Bayes juga memberikan performa yang sangat baik, terutama unggul dalam metrik Area Under the Curve (AUC) dengan nilai tertinggi 0.985, serta presisi yang tinggi sebesar 95,8%. Namun, nilai MCC untuk Naïve Bayes sedikit lebih rendah, yakni 0.487, dibandingkan dengan kNN.

Selain itu, visualisasi data dalam bentuk boxplot dan barplot memperlihatkan bahwa baik Naïve Bayes maupun kNN sama-sama didominasi oleh klasifikasi sentimen negatif. Distribusi sentimen negatif jauh lebih banyak dibandingkan sentimen netral dan positif. Hal ini sesuai contoh tweet seperti "sumpah gw kaya yang udah di titik depresi berat mau mati aja" yang dikategorikan sebagai negatif, sedangkan contoh tweet seperti "Aku bangga pada diriku sendiri, pelan-pelan

aku bisa melewati semuanya" dikategorikan sebagai positif. Seperti tabel berikut ini:

Hasil dari masing masing algoritma yang digunakan divisualisasikan menggunakan *Box-Plot* untuk menampilkan hasil dari sentiment data. Selain itu widget *Confusion Matrix* untuk mengevaluasi kinerja model klasifikasi dan matriks ini membandingkan nilai negatif dengan nilai prediksi.

		Predicted			Σ
		Negatif	Netral	Positif	
Actual	Negatif	787	0	43	830
	Netral	31	1	1	33
	Positif	0	0	10	10
Σ		818	1	54	873

Figure 3. Confusion Matrix Naïve Bayes

Model Naïve Bayes menunjukkan performa yang cukup baik untuk kelas egative, dengan 817 data dari 830 diklasifikasikan dengan benar, untuk kelas netral hanya 12 dari 33 data berhasil diprediksi dengan benar, sisanya banyak bertukar ke kelas negatif (16) dan positif (5), kelas positif menunjukan hasil yang seimbang 5 diprediksi dengan benar dan 5 salah ke kelas negatif.

		Predicted			Σ
		Negatif	Netral	Positif	
Actual	Negatif	816	14	0	830
	Netral	7	25	1	33
	Positif	2	7	1	10
Σ		825	46	2	873

Figure 4. Confusion Matrix kNN

Model kNN menunjukkan hasil sangat baik untuk kelas netral, dengan 25 dari 33 data diklasifikasikan dengan benar, pada kelas negatif 816 dari 830 data diprediksi benar, meskipun ada pergeseran sebesar 14 data ke kelas netral, untuk kelas positif model mengalami kesulitan hanya 1 yang benar di klasifikasikan.

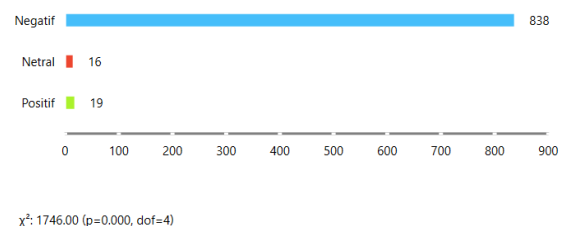


Figure 5. Boxplot Naïve Bayes

Meskipun prediksi untuk kelas negatif masih sangat dominan, model ini menunjukkan kemampuan yang lebih baik dalam mendeteksi kelas netral dan positif dibanding kNN. Prediksi yang menyebar ini mengindikasikan bahwa Naïve Bayes lebih sensitif terhadap variasi sentiment dalam data, walaupun kelas minoritas masih belum terdeteksi secara optimal.

