

Analisis Sentimen Komentar Twitter (X) tentang Kesehatan Mental menggunakan Naïve Bayes untuk Mengukur Opini Publik

^{1*} Umul Latifah, ² Tasbi Khatuz Zuhriya, ³ Erna Daniati

¹²³ Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail : ¹ifahlatifa01@gmail.com, ²tasbiriya29@gmail.com, ³ernadaniati@unpkediri.ac.id

Penulis Korespondens : Erna Daniati

Abstrak— Penelitian ini dilakukan dengan tujuan untuk menganalisis komentar pengguna Twitter (X) terhadap isu kesehatan mental serta dampaknya terhadap opini publik. Dalam proses analisis komentar pengguna, metode yang digunakan adalah algoritma Naïve Bayes dengan pendekatan Analisis Sentimen untuk mengklasifikasi komentar menjadi sentimen positif dan negatif. Data yang digunakan dalam penelitian ini merupakan dataset Twitter (X) bertema kesehatan mental yang diperoleh melalui platform Kaggle. Dengan adanya penelitian ini diharapkan dapat memetakan opini publik yang bisa dijadikan dasar pemahaman terhadap persepsi masyarakat tentang kesehatan mental di media sosial khususnya Twitter (X).

Kata Kunci— Analisis Sentimen, Kesehatan Mental, Twitter (X)

Abstract— This research is conducted with the aim of analyzing Twitter (X) user comments on mental health issues and their impact on public opinion. In the process of analyzing user comments, the method used is the Naïve Bayes algorithm with a Sentiment Analysis approach to classify comments into positive and negative sentiments. The data used in this research is a Twitter (X) dataset themed around mental health obtained through the Kaggle platform. With this research, it is hoped that it can map public opinions that can serve as a basis for understanding community perceptions of mental health on social media, especially Twitter (X).

Keywords— Sentiment Analysis, Mental Health, Twitter (X)

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Perkembangan teknologi digital saat ini yang cukup pesat [1], dilihat dari beredarnya platform media sosial yang tidak terlepas dari kehidupan manusia. Media sosial memiliki berbagai macam fungsi, tidak hanya sebagai sarana komunikasi, tetapi juga bisa dijadikan tempat masyarakat untuk menyampaikan opini.

Twitter (X) menjadi salah satu platform media sosial yang memberikan pengaruh signifikan [2]. Twitter (X) memfasilitasi pengguna untuk mengutarakan pendapat mereka mengenai beragam persoalan sosial, termasuk isu kesehatan mental. Melalui Twitter (X), opini publik dapat memengaruhi pandangan masyarakat terhadap isu-isu tertentu. Di sisi lain, kondisi kesehatan

mental di Indonesia menunjukkan tantangan serius, dengan meningkatnya kasus gangguan jiwa yang belum diimbangi dengan layanan kesehatan yang memadai [3].

Untuk memahami bagaimana opini publik terbentuk di media sosial terkait isu kesehatan mental, analisis sentimen merupakan salah satu metode yang efektif. Analisis sentimen adalah cara mengumpulkan pendapat masyarakat melalui media sosial tentang berbagai isu terkini [4]. Melalui analisis ini akan diidentifikasi kecenderungan opini apakah bersifat positif atau negatif.

Seiring berkembangnya teknologi, data mining dan teknik pengolahan data teks seperti text mining semakin banyak digunakan untuk mengeksplorasi serta menganalisis data teks tidak terstruktur [5]. Banyak sekali pilihan metode yang digunakan dalam proses klasifikasi teks Analisis Sentimen, salah satu diantaranya adalah metode algoritme Naïve Bayes. Algoritme Naïve Bayes banyak digunakan dalam klasifikasi teks untuk analisis sentimen karena efisiensinya dalam menganalisis data dengan jumlah yang besar dan akurasi yang cukup baik dalam analisis teks [6].

Sementara penelitian terdahulu menurut [7] menggunakan platform media sosial Facebook dalam meneiliti analisis sentimen terhadap isu kesehatan mental, fokus kategori hanya sentimen negatif, dan menghasilkan klasifikasi komentar negatif saja. Penelitian ini menghadirkan kebaruan dengan menggunakan platform media sosial Twitter (X) untuk meneliti analisis sentimen, fokus kategori sentimen bisa positif atau negatif, dan menghasilkan analisis dampak terhadap opini publik. Penelitian ini dilakukan dengan tujuan untuk menganalisis komentar netizen Twitter (X) menggunakan metode Analisis Sentimen dengan algoritma Naïve Bayes untuk memahami kecenderungan opini publik dan dampaknya terhadap isu kesehatan mental.

