

IMPLEMENTASI ALGORITMA NAÏVE BAYES DALAM ANALISIS SENTIMEN TERHADAP TREND TIKTOK

Wahyu Tisno Atmojo¹⁾; Ericka Keisya²⁾; Afifah Trista Ayunda³⁾

^{1,2,3)} Sistem Informasi, Universitas Pradita Tengerang, Indonesia

¹⁾wahyu.tisno@pradita.ac.id; ²⁾ericka.keisya@student.pradita.ac.id; ³⁾afifah.trista@pradita.ac.id

ABSTRACT

Social networking is becoming more and more important. Social media's purpose has evolved from its first appearance as a place just for self-actualization to include online buying and selling, self-actualization, and other functions. TikTok is one of the social media platforms that is currently in high demand; opinions about its rise are mixed and include both positive and negative aspects. The goal of this study is to closely examine and comprehend how people react to the phenomena of TikTok's development by keeping an eye on user-generated material in tweets and the evolution of sentiment over time. This experimental study suggests using the Naïve Bayes Algorithm as a sentiment analysis method to examine how Twitter users are responding to the TikTok craze. In-depth insights into the dynamics of Twitter users' reactions to the TikTok trend are sought by this research, which combines sentiment analysis technology with Confusion Matrix performance evaluation. According to the sentiment analysis results, the majority of user comments are neutral (57.03%), followed by critical (33.20%) and affirmative (9.77%) remarks. This illustrates the nuanced reactions that people have had to the TikTok movement, in which the majority of users share their ideas in an unbiased manner. The significance of this research lies in its ability to provide an answer.

Keywords: Naïve Bayes, Twitter, TikTok Trend, Sentiment Analysis

I. PENDAHULUAN

Media Sosial saat ini berkembang dengan sangat pesat, fungsi dari sosial media saat ini mulai berubah dari yang awalnya untuk aktualisasi diri mulai bergeser untuk keperluan bisnis misalnya untuk jual beli barang dan jasa secara online. Media sosial sendiri dijelaskan sebagai suatu platform Online yang memungkinkan pengguna untuk bertukar berita, berinteraksi, berbagi konten, dan juga dapat dijadikan sebagai forum diskusi [1]. Berbagai informasi dapat diperoleh melalui media sosial bahkan hingga berbagai privasi dan data pribadi. Tercatat pengguna media sosial aktif di Indonesia dalam 3 tahun terakhir meningkat pesat dengan jumlah tertinggi pada tahun 2022 di 191 juta pengguna yang dikutip dari (<https://dataindonesia.id/>). Tercatat pada tahun 2021 Indonesia sebagai 10 besar negara pada urutan ke 9 sebagai negara dengan masyarakat yang kecanduan media sosial [2]. Saat ini kehadiran media sosial dijadikan juga sebagai alat untuk memobilisasi masyarakat untuk protes kepada pemerintah. Media sosial sendiri hadir dalam berbagai bentuk, teks

yang dapat digunakan untuk menyampaikan dan mengungkapkan pendapat, yang dikutip dari (<https://dataindonesia.id/>). Sebagai contoh Trend kampanye yang dilakukan oleh CAPRES melalui media sosial seperti TikTok.

Media sosial telah menjadi platform yang tak terelakkan dalam interaksi online, salah satunya adalah Tiktok yang saat ini sedang meningkat pesat penggunaannya. DataIndonesia.id memberikan gambaran statistik peningkatan penggunaan Tiktok 4 tahun terakhir hingga mencapai puncak di tahun 2023 dengan total 1,62 miliar pengguna dan total pendapatan sebesar US\$3,54 miliar [3]. Fenomena pesatnya pertumbuhan TikTok menimbulkan minat besar, bukan hanya dari segi hiburan, tetapi juga dalam konteks analisis sentimen. Dalam rangka memahami cara pengguna merespons tren yang berkembang di TikTok, penelitian ini akan menjelajahi analisis sentimen dengan menggunakan algoritma Naïve Bayes.

Analisis sentimen merupakan kunci untuk menggali pandangan, perasaan, dan respons pengguna terhadap konten yang tersebar luas

DOI : <https://doi.org/10.30646/tikomsin.v13i2.1015>

ISSN Online : 2620-7532

di TikTok. Untuk mendukung pendekatan ini, kita merujuk pada sejumlah jurnal kunci yang membahas berbagai aspek analisis sentimen, *machine learning*, dan perkembangan media sosial.

