Jurnal Informatika dan Sistem Informasi (JIFoSI)
Vol. 6, No. 1. April 2025
e-ISSN : 2722-130
DOI: https://doi.org/10.33005/jifosi.v6i1.466

ANALISIS SENTIMEN KESEHATAN MENTAL DI TWITTER MENGGUNAKAN *NAIVE BAYES CLASSIFIER* DAN *K-NEAREST NEIGHBOR*

Rizqa Annisa Dwiatmoko[1*], Imelda[2]
E-mail : [1)] 2011501877@student.budiluhur.ac.id , [2)] imelda@budiluhur.ac.id
[1,2] Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

(Naskah masuk: 31 Desember 2024, diterima untuk diterbitkan: 30 April 2025)

Abstrak

Isu kesehatan mental semakin mendapat perhatian di media sosial, khususnya di Twitter, dan memengaruhi persepsi publik serta kebijakan kesehatan. Dalam konteks digital saat ini, penting untuk menganalisis persepsi masyarakat yang tercermin melalui unggahan di media sosial. Penelitian ini bertujuan menganalisis sentimen tweet terkait kesehatan mental menggunakan dua algoritma klasifikasi populer dalam pembelajaran mesin, yaitu Naïve Bayes Classifier dan K-Nearest Neighbor (KNN). Proses dilakukan dalam dua tahap: pelabelan otomatis menghasilkan 62,5% tweet positif, 19,64% negatif, dan 17,86% netral; kemudian diverifikasi secara manual yang menghasilkan 67,56% positif, 24,4% netral, dan 8,04% negatif. Dataset terdiri dari 336 tweet yang dikumpulkan selama periode 26–31 Maret 2024. Hasil menunjukkan bahwa algoritma KNN memiliki akurasi lebih tinggi dalam mengklasifikasikan sentimen positif sebesar 88,24%, dibandingkan Naïve Bayes yang mencapai 60,78%. Penelitian ini berkontribusi pada pengembangan metode analisis sentimen yang lebih akurat dan dapat digunakan untuk mendukung perancangan intervensi kesehatan yang lebih tepat serta responsif terhadap persepsi masyarakat. Dengan demikian, hasil penelitian diharapkan dapat membantu penyusunan strategi komunikasi dan kebijakan yang lebih efektif dalam merespons isu kesehatan mental di media sosial.

Kata kunci: *Analisis Sentimen, Naïve Bayes Classifier, K-Nearest Neighbor, Kesehatan Mental, Twitter, Media Sosial.*

1. PENDAHULUAN

Analisis sentimen merupakan teknik yang digunakan untuk menilai dan menginterpretasikan emosi, opini, dan sikap dari teks yang diunggah oleh pengguna di berbagai platform digital. Teknik ini menggunakan algoritma pembelajaran mesin dan *text mining* untuk mengklasifikasikan teks menjadi sentimen positif, negatif, atau netral. Dengan mengolah data dalam jumlah besar, analisis sentimen mampu mengungkap pola pikir dan perasaan masyarakat terhadap berbagai topik. Sebagai contoh, analisis sentimen dapat diterapkan untuk memahami pandangan publik mengenai kesehatan mental yang diekspresikan melalui media sosial seperti Twitter.

Data di Indonesia menunjukkan bahwa sebanyak 6,1% penduduk Indonesia berusia 15 tahun ke atas mengalami gangguan kesehatan mental [1]. Fakta ini menegaskan urgensi isu kesehatan mental, terutama dalam era modern ini, di mana dampaknya terhadap kualitas hidup dan produktivitas individu semakin signifikan. Data dari *World Health Organization* (WHO) tahun 2016 menyebutkan bahwa terdapat sekitar 35 juta orang yang mengalami depresi, 60 juta orang dengan bipolar, 21 juta orang dengan skizofrenia, serta 47,5 juta orang dengan demensia [2]. Dengan adanya data ini, menjadi semakin penting untuk

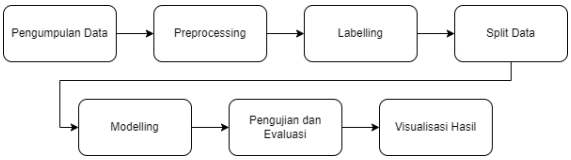
menganalisis sentimen terhadap pembicaraan tentang kesehatan mental di platform-media sosial, seperti Twitter. Peningkatan metode analisis sentimen pada tweet tentang kesehatan mental menjadi hal yang krusial dalam menghadapi tantangan ini. Kesalahan interpretasi atau pemahaman yang kurang tepat terhadap sentimen yang disampaikan dapat memperburuk stigma terhadap kesehatan mental, bahkan berpotensi merugikan individu yang memerlukan dukungan psikologis yang tepat.

