

ANALISIS SENTIMEN MASYARAKAT TERHADAP KESEHATAN MENTAL PADA TWITTER MENGGUNAKAN ALGORITME *K-NEAREST NEIGHBOR*

Ahmad Ilham¹, Wahyu Pramusinto²

¹Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia
²Manajemen Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia
Email: ^{1*}1911502134@student.budiluhur.ac.id, ²wahyu.pramusinto@budiluhur.ac.id
(* : corresponding author)

Abstrak- Masalah kesehatan mental akhir-akhir ini menjadi masalah kesehatan yang penting di seluruh dunia, banyak masyarakat khawatir dengan adanya penyakit tersebut. Upaya penurunan dalam jumlah penderita gangguan jiwa tidak akan berhasil jika masyarakat disertai diskriminasi terhadap penderita gangguan jiwa di lingkungan sekitarnya. Kesalahpahaman dan stigma serta rendahnya kesadaran masyarakat akan kesehatan jiwa turut andil dalam upaya penyembuhan penyakit jiwa. Dari penjelasan permasalahan di atas maka sangat penting dengan adanya analisis sentimen positif dan negatif untuk mengetahui opini masyarakat terhadap kesehatan mental di media sosial twitter. Pada penelitian ini, peneliti menggunakan metode yang dinamakan *text mining*, analisis sentimen dan algoritme *K-Nearest Neighbor*, tingkat akurasi pada analisis algoritme *K-Nearest Neighbor* sangatlah mudah dan efektif jika digunakan pada data latih yang jumlahnya besar dan terdapat informasi kurang atau bahkan tidak berarti. Selain itu data yang sudah didapatkan kemudian dianalisis, pada penelitian kali ini data berjumlah 400 tweet dengan kata kunci 'kesehatan mental' setelah itu data melalui tahap pembagian dengan perhitungan 1:9, 10% data uji dan 90% data latih. Kemudian data tersebut diolah menggunakan metode *CountVectorizer* untuk selanjutnya ke tahap klasifikasi menggunakan algoritme *K-Nearest Neighbor*. Hasil akurasi terbesar didapatkan menggunakan nilai $K=9$ dengan akurasi 93%, presisi 94% dan *recall* 99%. Adapun hasil presentasi masyarakat berkomentar positif pada periode Juni 2023 dengan jumlah 88,75% dan masyarakat berkomentar negatif 11,25%.

Kata Kunci: sentimen analisis, twitter, k-nearest neighbor

ANALYSIS OF PUBLIC SENTIMENT ON MENTAL HEALTH ON TWITTER USING *K-NEAREST NEIGHBOR* ALGORITHM

Abstract- Mental health problems have recently become an important health problem around the world, many people are worried about the presence of the disease. Efforts to reduce the number of people with mental disorders will not succeed if the community is accompanied by discrimination against people with mental disorders in the surrounding environment. Misunderstanding and stigma as well as low public awareness of mental health contribute to the healing of mental illness. From the explanation of the above problems, it is very important with the analysis of positive and negative sentiment to determine public opinion on mental health on social media twitter. In this study, researchers used a method called text mining, sentiment analysis and *K-Nearest Neighbor* algorithm, the level of accuracy in the analysis of the *k-Nearest Neighbor* algorithm is very easy and effective if used on large amounts of training data and there is less or even meaningless information. In addition, the data that has been obtained and then analyzed, in this study the data amounted to 400 tweets with the keyword 'mental health' after the data through the division stage with a calculation of 1:9, 10% test data and 90% training data. Then the data is processed using the *CountVectorizer* method to proceed to the classification stage using the *k-Nearest Neighbor* algorithm. The greatest accuracy results obtained using the value of $K=9$ with 93% accuracy, 94% precision and 99% recall. The results of the presentation of the community commented positively in the June 2023 period with a total of 88.75% and the community commented negatively 11.25%.