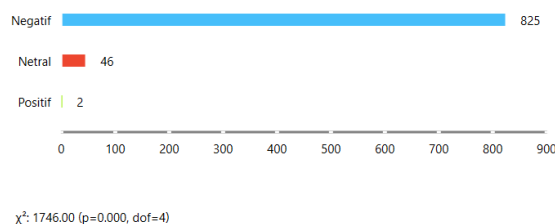


Figure 6. Boxplot kNN

Hasil ini memperlihatkan bahwa model kNN memiliki ketidakseimbangan prediksi yang sangat tinggi dengan dominasi kuat terhadap kelas negatif. Meskipun terdapat sedikit prediksi untuk kelas netral, model hampir tidak mengenali kelas positif. Hal ini mencerminkan bias model terhadap kelas mayoritas.

3.2. Pembahasan

Temuan ini menunjukkan bahwa algoritma pembelajaran mesin efektif dalam mendeteksi sentimen di media sosial dan berpotensi digunakan sebagai indikator awal dalam memantau gejala depresi di masyarakat. Mayoritas tweet menunjukkan muatan sentimen negatif, sejalan dengan data epidemiologis bahwa tingkat depresi di Indonesia, khususnya di kalangan remaja dan dewasa muda, cukup tinggi.

Selain angka-angka metrik tersebut, pembahasan ini juga perlu meninjau faktor lain seperti interpretabilitas, waktu komputasi, dan potensi penerapan model di dunia nyata. Naïve Bayes memiliki keuntungan komputasi lebih cepat dan lebih mudah diimplementasikan dalam skenario real-time monitoring. Sementara kNN, meski lebih akurat dalam mengenali kasus depresi secara keseluruhan, memerlukan sumber daya komputasi yang lebih besar, terutama saat jumlah data meningkat, karena harus menghitung jarak ke setiap contoh latih.

Visualisasi data dalam bentuk boxplot dan barplot memperlihatkan bahwa baik Naïve Bayes maupun kNN sama-sama didominasi oleh klasifikasi sentimen negatif. Distribusi sentimen negatif jauh lebih banyak dibandingkan sentimen netral dan positif. Ini memperkuat hipotesis bahwa Twitter sering digunakan sebagai media untuk melampiaskan perasaan negatif dan tekanan emosional, terutama oleh remaja dan dewasa muda. Contoh tweet seperti "sumpah gw kaya yang udah di titik depresi berat mau mati aja" memperlihatkan betapa gamblangnya ekspresi depresi di platform ini, sedangkan tweet

seperti "Aku bangga pada diriku sendiri, pelan-pelan aku bisa melewati semuanya" menunjukkan adanya sentimen positif sebagai bentuk afirmasi diri dan semangat untuk pulih.

Hasil penelitian ini mendukung data epidemiologis seperti Riset Kesehatan Dasar (Riskesdas) dan Survei Kesehatan Indonesia (SKI), di mana prevalensi gejala depresi tertinggi terdapat pada kelompok usia 15-24 tahun. Dengan kata lain, media sosial bisa menjadi cermin untuk menggali fenomena kesehatan mental yang terjadi di masyarakat secara lebih dinamis dan cepat. Analisis ini juga membuka peluang intervensi dini berbasis data real-time untuk menjangkau mereka yang membutuhkan bantuan, bahkan sebelum mereka mengakses layanan kesehatan formal.

Dari segi kinerja, kNN lebih sesuai bila prioritasnya adalah mengenali sebanyak mungkin kasus depresi (karena recall lebih tinggi), sedangkan Naïve Bayes lebih andal dalam menjaga ketepatan prediksi dan menangani ketidakseimbangan data. Dengan memanfaatkan platform Twitter dan pendekatan pembelajaran mesin ini, terdapat potensi untuk mengembangkan alat pemantauan kesehatan mental berbasis AI. Alat ini bisa digunakan pemerintah, lembaga kesehatan, dan LSM untuk merancang program pencegahan dan intervensi berbasis data.

Secara keseluruhan, studi ini berkontribusi dalam memperkuat literatur mengenai penggunaan data publik dari media sosial untuk deteksi masalah kesehatan mental. Penelitian lebih lanjut bisa memperkaya dataset agar lebih representatif, menambahkan fitur linguistik dan semantik (misalnya penggunaan analisis emosi dan topik), serta menggunakan algoritma deep learning seperti LSTM dan BERT untuk meningkatkan akurasi dan ketahanan model terhadap variasi bahasa. Selain itu, evaluasi longitudinal dan implementasi dalam bentuk dashboard monitoring secara real-time bisa menjadi pengembangan di masa mendatang agar dampaknya lebih nyata dan praktis.