Beberapa penelitian sebelumnya telah sukses menerapkan analisis sentimen kesehatan mental di Twitter dengan menggunakan algoritma seperti *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Sebagai contoh, penelitian yang dilakukan oleh Karina Aulia dan Lia Amelia (2020) berhasil mencapai akurasi 89% dengan mengaplikasikan algoritma Naïve Bayes, menyoroti dominasi sentimen positif dalam percakapan seputar kesehatan mental [3]. Sementara itu, studi yang dilakukan oleh Yunis Femilia Nugraini dan timnya (2021) menggunakan metode yang serupa dengan akurasi yang sebanding pada pembagian data 70:30 [4]. Di sisi lain, penelitian oleh Mahesworo Langgeng Wicaksono dan rekan (2022) memperkenalkan algoritma *K-Nearest Neighbor* (KNN) dengan akurasi 58.39% pada data split 70:30, menyoroti keunggulan dan perbedaan pendekatan dalam pengolahan data [5].

Namun, penelitian-penelitian tersebut hanya menggunakan satu algoritma, sehingga kurang memberikan perspektif yang luas tentang bagaimana kombinasi beberapa algoritma dapat memperkaya hasil analisis sentimen. Penelitian ini menerapkan dua algoritma sekaligus, yaitu Naïve Bayes dan K-Nearest Neighbor (KNN), untuk menganalisis sentimen terkait kesehatan mental di Twitter. Selain itu, penelitian ini juga menggunakan dua metode labelling, yaitu otomatis dan manual, dengan hasil akhir yang menggunakan labelling manual oleh ahli pakar. Pendekatan ini memberikan perspektif baru yang belum banyak dieksplorasi dalam penelitian sebelumnya dan diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai opini publik terhadap kesehatan mental di Indonesia.

2. METODOLOGI

Tahapan metode penerapan penelitian dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Tahap Metode

2.1 Pengumpulan Data

Dataset penelitian ini terdiri dari 348 tweet yang dikumpulkan dari Twitter selama periode 26 Maret hingga 31 Maret 2024. Pengumpulan data dilakukan melalui proses *crawling* menggunakan *tweet harvest* dengan menggunakan kata kunci 'kesehatan mental' sebagai parameter utama.

2.2 *Preprocessing*

Selama tahap *preprocessing*, data *tweet* diproses melalui beberapa langkah penting untuk memudahkan analisis lanjutan. Langkah-langkah ini mencakup penyesuaian huruf besar dan kecil, pembersihan data, konversi kata slang, penghapusan *stop word*, tokenisasi, dan *stemming*. Ilustrasi dari proses *preprocessing* data dapat dilihat pada Gambar 2.

Gambar 2. Langkah *Preprocessing*

a. *Case Folding*, dilakukan dengan mengonversi semua huruf menjadi huruf kecil. Misalnya, kata ‘Kesehatan’ atau ‘KESEHATAN’ akan diubah menjadi ‘kesehatan’.
b. *Cleaning*, Teks diproses melalui tahap pembersihan untuk menghilangkan elemen-elemen yang tidak diinginkan. Langkah-langkah pembersihan ini mencakup penghapusan URL, mention, dan hashtag, menghilangkan karakter-karakter khusus, serta merapikan spasi yang berlebihan.
c. *Slang word*, Normalisasi slang dilakukan untuk mengonversi kata-kata tidak formal, seperti istilah gaul atau singkatan, menjadi bentuk bakunya. Sebagai ilustrasi, kata 'gak' diubah menjadi 'tidak', 'yg' menjadi 'yang', dan 'sm' menjadi 'sama'. Proses ini menggunakan kamus slang yang tersimpan dalam basis data.
d. *Stop Word*, Istilah-istilah yang tidak memiliki arti penting namun sering muncul, seperti 'untuk', 'yang', dan 'apa', dihilangkan melalui proses pembersihan stop word. Tahapan ini didasarkan pada kamus *stopword* dari Pustaka Sastrawi, ditambah dengan entri kata relevan dalam basis data.
e. *Tokenizing*, Teks dipecah menjadi unit-unit kata melalui proses tokenisasi. Misalnya, frasa 'kesehatan mental' akan dipisahkan menjadi dua kata yaitu 'kesehatan' dan 'mental' setelah tokenisasi dilakukan.
f. *Stemming*, bertujuan menghapus imbuhan dari kata dan mengembalikannya ke bentuk dasar. Misalnya, kata 'menghadiri' disederhanakan menjadi 'hadir', dan 'belajar' menjadi 'ajar'.

