



Klasifikasi Sentimen Masyarakat Terhadap Revisi Undang-Undang Tentara Nasional Indonesia Menggunakan Naïve Bayes Classifier

Abdul Haris Kurnia Sandi Harahap, Elin Haerani*, Lola Oktavia, Okfalisa, Fitra Kurnia

Fakultas Sains Dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹12050116127@students.uin-suska.ac.id, ^{2,*}elin.haerani@uin-suska.ac.id, ³lola.oktavia@uin-suska.ac.id,

⁴okfalisa@gmail.com, ⁵fitra.kurnia.hasbi@gmail.com

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Abstrak—Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI) menjadi wacana hangat dalam kebijakan publik di Indonesia dan menimbulkan kontroversi di kalangan masyarakat, karena kemunculannya yang mendadak dan tidak melalui proses perencanaan yang terbuka. Hal ini menimbulkan kekhawatiran terhadap potensi dominasi militer dan kembali aktifnya dwifungsi ABRI (Angkatan Bersenjata Republik Indonesia). Klasifikasi sentimen masyarakat terhadap RUU TNI menjadi fokus penelitian. Komentar dikategorikan ke dalam dua bentuk kelas sentimen, yaitu positif dan negatif. Adapun tahapan penelitian meliputi pengumpulan data, proses pelabelan sentimen, pembersihan data, penyeragaman teks menjadi huruf kecil, pemecahan kalimat atau dokumen menjadi bagian-bagian yang lebih kecil, normalisasi data teks, penanganan negasi, *Stopword Removal*, dan *stemming*, pembobotan menggunakan teknik TF-IDF, pembangunan model klasifikasi, hingga evaluasi kinerja model yang dihasilkan. Metode *Naïve Bayes Classifier* mengklasifikasikan 1547 data komentar yang diambil dari 2 akun media sosial Instagram. Model *Naïve Bayes Classifier* memperoleh akurasi sebesar 83,74%, precision 81,17%, recall 87,86%, dan F1-score 84,38%. Penelitian ini memiliki keterbatasan, antara lain data yang dikumpulkan masih terbatas. ketidakseimbangan jumlah data antar kategori sentimen, data hanya dari satu platform media sosial, serta belum optimalnya identifikasi terhadap sentimen positif dan negatif dan disarankan untuk penelitian selanjutnya supaya membandingkan dengan metode klasifikasi lainnya, memperluas dataset, memperluas ruang lingkup pengumpulan data dengan melibatkan berbagai platform media sosial dengan rentang waktu yang lebih luas, sehingga dapat memberikan gambaran pandangan masyarakat yang lebih menyeluruh, serta menguji kombinasi algoritma yang lebih luas. Studi ini dapat dijadikan sebagai indikator awal untuk evaluasi kebijakan secara cepat, dimana komentar positif atau negatif dari masyarakat di media sosial bisa menjadi masukan penting dalam meninjau efektivitas suatu kebijakan.

Kata Kunci: RUU TNI; Kebijakan Publik; Sentimen Masyarakat; Pra-Pemrosesan; Media Sosial; Naïve Bayes Classifier

Abstract—The revision of the Indonesian National Armed Forces Bill (RUU TNI) has become a hot topic in Indonesian public policy and has sparked controversy among the public due to its sudden emergence and lack of open planning process. This has raised concerns about the potential for military domination and the return of the dual function of the ABRI (Indonesian Armed Forces). The classification of public sentiment towards the RUU TNI is the focus of this study. Comments are categorized into two types of sentiment classes, namely positive and negative. The research stages include data collection, sentiment labeling, data cleaning, text normalization to lowercase letters, sentence or document segmentation into smaller parts, text data normalization, negation handling, stopword removal, and stemming, weighting using the TF-IDF technique, model classification development, and evaluation of the model's performance. The Naïve Bayes Classifier method classified 1,547 comment data points collected from two Instagram social media accounts. The Naïve Bayes Classifier model achieved an accuracy of 83.74%, precision of 81.17%, recall of 87.86%, and an F1-score of 84.38%. This study has limitations, including the limited amount of data collected. These include an imbalance in the amount of data between sentiment categories, data from only one social media platform, and the suboptimal identification of positive and negative sentiments. It is recommended that future research compare this method with other classification methods, expand the dataset, broaden the scope of data collection by involving various social media platforms over a wider time span, thereby providing a more comprehensive picture of public opinion, and test a wider range of algorithm combinations. This study can serve as an initial indicator for rapid policy evaluation, where positive or negative comments from the public on social media can provide important input in assessing the effectiveness of a policy.

