

Analisis Sentimen Terhadap Bullying Di Indonesia Pada Twitter Menggunakan Naïve Bayes dan SVM

Abdu Malik AlHakim¹, Harun Leonardo², Alifia Nursyahrani Putri³

^{1,2,3}Informatika, Ilmu Komputer, Universitas Bhayangkara Jakarta Raya

Email : 202110715209@mhs.ubharajaya.ac.id, 202020115041@mhs.ubharajaya.ac.id, 202010225106@mhs.ubharajaya.ac.id

ABSTRAKSI

Perundungan atau yang lebih dikenal dengan istilah bullying, telah menjadi permasalahan sosial yang serius di Indonesia. Dalam beberapa tahun terakhir, kasus bullying semakin meningkat, terlebih di kalangan anak-anak dan remaja. Bullying dapat terjadi dimana saja, termasuk di rumah, lingkungan kerja, komunitas, media sosial, dan sekolah, namun paling sering terjadi justru di lingkungan pendidikan. Twitter atau “X” merupakan sosial media yang paling banyak digunakan di Indonesia, kerap menjadi wadah bagi masyarakat untuk mengutarakan opini mereka terhadap bullying. Penelitian ini bertujuan untuk menganalisis sentimen terhadap bullying di Indonesia melalui komentar atau *tweet* yang dikumpulkan dari twitter dengan menggunakan metode Naïve Bayes dan Support Vector Machine (SVM). Dari hasil analisis terhadap 330 *tweet*, metode Naïve Bayes menunjukkan akurasi sebesar 77,27%, sedangkan metode SVM menunjukkan akurasi sebesar 72,72%.

Kata Kunci: Perundungan, Bullying, Twitter, Naïve Bayes, Support Vector Machine (SVM)

ABSTRACT

Bullying has become a serious social problem in Indonesia. In the past few years, bullying cases are increasing, especially among children and adolescents. Bullying can occur anywhere, including at home, work, community, social media, and school, but it is most common in educational settings. Twitter or "X" is the most used social media in Indonesia, often a place for people to express their opinions on bullying. This research aims to analyze sentiment towards bullying in Indonesia through comments or tweets collected from Twitter using Naïve Bayes and Support Vector Machine (SVM) methods. From the analysis of 330 tweets, the Naïve Bayes method showed an accuracy of 77.27%, while the SVM method showed an accuracy of 72.72%.

Keywords: Bullying, Sentiment Analysis, Twitter, Naïve Bayes, Support Vector Machine (SVM)

Penulis Korespondensi

Abdu Malik AlHakim

Tanggal Submit : 16/06/2024

Tanggal Diterima : 17/07/2025

Tanggal Terbit : 25/07/2025

This is an open access article under the [CC-BY-NC-SA](#) license



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

I. PENDAHULUAN

Perundungan merupakan terjemahan dari kata *bullying* dalam Bahasa Inggris, sudah menjadi fenomena yang semakin mengkhawatirkan di berbagai negara, termasuk Indonesia. Bullying adalah jenis perilaku agresif yang di mana perilakunya bertujuan untuk menyakiti atau mengganggu, hal tersebut terjadi secara berulang dan terjadi ketimpangan kekuatan, di mana seseorang atau kelompok yang lebih kuat menyerang orang atau kelompok yang lebih lemah[1]. Bullying dapat terjadi di berbagai tempat, seperti di

lingkungan sekolah sampai media sosial, dan dapat berdampak serius pada kondisi mental para korbannya.

Dalam era digital saat ini, Twitter adalah salah satu platform media sosial yang paling sering diakses di Indonesia. Belum lama ini twitter bertransformasi menjadi “X”. Sebagian besar masyarakat menggunakan twitter untuk berbagi pendapat tentang berbagai topik, termasuk bullying. Komentar atau *tweet* dari para pengguna Twitter menunjukkan berbagai macam opini baik itu opini positif atau negatif dan sikap mereka terhadap isu bullying. Banyak pengguna yang

membagikan cerita tentang pengalaman mereka sendiri atau pengalaman teman mereka yang pernah menjadi korban perundungan. Kumpulan cerita ini menunjukkan berbagai perspektif dan sikap terhadap fenomena bullying, sekaligus menggambarkan dampak jangka panjang yang dialami para korban.

Penelitian ini bertujuan untuk mengeksplorasi penerapan metode Naïve Bayes dan Support Vector Machine (SVM) dalam menganalisis sentimen terhadap isu bullying di Indonesia, khususnya melalui data yang diperoleh dari twitter. Dengan demikian, diharapkan analisis sentimen ini dapat memberikan kontribusi dalam pemahaman terhadap fenomena bullying di Indonesia dan menjadi dasar dalam upaya pencegahan dan strategi penanggulangan yang lebih efektif di kemudian waktu.

