

ANALISIS SENTIMEN TERHADAP *CYBERBULLYING* DI MEDIA SOSIAL MENGGUNAKAN METODE KUANTITATIF *NAÏVE BAYES*

Siti Hani Fadilah¹⁾, Adiya Fanial Rizki²⁾, Khairunnisa Regita Cahyani³⁾,
Giantika Chrisnawati⁴⁾

Prodi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

Correspondence author: AF Rizki, adityafr84@gmail.com, Jakarta, Indonesia

Abstract

TikTok is a digital interaction platform with a high user base. Cyberbullying is a technology-based form of bullying that can have serious impacts on the mental and social well-being of victims; therefore, understanding public perception of the issue is crucial. This study aims to identify public opinion trends regarding the current prevalence of cyberbullying. This study uses data from comments on a TikTok account and classifies these comments into positive, negative, and neutral sentiment classes. The method used is quantitative, with the naïve bayes algorithm, supported by text mining processes including data cleaning, tokenisation, stopword removal, stemming, and extraction using TF-IDF. The research dataset consisted of 1,100 comments, and after filtering, 779 comments were finalised. The test results show that naïve bayes is capable of providing good sentiment classification based on accuracy, precision, recall, and F1-score values. These findings are expected to raise awareness and minimise the occurrence of cyberbullying on social media.

Keyword : *sentiment analysis, cyberbullying, naïve bayes, social media, tiktok*

Abstrak

Media Sosial Tiktok menjadi salah satu ruang interaksi digital dengan tingkat pengguna yang tinggi. *Cyberbullying* merupakan perilaku perundungan berbasis teknologi yang dapat berdampak serius pada kondisi mental dan sosial korban, sehingga diperlukan pemahaman mengenai persepsi publik mengenai isu tersebut. Penelitian ini bertujuan untuk mengidentifikasi kecenderungan opini masyarakat terhadap kasus *cyberbullying* yang sedang marak terjadi. Penelitian ini menggunakan data dari komentar salah satu akun di Tik Tok, serta mengklasifikasi komentar tersebut kedalam kelas sentiment positif, negatif, maupun netral. Metode yang digunakan adalah kuantitatif dengan algoritma *naïve bayes*, didukung oleh proses *text mining* mulai dari pembersihan data, tokenisasi, *stopword removal*, *stemming*, dan ekstraksi menggunakan TF-IDF. *Dataset* penelitian ini terdiri dari 1.100 komentar dan telah dilakukan penyaringan komentar yang berubah menjadi 779 komentar. Hasil pengujian menunjukkan bahwa *naïve bayes* mampu memberikan klasifikasi sentimen dengan baik berdasarkan nilai akurasi, *precision*, *recall*, maupun *F1-score* yang diperoleh. Temuan ini diharapkan mampu meningkatkan kesadaran dan meminimalkan terjadinya *cyberbullying* di media sosial.

Kata Kunci : *analisis sentimen, cyberbullying, naïve bayes, media sosial, tiktok*

A. PENDAHULUAN

Perkembangan teknologi yang pesat membuat media sosial menjadi alat penting untuk mempermudah interaksi sehari-hari, termasuk komunikasi dan akses informasi. Berbagai platform populer di antaranya Youtube, Facebook, TikTok, Instagram, LinkedIn, maupun Twitter memungkinkan setiap individu di dunia membangun jaringan sosialnya secara daring (Damayanti et al., 2023).

TikTok sebagai platform media sosial yang memungkinkan pembuatan video singkat dengan durasinya 15–60 detik dengan dukungan beragam fitur di antaranya stiker, musik, filter, maupun efek lainnya. Kehadiran aplikasi ini juga memberi peluang bagi praktik *cyberbullying*, di mana pengguna dapat mengunggah foto, video, atau tulisan dengan bahasa yang menyakiti, umumnya melibatkan individu lain dalam konten yang dibagikan (Aser et al., 2022).