Analisis sentimen dijelaskan sebagai metode yang digunakan untuk melakukan klasifikasi terhadap suatu objek terhadap dua kategori sentimen yang terdiri dari sentimen negatif dan positif [4]. Dalam jurnal mengenai “Penerapan Model klasifikasi Metode *Naive Bayes* Terhadap Penggunaan Akses Internet” dijelaskan manfaat dari metode *naive bayes* sebagai *machine learning* yang dapat digunakan untuk menentukan tingkat kemungkinan [5].

Pentingnya Tik Tok sebagai fenomena budaya dan platform esensi ekspresi diri, *personal branding*, hingga media bisnis juga cara baru untuk memahami berbagai budaya lain [6]. Studi ini memberikan wawasan mengenai dampak positif dan negatif Tik Tok pada budaya remaja, dengan memberikan latar belakang yang penting untuk pemahaman lebih lanjut. Tiktok memiliki tidak sedikit trend yang banyak ditentang dan mendapat respon negatif dari masyarakat, bahkan tiktok banyak membawa pengaruh terhadap perubahan perilaku remaja saat ini dari cara berpakaian hingga perubahan gaya hidup [7].

Trend tiktok telah berkembang pesat sejak munculnya wabah COVID-19 sebagai salah satu media hiburan saat itu sebesar 64% pengguna TikTok menyatakan aplikasi TikTok dapat menurunkan tingkat stres yang dinyatakan pada penelitian jurnal di tahun 2021 dan 76% merasa bahwa trend TikTok membawa dampak positif, namun tidak sedikit masyarakat yang merasa bahwa trend TikTok membawa dampak negatif bagi masyarakat dengan total 64% juga dari masyarakat yang merasa Trend TikTok membawa pengaruh buruk [8].

Pentingnya memahami tren di TikTok dapat berdampak besar pada berbagai hal positif seperti perubahan pada sistem pembelajaran melalui sistem penyampaian materi yang lebih interaktif dan dapat mendorong pelajar untuk lebih mengekspresikan diri dalam pembelajaran [9], hal ini diperkuat dengan penelitian yang dilakukan oleh [10] yang menunjukkan tingkat relevansi materi pembelajaran dengan tiktok sebesar 55,36%. Pemanfaatan positif trend Tiktok lain adalah pemanfaatan trend untuk

media promosi [11], dengan tiktok yang memiliki jumlah pengguna banyak promosi produk atau brand menjadi salah satu hal yang sangat efektif untuk diterapkan.

Penelitian terkait dengan sentimen analisis aplikasi TikTok telah dilakukan sebelumnya oleh [18] dimana hasil penelitian tersebut menyatakan bahwa hasil analisis menunjukkan bahwa model memiliki akurasi 52,27%, nilai precision 19,60%, dan nilai recall 50,83%. Meskipun angka ini relatif rendah, temuan ini menunjukkan bagaimana pengguna melihat TikTok sebagai alat pembelajaran. Diharapkan penelitian ini akan memberikan pengetahuan tambahan kepada pengembang aplikasi pembelajaran berbasis media sosial dan membantu meningkatkan konten edukatif di TikTok. Penelitian lain dilakukan oleh [19] yang bertujuan untuk melakukan analisis sentimen, yang mengklasifikasikan komentar tiktok yang dapat menyebabkan depresi menjadi dua kategori: sentimen positif dan sentimen negatif. Studi ini menggunakan metode *Naive Bayes* untuk menentukan nilai presisi, akurasi, dan F1-Score. Selain itu, penelitian ini bertujuan untuk membuat pengguna lebih bijak di sosial media, terutama di media sosial TikTok.

Hal hal positif lain dari Trend dalam aplikasi TikTok dijelaskan dengan lebih detail dan mendalam dalam jurnal terkait “Perilaku Generasi Z terhadap Penggunaan Media Sosial Tiktok: Tiktok sebagai Media Edukasi dan Aktivisme” yang menyatakan penggunaan tiktok sebagai media untuk gerakan aktivisme dan edukasi yang di saksikan langsung hingga 23.8 miliar dengan tagar #blacklivesmatter[12].