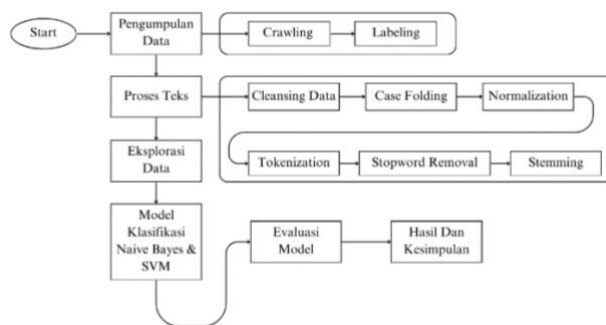
II. PENELITIAN YANG TERKAIT

Metode algoritma Naïve Bayes dan Support Vector Machine pada artikel terkait cukup banyak digunakan untuk melakukan analisa sentimen publik pada twitter. Salah satunya digunakan untuk menganalisa sentimen pada Boyband BTS dengan metode Naïve Bayes, Support Vector Machine (SVM). Hasil menunjukkan bahwa terdapat perbedaan tingkat akurasi pada Naïve Bayes yaitu sebesar 79%, sedangkan Support Vector Machine menghasilkan nilai akurasi yang lebih tinggi, yaitu sebesar 81%[2]. Artikel lain menunjukkan bahwa tingkat akurasi yang dihasilkan dengan metode metode SVM lebih unggul yaitu sebesar 88.8% sedangkan metode Naïve bayes menghasilkan tingkat akurasi sebesar 82.51%, yang di gunakan untuk menganalisis sentimen kebijakan vaksinasi Covid-19 pada twitter[3]. Dalam artikel lainnya yang menggunakan metode Naïve Bayes untuk menganalisa sentimen pengguna transportasi Jakarta terhadap transjakarta pada twitter menghasilkan tingkat akurasi 61.1% dan metode K-NN sebesar 75.7%[4].

III. METODE PENELITIAN

A. Tahapan Penelitian

Penelitian ini diawali dengan identifikasi, pengumpulan data dari Twitter dengan menggunakan data dari twitter, dimana data yang diperoleh sebanyak 330 data tweet dalam bentuk csv, yang selanjutnya proses pemberian label data positif dan negatif, melakukan pre-processing data, eksplorasi data dengan tampilan visualisasi, kemudian menguji algoritma Naïve bayes Support Vector Machine untuk menemukan hasil dan kesimpulan. Selanjutnya, langkah-langkah penelitian yang terlihat pada Gambar. 1



Gambar. 1 Tahapan Penelitian

B. Keyword

Pemilihan keyword atau kata kunci harus sesuai dengan topik bullying agar menghasilkan hasil yang relevan. Keyword perlu di tentukan karena untuk melakukan scraping pada twitter library membutuhkan keyword untuk memilih kata apa saja yang muncul dalam tweet yang akan di scraping, lalu data tersebut di masukan kedalam dataset. Adapun beberapa keyword yang di gunakan untuk proses scraping antara lain : “bully”, “kekerasan”, “siswa”, “perundungan”, “mahasiswa”, “anak”, “bullying”, “dibully”, “SMA”, “dikatain”, “dimaki”, “dimaki”.

C. Scraping

Twitter merupakan salah satu platform media sosial yang paling sering diakses di Indonesia. Kebanyakan masyarakat menggunakan platform tersebut untuk mengutarakan opini nya terhadap suatu hal atau fenomena. Namun platform membatasi jumlah karakter yang dapat diposting pengguna di setiap halaman profil pengguna maksimal 140 karakter[5].

Semua data yang dikumpulkan untuk penelitian ini di peroleh menggunakan teknik scraping pada Twitter. Scraping adalah teknik yang digunakan untuk mengumpulkan data dari internet secara otomatis[6]. Proses scraping ini menggunakan tools yang di jalankan pada node.js. Scraping data dilakukan berdasarkan pemilihan keyword, lalu dataset tweet yang masih mentah di ekspor dan di jadikan dalam bentuk csv file. dan bahasa yang dipakai dalam proses pengambilan data di twitter yaitu bahasa Indonesia.

D. Labeling (Pemberian Label)

Setelah proses pengumpulan data dengan scraping, tahap selanjutnya adalah memberikan label pada data tweet. setiap data tweet yang sudah di kumpulkan lalu diberikan label positif atau negatif berdasarkan sentimen yang diwakili oleh tweet atau komentar tersebut. tweet yang memiliki sentimen positif (baik) akan diberikan label positif, sementara komentar atau tweet yang memiliki sentimen negatif (buruk) akan diberikan label negatif. Proses labeling dilakukan secara manual.

E. Pra-Proses Data

Tujuan dari pre-processing data adalah untuk membersihkan data atau data cleaning, integrasi data, transformasi data, dan reduksi data[3]. Dataset yang diambil dari media sosial Twitter isinya masih tidak terstruktur karena masih memuat kata dan simbol mentah dari pengguna. Oleh karena itu di perlukan proses cleaning data dengan menggunakan beberapa aktivitas transformasi data seperti :

- Cleansing : atau pembersihan data dengan penghapusan mention, hashtag, karakter atau tanda baca, URL, dan karakter lain selain karakter[7].
- Case folding : mengubah semua data tweet menjadi huruf kecil. Proses ini dilakukan setelah tahapan cleansing selesai. Penggunaan huruf kapital dan huruf kecil yang tidak seragam banyak ditemukan dalam data komentar twitter.
- Tokenizing : adalah tahap pemisahan string input menjadi kata-kata penyusunnya[8]. pemisahan kata dilakukan dengan memisahkan kalimat menggunakan spasi [7].
- Stopword removal : proses menghilangkan kata-kata yang sering muncul pada data teks namun tidak mempunyai arti penting. misalnya terdapat kata sambung, kata depan, kata ganti,
- Stemming : Mengubah semua kata pada data teks menjadi kata dasar dengan menghapus awalan, sisipan, dan akhiran.

F. EDA

Exploratory Data Analysis (EDA) dilakukan pada data yang bertujuan untuk memahami sifat dan karakteristiknya, menemukan pola, mendeteksi anomali, mendapatkan wawasan, dan menerima umpan balik. Keputusan pemodelan akan dibuat berdasarkan temuan penelitian[9]. Dalam analisis ini ada beberapa aktivitas eksplorasi yang dilakukan. Yaitu visualisasi distribusi kelas sentimen, analisis kata yang sering muncul lalu di visualisasikan.