Keywords: TNI Bill; Public Policy; Public Sentiment; Pre-Processing; Social Media; Naïve Bayes Classifier

1. PENDAHULUAN

Revisi Undang-Undang Tentara Republik Indonesia (RUU TNI) menjadi wacana hangat dalam kebijakan publik di Indonesia. Peralihan undang-undang ini diharapkan mampu merespons tantangan keamanan nasional yang semakin rumit serta menyesuaikan peran TNI dengan dinamika sosial dan politik yang berkembang. Namun, disisi lain peralihan ini juga menimbulkan kekhawatiran di kalangan masyarakat mengenai potensi yang dapat mengubah dinamika hubungan antara sipil dengan militer di Indonesia[1].

RUU TNI menuai banyak pro dan kontra disebabkan karena isi pembahasan dari RUU TNI tersebut dipandang mampu membangkitkan kembali fungsi ganda ABRI dan membahayakan mutu demokrasi. Dalam RUU TNI tersebut juga menimbulkan kekhawatiran dapat mengawali langkah untuk pasukan aktif untuk menempati jabatan sipil[2]

Kebijakan yang muncul dari proses demokrasi harus mencerminkan keinginan masyarakat dalam proses yang demokratis. Untuk mengetahui penerimaan publik terhadap kebijakan RUU TNI, penting untuk melihat sentimen publik terhadap RUU TNI. Analisis sentimen juga digunakan untuk memahami bagaimana publik bereaksi terhadap keputusan kebijakan yang berkaitan dengan keamanan dan pertahanan[3]. Sentimen publik dapat mencerminkan kepercayaan atau keraguan terhadap TNI, yang pada gilirannya dapat memengaruhi stabilitas politik dan keamanan nasional[4].

Seiring dengan pesatnya perkembangan teknologi, penyebaran informasi melalui media sosial berlangsung sangat cepat, memunculkan beragam respons dari masyarakat, baik yang bersifat positif maupun negatif. Untuk menganalisis



opini publik yang umumnya belum terstruktur, diperlukan pemrosesan teks guna memahami sudut pandang masyarakat[5].

Studi terbaru menunjukkan bahwa media sosial telah menjadi sarana penting untuk mengutarakan opini publik tentang kebijakan publik. Reaksi masyarakat terhadap kebijakan strategis, termasuk RUU TNI[6], dapat diperoleh dari Instagram, tempat jutaan unggahan diunggah setiap hari. Klasifikasi sentimen berbasis data adalah cara ilmiah untuk menilai reaksi masyarakat.

Penelitian terdahulu telah memanfaatkan berbagai media sosial untuk melakukan klasifikasi sentimen terhadap isu-isu kebijakan publik. Seperti yang ditunjukkan oleh penelitian Fathoni dkk. pada tahun 2025, algoritma Support Vector Machine (SVM) di platform X digunakan untuk menganalisis sentimen publik terkait kebijakan pemerintah (studi kasus: RUU TNI) dan menghasilkan akurasi sebesar 94,66%.[5]. Penelitian Anggaraini dkk pada tahun 2021 menggunakan algoritma Naïve Bayes Classifier (NBC) untuk menilai persepsi masyarakat tentang Kebijakan Kartu Pekerja Indonesia pada platform Twitter dan mencapai nilai akurasi 91.06%, presisi 90.35%, dan recall 90.65%.[7]. Penelitian Rabbani dkk pada tahun 2023 mengklasifikasikan sentimen masyarakat menggunakan algoritma SVM terkait eskalasi harga BBM pada platform Twitter dan memperoleh kenaikan kinerja algoritma sebesar 2% kernel polynomial pada splitting data 70:30 dan 80:20 [8]. Penelitian Wibowo dkk pada tahun 2022 menggunakan algoritma Naive Bayes Classifier (NBC) dalam analisis sentimen masyarakat terkait kenaikan Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 pada platform Twitter dan memperoleh hasil sebesar 88,5% di mana sentimen negatif mendapatkan nilai precision 12%, recall 17%, dan f1-score 14%, sedangkan untuk sentimen positif mendapatkan nilai precision 95%, recall 93%, dan f1-score 94%[9]. Penelitian Naraswati dkk pada tahun 2021 menggunakan algoritma Naive Bayes Classifier (NBC) dalam analisis sentimen masyarakat terkait Kebijakan Penanganan Covid-19 di Indonesia pada platform Twitter dan memperoleh tingkat akurasi klasifikasi sebesar 87,34%, sensitivitas sebesar 93,43%, dan spesifisitas 71,76%[10].