2.3 *Labelling*
Pada tahap *labelling*, dokumen atau kalimat diberi label (kelas) berdasarkan ciri-ciri yang ada di dalamnya. *Tweet* yang telah melalui *preprocessing* dan menghasilkan *clean text* akan diberi label positif, netral, atau negatif. Proses *labelling* dilakukan dengan dua cara: manual dan otomatis. *Labelling* manual melibatkan intervensi manusia atau pakar untuk memberikan label berdasarkan interpretasi sentimen yang kompleks dan kontekstual. Sebaliknya, *labelling* otomatis menggunakan kamus kata dan algoritma untuk memberi label berdasarkan karakteristik teks.

2.4 *Split Data*
Pada tahap *Split Data*, hanya *tweet* yang berlabel Positif dan Negatif yang digunakan. *Tweet* yang telah diberi label akan dibagi menjadi dua bagian: data latih (80%) dan data uji (20%). Data latih digunakan untuk melatih model dalam mengenali pola dan membuat prediksi, sementara data uji digunakan untuk menguji sejauh mana model dapat memprediksi atau mengklasifikasikan data baru yang belum pernah dilihat sebelumnya.

2.5 TF-IDF
TF-IDF (Term Frequency-Inverse Document Frequency) adalah proses mengubah data teks menjadi data numerik untuk pembobotan setiap kata. Metode ini mengevaluasi pentingnya suatu kata dalam dokumen. TF mengukur frekuensi kemunculan kata dalam dokumen untuk menunjukkan seberapa penting kata tersebut, sementara DF mengukur frekuensi dokumen yang mengandung kata tersebut untuk menunjukkan seberapa umum kata tersebut. IDF merupakan kebalikan dari DF. Algoritma TF-IDF menilai bobot setiap kata kunci di berbagai kategori dengan mempertimbangkan tingkat kemiripan antar kategori [6].

Dengan demikian, hasil pembobotan kata menggunakan TF-IDF diperoleh dari perkalian antara TF dan IDF. Berikut Persamaan (1), (2), dan (3) adalah perhitungan dari TF-IDF:

$$TF\ (t)\ =\ \frac{f_{t,d}}{\Sigma_{t\,d}} \tag{1}$$

$$IDF\ (t)\ = log(\frac{|D|}{f_{t,\,d}}) \tag{2}$$

$$TF\ -\ IDF\ (t)\ =\ TFt,d\ *\ IDFd \tag{3}$$

Dimana:
TF-IDF (t) = Bobot TF-IDF
IDFd = *Inverse Document Frequency*
TFt,d = Frekuensi suatu kata

2.6 *Naïve Bayes Classifier*

Naïve Bayes merupakan sebuah metode pengklasifikasian dengan menggunakan probabilitas sederhana yang berakar pada Teorema Bayes dan memiliki asumsi ketidaktergantungan (*independent*) yang tinggi dari masing – masing kondisi atau kejadian [7], Bentuk umum teorema bayes adalah persamaan (4) sebagai berikut:

$$P\ (\ H\ |\ X\ )\ =\frac{P\ (\ X\ |\ H\ )\ .P\ (\ H\ )}{P\ (\ X\ )} \tag{4}$$

X = Data dengan kelas yang belum diketahui
H = Hipotesa data X merupakan suatu kelas spesifik
P(H | X) = Probabilitas hipotesis H berdasarkan kondisi X (*posterior probability*)
P(H) = Probabilitas hipotesis H (*prior probability*)

2.7 *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) adalah metode populer dalam *machine learning* untuk klasifikasi. Algoritma ini bertujuan untuk mengategorikan objek ke dalam salah satu kelas yang telah ditentukan berdasarkan data sampel yang ada. KNN bekerja dengan mengelompokkan data ke dalam kelas yang telah ditetapkan berdasarkan jarak atau kesamaan dengan data pelatihan yang tersedia [8]. Tahapan dalam proses K-NN yaitu:

a. Menentukan nilai K,
b. Menghitung jarak antara data yang akan diklasifikasikan dengan data berlabel,
c. Menentukan nilai K terkecil,

Mengklasifikasikan data berdasarkan metrik data.