G. Naïve Bayes dan Support Vector Machine (SVM)

Algoritma Bayesian adalah algoritma statistik. Mereka dapat memprediksi probabilitas keanggotaan kelas, seperti probabilitas bahwa sampel tertentu termasuk dalam kelas tertentu. Klasifikasi Bayesian, seperti yang diuraikan dalam teorema Bayes, menjadi dasar dari pendekatan yang terstruktur. Penelitian yang membandingkan berbagai algoritma klasifikasi telah menemukan bahwa Naïve Bayes sederhana menunjukkan kinerja yang sebanding dengan algoritma pohon keputusan dan jaringan saraf tiruan[10].

Selain itu Naive Bayes memiliki waktu klasifikasi yang cukup singkat, sehingga mempercepat proses sistem analisis sentimen[11]. Dalam penelitian ini, data uji berupa tweet dari hasil scraping di twitter. Kemudian klasifikasi digunakan untuk mencari kategori yang

paling mungkin untuk dokumen baru yang belum pernah dilihat sebelumnya.

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space, problem klasifikasi dapat diterjemahkan dengan usaha menemukan garis (hyperplane) yang memisahkan antara kedua kelompok tersebut dan berbagai alternatif garis pemisah (discrimination boundaries) tersedia, Hyperplane pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin hyperplane tersebut dan mencari titik maksimalnya[12].

H. Pengujian dan Evaluasi

Setelah model dilatih, proses selanjutnya adalah uji untuk melihat seberapa baik performanya. Hasil prediksi ini kemudian digunakan untuk menyempurnakan model. Untuk mengevaluasi kinerja model menggunakan *confusion matrix* dan tabel akurasi untuk mengukur tingkat akurasinya. Setelah pengujian selesai, akan dibuat daftar prediksi kelas dari data pengujian. Kemudian, prediksi ini dibandingkan dengan data asli yang sebelumnya disembunyikan untuk melihat seberapa akurat model tersebut. Proses ini dilakukan secara otomatis di python, namun jika perhitungan manual perhitungan nilai *accuration*, *precision*, *recall*, rumus masing-masing perhitungannya seperti berikut ini [13]:

Rumus *accuration* (acc) : Merupakan rumus untuk menghitung rasio prediksi benar (True Positive dan True Negative) terhadap total data.

$$\text{acc} = \frac{TN + TP}{FN + FP + TN + TP} \quad (1)$$

Rumus *precision* (pre) : Merupakan rumus untuk menghitung rasio prediksi True Positive terhadap total data yang di prediksi positif.

$$\text{pre} = \frac{TP}{FP + TP} \quad (2)$$

Rumus *recall* (rec) : Merupakan rumus untuk menghitung rasio prediksi True Positive dibandingkan dengan keseluruhan data True Positive.

$$\text{rec} = \frac{TP}{FN + TP} \quad (3)$$

Bentuk dari *confusion matrix* terlihat pada Tabel 2

Tabel 1. Bentuk Confusion Matrix

	True Positive (1)	True Negative (0)
Pred. Pos (1)	TP	FP
Pred. Neg (0)	FN	TN

- TP: Prediksi *True Positive*, yaitu data sentimen positif yang telah diklasifikasikan dengan hasil positif.
- TN: Prediksi *True Negative*, yaitu data sentimen negatif yang telah diklasifikasikan dengan hasil negatif.
- FP: Prediksi *False Positive*, yaitu data sentimen positif yang telah diklasifikasikan namun dengan hasil negatif.

IV. HASIL DAN PEMBAHASAN

Dateset penelitian ini didapatkan dari media sosial *Twitter* atau “X”, Proses pengambilan data dilakukan menggunakan teknik *scraping*. Hasil *scraping* ini menghasilkan data yang masih sangat mentah sebanyak 330 data, oleh karena itu di perlukan beberapa tahapan lagi untuk bisa di analisis. Hasil dari *scraping* dapat terlihat pada Gambar. 2