Keywords: sentiment analysis, twitter, k-nearest neighbors

1. PENDAHULUAN

Analisis sentimen adalah metode untuk mengumpulkan opini publik menggunakan media sosial yang didalamnya terdapat isu-isu terkini. Di era digital saat ini, analisis sentimen sering digunakan untuk menganalisis

pandangan masyarakat terhadap peristiwa yang sering terjadi di jejaring sosial, termasuk Twitter. Contoh analisis sentimen terapan antara lain menganalisis sikap orang terhadap peristiwa atau isu terkini melalui komentar Twitter. Pada kesempatan kali ini, penulis mengulas analisis sentimen masyarakat terhadap kesehatan mental.

Dalam era modern, kesehatan mental telah menjadi sorotan utama. Perubahan gaya hidup, tekanan sosial, dan perkembangan teknologi telah menciptakan tantangan baru dalam kesehatan mental individu. Isu ini meresap dalam berbagai aspek, termasuk pendidikan, pekerjaan, hubungan antarpribadi, dan kualitas hidup secara keseluruhan. Tingkat stres, kecemasan, depresi, dan gangguan mental lainnya semakin meningkat, memicu perhatian para profesional kesehatan, ilmuwan sosial, dan pemerintah. Perubahan pola hidup yang terhubung dengan urbanisasi, globalisasi, dan teknologi menciptakan beban baru yang mendorong tantangan psikologis yang belum pernah terjadi sebelumnya. Oleh karena itu, penelitian dan perhatian terhadap kesehatan mental menjadi sangat penting. Diperlukan pemahaman yang lebih dalam tentang faktor-faktor yang mempengaruhi kesehatan mental, strategi pencegahan gangguan mental, dan intervensi yang efektif. Selain itu, sangat penting untuk mengatasi stigma terhadap masalah kesehatan mental sehingga individu merasa nyaman mencari bantuan jika diperlukan.

Pada tahun 2018, Survei Kesehatan Mental mengatakan: “19 Juta orang mengalami gangguan mental, mayoritas seseorang di usia 15 tahun”. Menurut Organisasi Kesehatan Mental Dunia (WHO), lebih dari 264 juta orang dari segala usia menderita depresi, tanpa memandang jenis kelamin, status sosial, dan usia. [1]. Ada banyak pendapat pro dan kontra terhadap keberadaan kasus kesehatan jiwa di kalangan masyarakat. Oleh karena itu sentimen ini menimbulkan banyak pendapat tentang suatu hal, untuk itu diperlukan suatu sistem untuk mengungkapkan setiap kalimat sentimen tersebut.

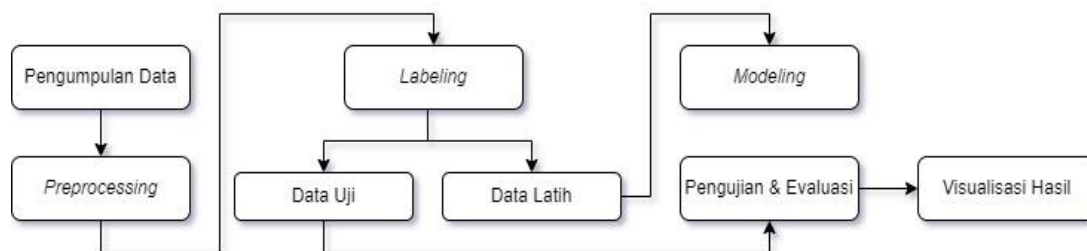
Dengan adanya pendapat pro dan kontra bertujuan untuk mengetahui pendapat positif atau negatif dari sebuah komentar yang dilontarkan oleh masyarakat di jejaring sosial Twitter. Twitter adalah jenis media sosial *microblogging* dan didalamnya memfasilitasi pengguna untuk mempublikasikan aktivitasnya dalam bentuk pendapat dan lain-lain. Sedangkan untuk versi terbaru sudah mendukung format video dan foto sebagai media tweet sehingga lebih menarik bagi pengguna.