Teknik kecerdasan buatan seperti Text Mining semakin banyak digunakan di era digital untuk analisis sosial. *Naïve Bayes Classifier* merupakan algoritma yang paling umum diterapkan dalam klasifikasi sentimen. Algoritma ini dapat mengkategorikan opini ke dalam kategori positif, negatif, dan netral[11]. Algoritma Naïve Bayes Classifier, yang dikenal juga sebagai metode NBC adalah metode yang menggunakan peluang dan data pengukuran untuk menaksir probabilitas dalam klasifikasi.

Mengidentifikasi kelas terbaik, mengkategorikan, dan menentukan kemungkinan bahwa suatu kelas berisi setiap kelompok tertentu adalah tujuan dari metode NBC, yang memerlukan jumlah data latihan yang minimal untuk proses klasifikasi teks[12]. Salah satu keunggulan dari metode Naïve Bayes Classifier adalah bahwa itu memerlukan jumlah data latihan yang minimal.

Meskipun banyak studi yang meneliti dampak kebijakan militer, hanya sedikit yang meneliti respons masyarakat terhadap RUU TNI[13]. Banyak penelitian yang sering kali berfokus pada elemen kebijakan dan hukum daripada respon masyarakat. Selain itu, pendekatan analisis yang digunakan dalam penelitian sebelumnya sering kali bersifat kualitatif, seperti survei dan wawancara.

Naïve Bayes digunakan untuk mengklasifikasikan pendapat publik terkait RUU TNI yang merupakan tujuan dari penelitian. Metode ini dipakai karena memiliki keunggulan dibandingkan dengan metode atau algoritma lainnya. Diharapkan penelitian ini akan membantu dalam pengembangan analisis kebijakan publik berbasis data. Studi ini dapat dijadikan sebagai indikator awal untuk evaluasi kebijakan secara cepat, dimana komentar dukungan atau penolakan dari masyarakat pada media sosial bisa sebagai masukan utama dalam meninjau efektivitas suatu kebijakan.