timestamp	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id	username
1.78E+18	Tue Apr 16 06:20:00	20	kenapa nak kahwin dengan anak yatim	1.78E+18		in	ĐYtĐYt%	0	4	9	https://tw	1.28E+18	goodlucksofea
1.78E+18	Tue Apr 16 06:20:00	0	setelah ini aku gamau bully anak org lg	1.78E+18	https://pbs.twimg.c	in		0	0	0	https://tw	1.27E+18	tugangngigoo
1.78E+18	Tue Apr 09 06:38:00	38	semua ok maafin NON MOOTS DAN N	1.78E+18		in	in your wi	0	5	5	https://tw	1.27E+18	youbrghit
1.78E+18	Mon Apr 15 06:01:00	1	Orang tua yang suka pakai hukuman fis	1.78E+18		in	Somewhe	0	0	1	https://tw	1.93E+08	ArienFianne
1.78E+18	Tue Apr 16 06:22:00	22	Hype ohm yang ketika umur 13 pernah	1.78E+18		in	Riau, Indo	1	1	6	https://tw	1.71E+18	ChandraLiv31677
1.78E+18	Mon Apr 15 06:08:00	8	Sending my virtual hugs. Pernah ada d	1.78E+18	https://pbs.twimg.c	in		1	2	0	https://tw	18584336	Remasterise
1.78E+18	Sun Apr 14 06:50:00	50	cw tw // tiktok mention of bully aku g	1.78E+18	https://pbs.twimg.c	in		0	11	5	https://tw	1.28E+18	convomfs
1.78E+18	Wed Apr 10 06:03:00	3	Hehe ada anak baru bully dulu ga nih?	1.78E+18	https://pbs.twimg.c	in	Yogyakarta	0	2	0	https://tw	1.31E+18	anggasuljaya
1.78E+18	Mon Apr 15 06:01:00	1	Masih gak percaya teh anak ba tiktokei	1.78E+18		in	â tãj ... â-'	0	12	0	https://tw	1.50E+18	jutefk
1.78E+18	Sun Apr 14 06:02:00	2	POV selalu di bully anak any @chepyrf	1.78E+18	https://pbs.twimg.c	in		0	2	0	https://tw	8.68E+08	alyaaveyou
1.78E+18	Sun Apr 14 06:13:00	13	cw // tiktok bullying guys wdyt about	1.78E+18	https://pbs.twimg.c	in		1	10	0	https://tw	1.31E+18	convomf
1.78E+18	Wed Apr 10 06:00:00	0	@afiqazni @IANZAINAL @saythename	1.78E+18		in	Northern	0	1	0	https://tw	1.50E+18	RobinTripleX
1.78E+18	Tue Apr 16 06:08:00	8	Coba mikir org yg sempet di bully sm c	1.78E+18		in	kasi ĐYã	0	6	2	https://tw	1.71E+18	imsintoutyou
1.78E+18	Sat Apr 13 06:40:00	4089	Waktu habib Rizieq kena kasus pornog	1.78E+18		in	Indonesia	29	222	1512	https://tw	1.36E+18	Anak_Ogi
1.78E+18	Mon Apr 15 06:00:00	0	kwkk bocah bully anak mami disengg	1.78E+18		in	ur place	0	0	0	https://tw	8.83E+17	clumsyicha
1.78E+18	Sun Apr 14 06:00:00	0	Pnya ortu jg cmn ya wujud doang bully	1.78E+18		in	Isekai	0	1	0	https://tw	1.34E+18	fisikanjeng
1.78E+18	Wed Apr 17 06:00:00	0	Pernah di bully abis abisan perkara du	1.78E+18		in	Bogor	0	1	0	https://tw	3.15E+08	galerideone
1.78E+18	Wed Apr 10 06:00:00	0	hook anakku gw mandiin tiap hari ini s	1.78E+18		in		0	1	0	https://tw	1.59E+18	morowinic
1.78E+18	Sat Apr 13 06:28:00	28	Hot News pekan ini : 1. Bang Rhoma di	1.78E+18		in	Tegal, Jaw	1	2	3	https://tw	1.50E+18	Abahpurwa1
1.78E+18	Mon Apr 08 06:00:00	0	jangan bully anak kabupaten njink	1.78E+18		in	ĐYtĐYt,	0	0	0	https://tw	1.64E+18	pemmbaik
1.78E+18	Tue Apr 09 06:03:00	3	LOL... anak kicik ini bully???? Wkwkwk	1.78E+18	https://pbs.twimg.c	in	EXO PLAN	0	0	0	https://tw	1.22E+18	ZzarPark61
1.78E+18	Tue Apr 16 06:00:00	0	bully anak orang https://t.co/pygSIRHI	1.78E+18	https://pbs.twimg.c	in	Merantau	0	0	0	https://tw	1.32E+18	fauadiarka
1.78E+18	Sun Apr 14 06:00:00	0	moots indo kalian lewat fyp ga sih vide	1.78E+18		in		0	0	0	https://tw	1.75E+18	goldenbreezeze
1.78E+18	Sat Apr 13 06:00:00	0	ini darii tweet ini ya stop bully anak gw	1.78E+18		in	they 18	0	1	0	https://tw	1.24E+18	kzutr
1.78E+18	Sun Apr 14 06:00:00	0	@arriie39 @milkteajae Ohh buntut b	1.78E+18	https://pk.arriie39	in	She/her â	0	2	0	https://tw	1.15E+18	Minn00ngie
1.78E+18	Fri Apr 12 06:00:00	0	Bandel dan ngeselin. Gw suka nge bull	1.78E+18		in		0	0	0	https://tw	1.24E+18	akujalalaa
1.78E+18	Mon Apr 15 06:00:00	1	aku pernah jadi korban bully pas umur	1.78E+18		in	jawa Timu	0	1	0	https://tw	1.59E+18	wizone_zlka
1.78E+18	Mon Apr 15 06:00:00	6	ini curhatan anak yg sering di-bully. Ak	1.78E+18	https://pbs.twimg.c	in		0	1	0	https://tw	7.95E+17	Sammisoh
1.78E+18	Mon Apr 15 06:00:00	0	@ipangrock154 bully anak yatim	1.78E+18		in	ipangrock tr	0	1	0	https://tw	1.41E+18	yoursunfloww__
1.78E+18	Mon Apr 08 06:00:00	1	ini bacaan anak ga sih? edukasinya bag	1.78E+18	https://pbs.twimg.c	in	ĐYt%: 37:5	0	0	0	https://tw	9.29E+17	revellain
1.78E+18	Sat Apr 13 06:50:00	50	Sampe saat ini wajah Lily (bnyk yg bila	1.78E+18		in		0	3	0	https://tw	1.26E+09	overcontrolneed
1.78E+18	Wed Apr 10 06:13:00	13	Bakit si Duterte Kasi magaling sya mag	1.78E+18	https://pbs.twimg.c	in		2	2	4	https://tw	1.56E+18	Dutsfantanyl
1.78E+18	Mon Apr 15 06:00:00	0	@pacalminaa @twice_mixx @aerielc	1.78E+18		in	pacalminaa	0	2	0	https://tw	1.37E+18	ufabqgori
1.78E+18	Mon Apr 15 06:00:00	0	ada anak muda yg berusaha mengenal	1.78E+18		in		0	0	0	https://tw	1.74E+18	ajj_zero7
1.78E+18	Tue Apr 09 06:10:00	105	@Intinyadeh Kenapa anak pejabat nih	1.78E+18		in	intinyadel	0	0	9	https://tw	1.32E+18	apaluliat2ajg
1.78E+18	Wed Apr 17 06:00:00	0	@kosmo_online pada ibu bala punya	1.78E+18		in	kosmo_or	0	0	0	https://tw	1.73E+18	EJ777778888888
1.78E+18	Sun Apr 14 06:00:00	0	Anak pintar semoga kedepannya suks	1.78E+18		in		0	0	0	https://tw	1.54E+18	BaleendahP

Gambar. 2 Contoh gambar satu.