Dengan menggabungkan literatur ini, penelitian ini bertujuan untuk menyajikan pemahaman mendalam terhadap dinamika sentimen pengguna di TikTok melalui pendekatan algoritma *Naive Bayes*. Daftar pustaka ini menjadi landasan teoritis dan metodologis yang kuat untuk menjelajahi kompleksitas respons pengguna terhadap tren media sosial yang terus berkembang pesat.

II. TINJAUAN PUSTAKA

2.1. *Sentiment Analysis*

Sentiment Analysis adalah cabang ilmu dari text mining, program bahasa alami, dan kecerdasan buatan yang digunakan untuk mendapatkan informasi atau pengetahuan

baru dengan mengekstrak, memahami, dan mengolah data teks secara otomatis.[13]

2.2. Naive Bayes Classifier

Naive Bayes Classifier merupakan suatu metode klasifikasi sederhana yang dapat menghitung semua kemungkinan dengan menggabungkan antara variasi dan frekuensi nilai dari basis data. Algoritma *Naive Bayes Classifier* menggunakan teorema Bayes untuk memperkirakan semua atribut yang saling bebas yang dapat diberikan oleh nilai pada variabel kelas [14]. Berikut bentuk rumus umum dari teorema bayes:

$$P(H \setminus X) = \frac{P(X \setminus H)P(H)}{P(H)} \dots\dots\dots(1)$$

2.3. Confusion Matrix

Proses validasi menggunakan metode *Confusion Matrix* untuk mengetahui sebaran kebenaran data prediksi terhadap data aktual. Serta pengujian algoritma klasifikasi dapat dilakukan menggunakan pengujian *Precision*, *Recall*, dan *F1-SCORE*, serta *Accuracy* dalam dataset.

Berikut untuk bentuk rumus dari Confusion Matrix [15]:

Tabel 1. Confusion Matrix

Label	Defenisi
True Positive (TP)	Jumlah positif yang dianggap positif
True Negative (TN)	Jumlah negatif yang dianggap negatif
False Positive (FP)	Jumlah positif yang dianggap negatif
False Negative (FN)	Jumlah negatif yang dianggap positif

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2)$$

$$Recall = \frac{TP}{TP+Fn} \dots\dots\dots(3)$$

$$F1 = \frac{2+Precision+Recall}{Precision+Recall} \dots\dots\dots(4)$$

$$Accuracy = \frac{TP+FN}{TP+TN+FP+FN} \dots\dots\dots(5)$$

2.4. Aplikasi TikTok

TikTok adalah aplikasi atau jaringan media sosial di mana pengguna dapat membuat video pendek yang dapat melakukan berbagai ekspresi, gaya, gerakan, dan tarian. TikTok menyediakan filter yang menarik bagi penggunanya [16]. TikTok juga tersedia di

Playstore, yang merupakan aplikasi fasilitas Google yang memungkinkan pengguna handphone untuk mengases atau mendownload aplikasi. [17]

III. METODE PENELITIAN

Adapun metode penelitian yang digunakan dalam melakukan penelitian ini meliputi tahap tahap yang diawali dengan proses pengumpulan data (*crawling data*) yang diperoleh dari platform social media Twitter (X), dengan dilanjutkan tahap selanjutnya cleaning data, tahap ketiga dilakukan tahap preprocessing data yang meliputi *case folding*, *stopword removal*, *Word normalizer*, *stemming*, dan *tokenizing*. Selanjutnya pelabelan data latih dilakukan dengan dilanjutkan pembobotan kata (*weighting*) menggunakan metode TF-IDF (*Term Frequency - Inverse Document Frequency*). Data yang telah diproses tersebut, akan dilakukan pelabelan dan proses klasifikasi menggunakan Algoritma Naive Bayes. Tahap validasi akan dilakukan setelah penerapan algoritma naive bayes berhasil dilakukan untuk memastikan tingkat akurasi dari penerapan algoritma sebelumnya.

3.1. Identifikasi Masalah

Penelitian dilakukan berdasarkan pada permasalahan yang telah dijelaskan sebelumnya yaitu:

1. Untuk mengetahui sentimen publik terhadap Trend pada aplikasi TikTok. Hasil dari Analisis sentimen dalam bentuk persentase seberapa besar tanggapan Masyarakat yang positif, negatif, atau netral.
2. Mengetahui kata-kata yang sering muncul pada setiap sentiment terkait Trend Tiktok positif, negatif, dan netral.