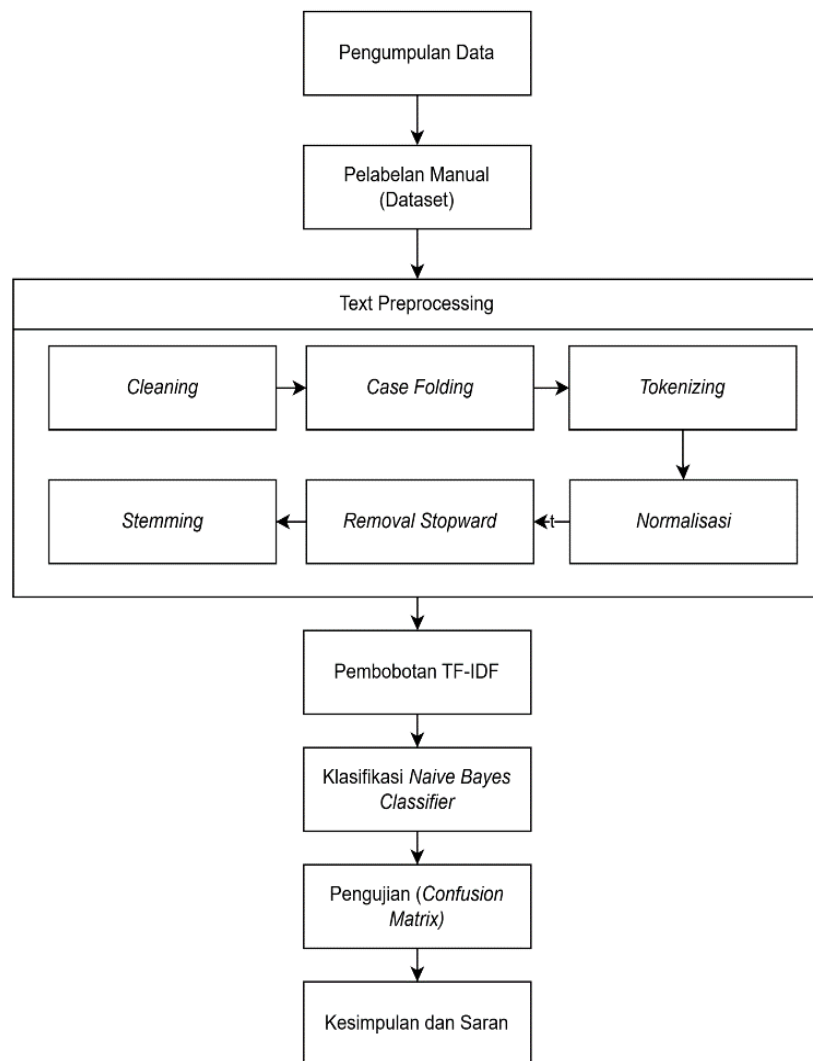
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dengan menggunakan data yang dikumpulkan dari platform media sosial Instagram, penelitian ini mencoba mengklasifikasikan sentimen publik terhadap Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI). Penelitian ini menggunakan algoritma *Naïve Bayes Classifier*, yang dikenal sangat efektif dalam memproses klasifikasi teks[14]. Fokus penelitian adalah sentimen masyarakat yang dikumpulkan dari kolom komentar pengguna Instagram mengenai RUU TNI. Pengumpulan data, pelabelan sentimen secara manual, *preprocessing*, pembobotan menggunakan metode TF-IDF, pembagian dataset, penerapan algoritma, evaluasi kinerja model, dan pengujian dengan *Confusion Matrix* adalah semua bagian dari proses penelitian. Setiap langkah direncanakan untuk memastikan bahwa data telah melalui proses transformasi yang memadai sebelum dianalisis melalui teknik klasifikasi.

Studi ini dimulai dengan pengumpulan data menggunakan alat Easycomment.ai untuk mengumpulkan komentar dari platform Instagram. Setelah data dikumpulkan, komentar dilabelkan secara manual berdasarkan isi, dengan kalimat positif dan negatif. Komentar yang mendukung kebijakan dilabelkan dengan label positif, sedangkan komentar yang menolak atau mengkritik dilabelkan dengan label negatif. Setelah pelabelan selesai, tahap selanjutnya adalah *preprocessing*. Tahap ini terdiri dari beberapa proses penting, seperti membersihkan data, menggabungkan case folding, tokenizing, menghapus stopword, normalisasi, dan stemming.

Proses ini bermaksud untuk membersihkan dan menstandarkan teks sehingga dapat untuk diteliti. Selanjutnya, metode TF-IDF digunakan untuk mengubah data yang telah diproses menjadi teks numerik dengan menghitung bobot kata. Data kemudian dibagi menjadi dua bagian, data latih dan data uji. Tahap berikutnya adalah pengujian menggunakan *Confusion Matrix*. Gambar 1 menunjukkan tahapan penelitian.



Gambar 1. Langkah Penelitian

2.2 Data Dikumpulkan

Dataset penelitian ini berbentuk teks sentimen dari komentar di unggahan akun Instagram resmi @detikcom dan @detiknewsofficial. Pengulangan yang dipilih adalah foto yang diunggah dari Maret 2025 hingga April 2025. Metode scraping digunakan untuk mengumpulkan dataset dengan bantuan alat easyscomment.ai. Scraping adalah proses mengambil data dari situs web[15].

2.3 Pelabelan Manual Dataset

Semua data yang dikumpulkan perlu dilabeli secara manual. Data divalidasi oleh ahli bahasa Indonesia. Komentar diklasifikasikan sebagai positif atau negatif berdasarkan isinya. Label positif diberikan untuk komentar yang mendukung kebijakan, seperti apresiasi, masukan konstruktif, atau respons positif lainnya. Sementara itu, label negatif diberikan untuk komentar yang menolak, mengkritik secara tidak konstruktif, atau berpotensi memicu perpecahan.