Comma Separated Values (.csv). Selanjutnya akan di pilih hanya *tweet* saja yang akan diberi label dan diolah.

Tabel 2. Pemberian Label

Tweet	Sentimen
@tanyarlfe Masih 16 tahun masih dibawah umur loh umur umur masih cari jati diri bawaannya pengen memberontak harusnya ya dibimbing sih kalo gamau anaknya makin ilang arah. Ni netizen juga ga ada otak bully anak dibawah umur	negatif
aku punya teman sejak SD. dulu kami sama-sama target perundungan. sempet salah paham lalu baikan. SMP pisah dia sering titip surat ke teman satu tempat les dia yang satu sekolah dengan aku. SMA bareng lagi kelas 11-12 satu kelas lagi. masih kontakan sampai lulus kuliah (cont.)	positif

Setelah memberikan label pada data, tahap selanjutnya adalah melakukan *pre processing* pada data *tweet*. Adapun di *pre processing* data ini di lakukan menggunakan bahasa *python* dengan aktivitas sebagai berikut :

a. Cleansing

Tabel 3. Cleansing Tweet

Tweet	Hasil Cleansing
Kenapa pelaku bully macam mario dandy layak dihukum mati? Karena selain kasus yang sama david dia juga masih ada 1 kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi https://t.co/gIxy*****	Kenapa pelaku bully macam mario dandy layak dihukum mati Karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi

b. Case Folding

Tabel 4. Case Folding Tweet

Tweet	Hasil Cleansing
Kenapa pelaku bully macam mario dandy layak dihukum mati Karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi	kenapa pelaku bully macam mario dandy layak dihukum mati karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi

c. Tokenize

Tabel 5. Tokenize Tweet

Tweet	Hasil Cleansing
kenapa pelaku bully macam mario dandy layak dihukum mati karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi	"kenapa" "pelaku" "bully" "macam" "mario" "dandy" "layak" "dihukum" "mati" "karena" "selain" "kasus" "yang" "sama" "david" "dia" "juga" "masih" "ada" "kasus" "hukum" "lagi" "yang" "sekarang" "kasusnya" "anyep" "ilang" "ditelan" "bumi"

d. Stop Removal

Tabel 6. Stop Removal Tweet

Tweet	Hasil Cleansing
kenapa pelaku bully macam mario dandy layak dihukum mati karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi	pelaku bully macam mario dandy layak dihukum mati selain kasus sama david masih ada kasus hukum lagi sekarang kasusnya anyep ilang ditelan bumi

e. Stemming

Tabel 6. Stop Removal Tweet

Tweet	Hasil Cleansing
pelaku bully macam mario dandy layak dihukum mati selain kasus sama david masih ada kasus hukum lagi sekarang kasusnya anyep ilang ditelan bumi	laku bully macam mario dandy layak hukum mati kasus david kasus hukum sekarang kasus anyep ilang tel bumi

Dalam tahap cleansing, data teks mentah dibersihkan dari elemen-elemen yang tidak relevan atau mengganggu analisis, seperti URL, tanda baca berlebih, dan karakter khusus. Sebagai contoh, kalimat “Kenapa pelaku bully macam mario dandy layak dihukum mati? Karena selain kasus yang sama david dia juga masih ada 1 kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi https://t.co/gIxy*****” diubah menjadi “Kenapa pelaku bully macam mario dandy layak dihukum mati Karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi”. Proses ini