II. METODE

Pada penelitian ini, data didapatkan dari platform kagle yaitu dataset yang berjudul “Sentiment Analysis Mental Health Tweets 2017-2023”. Dimana dalam dataset tersebut terdapat 146.394 komentar pengguna Twitter (X) yang akan dianalisi dengan metode Analisis Sentimen (sentiment analysis) berdasarkan algoritma Naïve Bayes untuk menghasilkan hasil klasifikasi dalam bentuk tweet positif dan negatif [8]. Berikut tahapan-tahapan prepocessing yang dilakukan sebelum melakukan analisis data menggunakan metode Analisis Sentimen terhadap komentar pengguna Tweeter :

a.) *Case Folding*

Proses untuk mengubah huruf kapital (Uppercase) menjadi huruf kecil (Lowercase), seperti teks “Awake but tired” akan diubah menjadi “awake but tired”. [9]

b.) *Removing Numbers*

Proses yang digunakan untuk menghapus semua digit angka yang ada dalam teks, seperti teks “it's just over 2 years” menjadi “it's just over years”. [10]

c.) *Punctuation Removal*

Proses yang digunakan untuk menghapus semua karakter *non-alfanumerik* dalam teks, seperti teks “rt @sewhq: #retro bears” menjadi teks “rt sewhq retro bears”. [7]

d.) *Stopwords Removal*

Proses yang digunakan untuk menghapus kata-kata umum seperti “the”, “is”, “in” dan lain sebagainya yang memiliki arti tidak penting dalam teks, seperti teks “it's just over years since” menjadi teks “years since”. [11]

e.) *Lemmatization*

Proses yang digunakan untuk mengubah kata ke dalam bentuk dasarnya, seperti kata “lists” diubah menjadi kata “list”.[12]

f.) *Word Tokenization*

Proses yang digunakan untuk memecah teks menjadi sebuah kata, seperti “year since diagnosed” menjadi “year, since, diagnosed”. [13]

g.) *Naïve Bayes*

Salah satu metode proses klasifikasi sederhana dan paling banyak digunakan dalam proses analisis data. Dalam proses klasifikasi Naïve Bayes dilakukan dengan menghitung probabilitas kelas berdasarkan distribusi kata pada data. Dengan menggunakan Naïve Bayes proses klasifikasi akan menjadi lebih sederhana, cepat dan akurat. [7]

h.) *Confusion Matrix*

Metode digunakan untuk mengukur tingkat kedekatan nilai prediksi dengan nilai sebenarnya atau biasa disebut dengan Akurasi, tingkat ketepatan dalam pengklasifikasian atau biasa disebut dengan Presisi, mengukur banyak proporsi positif data yang benar-benar positif atau biasa disebut Recall, dan yang terakhir untuk mengetahui nilai rata-rata harmonis serta memberikan keseimbangan antara presisi dan recall atau biasa disebut dengan F1-score.[7]

Tabel 1. Confusion Matrix

	Classified Positive	Classified Negative
Actual Positive	TP (True Positive)	FP (False Positive)
Actual Negative	FN (False Negative)	TN (True Negative)

Keterangan :

- TP (True Positive) : Banyaknya data positif yang terdeteksi dengan benar
- FT (False Positive) : Banyaknya data negatif yang terdeteksi sebagai data positif
- FN (False Negative) : Banyaknya data positif namun terdeteksi sebagai data negatif
- TN (True Negative) : Banyaknya data negatif yang terdeteksi dengan benar

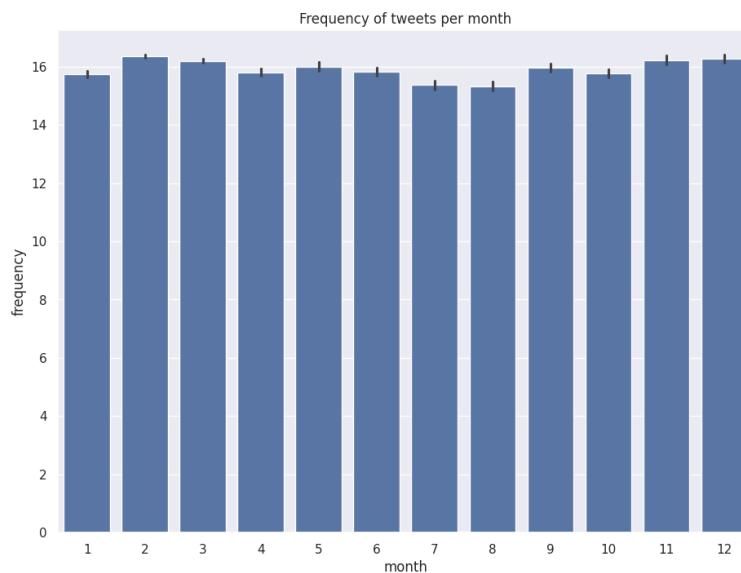
$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