Cyberbullying sebagai perilaku pelecehan atau intimidasi yang memanfaatkan media elektronik atau teknologi komunikasi (Nugraha & Astuti, 2023). Dampaknya dari tindakan ini sifatnya serius bagi kesejahteraan maupun kesehatan mental korban, sekaligus menjadi tantangan yang kian mengemuka di era digital. Oleh karenanya, upaya pencegahan yang efektif perlu dilaksanakan pembuat kebijakan, orang tua serta pendidik (Marlef et al., 2024). Dalam memahami pandangan publik secara sistematis, diperlukan pendekatan analisis berbasis data. Analisis sentimen sebagai metode yang dipergunakan dalam mengelolah data teks dan mengklarifikasikan opini menjadi kategori seperti positif, negatif, ataupun netral (Syahira & Kurniawan, 2024).

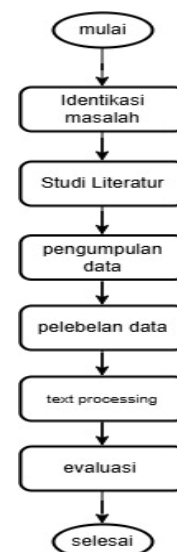
Penelitian ini merujuk pada beberapa studi sebelumnya yang telah dikaji beberapa kasus *cyberbullying* di media sosial serta pendekatan analisis sentimen terhadap opini publik di media sosial. Penelitian (Wibisono et al., 2025) menggunakan metode *Naïve bayes* untuk secara efektif dan efisien

mengklasifikasikan komentar terkait *cyberbullying*. Metode *Naïve bayes* merupakan salah satu metode klasifikasi. Metode *Naïve bayes* dilakukan dengan membandingkan nilai posterior dari kelas-kelas yang ada. Nilai posterior yang paling tinggi yang terpilih sebagai hasil klasifikasi (Khadafi et al., 2022).

Dari latar belakang di atas, penelitian ini dilaksanakan dengan maksud untuk memahami persepsi publik terhadap *cyberbullying* yang sampai saat ini masih terjadi di media sosial. Data dalam penelitian ini diperoleh dari komentar salah satu unggahan aku di media sosial yang membahas tentang bulling di media sosial, dengan menggunakan metode kuantitatif dan metode *Naïve bayes*. Penelitian ini diharapkan mampu untuk menghasilkan analisis dan bermanfaat secara sosial dan dapat menjadi pertimbangan masyarakat agar bijak dan *aware* terhadap kasus *cyberbullying* di media sosial.

B. METODE PENELITIAN

Penelitian ini mengadopsi model kuantitatif dengan memanfaatkan metode analisis sentimen berbasis *Naïve bayes*. Tahapan maupun prosedur yang diterapkan dijabarkan:



Gambar 1. Alur proses penelitian

Identifikasi Masalah

Pada tahap awal ini menggunakan pendekatan kuantitatif dengan menggunakan analisis sentimen pada masyarakat terhadap *cyberbullying* dalam kehidupan beropini di media sosial.

Studi Literatur

Studi literatur melibatkan data pendukung lainnya seperti jurnal dan penelitian terkait, sebagai acuan dalam penelitian ini.

Pengumpulan Data

Pada langkah ini dikumpulkan secara spesifik penetapan data sesuai batasan masalah, yaitu komentar pengguna salah satu unggahan pada platform TikTok dari akun @bulelengterkini pada tahun 2025, tentang pembulian yang dilakukan oleh mahasiswa senior di unud. Akun ini dipilih karena berisi konten yang bersifat informatif dan aktual, sehingga banyak mampu memicu beragam respons emosional dari pengguna dalam bentuk komentar. Variasi komentar yang muncul mulai dari sentimen positif, negatif, hingga netral.

Pelabelan Data

Setelah dapat data dari @bulelengterkini dari postingan di TikTok selanjutnya adalah pelabelan dengan label komentar tentang opini masyarakat terhadap *cyberbullying* dengan label positif, negatif dan netral. Dan kami juga menghapus komentar yang tidak sesuai dengan penelitian ini. Pelabelan ini dilakukan secara manual ke dalam *google sheet*.

Text Processing

Tahap *Text processing* bertujuan untuk memproses data mentah agar siap digunakan (Putra et al., 2025). Tahap *Text processing* meliputi langkah-langkah berikut : *cleaning*, *tokenize*, *transform case*, *stopword*, *filter*, *stemming*.