Beberapa penelitian sebelumnya tentang Analisis Sentimen terhadap kesehatan mental diantaranya penelitian yang pertama adalah Implementasi Data Mining Dalam Kasus *Mental Health* Menggunakan Metode *Naive Bayes* [2], yang menyatakan bahwa, menggunakan algoritme *Naive Bayes* memiliki nilai macro average dan didapatkan hasil untuk presisi, *recall* dan *f1-score* dengan nilai 63%-74%, dan *weighted average* untuk presisi, *recall* dan *f1-score* dengan hasil nilai 89%-92%. Dalam penelitian lain yang berjudul Analisis Sentimen Kesehatan Mental Menggunakan *K-Nearest Neighbor* [3], yang menyatakan bahwa berdasarkan ekspreimennya dengan dataset ulasan pendapat positif sebanyak 639 dan dengan ulasan sentimen negatif sebanyak 193, maka hasil pemrosesan modeling dengan algoritme KNN didapatkan hasil terbaik menggunakan metode split data 70:30 dengan hasil nilai $K=5$, yaitu presisi 60,87%, *recall* 44,03% dan akurasi 58,39%. Dan dalam penelitian lain terhadap kesehatan mental di Indonesia yaitu dengan judul Analisis Sentimen Untuk Pengukurang Tingkat Depresi Menggunakan *Deep Learning* [4], menyatakan berdasarkan analisis dan evaluasi model CNN menggunakan metode *confusion matrix* menghasilkan nilai akurasi 82,90%.

Kontribusi penelitian ini bertujuan untuk mengklasifikasikan opini masyarakat pada sosial media Twitter terhadap kesehatan mental. Metode yang digunakan adalah melakukan analisis sentimen dengan menggunakan pendekatan machine learning sekaligus disertai dengan adanya fitur kamus sentimen, dengan ekstrasi fitur *CountVectorizer* dengan algoritme *K-Nearest Neighbor*. Mengingat keterbatasan sumber daya manusia, maka dari itu *Dataset* yang digunakan sebanyak 400 tweets yang bersumber dari twitter dengan kata kunci ‘kesehatan mental’. Data yang digunakan diambil pada periode 05 Juni 2023 sampai 09 Juni 2023. Analisis tersebut menghasilkan 2 pendapat yaitu positif dan negatif.

2. METODE PENELITIAN

Alur program analisis sentimen ialah sebuah gambaran umum yang menunjukkan alur dari penelitian yang dilakukan. Alur penelitian tersebut dapat dilihat pada Gambar 1 dibawah ini:



Gambar 1 Alur Penelitian

Berdasarkan Gambar 1, Pengumpulan data dimulai dengan proses *crawling* untuk mengumpulkan kicauan (tweet) sebagai dataset. Selanjutnya, tweet-tweet tersebut disimpan dalam bentuk excel dan dimasukkan ke dalam database untuk proses *preprocessing*. Pada tahapan *preprocessing*, dilakukan penyaringan, pembuangan, dan perbaikan kata-kata agar tweet menjadi lebih terstruktur. Hasil dari tahap *preprocessing* berupa *clean text* yang akan digunakan pada tahap selanjutnya, yaitu tahap *labeling*. Pada tahap *labeling*, tweet-tweet tersebut akan diberi kelas atau label berupa sentiment positif atau negatif berdasarkan isinya. Setelah diberi label, tweet-tweet yang telah berlabel akan dibagi menjadi dua bagian: data uji dan data latih. Data latih digunakan untuk membangun pengetahuan dan melatih model klasifikasi. Proses pembangunan pengetahuan ini dilakukan melalui proses *modeling*, yang menghasilkan sebuah model latih menggunakan data latih yang sudah ada. Data uji, di sisi lain, digunakan untuk menguji seberapa akurat model latih yang telah dibuat. Proses pengujian dan evaluasi dilakukan untuk mengukur tingkat keakuratan model tersebut. Setelah pengujian dan evaluasi selesai, hasil metode yang diterapkan akan divisualisasikan dalam bentuk presentase dan grafik untuk memudahkan pemahaman atas hasil penelitian..