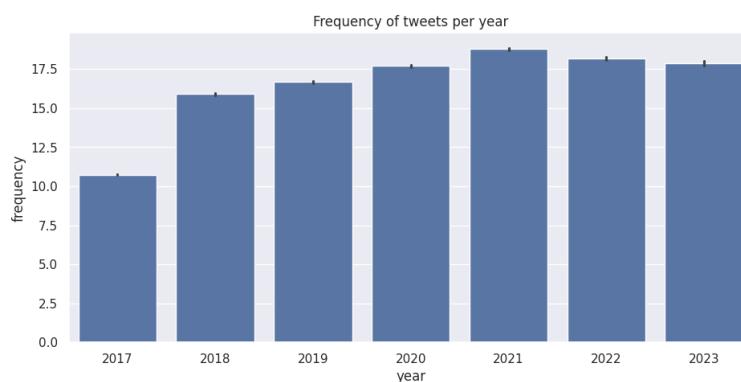
$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$F1\text{-score} = 2 \times \frac{TP}{TP + FP} \times 100\% \quad (4)$$

III. HASIL DAN PEMBAHASAN



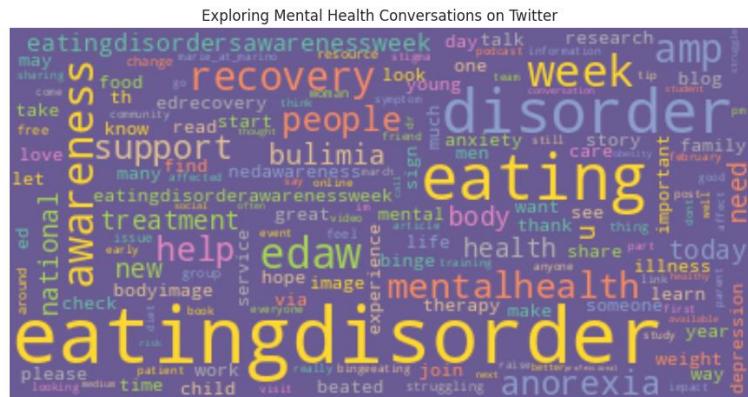
Gambar 1. Frekuensi Twitter (X) per Bulan



Gambar 2. Frekuensi Twitter (X) per Tahun

Hasil analisis pada komentar Twitter (X) tahun 2017-2023 berdasarkan gambar diatas diketahui bahwa komentar Twitter (X) cenderung stabil di setiap bulannya, artinya banyak pengguna yang aktif membuat komentar di platform Twitter (X). Meskipun sempat terjadi penurunan di tahun 2022. Dengan frekuensi komentar Twitter (X) diatas, perlu dilakukan analisis sentimen menggunakan metode Naïve Bayes untuk mengelompokan komentar Twitter (X) kedalam kategori *positive* dan *negative* serta menentukan Akurasi, Presisi, Recall dan F1-score

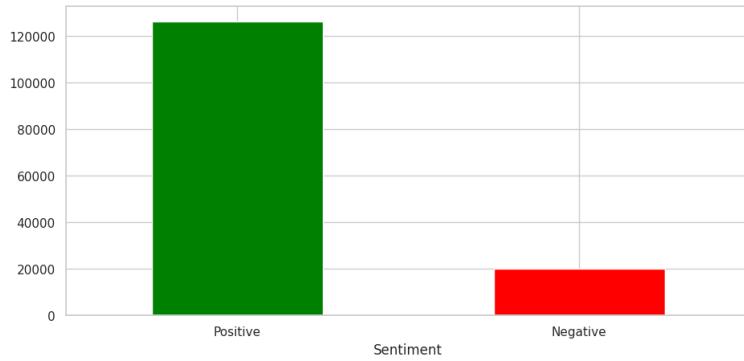
dengan bantuan Confusion Matrix. Dimana proses-proses analisis tersebut dilakukan dengan menggunakan bantuan *software* Python. Setelah melakukan semua proses analisis tersebut akan diketahui dampak opini publik mengenai isu kesehatan mental pada komentar platform Twitter (X).