2.8 *Euclidean Distance*

Euclidean Distance merupakan metode perhitungan jarak antara dua titik dalam ruang *Euclidean* yang diperkenalkan oleh matematikawan Yunani, Euclid [9]. Dalam ruang dua dimensi, perhitungan jarak menggunakan *Euclidean Distance* dapat dituliskan dengan rumus yang ditunjukkan pada persamaan (5).

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x^i - y^i)^2} \tag{5}$$

Dimana :
d(x,y) = Jarak
xi = Koordinat ke-i data latih y_i = Koordinat ke-i data uji
n = Jumlah Koordinat

2.9 Pengujian dan Evaluasi

Hasil kedua algoritma akan dievaluasi dengan *Confusion Matrix*. *Confusion Matrix* adalah sebuah metode yang dapat mengevaluasi seberapa akurat klasifikasi yang telah dilakukan dengan memberikan gambaran tentang kinerja model dalam mengklasifikasikan data [10].

Tabel 1. Rumus *Confusion Matrix*

Kelas Prediksi	Data Aktual	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan :

True Positive (TP) = Jumlah data yang sebenarnya Positif dan diprediksi sebagai Positif,
True Negative (TN) = Jumlah data yang sebenarnya Negatif dan diprediksi sebagai Negatif,
False Positive (FP) = Jumlah data yang sebenarnya Negatif tetapi diprediksi sebagai Positif,
False Negative (FN) = Jumlah data yang sebenarnya Positif tetapi diprediksi sebagai Negatif.

Akan dilakukan perhitungan terhadap tiga parameter, yaitu akurasi persamaan (6), presisi persamaan (7), dan *recall* persamaan (8), menggunakan rumus *Confusion Matrix* untuk mengukur ketiga metrik tersebut.

$$akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

$$presisi = \frac{TP}{TP + FP} \quad (7)$$

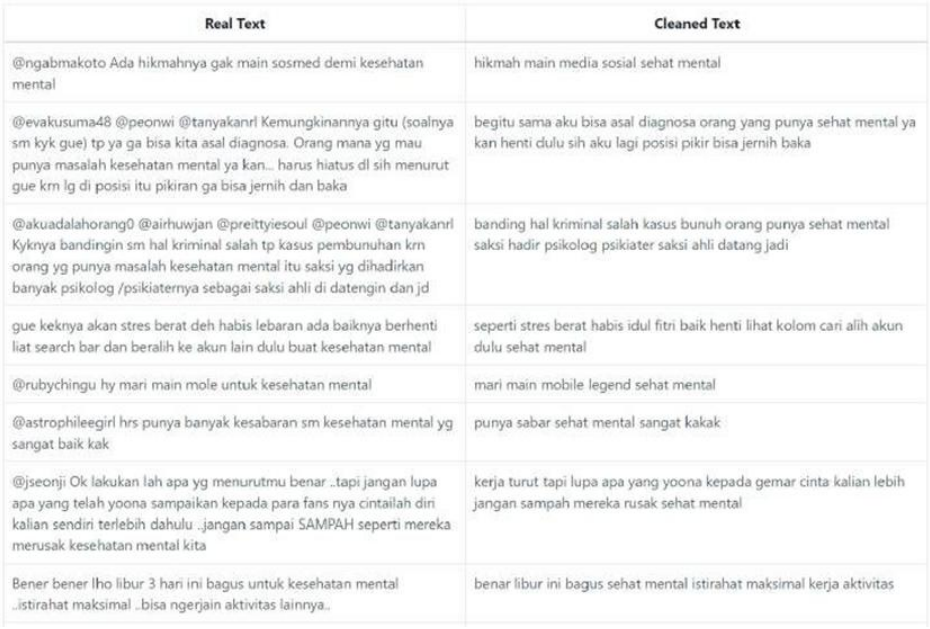
$$recall = \frac{TP}{TP + FN} \quad (8)$$