2.1 Pengumpulan Data

Pengumpulan data dilakukan melalui proses *crawling* menggunakan alat bernama *Tweet Harvest*. *Crawling* ini merupakan tahapan krusial dalam analisis data, di mana data Twitter dianalisis dengan memanfaatkan *Aplikasi Programming Interface* (API) yang disediakan oleh Twitter. Dengan menggunakan API ini, data berupa teks yang diunggah oleh pengguna Twitter dapat diakses dan ditambang.[5]. Proses *crawling* tersebut memungkinkan kita untuk mengakses dan mengumpulkan data (tweet) dari media sosial Twitter, maka dari itu penulis berhasil mengumpulkan data dengan total 400 data (tweet) sebagai dataset.

2.2 Preprocessing

Preprocessing adalah tahap awal dalam analisis di mana data teks, khususnya data tweet dari Twitter, diolah sebelum melanjutkan ke tahap berikutnya. Pada tahap *preprocessing*, data tweet yang diperoleh dari Twitter diproses tanpa modifikasi, sehingga masih berupa data mentah. Tujuan dari *preprocessing* ini adalah untuk membersihkan data tersebut agar menjadi data yang lebih terstruktur dan dapat lebih mudah diolah pada tahap selanjutnya. [6]. Pada penelitian ini tahapan yang ada pada *preprocessing* sebagai berikut:

a. Case Folding

Case folding adalah proses mengubah seluruh huruf yang terdapat dalam data ulasan berupa teks menjadi huruf kecil menggunakan fungsi *lowercase*. Dengan melakukan *case folding*, semua huruf dalam teks akan diubah menjadi huruf kecil tanpa mengubah konten atau makna dari teks tersebut. Tujuan dari tahapan ini adalah untuk mengkonsistensikan teks dan memudahkan analisis karena perbedaan huruf besar dan kecil tidak lagi menjadi masalah dalam pemrosesan data. [7].

b. Cleansing

Pada tahapan *cleansing*, karakter-karakter yang memiliki pengaruh yang kurang penting akan dihapus atau dihilangkan. Karakter-karakter tersebut meliputi tanda baca, simbol-simbol dalam *emoticon*, *URL* atau tautan yang ada dalam data, *hashtag*, karakter, dan angka. Tindakan ini dilakukan untuk membersihkan data dari elemen-elemen yang tidak relevan atau tidak berkontribusi pada analisis lebih lanjut. Dengan menghapus karakter-karakter tersebut, data akan menjadi lebih bersih dan lebih mudah diolah dalam tahap selanjutnya. [7].

c. Mengubah Slang Word

Slang word adalah istilah yang digunakan untuk merujuk pada kata-kata yang tidak sesuai dengan ejaan baku dalam bahasa Indonesia (EYD). Termasuk dalam kategori ini adalah kata-kata singkatan, kata-kata gaul atau modern, dan juga kesalahan dalam penulisan kata. *Slang word* umumnya merupakan variasi atau

bentuk alternatif dari kata-kata yang digunakan dalam percakapan informal atau di kalangan tertentu. Karena tidak sesuai dengan ejaan baku, *slang word* tidak ditemukan dalam kamus resmi atau norma bahasa yang diatur oleh lembaga bahasa resmi. [8].

d. **Menghapus *Stop Word***

Stop word adalah sekelompok kata yang memiliki frekuensi tinggi dalam teks, namun memiliki sedikit atau bahkan tidak memiliki makna atau pengaruh yang signifikan dalam analisis atau pemrosesan teks. Kata-kata ini sering ditemukan dalam kalimat dan dapat ditemui secara umum dalam bahasa tertentu. Beberapa contoh *stop word* dalam bahasa Indonesia adalah "saya", "dan", "atau", dan lain sebagainya. Pada proses pengolahan teks, *stop word* biasanya dihapus atau diabaikan untuk mengurangi ukuran data dan meningkatkan efisiensi dalam analisis teks.. [8].

e. ***Stemming***

Tahapan *stemming* adalah langkah dalam *preprocessing* data di mana huruf-huruf imbuhan pada kata-kata dalam ulasan dihilangkan untuk mendapatkan kata dasar. Imbuhan tersebut dapat berupa awalan, akhiran, atau kombinasi keduanya. Tujuan dari *stemming* adalah untuk mengubah bentuk kata yang bervariasi menjadi bentuk dasarnya sehingga kata-kata yang memiliki akar kata yang sama dapat diidentifikasi sebagai satu kesatuan. Dengan demikian, analisis dan pengolahan lebih lanjut pada data teks dapat menjadi lebih efisien dan konsisten. [9].