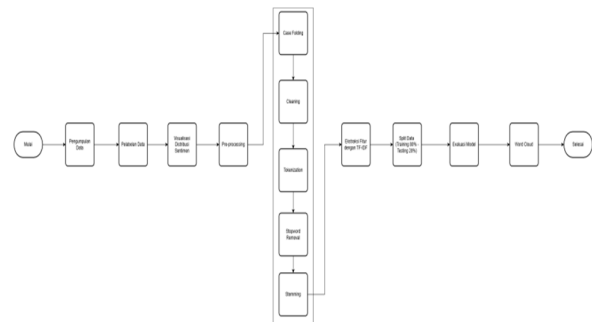
Evaluasi

Dalam tahap evaluasi ini pengujian menggunakan model yang di hasilkan oleh algoritma *naïve bayes* dengan mengadopsi

confusion matrix. Adapun tujuannya dari pengujian ini guna mengetahui seberapa baik model *naïve bayes* dalam melakukan klasifikasi komentar di TikTok .

Metode Pengolahan dan Analisis Data

Metode pengelolaan maupun analisis data dalam temuan ini mengadopsi *text mining* maupun *machine learning* untuk mengklafikasi sentimen.



Gambar 2. Alur Metode Pengolahan dan Analisis Data

Visualisasi Distribusi Sentimen

Kami membuat *bar chart* dan *pie chart* untuk melihat sebaran komentar positif, netral, dan negatif untuk melihat apakah data seimbang atau condong ke salah satu kelas.

Pre-processing

Case Folding: keseluruhan huruf diubah menjadi teks kecil. *Cleaning*: seluruh *mention*, *hashtag*, URL, angka, emotikon, dan tanda baca dihapus.

Tokenization: teks dipecah menjadi potongan kata-kata terpisah.

Stopword removal: kata-kata yang tidak memberikan informasi penting dibuang dari data.

Stemming: kata-kata dikembalikan ke bentuk dasar.

Ekstraksi fitur dengan TF-IDF: mengubah komentar menjadi vektor numerik sehingga model *naïve bayes* bisa menghitung bobot pentingnya tiap kata.

Split Data: Membagi data dengan dua bagian yakni latih maupun uji agar distribusi tiap kelas sentimen tetap proporsional.

Evaluasi Model

Melakukan pengukuran performa model dengan menghitung akurasi, *precision*, *recall*, *F1-score*, menggunakan *confusion matrix* dan memvisualkan hasil grafiknya.

1. Akurasi (*Accuracy*): Menggambarkan kemampuan model dalam mengklasifikasikan seluruh kategori dengan tepat. $Accuracy: (TP+TN) / (TP+FP+FN+TN)$
2. Presisi (*Precision*): Menunjukkan proporsi prediksi positif yang akurat. $Precision: (TP) / (TP+FP)$
3. *Recall* (Sensitivitas): Menilai seberapa efektif model dalam mengenali semua kasus positif yang sebenarnya. $Recall: TP / (TP+FN)$
4. *F1-score*: Mengombinasikan presisi maupun *recall* guna memberikan ukuran kinerja model yang seimbang. $F1-score: (2*Recall*Precision) / (Recall+Precision)$

Keterangan: TP merupakan data yang positif maupun berhasil diidentifikasi sebagai positif oleh model. TN: merupakan data yang sebenarnya negatif maupun diprediksi sebagai negatif oleh model. FP merupakan data yang sesungguhnya negatif namun terklasifikasi sebagai positif. FN merupakan data yang seharusnya positif tetapi terdeteksi sebagai negatif (Baehaqi & Cahyono, 2024).

Word Cloud

Membuat *Word Cloud* untuk setiap kelas sentimen agar terlihat kata apa yang sering muncul di masing-masing kelas.

Kesimpulan Penelitian

Menampilkan total data, proporsi *training/testing*, akurasi model, distribusi sentimen yang digunakan sebagai bagian akhir laporan penelitian.

C. HASIL DAN PEMBAHASAN

Penelitian menggunakan data komentar dari postingan di sosial media TikTok pada akun @bulelengterkini yang berjumlah 1.100 komentar. Setelah pemfilteran data, jumlah data komentar menjadi 779 komentar. Setelah itu data komentar ini di *split* menjadi 80% data training dan 20% data testing.