Gambar 3. Worldcloud Komentar Tweeter

Gambar diatas merupakan visualisasi *wordcloud* yang dihasilkan setelah melakukan semua proses *preprocessing* dataset. Dari gambar tersebut dapat diketahui bahwa kata yang paling banyak muncul di komentar Twitter (X) adalah eatingdisorder (gangguan makan), disorder (gangguan), eating (makan), recovery (pemulihan), awareness (kesadaran) dan lain-lain.

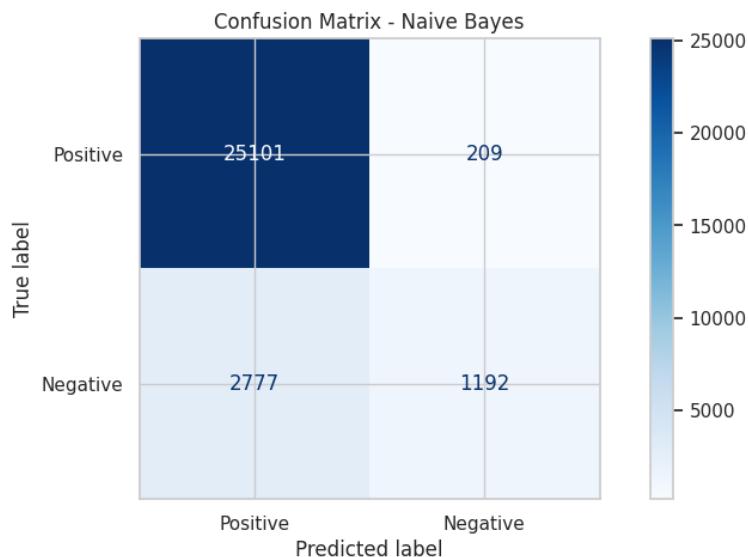
Meskipun kata yang paling banyak muncul pada visual wordcloud merupakan kata negative, setelah melakukan analisis dengan metode Analisis Sentimen menggunakan algoritma Naïve Bayes. Diketahui bahwa sebagian besar komentar Twitter (X) merupakan sentimen positive hal ini dapat dibuktikan melalui diagram chart berikut ini :



Gambar 4. Hasil Analisis Sentimen

Dari diagram tersebut diketahui bahwa komentar Twitter (X) yang memiliki sentimen *positive* sebanyak lebih dari 120.000 sedangkan komentar Twitter (X) yang memiliki sentimen *negative* sebanyak 20.000.

Setelah melakukan analisis dengan metode Analisis Sentimen menggunakan algoritma Naïve Bayes, langkah selanjutnya adalah menguji seberapa akurat metode Analisis Sentimen dengan algoritma Naïve Bayes yang digunakan dalam proses analisis komentar Twitter (X) dengan bantuan Confusion Matrix.



Gambar 5. Confusion Matrix Naïve Bayes

Dari gambar Confusion Matrix diatas dapat diketahui nilai dari TP, FP, FN dan TN sehingga dengan nilai-nilai tersebut dapat ditentukan Akurasi, Presisi, Recall dan F1-score.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{25101 + 1192}{25101 + 1192 + 2777 + 209} \times 100\% \\
 &= 89.80\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{25101}{25101 + 2777} \times 100\% \\
 &= 90.03\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{25101}{25101 + 209} \times 100\%
 \end{aligned}$$

$$= 99.17\%$$

$$\begin{aligned} F1 - Score &= 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \\ &= 2 \times \frac{90.03\% \times 99.17\%}{90.03\% + 99.17\%} \\ &= 94.37\% \end{aligned}$$