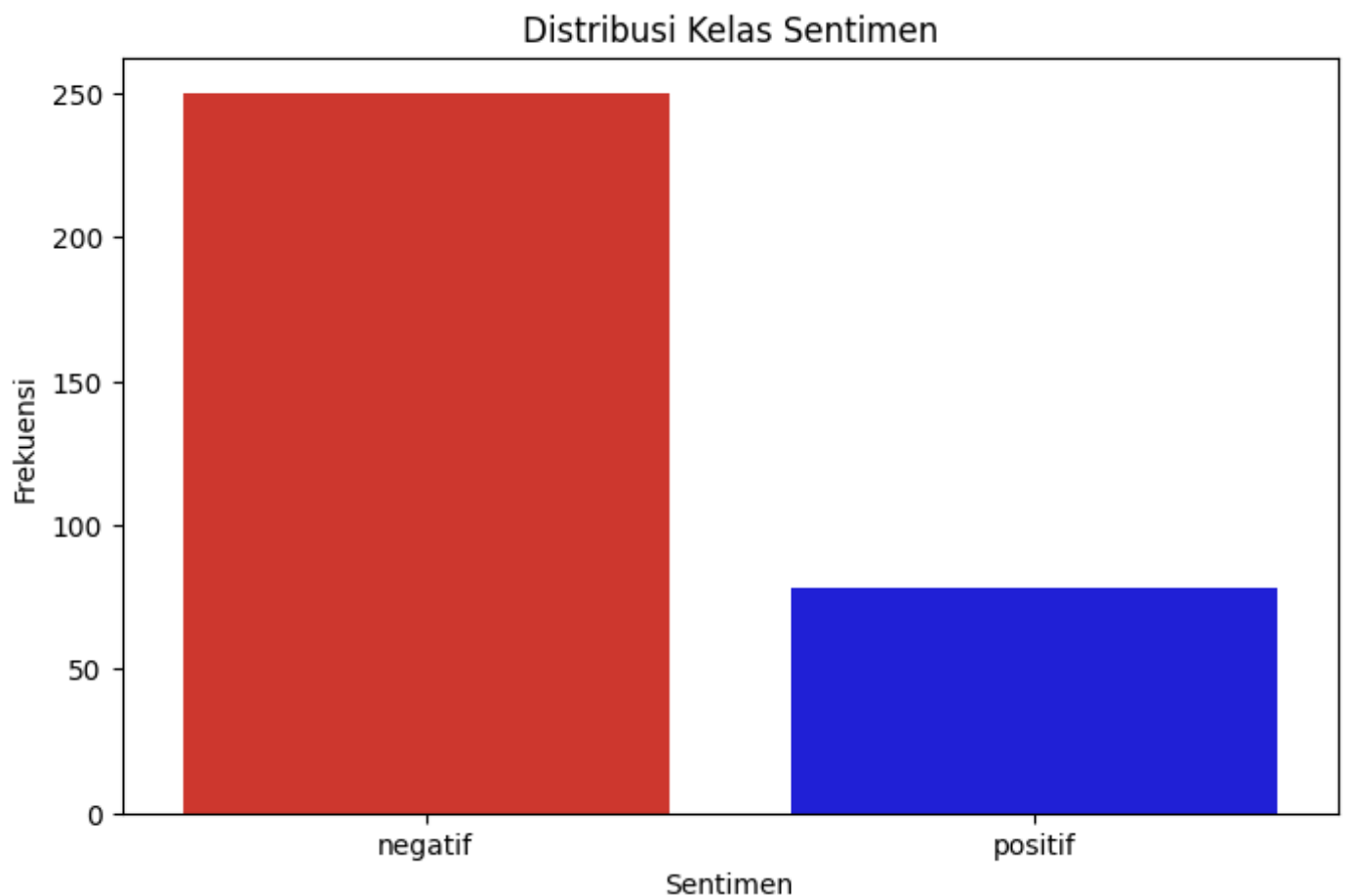
memastikan bahwa seluruh teks yang dianalisis hanya terdiri dari kata-kata bermakna, memudahkan langkah pemrosesan selanjutnya.

Setelah data bersih, langkah case folding menstandarkan huruf menjadi satu format, yakni semua huruf menjadi huruf kecil. Dengan demikian, variasi “Kenapa”, “kenapa”, dan “KENAPA” diperlakukan sama. Dari hasil cleansing sebelumnya, kalimat “Kenapa pelaku bully macam mario dandy layak dihukum mati karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi” diubah menjadi

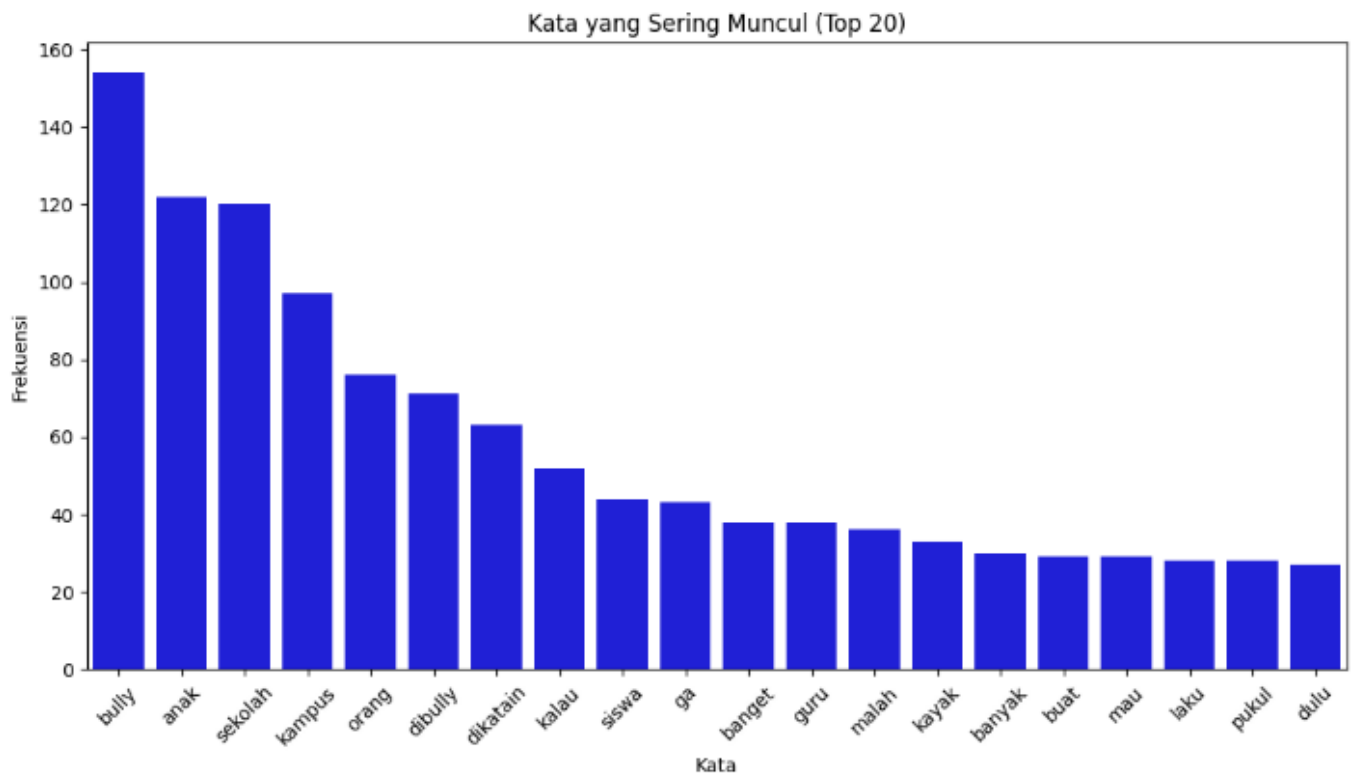
“kenapa pelaku bully macam mario dandy layak dihukum mati karena selain kasus yang sama david dia juga masih ada kasus hukum lagi yang sekarang kasusnya anyep ilang ditelan bumi”. Ini mencegah duplikasi entri berbasis perbedaan kapitalisasi.