2.3 ***Labeling***

Pelabelan merupakan proses pemberian label atau klasifikasi pada data yang telah terkumpul. Dalam tahap ini, data diberikan label untuk menentukan apakah setiap kalimat dalam dataset tersebut memiliki makna yang positif, negatif atau netral. [10]. Dalam penelitian ini, proses pelabelan dilakukan untuk memberikan klasifikasi pada setiap tweet dengan label positif atau label negatif. Label positif menunjukkan bahwa tweet tersebut cenderung bersifat setuju, mendukung, atau menerima terkait isu kesehatan mental. Sedangkan label negatif menunjukkan bahwa tweet tersebut cenderung bersifat menyangkal, menolak, atau menampik terkait isu kesehatan mental.

2.4 **Pembagian Data**

Pada tahap pembagian data, tweet yang telah diberi label akan dibagi menjadi dua bagian, yaitu data uji dan data latih. Proses pembagian data ini dilakukan dengan membagi *dataset* secara proporsional, di mana 90% dari total data akan menjadi data latih dan sisanya, 10%, akan menjadi data uji. Dengan pembagian ini, sebagian besar data akan digunakan sebagai data latih untuk melatih model, sementara sebagian kecil akan digunakan sebagai data uji untuk menguji tingkat akurasi model yang telah dibangun.

2.5 ***Modeling***

Modeling adalah proses pembangunan pengetahuan berdasarkan data yang sudah tersedia. Dalam proses ini, data latih yang akan digunakan untuk membangun model dipilih menggunakan teknik *sampling kuota*. Teknik *sampling kuota* merupakan metode *sampling* yang menentukan jumlah sampel dari populasi yang memenuhi kriteria tertentu, sehingga jumlah sampel yang diinginkan atau ditetapkan dapat tercapai. Dengan menggunakan teknik ini, data latih yang akan digunakan dalam proses *modeling* dipilih berdasarkan kuota tertentu, sehingga model yang dihasilkan dapat mewakili karakteristik dari data tersebut. [8]. Proses *modeling* menggunakan ekstraksi fitur *CountVectorizer* untuk menghitung frekuensi kata dalam dokumen atau kalimat. Data kemudian direpresentasikan dalam bentuk vektor dan disimpan sebagai model latih dalam format file JSON (.json).

2.6 **Pengujian dan Evaluasi**

Pada tahap evaluasi, sistem diuji untuk menilai performansi hasil klasifikasi yang telah dibuat. Proses ini melibatkan perhitungan nilai akurasi, presisi, *recall*, dan *f-measure*. Evaluasi performansi ini digunakan sebagai parameter untuk mengukur seberapa akurat implementasi metode yang digunakan. [11]. Untuk menghitung jarak antara data digunakan perhitungan *euclidean distance* dengan rumus sebagai berikut:

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

Keterangan:

$d(x, y)$ = jarak

x_i = data training

y_i = data testing

i = data training

n = dimensi data

Pengujian dilakukan dengan membandingkan hasil prediksi sistem (aplikasi) dengan hasil pelabelan oleh pakar. Data prediksi adalah hasil klasifikasi dari aplikasi yang dibangun, sedangkan data aktual diperoleh melalui proses pelabelan oleh pakar. Dalam penelitian ini, pengujian dilakukan pada model dengan data uji yang berbeda, dengan variasi nilai K yaitu: K=3, K=5, K=7, dan K=9. Hasil pengujian dievaluasi menggunakan *confusion matrix* untuk mengukur tingkat akurasi, presisi, dan *recall*.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset penelitian berasal dari media sosial Twitter berupa teks kicauan (tweet). Data tersebut diambil secara rutin melalui proses *crawling* menggunakan pustaka *Tweet Harvest*, dimulai dari tanggal 5 Juli 2023 hingga 9 Juli 2023, sehingga total *dataset* terdiri dari 400 data (tweet). Kata kunci yang digunakan dalam pengumpulan data antara lain adalah "kesehatan mental".