3. HASIL DAN PEMBAHASAN

Penelitian ini mengevaluasi akurasi kedua algoritma dalam mengidentifikasi kategori pada data uji, mengikuti tahapan yang digambarkan dalam Gambar 1. Proses dimulai dengan pengumpulan data dari Twitter menggunakan pustaka Tweet Harvest antara 26 Maret hingga 31 Maret 2024 dengan kata kunci terkait kesehatan mental. Data yang diperoleh dari crawling disimpan dalam format Excel (.CSV) dan diimpor ke sistem basis data untuk analisis lebih lanjut. Contoh data sampel mengenai kesehatan mental dapat dilihat pada Tabel 2 di bawah ini.
Data terbaru dari Twitter sering kali mengandung simbol, URL, dan elemen tidak relevan. Oleh karena itu, proses pembersihan data (preprocessing) diperlukan untuk menghapus elemen-elemen tersebut dan memastikan konsistensi data sebelum analisis. Hasil pembersihan ditunjukkan pada Gambar 3, yang menampilkan data yang telah dibersihkan dan siap untuk analisis lebih lanjut.

Tabel 2. Dataset

Tweet	*Username*	*Created_at*
ada hikmahnya gak main sosmed demi kesehatan mental	honey17desu	Sun Mar 31 08:39:38 +0000 2024
cut and go kesehatan mental lebih penting dari sekedar teman	nuna_boys	Sat Mar 30 23:14:46 +0000 2024
maybe ttp jaga kesehatan mental jangan lupa take profit	eltororado	Fri Mar 29 16:01:00 +0000 2024



Real Text	Cleaned Text
@ngabmakoto Ada hikmahnya gak main sosmed demi kesehatan mental	hikmah main media sosial sehat mental
@evakusuma48 @peonwi @tanyakanrl Kemungkinannya gitu (soalnya sm kyk gue) tp ya ga bisa kita asal diagnosa. Orang mana yg mau punya masalah kesehatan mental ya kan... harus hiatus dl sih menurut gue krn lg di posisi itu pikiran ga bisa jernih dan baka	begitu sama aku bisa asal diagnosa orang yang punya sehat mental ya kan henti dulu sih aku lagi posisi pikir bisa jernih baka
@akuadalahorang0 @airhuwjan @preittyiesoul @peonwi @tanyakanrl Kyknya bandingin sm hal kriminal salah tp kasus pembunuhan krn orang yg punya masalah kesehatan mental itu saksi yg dihadirkan banyak psikolog /psikiaternya sebagai saksi ahli di datengin dan jd	banding hal kriminal salah kasus bunuh orang punya sehat mental saksi hadir psikolog psikiater saksi ahli datang jadi
gue keknya akan stres berat deh habis lebaran ada baiknya berhenti liat search bar dan beralih ke akun lain dulu buat kesehatan mental	seperti stres berat habis idul fitri baik henti lihat kolom cari alih akun dulu sehat mental
@rubychingu hy mari main mole untuk kesehatan mental	mari main mobile legend sehat mental
@astrophileegirl hrs punya banyak kesabaran sm kesehatan mental yg sangat baik kak	punya sabar sehat mental sangat kakak
@jseonji Ok lakukan lah apa yg menurutmu benar ..tapi jangan lupa apa yang telah yoona sampaikan kepada para fans nya cintailah diri kalian sendiri terlebih dahulu ..jangan sampai SAMPAH seperti mereka merusak kesehatan mental kita	kerja turut tapi lupa apa yang yoona kepada gemar cinta kalian lebih jangan sampah mereka rusak sehat mental
Bener bener lho libur 3 hari ini bagus untuk kesehatan mental ..istirahat maksimal ..bisa ngerjain aktivitas lainnya..	benar libur ini bagus sehat mental istirahat maksimal kerja aktivitas

Gambar 3. *Preprocessing*

Setelah *preprocessing*, data bersih siap untuk tahap *labelling*. Ada dua pendekatan untuk menentukan kelas sentimen dari teks yang telah dibersihkan: pertama, menggunakan skor dari kamus sentimen, dan kedua, melibatkan pakar atau ahli. Hasil *labelling* ditunjukkan pada Gambar 4, yang memperlihatkan data yang telah diberi label sesuai dengan pendekatan yang diterapkan.

Setelah tahap *labelling* selesai, proses selanjutnya adalah *split data* seperti contoh Tabel 3. Hanya *tweet* berlabel Positif dan Negatif yang digunakan. Dari total 336 *tweet* yang dilabeli, setelah mengeluarkan *tweet* berlabel Netral, tersisa 254 *tweet*. Data ini dibagi dengan rasio 80:20, menghasilkan 203 *tweet* sebagai data latih dan 51 *tweet* sebagai data uji.