Langkah tokenize memecah kalimat menjadi unit terkecil yaitu token (kata). Setiap kata diapit oleh tanda petik untuk menandakan pemisahan, contohnya “kenapa” “pelaku” “bully” dan seterusnya. Tokenisasi memudahkan algoritma untuk menghitung frekuensi kata, membangun representasi vektor, atau menerapkan metode pembobotan seperti TF-IDF. Hasil tokenisasi menjadi dasar pengolahan lebih lanjut, seperti filtering atau pembuatan model pembelajaran mesin.

Setelah tokenisasi, dilakukan stop removal untuk menghilangkan kata-kata fungsi (stop words) yang umum dan sedikit memberi informasi, misalnya “yang”, “dan”, “di”. Kalimat token yang panjang dipadatkan menjadi “pelaku bully macam mario dandy layak dihukum mati selain kasus sama david masih ada kasus hukum lagi sekarang kasusnya anyep ilang ditelan bumi”. Akhirnya, pada tahap stemming, setiap kata dikembalikan ke bentuk dasarnya, sehingga “layak” menjadi “lak”, “dihukum” menjadi “hukum”, dan “ditelan” menjadi “tel”. Proses ini mereduksi variasi bentuk kata sehingga model dapat lebih fokus pada makna inti tanpa terpengaruh turunan morfologis. Setelah dilakukan *pre processing* data, selanjutnya data akan di eksplorasi dalam visualisasi.



Gambar. 3 Distribusi Kelas Sentimen



Pada tahap akhir evaluasi, fokus diletakkan pada analisis kinerja model Naïve Bayes melalui Confusion Matrix (Gambar 6). Matriks ini memetakan jumlah prediksi benar positif (TP), benar negatif (TN), salah positif (FP), dan salah negatif (FN), sehingga memberikan gambaran mendetail tentang bagaimana model membedakan antara kelas positif dan negatif. Dengan melihat distribusi nilai pada keempat sel

matriks, peneliti dapat mengidentifikasi kecenderungan model—apakah lebih sering melakukan false alarm (FP) atau melewatkan kasus positif (FN).

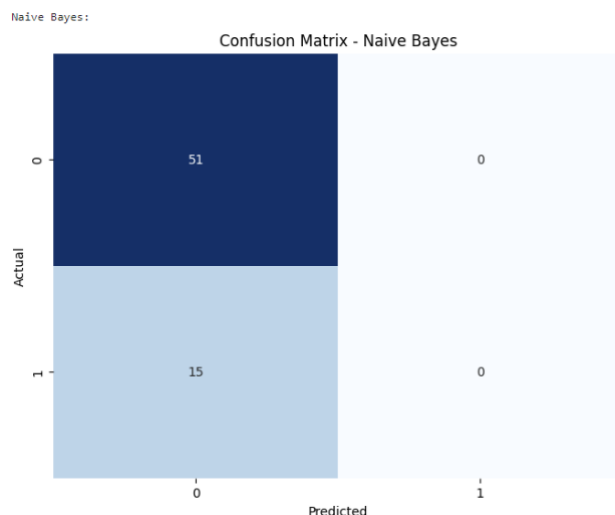
Berdasarkan Confusion Matrix tersebut, dihitung metrik utama yaitu Precision, Recall, F1-Score, dan akurasi (Gambar 7). Precision mengukur proporsi prediksi positif yang benar-benar akurat ($TP / (TP + FP)$), sedangkan Recall menunjukkan

sejauh mana model berhasil menangkap total kasus positif yang sebenarnya ($TP / (TP + FN)$). F1-Score merupakan harmonisasi antara Precision dan Recall, dihitung sebagai :

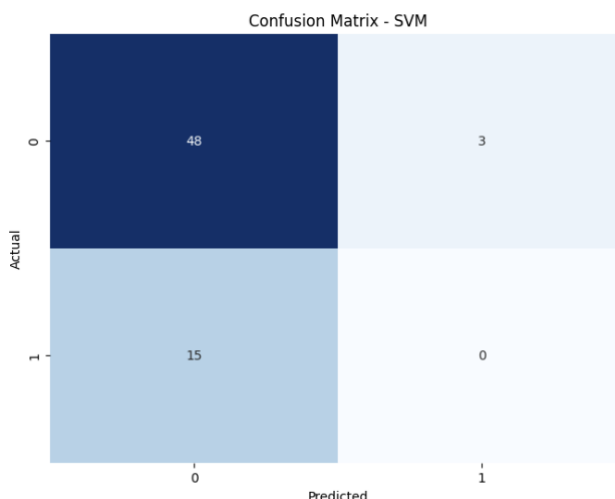
$$2 \times (Precision \times Recall) / (Precision + Recall)$$

Sehingga menyeimbangkan trade-off keduanya. Sementara itu, akurasi mengindikasikan persentase keseluruhan prediksi yang tepat ($(TP + TN) / (TP + TN + FP + FN)$).