3.3. Evaluasi

Evaluasi terhadap hasil analisis sentimen ini dilakukan dengan mempertimbangkan kinerja model, distribusi data, serta relevansi temuan terhadap kondisi sosial yang sebenarnya

3.3.1. Evaluasi Kinerja Model

Algoritma K-Nearest Neighbours (kNN) menunjukkan performa yang sangat baik dan unggul secara keseluruhan dalam analisis ini, menghasilkan Classification Accuracy (CA) sebesar 96,4%, F1-Score 96,3%, Precision 96,6%, Recall 96,4%, dan Matthews Correlation Coefficient (MCC) sebesar

0,645. Nilai MCC yang tinggi ini mengindikasikan bahwa kNN tidak hanya akurat, tetapi juga sangat seimbang dalam klasifikasi antar kelas sentimen, menjadikannya sangat cocok untuk kasus klasifikasi dengan distribusi data yang cenderung tidak merata. Sementara itu, Naïve Bayes, meskipun memiliki akurasi sedikit lebih rendah sebesar 95,5%, mencatatkan nilai Area Under the Curve (AUC) tertinggi, yaitu 98,5%. Performa lainnya meliputi Precision 95,8%, F1-Score 95,3%, Recall 95,5%, dan MCC 0,487. AUC yang tinggi pada Naïve Bayes menunjukkan kemampuannya yang sangat baik dalam memisahkan ketiga kelas sentimen, dan keunggulannya dalam efisiensi komputasi menjadikannya pilihan ideal untuk aplikasi real-time yang memerlukan kecepatan tinggi.

3.3.2. Evaluasi Distribusi Sentimen

Distribusi data yang sangat condong pada sentimen negatif menunjukkan bahwa pengguna Twitter lebih banyak mengungkapkan perasaan negatif atau tekanan emosional, terutama terkait isu depresi. Hal ini selaras dengan kenyataan bahwa media sosial sering menjadi tempat pelampiasan emosional, terutama bagi kelompok usia muda.

Ketimpangan distribusi ini juga menimbulkan tantangan dalam proses klasifikasi, terutama untuk sentimen netral dan positif yang lebih jarang muncul. Model yang tidak cukup tangguh akan cenderung bias terhadap kelas mayoritas (negatif).

4. CONCLUSION

Penelitian ini berhasil mengimplementasikan algoritma Naïve Bayes dan kNN untuk menganalisis sentimen terkait isu depresi di media sosial Twitter. Dengan dataset berjumlah 873 tweet, dilakukan klasifikasi sentimen menjadi tiga kategori: negatif, netral, dan positif.

K-Nearest Neighbours (kNN) menunjukkan performa terbaik dari hampir seluruh metrik evaluasi, dengan akurasi mencapai 96,4%, F1-score 96,3%, serta nilai Matthews Correlation Coefficient (MCC) tertinggi sebesar 0.645. Sementara itu, Naïve Bayes unggul dalam metrik Area Under the Curve (AUC) dengan catatan tertinggi 0.985, dan secara keseluruhan menunjukkan performa yang stabil, menjadikannya alternatif yang efisien untuk implementasi cepat.

Distribusi sentimen yang cenderung negatif mencerminkan realitas sosial bahwa Twitter kerap menjadi ruang ekspresi tekanan emosional, terutama di kalangan usia muda. Media sosial terbukti memiliki potensi sebagai alat bantu skrining dini isu kesehatan mental di masyarakat Indonesia.