Pengumpulan Data

Data komentar yang ada di *google sheet* sudah dilakukan pemfilteran data, kemudian diupload ke google colab. Setelah pemfilteran, total data yang ada menjadi 779 baris.

```
=====
MEMUAT DATA DARI GOOGLE SHEETS
=====

Data berhasil dimuat dari Google Sheets!
Total data: 779 baris
Sumber: Komentar TikTok @bulelengterkini
Topik: Cyberbullying di Media Sosial (Kasus Timothy Anugerah)


5 data pertama:
```

	komentar	Sentiment
0	dunia kampus emang bully nya parah guys...	(netral)
1	selucu ini di bully	(netral)
2	Kenapa sih orang baik selalu diambil duluan 🙄🙄	(netral)
3	💩💩💩💩💩💩💩💩orang2 knp ga punya hati	(netral)
4	Kasihani i, gak kebayang pas pertama berkenalan...	(positif)

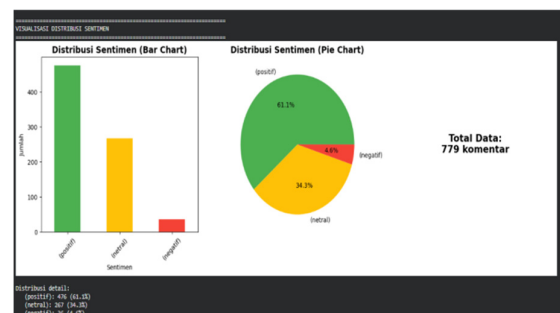
Gambar 4. Hasil Pemfilteran Data

Pelabelan Data

Pelabelan data terbagi atas 3 kelas sentimen (positif, negatif, maupun netral).

Visualisasi Distribusi Sentimen

Pada tahap ini dibuat diagram bar dan diagram pie untuk melihat kalau setiap kelas tidak condong ke satu kelas saja.



Gambar 5. Visualisasi Data

Pre-processing

Setelah melakukan pelabelan data, selanjutnya kami melakukan *pre-processing* terhadap *dataset*.

Case Folding

Di tahap ini seluruh huruf kapital akan dirubah menjadi huruf kecil.

Sebelum	Sesudah
Kenapa sih orang baik selalu diambil dulu	kenapa sih orang baik selalu diambil dulu

Cleaning

Membersihkan data teks dari mention, *hashtag*, URL, angka, *emoticon*, dan tanda baca.

Sebelum	Sesudah
Kenapa sih orang baik selalu diambil dulu	kenapa sih orang baik selalu diambil dulu

Tokenization

Tokenization merupakan teknik yang bertujuan untuk memecah teks menjadi unit-unit yang lebih kecil, seperti kata, frasa atau kalimat (Suhartoyo et al., 2025).

Sebelum	Sesudah
kenapa sih orang baik selalu diambil dulu	'kenapa', 'sih', 'orang', 'baik', 'selalu', 'diambil', 'dulu'

Stopword removal

Stopword removal merupakan salah satu metode yang bisa diterapkan dalam tahapan *text pre-processing*, dimana akan dilakukan penghapusan pada kata-kata yang dianggap sering muncul namun tidak bermakna dan berpengaruh signifikan terhadap makna kalimat atau teks (Angelina et al., 2023).

Sebelum	Sesudah
'kenapa', 'sih', 'orang', 'baik', 'selalu', 'diambil', 'dulu'	'kenapa', 'diambil', 'dulu'

Stemming

Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya (Zulfiqri et al., 2024).

Sebelum	Sesudah
'kenapa', 'dulu'	'diambil', kenapa ambil dulu

Ekstraksi Fitur TF-IDF

Metode representasi teks dalam pemrosesan bahasa alami yang menghitung bobot kata berdasarkan frekuensinya.

```
FEATURE EXTRACTION - TF-IDF VECTORIZATION
=====
TF-IDF selesai!
Ukuran data (matrix): (774, 659)
- 774 komentar
- 659 fitur (kata unik)

Top 20 kata dengan bobot TF-IDF tertinggi:
1. bully: 47.89
2. in: 29.21
3. rest: 28.53
4. lucu: 27.64
5. dibully: 24.29
6. jahat: 23.61
7. peace: 20.94
8. padahal: 18.62
9. kenapa: 18.44
10. liat: 17.15
11. ceria: 15.26
12. hati: 13.72
13. tenang: 13.69
14. suara: 12.42
15. allah: 12.29
16. nang: 12.21
17. ngebully: 12.05
18. love: 11.29
19. baik: 11.20
20. pasti: 11.15
```