3.2 Preprocessing

Pada penelitian ini, data asli yang diperoleh dari media sosial Twitter mengalami beberapa tahap *preprocessing*. Tahap-tahap tersebut meliputi *case folding*, *cleansing*, mengubah *slang word*, menghapus *stop word*, dan *stemming*.

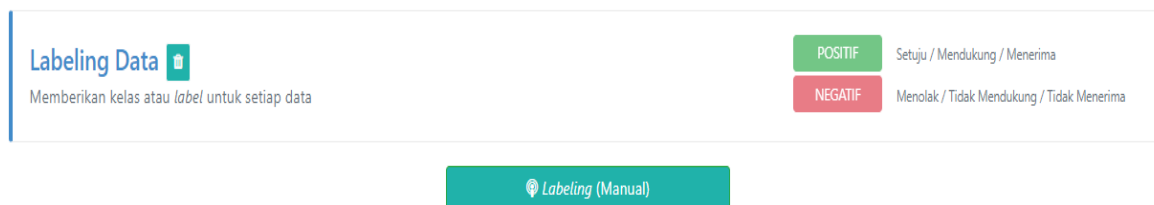
Tabel 1. Proses Preprocessing Data

Tahap	Input	Output
Case Folding	Sangat Penting “Kesehatan Mental” Karna Sangat Berdampak Bagi Manusia Kedepannya. Lebih Dari 264 Juta Orang Mengalami Depresi di Berbagai Usia. #KesehatanMental #MentalHealt https://www.halodoc.com/kesehatan/kesehatan-mental	sangat penting “kesehatan mental” karna sangat berdampak bagi manusia kedepannya. lebih dari 264 juta orang mengalami depresi di berbagai usia. #kesehatanmental #mentalhealt https://www.halodoc.com/kesehatan/kesehatan-mental
Cleansing	sangat penting “kesehatan mental” karna sangat berdampak bagi manusia kedepannya. lebih dari 264 juta orang mengalami depresi di berbagai usia. #kesehatanmental #mentalhealt https://www.halodoc.com/kesehatan/kesehatan-mental	sangat penting kesehatan mental karna sangat berdampak bagi manusia kedepannya. lebih dari juta orang mengalami depresi di berbagai usia.
Mengubah Slang Word	sangat penting kesehatan mental karna sangat berdampak bagi manusia kedepannya. lebih dari juta orang mengalami depresi di berbagai usia.	sangat penting kesehatan mental karena sangat berdampak bagi manusia kedepannya. lebih dari juta orang mengalami depresi di berbagai usia.

Menghapus Stop Word	sangat penting kesehatan mental karena sangat berdampak bagi manusia kedepannya. lebih dari juta orang mengalami depresi di berbagai usia.	sangat penting kesehatan mental sangat berdampak bagi manusia kedepannya. lebih juta mengalami depresi berbagai usia.
Stemming	sangat penting kesehatan mental sangat berdampak bagi manusia kedepannya. lebih juta mengalami depresi berbagai usia.	sangat penting sehat mental sangat dampak bagi manusia depan. lebih juta alami depresi bagai usia.

3.3 Labeling Data

Tahapan labeling merupakan tahapan yang hanya dapat dilakukan setelah tersedianya satu atau lebih data *clean text* pada basis data (*database*) hasil dari tahapan *preprocessing*. Labeling data dibagi menjadi dua (2) buah kategori: positif dan negatif. Pelabelan dilakukan dengan cara memberikan label secara manual kepada data yang sudah tersedia.



Gambar 1 Labeling Manual

3.4 Pengujian dan Evaluasi

Pada tahap pengujian dan evaluasi penelitian ini, digunakan model latih yang terdiri dari 360 data latih yang sudah diberi label, dengan 319 di antaranya positif dan 41 negatif. Selain itu, terdapat data uji sebanyak 80 data. Pengujian dilakukan dengan algoritme KNN menggunakan nilai $K=3$. Hasil sampel prediksi dari pengujian tersebut ditampilkan dalam Tabel 2.