Tweet	Automation	Manual Sentiment
hikmah main media sosial sehat mental	Positif	Positif
begitu sama aku bisa asal diagnosa orang yang punya sehat mental ya kan henti dulu sih aku lagi posisi pikir bisa jernih baka	Positif	Netral
banding hal kriminal salah kasus bunuh orang punya sehat mental saksi hadir psikolog psikiater saksi ahli datang jadi	Positif	Netral
seperti stres berat habis idul fitri baik henti lihat kolom cari alih akun dulu sehat mental	Positif	Negatif
mari main mobile legend sehat mental	Positif	Positif
punya sabar sehat mental sangat kakak	Positif	Positif
kerja turut tapi lupa apa yang yoona kepada gemar cinta kalian lebih jangan sampah mereka rusak sehat mental	Negatif	Positif
benar libur ini bagus sehat mental istirahat maksimal kerja aktivitas	Positif	Positif
buku bangun sistem ramah sehat mental sebentar unggah kapan	Positif	Positif

Gambar 4. *Labelling*

Tabel 3. Data *Training* dan Data *Testing*

Data *Training*	Data *Testing*
yakan sindir pedas baik sehat mental korbanin sehat mental fisik kakak sehat mental pikir karena lihat karya orang turut sukses	seru ada survei sehat mental pegawai layak periksa sehat mental karena mudah bohong maksud konseling sehat mental mau guru bimbing konseling dua pikir sehat mental hampir mati

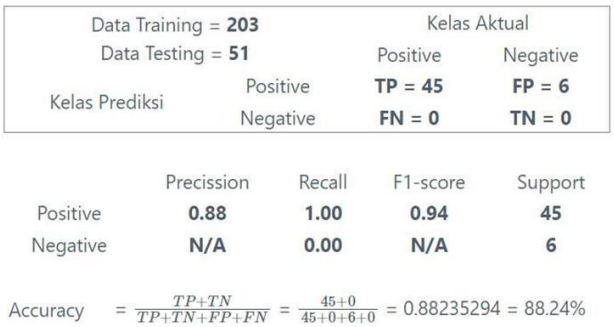
Untuk menentukan kelas sentimen, algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor* akan digunakan untuk memproses data uji. Sementara itu, data latih akan dianalisis dengan metode TF-IDF untuk menghitung bobot setiap kata. Hasil dari kedua algoritma ini selanjutnya akan dievaluasi menggunakan *Confusion Matrix* untuk mengukur akurasi. Gambar 5 dan 6 menunjukkan hasil evaluasi dalam bentuk *Confusion Matrix*, yang memperlihatkan jumlah prediksi yang benar dan salah untuk masing-masing algoritma, serta memberikan gambaran menyeluruh tentang performa dan tingkat akurasi mereka.

Data Training = **203**		Kelas Aktual	
Data Testing = **51**		Positive	Negative
Kelas Prediksi	Positive	**TP = 28**	**FP = 3**
	Negative	**FN = 17**	**TN = 3**

	Precission	Recall	F1-score	Support
Positive	**0.90**	**0.62**	**0.74**	**45**
Negative	**0.15**	**0.50**	**0.23**	**6**

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{28+3}{28+3+3+17} = 0.60784314 = 60.78\%$$

Gambar 5. *Confusion Matrix Naive Bayes*



Data Training = 203		Kelas Aktual	
Data Testing = 51		Positive	Negative
Kelas Prediksi	Positive	TP = 45	FP = 6
	Negative	FN = 0	TN = 0

	Precission	Recall	F1-score	Support
Positive	0.88	1.00	0.94	45
Negative	N/A	0.00	N/A	6

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{45+0}{45+0+6+0} = 0.88235294 = 88.24\%$$

Gambar 6. ***Confusion Matrix K-Nearest Neighbor***

Berdasarkan *confusion matrix* yang ditampilkan, terdapat perbedaan signifikan dalam akurasi antara model *Naïve Bayes Classifier* dan *K-Nearest Neighbor*. Model *K-Nearest Neighbor* mencapai akurasi 88,24%, sedangkan *Naïve Bayes Classifier* hanya 60,78%. Hasil pengujian menunjukkan bahwa *K-Nearest Neighbor* memiliki performa yang superior dalam klasifikasi, terutama dalam kategori positif. Meskipun *Naïve Bayes* unggul dalam mengidentifikasi kelas negatif, secara umum tidak optimal untuk kelas lain berdasarkan metrik yang diukur. Oleh karena itu, pemilihan algoritma yang paling cocok harus mempertimbangkan kebutuhan spesifik dan karakteristik data yang lebih mendalam berdasarkan hasil evaluasi ini.