2.4 Text Preprocessing

Preprocessing data adalah fase di mana data teks dibersihkan dan disederhanakan menjadi data yang lebih terstruktur [16]. Proses mengubah data mentah menjadi data yang bersih dan siap untuk digunakan dalam proses klasifikasi dikenal sebagai *preprocessing* data.[17]. Mengurangi noise pada dataset merupakan fungsi dari *text preprocessing*[18]. Ada tujuh tahapan *text preprocessing* yang harus dilakukan

a. Cleaning

Cleaning adalah proses menghilangkan tanda baca dan karakter yang tidak diperlukan dari teks. Menghapus mention, hastag, emoticon, URL, tanda seru, tanda tanya, tanda koma, dan tanda titik adalah cara untuk menghindari gangguan selama proses klasifikasi.

b. Case Folding

Case Folding merupakan proses penyeragaman teks menjadi huruf kecil.



c. *Tokenizing*

Tokenizing adalah teknik yang menganalisis kumpulan kata dalam sebuah kalimat untuk membaginya menjadi bagian yang lebih kecil, seperti kata atau karakter.

d. *Normalisasi*

Normalisasi adalah proses yang digunakan untuk mengubah kata-kata yang tidak baku menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia

e. *Negation Handling*

Merupakan teknik yang berfungsi untuk mendeteksi kata yang mengalami negasi dan mengubahnya menjadi kata dengan makna yang berlawanan.

f. *Removal Stopword*

Removal Stopword adalah tahapan menghilangkan kata yang tidak penting dan sering muncul dalam analisis teks. Tahapan ini dilakukan untuk mempercepat pemrosesan teks dengan mengurangi jumlah kata yang harus diproses

g. *Stemming*

Untuk mempermudah pengolahan dokumen, stemming menghilangkan berbagai variasi kata yang berbeda dalam dokumen yang sama. Proses stemming menghilangkan imbuhan yang melekat pada kata dan menjadikannya kata dasar.

2.5 Pembobotan TF-IDF

Frequency Term (TF), yang menunjukkan berapa kali sebuah kata muncul dalam sebuah dokumen, dan *Inverse Document Frequency* (IDF), yang menunjukkan seberapa jarang sebuah kata muncul di seluruh dokumen, dapat digunakan untuk menentukan seberapa banyak token yang muncul dalam sebuah dokumen.[19]. Jumlah kata yang ada dalam dokumen ditunjukkan oleh frekuensi kata, sedangkan frekuensi balik dokumen menunjukkan jumlah kata yang muncul secara keseluruhan dalam dokumen[20]. Nilai TF suatu kata meningkat bersamaan dengan frekuensi kemunculannya, sedangkan nilai IDF suatu istilah meningkat seiring dengan frekuensi kemunculannya dalam sekumpulan dokumen[21]. Penelitian ini merubah teks menjadi vector menggunakan library TfidfVectorizer. Setiap kata dalam teks dievaluasi menurut bobotnya menggunakan algoritma TF-IDF, yang diimplementasikan oleh modul Scikit-learn TfidfVectorizer[22].

$$TF = \frac{\text{Jumlah Kemunculan Term Pada Dokumen}}{\text{Total Jumlah Kata Dalam Dokumen}} \quad (1)$$

$$IDF = \log \left(\frac{\text{jumlah dokumen dalam korpus}}{\text{jumlah dokumen yang mengandung term}} \right) \quad (2)$$

$$W_{ij} = TF_{ij} \times \log \left(\frac{D}{DF_j} \right) \quad (3)$$

Rumus TF-IDF untuk menilai seberapa penting sebuah kata dalam suatu dokumen dibandingkan dengan keseluruhan korpus. TF (Term Frequency) mengukur seberapa sering kata muncul dalam dokumen, sedangkan IDF (Inverse Document Frequency) menilai seberapa jarang kata tersebut muncul di seluruh dokumen. Dengan mengalikan TF dan IDF, diperoleh nilai W_{ij} , yaitu bobot dari kata i dalam dokumen j . Rumus ini membantu menyoroti kata-kata yang benar-benar relevan dan mengurangi pengaruh kata-kata yang terlalu umum, sehingga analisis teks menjadi lebih tajam dan bermakna.