Hasil dari perhitungan ini memberikan gambaran seberapa andal Naïve Bayes dalam konteks dataset yang digunakan. Nilai Precision tinggi menunjukkan sedikitnya kesalahan prediksi positif, sedangkan Recall tinggi menandakan model jarang melewatkan kasus sebenarnya. F1-Score yang seimbang menegaskan kinerja stabil antara kedua aspek tersebut, dan akurasi keseluruhan memperlihatkan efisiensi model dalam memproses semua instansi. Dengan demikian, evaluasi numerik ini membantu menentukan apakah Naïve Bayes sudah layak diterapkan atau perlu perbaikan lebih lanjut.



Gambar. 6 Confusion Matrix Naïve Bayes



Gambar. 6 Confusion Matrix Naïve Bayes

Naive Bayes Accuracy: 0.7727272727272727

Classification Report:				
	precision	recall	f1-score	support
negatif	0.77	1.00	0.87	51
positif	1.00	0.00	0.00	15
accuracy			0.77	66
macro avg	0.89	0.50	0.44	66
weighted avg	0.82	0.77	0.67	66

Gambar. 7 Perhitungan Klasifikasi Naïve Bayes

Proses klasifikasi Naive Bayes menggunakan fungsi MultinomialNB. Setelah melakukan pengujian akurasi, berdasarkan perhitungan akurasi yang dilakukan, didapatkan hasil bahwa model ini dapat memprediksi dengan akurasi MultinomialNB yaitu mencapai 77.27%, yang dapat dilihat pada Gambar 7.

SVM Accuracy: 0.7272727272727273

Classification Report:				
	precision	recall	f1-score	support
negatif	0.76	0.94	0.84	51
positif	0.00	0.00	0.00	15
accuracy			0.73	66
macro avg	0.38	0.47	0.42	66
weighted avg	0.59	0.73	0.65	66

Gambar. 8 Perhitungan Klasifikasi SVM

Proses klasifikasi SVM ini memakai metode kernel linear untuk melihat nilai akurasi yang didapatkan. Berdasarkan perhitungan akurasi yang dilakukan, didapatkan hasil bahwa SVM dapat memprediksi dengan akurasi yaitu mencapai 72.72%, yang dapat dilihat pada Gambar. 8

V. KESIMPULAN

Analisis sentimen terhadap bullying di Indonesia telah berhasil dilakukan dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM). Pengumpulan data dilakukan melalui Twitter, yang merupakan platform sosial media paling populer di Indonesia, untuk mengumpulkan opini masyarakat mengenai bullying. Dari 330 tweet yang dianalisis, metode Naïve Bayes menunjukkan akurasi sebesar 77.27%, sedangkan metode SVM menunjukkan akurasi sebesar 72.72%. Penelitian ini menunjukkan bahwa kedua algoritma tersebut cukup efektif dalam menganalisis sentimen, meskipun Naïve Bayes memberikan hasil akurasi yang lebih tinggi dibandingkan SVM.

DAFTAR PUSTAKA

- [1] S. Raquel Seixas, J. Pinto Coelho, and G. Nicolas Fischer, "Bullies, Victims And Bully-Victims: Impact On Health Profile," 2013.
- [2] R. Noviana and I. Rasal, "Penerapan Algoritma Naive Bayes Dan SVM Untuk Analisis Sentimen

- Boy Band BTS Pada Media Sosial Twitter,” *JTS*, vol. 2, no. 2, pp. 51–60, Jun. 2023.
- [3] Styawati, A. Rahman Isnain, N. Hendrastuty, and L. Andraini, “Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis,” vol. 6, no. 1, pp. 56–60, Jan. 2021.
 - [4] I. Iwandini, A. Triayudi, and G. Soepriyono, “Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor,” *Journal of Information System Research (JOSH)*, vol. 4, no. 2, pp. 543–550, Jan. 2023, doi: 10.47065/josh.v4i2.2937.
 - [5] N. Tri Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes, KNN, dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line,” *Jurnal IKRA-ITH Informatika*, vol. 3, no. 2, pp. 1–9, Jul. 2019.
 - [6] R. Lawson, *Web Scraping with Python: Scrape Data From Any Website With The Power Of Python*. Packt Publishing Ltd., 2015.
 - [7] I. Bagus Nyoman Wijana Manuaba, G. Rasben Dantes, and G. Indrawan, “Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine (SVM) Dengan Penambahan Algoritma Levenshtein Distance,” *Jurnal Sistem Komputer dan Kecerdasan Buatan*, vol. 5, no. 2, pp. 9–17, 2022.
 - [8] I. Fahrur Rozi, E. Nur Hamdana, and M. Balya Iqbal Alfahmi, “Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus SAMSAT Kota Malang),” *Jurnal Informatika Polinema*, pp. 149–154, 2018, [Online]. Available: <https://path.com/p/3pB4Qs>
 - [9] S. Yadav, A. Kaushik, and K. Mc Daid, “Exploratory Data Analysis on Code-mixed Misogynistic Comments,” Mar. 2024, [Online]. Available: <http://arxiv.org/abs/2403.09709>
 - [10] J. Han and M. Kamber, “Data Mining: Concepts And Techniques,” 2006.
 - [11] B. Gunawan, H. P. Sasty, and E. P. Esyudha, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” vol. 4, no. 2, pp. 17–29, 2018, [Online]. Available: www.femaledaily.com
 - [12] Aziz Muslim M *et al.*, “Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer,” 2019.
 - [13] E. Dwianto and M. Sadikin, “Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine,” *Jurnal Format*, vol. 10, no. 1, pp. 94–100, 2021.