Tabel 2. Sampel data hasil prediksi

No	Tweet	Label aktual	Label prediksi
1	@ReceinAja Akan ada fase dimana lebih baik kita tidak mengetahui/mencari tahu perihal yang dapat meruntuhkan dunia kita. Maka dari itu, ada kalanya orang harus melakukan yang namanya puasa sosmed untuk menjaga kesehatan mental agar terhindar dari overthinking, berprasangka, & rasa kecewa.	Positif	Positif
2	Secara umum, konsep seperti ini ada benarnya. Tapi jika dilihat dari sisi keseimbangan energi, variasi nutrisi, dan kesehatan fisik maupun mental, semua jenis makanan itu memiliki fungsinya masing-masing.	Negatif	Positif
3	Sebagai contoh, makanan tinggi nutrisi berbasis whole food memang seharusnya menjadi mayoritas jenis makanan yang kita konsumsi. Tapi dari sisi kesehatan mental, sangat sulit dan “menyiksa” jika harus mengonsumsi hanya jenis makanan seperti itu.	Negatif	Positif
....
80	@sayhun49 lah ga tau? selain akrab, kalo fallback gua dapet jaminan kesehatan mental juga?	Negatif	Positif

Tabel 3 menunjukkan representasi *confusion matrix* hasil prediksi dari 80 data tweet pada pengujian KNN dengan nilai $K=3$. Kolom label aktual berisi data label yang diperoleh melalui proses pelabelan, sedangkan kolom label prediksi berisi data label hasil dari proses klasifikasi menggunakan KNN. *Confusion matrix* ini merepresentasikan hasil prediksi keseluruhan dari data uji (80 data tweet).

Tabel 3. Confusion matrix pengujian K=3

		Nilai Aktual	
		TRUE (<i>positive</i>)	FALSE (<i>negative</i>)
Nilai Prediksi	TRUE (<i>positive</i>)	71	6
	FALSE (<i>negative</i>)	3	0

Tabel 4 menunjukkan perolehan nilai akurasi, presisi, dan *recall* berdasarkan *Confusion matrix* dari pengujian K=3, dan adanya nilai 0 dikarenakan tidak terdapatnya tweet berlabel negatif dan diprediksi sebagai negatif. Nilai-nilai ini menggambarkan seberapa baik kinerja model dalam melakukan klasifikasi pada data uji.

Tabel 4. Nilai pengujian K=3

	Pengujian	
Akurasi	$\frac{71 + 1}{71 + 0 + 3 + 6}$	0.89 (89%)
Presisi	$\frac{71}{71 + 6}$	0.92 (92%)
<i>Recall</i>	$\frac{71}{71 + 3}$	0.96 (96%)

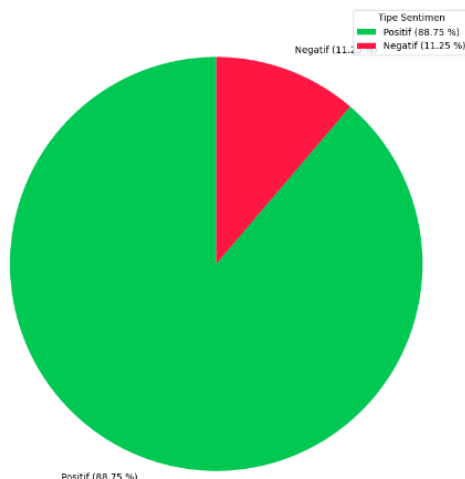
Pengujian dilakukan secara berulang dengan variasi nilai K yang berbeda-beda. Hasil pengujian secara keseluruhan dari berbagai nilai K dapat dilihat pada Tabel 5 berikut. Tabel ini memberikan gambaran komprehensif tentang performa model KNN dalam melakukan klasifikasi pada data uji dengan berbagai parameter K yang berbeda.