2.6 Pembagian Dataset

Dataset dibagi menjadi dua bagian: data pelatihan dan data pengujian. Data pelatihan digunakan untuk membangun model dengan mempelajari pola-pola dari data yang tersedia, sementara data pengujian digunakan untuk mengevaluasi kinerja model terhadap data yang belum pernah dikenali sebelumnya. Dalam penelitian ini, tiga rasio pembagian digunakan: 90:10, 80:20, dan 70:30.

2.7 Klasifikasi Naïve Bayes Classifier

Metode *Naïve Bayes* menggunakan hipotesis bebas yang kuat untuk mengkategorikan data berdasarkan kemungkinan sederhana[23]. Metode Naïve Bayes untuk klasifikasi biasanya digunakan dengan pendekatan probabilitas[24]. Proses klasifikasi teks dengan menggunakan. Rumus untuk metode klasifikasi Pengklasifikasi Naïve Bayes ditunjukkan dalam persamaan (4)[25].

$$P(C|X) = \frac{P(X|C) P(C)}{P(X)} \quad (4)$$

Menggunakan Teorema Bayes untuk menghitung kemungkinan suatu data termasuk ke dalam kelompok tertentu, dimana $P(C|X)$ adalah probabilitas bahwa data X termasuk dalam kelompok C setelah melihat data tersebut (posterior). $P(X|C)$ menunjukkan seberapa besar kemungkinan data X muncul jika berasal dari kelompok C (likelihood). $P(C)$ adalah peluang awal data berada dalam kelompok C (prior), dan $P(X)$ adalah total kemungkinan munculnya data X (evidence).

2.8 Pengujian

Confusion Matrix adalah tabel yang menaksir kinerja klasifikasi dengan mengontraskan perkiraan model dengan nilai kelas sebenarnya.[26]. Confusion Matrix digunakan untuk menilai kinerja sistem dengan menaksir nilai akurasi, recall, precision, dan F1-Score[27]. Penggunaan matriks ini ditunjukkan pada tabel 1.

**Tabel 1.** Asumsi *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Negatif	Positif
Negatif	TN	FP
Positif	FN	TP

Berdasarkan Tabel 1, didapatkan empat nilai

a. Akurasi

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

b. Precision

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

c. Recall

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

d. F1-Score

$$F1 - Score = 2x \frac{Precision \times Recall}{Precision+Recall} \quad (8)$$

3. HASIL DAN PEMBAHASAN

3.1 Data Dikumpulkan

Dengan menggunakan command exporter dari website easycomment.ai, data penelitian ini dikumpulkan dari tiga akun instagram, @detiknewsofficial, @detik.com. Komentar komentar ini berasal dari postingan yang membahas RUU TNI. Baik pada saat mengumpulkan data maupun saat menyimpannya. Data tersedia dalam format CSV.

3.2 Pelabelan Manual Dataset

Seorang pakar linguistik yang mengajar Bahasa Indonesia memvalidasi proses pelabelan data secara manual. Dataset terdiri dari 1755 komentar yang siap digunakan untuk klasifikasi sentimen; 573 di antaranya dianggap sebagai sentimen positif dan 1182 lainnya dianggap sebagai sentimen negatif. Dataset dikategorikan menjadi kelas positif dan negatif tanpa adanya netral, dikarenakan topik penelitian ini berkaitan dengan kebijakan publik yang bersifat kontroversial, di mana respons masyarakat cenderung terbagi secara tegas antara mendukung dan menolak. Keberadaan komentar netral sangat minim dan memasukkan kelas netral justru berpotensi menciptakan ambiguitas dalam anotasi data dan mengganggu keseimbangan distribusi kelas dalam model klasifikasi. Kebijakan yang kontroversial sering kali memunculkan kekhawatiran, kecemasan, atau bahkan rasa ketidakadilan di tengah masyarakat. Emosi-emosi tersebut lebih mudah diekspresikan melalui komentar negatif karena sifatnya yang spontan dan emosional. Hasil pelabelan dan klasifikasi data ditunjukkan pada Gambar 2.