Visualisasi data menunjukkan dominasi sentimen negatif, memperkuat temuan bahwa Twitter kerap digunakan sebagai medium untuk mengekspresikan tekanan psikologis. Hal ini selaras dengan data

epidemiologis yang menunjukkan tingginya angka gangguan depresi pada usia muda (15–24 tahun). Dengan demikian, media sosial berpotensi menjadi alat skrining awal yang penting dalam mendeteksi permasalahan kesehatan mental secara cepat dan dinamis.

Untuk studi selanjutnya, disarankan untuk memperluas cakupan dataset dari berbagai platform media sosial dan mengeksplorasi algoritma klasifikasi serta pendekatan deep learning yang lebih canggih. Validasi hasil dengan pakar kesehatan mental dan pengembangan prototipe sistem skrining dini berbasis media sosial juga menjadi langkah penting ke depan.

REFERENCES

- [1] K. Rahayu, V. Fitria, D. Septhya, R. Rahmadden, and L. Efrizoni, "Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 108–114, 2023.
- [2] I. Pengaruh et al., "ANSIETAS," no. January, 2022.
- [3] D. Daryanto and M. T. Sari, "Pemberdayaan Kader Posyandu Lansia dalam Deteksi Dini Risiko dan Pencegahan Depresi pada Lanjut Usia di Desa Penyengat Olak Muaro Jambi," *J. Abdimas Kesehat.*, vol. 3, no. 1, p. 81, 2021.
- [4] I. R. Ramadani et al., "Depresi, Penyebab Dan Gejala Depresi Tryana Fauziyah," *Bersatu J. Pendidik. Bhinneka Tunggal Ika*, vol. 2, no. 2, pp. 89–99, 2024.
- [5] A. T. Kaeng and R. Siby, "Mewaspadai Dampak Depresi pada Generasi Z," *Sumikolah J. Ilmu Pendidik.*, vol. 1, no. 2, pp. 50–58, 2023.
- [6] V. C. Agustiyani, "Analisis Wacana Kritis terhadap Representasi Depresi dan Pemikiran Bunuh Diri di Kalangan Gen Z dalam Artikel CNN Indonesia " Depresi hingga Suicidal Thought ' Hantui ' Gen Z Indonesia ", vol. 19, 2024.
- [7] R. V. Pangerapan, H. Munayang, and B. H. R. Kairupan, "Hubungan antara Aktivitas Fisik dan Depresi pada Remaja Sekolah Menengah Pertama," *Med. Scope J.*, vol. 5, no. 1, pp. 45–49, 2023.
- [8] F. S. Putri, Z. Nazihah, D. P. Ariningrum, S. Celesta, and C. Kharin Herbawani, "Depresi Remaja di Indonesia: Penyebab dan Dampaknya Adolescent Depression in Indonesia: Causes and Effects," *Jurnal kesehatan poltek keskesmenkes ripang kalpinang*, vol. 10(2), no. 2, pp. 99–108, 2022.
- [9] R. N. Handayani, "Optimasi Algoritma Support Vector Machine untuk Analisis Sentimen pada Ulasan Produk Tokopedia Menggunakan PSO," *Media Inform.*, vol. 20, no. 2, pp. 97–108, 2021.
- [10] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022.
- [11] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector

Machine Pada Kasus Kenaikan Harga BBM,” *Jambura J. Math.*, vol. 5, no. 2, pp. 431–445, 2023.

[12]L. Hidayati, L. P. Kusuma, D. Agustini, and V. Y. P. Ardhana, “Implementasi Web Scraping Untuk Pengumpulan Data Media Sosial Lingkup Pemerintah Provinsi Ntb,” *J. Sist. Inf. dan Inform.*, vol. 7, no. 1, pp. 63–72, 2024.

[13]E. Pranadjaya, E. S. Pangestu, C. O. Sereati, S. Octaviani, and M. Darmawan, “Perbandingan Algoritma Machine Learning menggunakan Orange Data Mining untuk Klasifikasi Jenis Kendaraan pada Sistem Tilang Digital,” *J. Elektro*, vol. 17, no. 1, pp. 41–47, 2024.