Gambar 6. Ekstraksi Fitur TF-IDF

Split Data

Melakukan pembagian *dataset* menjadi subset Data Training maupun Testing untuk mencegah *overfitting* dengan memastikan model menggeneralisasi pada data baru.

```
SPLIT DATA - TRAINING & TESTING
=====
Data berhasil dibagi!
Data Training: 619 komentar (80.0%)
Data Testing : 155 komentar (20.0%)

Distribusi sentimen di Training Set:
Sentiment
positif 378
netral 212
negatif 29
Name: count, dtype: int64

Distribusi sentimen di Testing Set:
Sentiment
positif 95
netral 53
negatif 7
Name: count, dtype: int64
```

Gambar 7. Proses Split Data

Evaluasi Model

Melakukan pengukuran performa model dengan menghitung akurasi, *precision*, *recall*, *F1-score*, menggunakan *confusion matrix* dan memvisualkan hasil grafiknya.

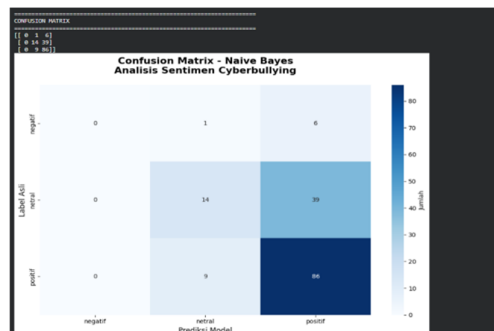
EVALUASI MODEL - HASIL PENGUJIAN

ACCURACY: 64.52%

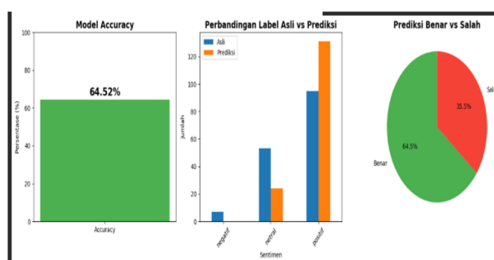
CLASSIFICATION REPORT:

	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	7
netral	0.58	0.26	0.36	53
positif	0.66	0.91	0.76	95
accuracy			0.65	155
macro avg	0.41	0.39	0.37	155
weighted avg	0.60	0.65	0.59	155

Gambar 8. Hasil Pengujian Evaluasi Model



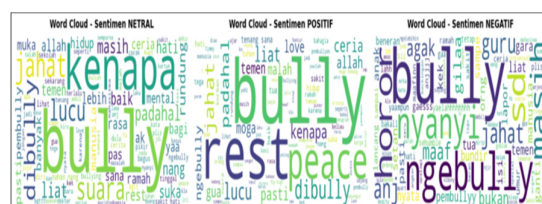
Gambar 9. Hasil *Confusion matrix*



Gambar 10. Visualisasi Hasil Pengujian

Analisis Word Cloud

Analisis *word cloud* berupa kata-kata yang sering muncul di setiap kelas sentimennya.



Gambar 11. Word Cloud

Word cloud per kelas mengungkap kata kunci dominan “*bullying*” dan “korban” pada negatif, “sadar” pada positif, dan “info” pada netral, memperkuat wawasan pola opini

masyarakat. Hasil ini menegaskan *naïve bayes* unggul untuk deteksi sentimen *cyberbullying* di sosial media.

D. PENUTUP

Pengujian ini dilaksanakan dengan mengadopsi *confusion matrix* guna mengukur performa klasifikasi sentimen pada 779 komentar TikTok yang telah di split data yakni 80% training maupun 20% testing. *Confusion matrix* menghasilkan evaluasi utama seperti akurasi, *precision*, *recall*, maupun *F1-score*, menunjukkan bahwa *naïve bayes* mampu membedakan sentimen positif, negatif, dan netral dengan baik, membuktikan bahwa *naïve bayes* efektif untuk penelitian sentimen ini.