3.2. Crawling Data

Crawling data atau dengan kata lain pengumpulan data merupakan tahap awal yang diperlukan dalam proses sentimen analisis. Tahap ini merupakan tahap untuk melakukan pengumpulan data yang berasal dari Twitter dan kemudian dengan menggunakan *tools* dari sistem Google Colab juga Twitter API dengan tujuan menarik data. Setelah proses penarikan data, maka data akan berada dalam file dan download kedalam format csv.

3.3. Cleaning Data

Tahap ini merupakan tahap pengolahan data hasil *crawling* yang masih terdapat banyak data yang berantakan dan sulit terbaca dengan adanya data duplikat, dan simbol simbol lain yang tidak relevan diterapkan dalam penelitian ini. Sehingga, diperlukan cleaning data untuk memastikan data bersih dan dapat dengan mudah di baca agar dapat dilakukan pemrosesan selanjutnya.

1. *Case Folding*: Tahap ini merupakan tahap yang bertujuan untuk melakukan penyeragaman karakter data.
2. *Word Normalizer*: Bagian ini merupakan bagian yang bertujuan untuk memperbaiki kata dalam ulasan sehingga, dapat menghasilkan kalimat yang sesuai dengan kaidah bahasa Indonesia.
3. *Stemming*: Bagian ini merupakan bagian dalam melakukan penguraian kata menjadi suatu kata dasar tanpa adanya imbuhan baik pada bagian awal maupun akhir kata.
4. *Stopwords Removal*: Bagian ini merupakan bagian yang bertujuan untuk menghilangkan kata yang memiliki arti atau makna yang kurang jelas dalam data.
5. *Tokenizing*: bagian ini memiliki fungsi dalam melakukan pembagian teks input menjadi array token.
6. *Weighting*: Merupakan suatu proses pembobotan pada tiap kata dalam data.

3.4. Pelabelan Data

Proses selanjutnya adalah proses pelabelan data, proses ini dilakukan jika proses sebelumnya telah terealisasi. Pelabelan ini dilakukan secara manual dengan melabelkan setengah dari data (*positif, negatif, dan netral*). Proses ini bertujuan untuk melakukan pelatihan pada algoritma *Naive Bayes* dalam membentuk model yang akan menentukan label sentimen pada data.

3.5. Naive Bayes Classifier

Dengan menggabungkan variasi dan frekuensi nilai dalam basis data, *Naive Bayes Classifier* dapat menghitung semua kemungkinan.

3.6. Validasi

Dalam proses validasi, metode Confusion Matrix digunakan untuk menentukan sebaran kebenaran data prediksi terhadap data aktual.

IV. HASIL DAN PEMBAHASAN

Penelitian dilakukan dengan pengumpulan data *tweets* terkait Trend tiktok di tahun 2023. Pengambilan data dilakukan pada bulan Oktober dengan jumlah 300 data, dengan kata kunci yang digunakan dalam penarikan data adalah "Trend Tiktok". Agar data yang diperoleh berasal dari opini masyarakat Indonesia, maka *filter* dilakukan saat penarikan data *tweets* dengan mengatur bahasa dari data yang diperoleh hanya data dengan bahasa Indonesia.

4.1. Data Mentah

Data mentah diperoleh melalui pengambilan data dari hasil teknik *Crawling* data melalui Google Colab yang menggunakan bahasa pemrograman Python dengan dasar perintah penarikan data yang digunakan berasal dari *library* pandas. Data mentah masih memiliki banyak elemen atau simbol hingga URL yang tidak diperlukan dalam tulisan, dengan contoh seperti pada tabel data mentah berikut:

Tabel 2. Data Mentah

N o	Teks
1	Tue Sep 26 00:11:04 +0000 2023 1706461349608660000 Itu OP ZOM100 pdhl klo kelar on time dgn jadwal produksi yg sehat bisa jd trend Tiktok dance anjir buat alat marketing ðŸ¥² 884230313395724000 1706461349608660000 ryuuarm https://twitter.com/ryuuarm/status/1706461349608669184
2	Mon Sep 25 23:33:42 +0000 2023 1,70645E+18 @nastarjpeg @convomfs Udah kok, cuma lagi trend aja itu di tiktok 290919258 1,70615E+18 Moonlvr_ https://twitter.com/Moonlvr_/status/1706451944724914321
3	Mon Sep 25 22:08:42 +0000 2023 1,70643E+18 invitation com trend no tiktok agr o junny hita 1,4969E+181,70643E+18 minghaocria https://twitter.com/minghaocria/status/1706430552893472939
4	Mon Sep 25 22:01:02 +0000 2023 1,70643E+18 Terjebak #nostalgia #fypã, #trend #TikTok https://t.co/ec4RkXHXyd 251412229 1,70643E+18 cikgunuri https://twitter.com/cikgunuri/status/1706428624428278134