Dari proses perhitungan Akurasi, Presisi, Recall dan F1-score dengan bantuan peta Confusion Matrix diketahui bahwa nilai dari Akurasi sebesar 89.90% yang menunjukkan bahwa algoritma Naïve Bayes dapat mengklasifikasikan data dengan benar hampir 90% dari seluruh data yang diuji. Nilai dari Presisi sebesar 90.03% ini menunjukkan bahwa semua prediksi positif yang dibuat dengan menggunakan metode Naïve Bayes 90% diantaranya benar-benar positif. Nilai dari Recall sebesar 99.17% yang menunjukkan bahwa algoritma Naïve Bayes hampir berhasil menemukan seluruh data positif yang sebenarnya dan nilai dari F1-score sebesar 94.37% yang memperlihatkan keseimbangan performa metode Naïve Bayes dalam menghasilkan prediksi yang akurat.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan diatas, hasil dari analisis komentar pengguna Twitter (X) mengenai isu kesehatan mental menggunakan metode Analisis Sentimen, mayoritas menunjukkan sentimen positif. Dalam hal ini menjelaskan bahwa opini publik di media sosial Twitter (X) cenderung memberikan simpati, mendukung, serta menyuarakan pemulihian dan kesadaran terhadap permasalahan kesehatan mental. Dengan demikian, dapat disimpulkan bahwa persepsi publik terhadap isu kesehatan mental mayoritas bersifat positif.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Prodi Sistem Informasi dan Universitas Nusantara PGRI Kediri yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Frengki Apryanto, “PERAN GENERASI MUDA TERHADAP PERKEMBANGAN TEKNOLOGI DIGITAL DI ERA SOCIETY 5.0,” *Media Husada Journal of Community Service*, vol. 2, no. 2, pp. 130–134, 2022, doi: 10.33475/mhjcs.v2i2.35.
- [2] J. Saragih, M. Purba, M. Manik, N. D. Aulia, W. Wulandari, and O. A. Sihaloha, “PERAN INFLUENCER INSTAGRAM DALAM MEMBENTUK OPINI PUBLIK DAN PARTISIPASI POLITIK,” *Journal Education and Government Wiyata*, vol. 2, no. 4, pp. 396–406, Nov. 2024, doi: 10.71128/e-gov.v2i4.157.

- [3] A. N. Haryanti, M. B. S. Putra, N. Larasati, V. N. Khairunnisa, and L. D. D. A, "Analisis Kondisi Kesehatan Mental di Indonesia Dan Strategi Penanganannya," *Student Research Journal*, vol. 2, no. 3, pp. 28–40, Jun. 2024, doi: 10.55606/srjyappi.v2i3.1219.
- [4] M. Syarifuddinn, "ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.
- [5] M. M. M. Olhang, S. Achmadi, and F. X. Ariwibisono, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP COVID-19 DI INDONESIA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC)," *Jurnal Mahasiswa Teknik Informatika*, vol. 4, no. 2, pp. 214–221, Sep. 2020, doi: 10.36040/jati.v4i2.2695.
- [6] A. Safira and F. N. Hasan, "ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 59–70, Jan. 2023, doi: 10.31849/zn.v5i1.12856.
- [7] Zaenal, Y. Salim, and L. B. Ilmawan, "Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 1, no. 4, pp. 259–265, Nov. 2020, doi: 10.33096/busiti.v1i4.666.
- [8] E. Daniati and H. Utama, "TOPSIS in Decision-Making Framework Based on Twitter Sentiment Analysis," *International Conference on Information and Communications Technology*, pp. 268–273, Aug. 2021, doi: 10.1109/ICOIACT53268.2021.9564015.
- [9] A. P. Yoga, Y. Umaidah, and A. Voutama, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 897–910, Sep. 2021, doi: 10.30645/j-sakti.v5i2.386.
- [10] O. S. D. Silaen, H. Herlawati, and R. Rasim, "Analisis Sentimen Mengenai Gangguan Bipolar Pada Twitter Menggunakan Algoritma Naïve Bayes," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 6, no. 2, pp. 63–73, Nov. 2022, doi: 10.31603/komtika.v6i2.8198.
- [11] Y. A. N. Jannah and R. B. Prasetyo, "Analisis Sentimen dan Emosi Publik pada Awal Pandemi COVID-19 Berdasarkan Data Twitter dengan Pendekatan Berbasis Leksikon," *SEMINAR NASIONAL OFFICIAL STATISTICS*, vol. 2022, no. 1, pp. 597–608, Nov. 2022, doi: 10.34123/semnasoffstat.v2022i1.1483.
- [12] G. Z. Dhamara, D. N. Alamsyah, P. W. Saputro, E. Daniati, and A. Ristyawan, "Analisis Sentimen Aplikasi Mybca Melalui Review Pengguna Di Google Play Store Menggunakan Algoritma Naive Bayes," *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 8, no. 2, pp. 1102–1111, Aug. 2024, doi: 10.29407/inotek.v8i2.5044.
- [13] I. N. Bhakti, A. Z. Sholikhin, M. A. Lukman, E. Daniati, and A. Ristyawan, "Klasifikasi Kategori Berita Menggunakan Naive Bayes," *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 8, no. 2, pp. 1155–1164, Aug. 2024, doi: 10.29407/inotek.v8i2.5051.