Tabel 5. Hasil pengujian dan Evaluasi

	K=3	K=5	K=7	K=9
Akurasi	0.89	0.89	0.91	0.93
Presisi	0.92	0.92	0.92	0.94
<i>Recall</i>	0.96	0.96	0.96	0.99

Berdasarkan Tabel 5, penggunaan nilai K di atas 9 tidak digunakan karena hasilnya cenderung konstant, maka dari itu ditemukan bahwa nilai akurasi terendah terjadi pada K=3 dan K=5 dengan akurasi 89%, presisi 92%, dan *recall* 96%, yang sama untuk kedua nilai K tersebut. Sementara itu, nilai akurasi tertinggi diperoleh pada K=9 dengan akurasi 93%, presisi 94%, dan *recall* 99%. Hasil menunjukkan bahwa sentimen positif memiliki persentase yang lebih tinggi dibandingkan sentimen negatif. Gambaran hasil sentimen dapat dilihat pada gambar berikut:

Pie Chart persentase sentimen pada 400 Data *tweet*.
355 *tweet* positif (88.75 %) & 45 *tweet* negatif (11.25 %)



Gambar 2 *Pie Chart*

4. KESIMPULAN

Berdasarkan *dataset* yang telah diproses, pandangan (sentimen) masyarakat Indonesia terhadap kesehatan mental cenderung positif sebesar 88.75%. *Dataset* terdiri dari 400 *tweet*, di mana 301 data memiliki sentimen positif dan 99 data sisanya bersentimen negatif. Penggunaan ekstraksi fitur *CountVectorizer* dan algoritme *K-Nearest Neighbor* (KNN) dalam analisis sentimen berjalan dengan baik. Hasil pengujian dan evaluasi menunjukkan nilai tertinggi yang diperoleh adalah akurasi 93%, presisi 94%, dan recall 99% dengan menggunakan nilai $K=9$.

DAFTAR PUSTAKA

- [1] S. Bulan and zahro varisna Rohmadani, "DECODE : Jurnal Pendidikan Teknologi Informasi," *Decod. J. Pendidik. Teknol. Inf.*, vol. 2, no. 2, pp. 59–65, 2022.
- [2] Y. Familia Nugraini, R. Rohmat Saedudin, and R. Andreswari, "Implementasi Data Mining Dalam Kasus Mental Health Pada Sosial Media Twitter Menggunakan Metode Naive Bayes Implementation of Data Mining in the Case of Mental Health on Social Media Twitter Using Naive Bayes Method," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9260–9265, 2021.
- [3] 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.
- [4] A. Rizki and Y. Sibaroni, "Analisis Sentimen Untuk Pengukuran Tingkat Depresi Pengguna Twitter Menggunakan Deep Learning," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 11367–11375, 2021.
- [5] A. Wandani, "Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 651–665, 2021.
- [6] A. A. Permana, arsanah arsanah, D. A. Kristiyanti, and M. Sihotang, "Analisis Sentimen Pendapat Masyarakat Terhadap Ppkm Darurat Pada Media Sosial Twitter Menggunakan Metode Naive Bayes," *JIKA (Jurnal Inform.)*, vol. 5, no. 3, pp. 400–404, 2021, [Online]. Available: <http://jurnal.umt.ac.id/index.php/jika/article/view/5190>
- [7] A. Yoga 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.
- [8] M. Priandi and Painem, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor,"

- Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 311–319, 2021.
- [9] S. Ernawati and R. Wati, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel,” *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [10] H. A. R. Harpizon, R. Kurniawan, Iwan Iskandar, R. Salambue, E. Budianita, and F. Syafria, “Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes,” ... *Di YouTube ...*, vol. 5, no. 1, pp. 131–140, 2022, [Online]. Available: <http://repository.uin-suska.ac.id/59746/>
- [11] T. Akhir, W. P. Ali, Y. Sibaroni, and S. Si, “Analisis Sentimen Masyarakat Terhadap Kinerja Presiden Indonesia Dalam Aspek Ekonomi , Kesehatan , dan Pembangunan Berdasarkan Opini dari Twitter Program Studi Sarjana Informatika Fakultas Informatika Universitas Telkom Bandung Analisis Sentimen Masyarakat,” vol. 6, no. 2, pp. 8637–8649, 2019.