E. DAFTAR PUSTAKA

- Angelina, S. J., Negara, A. B. P., & Muhardi, H. (2023). Analisis Pengaruh Penerapan Stopword Removal Pada Performa Klasifikasi Sentimen Tweet Bahasa Indonesia. *Jurnal Aplikasi Dan Riset Informatika*, 1(2), 165–173. <https://doi.org/10.26418/jari.v2i1.69680>
- Aser, F. G., Paramita, S., & Sudarto. (2022). Fenomena Cyberbullying di Media Sosial TikTok. *KIWARI: Jurnal Fakultas Ilmu Komunikasi Universitas Tarumanegara*, 1(3), 449–453. <https://doi.org/10.24912/ki.v1i3.15763>
- Baehaqi, F., & Cahyono, N. (2024). Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan Algoritma Naïve Bayes. *The Indonesian Journal of Computer Science*, 13(1), 1051–1063. <https://doi.org/10.33022/ijcs.v13i1.3301>
- Damayanti, A., Delima, I. D., & Suseno, A. (2023). Pemanfaatan Media Sosial Sebagai Media Informasi dan Publikasi (Studi Deskriptif Kualitatif pada Akun Instagram @rumahkimkotatangerang).

- PIKMA : Publikasi Ilmu Komunikasi Media Dan Cinema*, 6(1), 173–190.
<https://doi.org/10.24076/pikma.v6i1.1308>
- Khadafi, M. Al, Kartika, K. P., & Febrinita, F. (2022). Penerapan Metode Naive Bayes Classifier dan Lexicon Based Untuk Analisis Sentimen Cyberbullying Pada BPJS. *JATI : Jurnal Mahasiswa Teknik Informatika*, 6(2), 725–733.
<https://doi.org/10.36040/jati.v6i2.5633>
- Marlef, A., Masyhuri, M., & Muda, Y. (2024). Mengenal dan Mencegah Cyberbullying: Tantangan Dunia Digital. *Journal of Education Research*, 5(3), 4002–4010.
<https://doi.org/10.37985/jer.v5i3.1295>
- Nugraha, D., & Astuti, P. (2023). Analisis Sentimen Cyberbullying Pada Sosial Media Instagram Menggunakan Metode Support Vector Machine. *Information System for Educators and Professionals*, 8(2), 153–164.
<https://doi.org/10.51211/isbi.v8i2.2535>
- Putra, D. M. D. U., Suryawan, I. W. D., Andayani, S., & Rusadi, E. Y. (2025). Analisis Sentimen terhadap Fenomena Artis Berjualan di Tiktok Shop pada Media Sosial X Menggunakan Model Naïve Bayes. *JUSITIK : Jurnal Sistem Dan Teknologi Informasi Komunikasi*, 8(2), 108–113.
<https://doi.org/10.32524/jusitik.v8i2.1452>
- Suhartoyo, R., Julyo, V., Irsyad, H., & Rahman, A. (2025). Keyword Extraction of Scientific Journal Abstracts Using TF-IDF and KeyBERT Methods. *AICOMS : Applied Information Technology and Computer Science*, 4(2), 1–9.
<https://doi.org/10.58466/aicoms.v4i2.1814>
- Syahira, M. A., & Kurniawan, R. (2024). Analisis Sentimen Cyberbullying Pada Media Sosial X Menggunakan Metode Support Vector Machine. *Media Informatika Budidarma*, 8(3), 1724–1733.
<https://doi.org/10.30865/mib.v8i3.7926>
- Wibisono, B., Machmud, A., Suryani, N., & Yunita, Y. (2025). Analisis Sentimen Cyberbullying Pada Komentar X Menggunakan Metode Naïve Bayes. *Computer Science (CO-SCIENCE)*, 5(1), 8–16.
<https://doi.org/10.31294/coscience.v5i1.5152>
- Zulfiqri, R., Sari, B. N., & Padilah, T. N. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Media Sosial Instagram Pada Situs Google Play Store Menggunakan Naive Bayes Classifier. *JITET : Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3), 2965–2973.
<https://doi.org/10.23960/jitet.v12i3.4995>