4. KESIMPULAN

Analisis terhadap 336 *tweet* yang telah diberi label sentimen menunjukkan perbedaan antara pendekatan *labelling* otomatis dan manual. Dari *labelling* otomatis, 210 *tweet* (62.5%) dikategorikan sebagai positif, 60 *tweet* (17.86%) sebagai netral, dan 66 *tweet* (19.64%) sebagai negatif. Sementara itu, *labelling* manual mengidentifikasi 227 *tweet* (67.56%) sebagai positif, 82 *tweet* (24.4%) sebagai netral, dan hanya 27 *tweet* (8.04%) sebagai negatif. Meskipun terdapat variasi signifikan dalam proporsi *tweet* netral dan negatif antara kedua metode, keduanya menunjukkan bahwa mayoritas *tweet* cenderung mengungkapkan sentimen positif. Dalam konteks analisis opini masyarakat di Twitter tentang kesehatan mental di Indonesia, metode *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* diaplikasikan dengan hasil yang berbeda, di mana KNN menunjukkan akurasi sebesar 88.24% dan *Naïve Bayes* mencapai akurasi 60.78%.

5. DAFTAR RUJUKAN

[1] Rokom, "Menjaga Kesehatan Mental Para Penerus Bangsa," sehatnegeriku.kemkes.go.id,2023. https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20231012/3644025/menjaga-kesehatan-mental-para-penerus-bangsa/

[2] "Peran Keluarga Dukung Kesehatan Jiwa Masyarakat," kemkes.go.id, 2016. https://www.kemkes.go.id/id/rilis-kesehatan/peran-keluarga-dukung-kesehatan-jiwa-masyarakat (accessed Apr. 30, 2024).

[3] K. Aulia and L. Amelia, "Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes," Siliwangi J. (Seri Sains Teknol.), vol. 6, no. 2, pp. 60–65, 2020.

[4] Y. F. Nugraini, R. R. Saedudin, and R. Andreswari, "Implementasi Data Mining Dalam Kasus Mental Health Pada Sosial Media Twitter Menggunakan Metode Naive Bayes," e-Proceeding Eng., vol. 8, no. 5, pp.

9260–9265,2021.[Online].Available: https://repository.telkomuniversity.ac.id/pustaka/files/170554/jurnal_eproc/implementasi-data-mining-dalam-kasus-mental-health-pada-sosial-media-twitter-menggunakan-metode-naive-bayes.pdf

[5] M. L. Wicaksono, R. Rusdah, and D. Apriana, "Sentiment Analysis Of Mental Health Using K-Nearest Neighbors On Social Media Twitter," BIT (Fakultas Teknol. Inf. Univ. Budi Luhur), vol. 19, no. 2, p. 98, 2022. doi: 10.36080/bit.v19i2.2042.

[6] H. Aditya, Ardiansyah, Sidik, and W. Gata, "Analisis Sentimen Penggunaan Twitter Terhadap Penggunaan Cairan Desinfektan Menggunakan Metode Term Frequency – Inverse Document Frequency Dan Support Vector Machine," Informan's - J. Ilmu-Ilmu Inform. dan Manaj., vol. 14, no. 2, pp. 167–174, 2020.

[7] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," J. SIMETRIS, vol. 10, no. 2, pp. 681–686, 2019.

[8] A. Y. Pratama et al., "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," J. Sains Komput. Inform. (J-SAKTI), vol. 5, no. 2, pp. 897–910, 2021.

[9] R. A. Shafira, Yahfizham, and A. M. Harahap, "Menentukan Jarak Terpendek Dalam Pengiriman Barang Dengan Perbandingan Euclidean Distance Dan Manhattan Distance," J. Sci. Soc. Res., vol. VI, no. 3, pp. 678–685, 2023.

[10] K. Yan, D. Arisandi, and T. Tony, "Analisis Sentimen Komentar Netizen Twitter Terhadap Kesehatan Mental Masyarakat Indonesia," *J. Ilmu Komput. dan Sist. Inf.*, vol. 10, no. 1, 2022. doi: 10.24912/jiksi.v10i1.17865.