5	Mon Sep 25 18:55:36 +0000 2023 1,70638E+18 gagal mengikuti trend tiktok karena sahabatku sedang mabok https://t.co/nrHTCOTl7F1,53777E+181,70638E+18alodairy https://twitter.com/alodairy/status/17063819560979050699
---	---

4.2. Cleaning Data

Data yang telah diperoleh dari hasil *Crawling* data sebelumnya, kemudian melalui proses *Cleaning* data atau proses pembersihan data dari elemen dan simbol simbol yang tidak diperlukan sebelumnya. *Cleaning* data dilakukan dengan menggunakan rapidminer. Dari total data awal 300 data setelah melalui proses cleaning tersisa 250 data. Contoh dari hasil data cleaning adalah seperti gambar pada tabel berikut:

Tabel 3. Cleaning Data

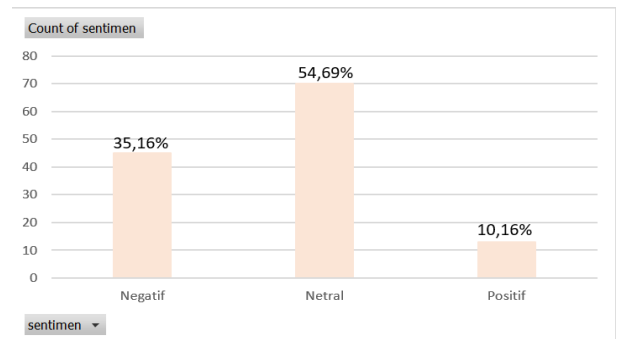
No	Teks
1	Itu OP ZOM100 pdhl klo kellar on time dgn jadwal produksi yg sehat bisa jd trend Tiktok dance anjir buat alat marketing
2	Udah kok cuma lagi trend aja itu di tiktok
3	gagal mengikuti trend tiktok karena sahabatku sedang mabok
4	tiktok trend reference sorry
5	Review 5 mood pemilik dn mihya 1 Nangis sampe sesedih sedihnya orang sedih 2 Ngakakin typo ngomong 3 Depressed janlup mikirin sebebana apa dia 4 Ngantuk 5 Misuh misuh 6 Bikin trend tiktok viral dr sexci sampe jamed additional

Tabel 4. Pelabelan Data

No	Teks	Sentimen
1	Review 5 mood pemilik dn mihya 1 Nangis sampe sesedih sedihnya orang sedih 2 Ngakakin typo ngomong 3 Depressed janlup mikirin sebebana apa dia 4 Ngantuk 5 Misuh misuh 6 Bikin trend tiktok viral dr sexci sampe jamed additional	Negatif
2	mileapo trend bukankah semesta yg mempertemukan kita ivan gunawan bold glamour effect tiktok reaction video	Netral
3	anjir ini trend tiktok bnyk bgt sih	Negatif
4	sumpah gwe benci banget sama sirkel di kelas gua yang ngebet pengen ikutan trend	Negatif

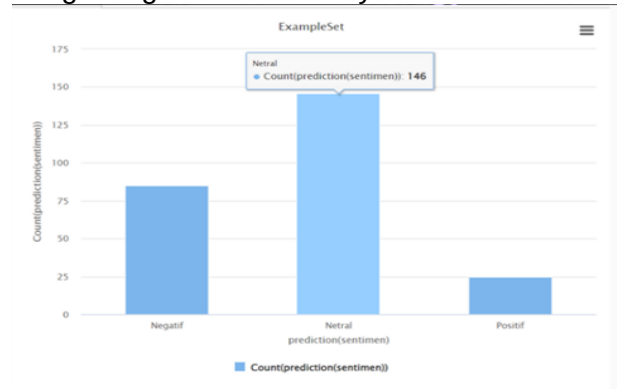
	tiktok rame rame lu sama sirkel lu aja sono yang bikin anjinggg	
5	lucu bgt dah trend tiktok me as me as ini	Positif

Hasil dari pelabelan data manual digambarkan dalam grafik pada gambar berikut:



Gambar 1. Grafik Presentase Sentimen

Berikut untuk keseluruhan data pelabelan dengan algoritma naive bayes:



Gambar 2. Grafik Sentimen

Hasil dari data pelabelan *tweets* dinyatakan bahwa sebanyak 146 data *tweets* menyatakan pernyataan komentar Netral dengan presentase 57,03%, 85 data *tweets* dengan label Negatif pada presentase 33,20% dan 25 data di presentase 9,77% menyatakan komentar positif.

Kemudian, hasil analisis juga mengidentifikasi kata-kata yang sering muncul dalam setiap kategori sentimen. Hal ini memberikan wawasan tambahan tentang pandangan dan perasaan pengguna terkait dengan tren TikTok. Analisis kata-kata ini dapat membantu dalam memahami argumen atau isu tertentu yang mencuat dalam diskusi penggunaan. Kata kata yang sering muncul dan diperbincangkan adalah good app, nice app dapat terlihat dalam word cloud berikut:

- Denpasar, "PERSEPSI MASYARAKAT TERHADAP APLIKASI TIKTOK SEBAGAI MEDIA MENURUNKAN TINGKAT STRES DI ERA PANDEMI COVID-19."
- [9] N. S. Ramdani, H. Nugraha, and A. Hadiapurwa, "POTENSI PEMANFAATAN MEDIA SOSIAL TIKTOK SEBAGAI MEDIA PEMBELAJARAN DALAM PEMBELAJARAN DARING," *Akademika*, vol. 10, no. 02, pp. 425–436, Dec. 2021, doi: 10.34005/akademika.v10i02.1406.
- [10] P. Seminar, N. Penelitian, and D. Pengabdian, *Penelitian dan Pengabdian Inovatif pada Masa Pandemi Covid-19*. 2021.
- [11] E. T. Endarwati and Y. Ekawarti, "EFEKTIFITAS PENGGUNAAN SOSIAL MEDIA TIK TOK SEBAGAI MEDIA PROMOSI DITINJAU DARI PERSPEKTIF BUYING BEHAVIORS," *MANDAR: Management Development and Applied Research Journal*, vol. 4, pp. 112–120, 2021.
- [12] F. I. R. Firamadhina and H. Krisnani, "PERILAKU GENERASI Z TERHADAP PENGGUNAAN MEDIA SOSIAL TIKTOK: TikTok Sebagai Media Edukasi dan Aktivisme," *Share: Social Work Journal*, vol. 10, no. 2, p. 199, Feb. 2021, doi: 10.24198/share.v10i2.31443.
- [13] J. Homepage et al., "Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, pp. 24–33, 2021.
- [14] R. Rachman, R. N. Handayani, and I. Artikel, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *JURNAL INFORMATIKA*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [15] W. Hidayat, M. Ardiansyah, and A. Setyanto, "Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb," *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 1, pp. 11–20, Jun. 2021, doi: 10.29408/edumatic.v5i1.3125.
- [16] S. D. Fatimah, C. Hasanudin, and A. K. Amin, "Pemanfaatan Aplikasi Tik Tok Sebagai Media Pembelajaran Mendemonstrasikan Teks Drama," *Indonesian Journal Of Education And Humanity*, vol. 1, 2021.
- [17] S. M. Pratama and Muchlis, "PENGARUH APLIKASI TIK TOK TERHADAP EKSPRESI KOMUNIKASI MAHASISWA UNIVERSITAS ISLAM NEGERI (UIN) SUNAN AMPEL SURABAYA TAHUN 2020," *Internasional Journal Of Education Resources*, vol. 1, no. 2, pp. 102–115, 2020.
- [18] Apriani, Elsa et al, "Analisis Sentimen Penggunaan TikTok Sebagai Media Pembelajaran Menggunakan Algoritma Naïve Bayes Classifier", *MALCOM: Indonesian Journal of Machine Learning and Computer Science* vol 4 July 2024.
- [19] D.F. Mey Dina, T. Haryanti and M.A Haq, "ANALISIS SENTIMEN TERHADAP KOMENTAR PADA MEDIA SOSIAL TIKTOK YANG BERPOTENSI MENYEBABKAN DEPRESI MENGGUNAKAN METODE NAIVE BAYES", *Jurnal Ilmiah Computing Insight* Vol.7 No.1, pp. 1–9, 2025.