	A	B	C	D	E	F	G
1	Username	Text	Sentimen				
2	juanditojv	Ini di sahkan nya di dpr atau hotel	Positif				
3	__almarh	Waktunya revolusi	Negatif				
4	pipxyz	Guna mereka apa ya buat rakyat	Negatif				
5	hrynp	Gw jujur miris bgt dari ihsg turun kasus korupsi	Negatif				
6	moh.syukr	Sudahlah	Positif				
7	uchia_20	Komisi kontrol	Negatif				
8	jonhendri.j	HANYA TNI KOE,, YG BISA DI ANDAL KAN,,	Positif				
9	jonhendri.j	SEMOGA MENJADI LEBIH BAIK,, KARNA TNI,,	Positif				
10	to_desii	DPR gak guna!!	Negatif				
11	diandhisty	Apa itu prioritas ???? mereka menjadi ambigu	Positif				
12	herdi.willie	Untung gua ga ikut pemilu samsek	Negatif				
13	akudi2022	RUU PERAMPASAN ASET KORUPTOR KAPAN !	Positif				
14	jasonn_71	DPR itu wakil pemerintah ya, bukan wakil rak	Negatif				
15	babaspeot	Berapa harga per pasal?	Negatif				
16	syifaaafw	padahal rakyat banyak yg menolak loh???	Negatif				
17	akhmad10	RuU perampasan as3t mana	Positif				
18	ars_saktip	RAPAT RAPAT MULU HASILNYA BIKIN SENGSA	Negatif				
19	mayanajw	Konoha	Negatif				
20	tempe_sar	Bolehkan seseorang memasang bom ke temp	Negatif				
21	tommy_ta	Ada juga yg lebih penting percepat di sah kan	Negatif				
22	komang97	UU PERAMPASAN ASET KALIAN LUPAKAN... F	Negatif				
23	firdaws_ar	TIDAK BAKALAN ADA UU YG MENGUNTUNG	Negatif				
24	khoerulale	Lebih penting RUU PERAMPASAN ASET KORU	Positif				
25	mu_mu.ol	#NIPERADILANUMUM	Negatif				
26	dadit_101	Kok RUU Perampasan Aset enggak di bahas-b	Positif				
27	nurb_brun	udah ga usah tanya UU perampasan aset gmr	Negatif				

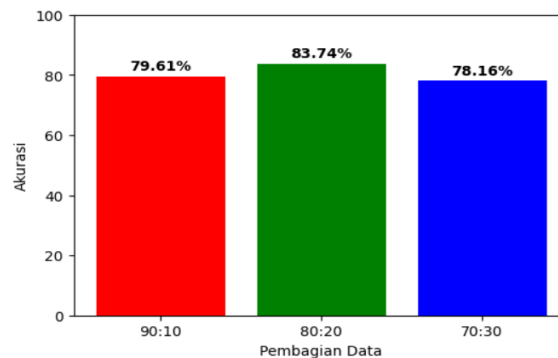
Gambar 2. Pelabelan Manual Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	aaaissuuuuda	aahhh	abad	abadi	abab	abai	abang	abdi	abri	absen	acak	account	aceh	aco	acungin	ada	adaaa	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	10	0	0	0	0	0	0,46895	0	0	0	0	0	0	0	0	0	0	0
13	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 5. Frame Data Pembobotan TF-IDF

3.5 Klasifikasi Naïve Bayes Classifier

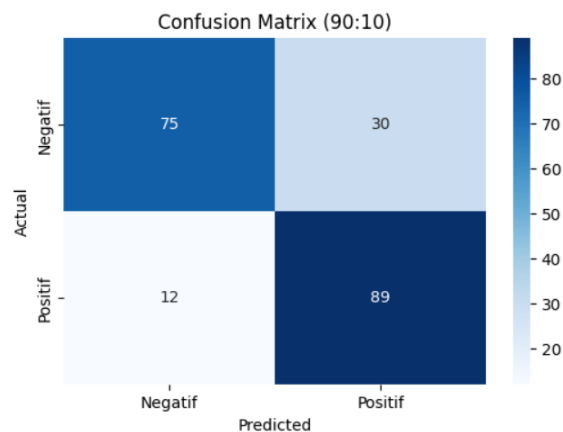
Menggunakan strategi berbagi data yang berbeda, tiga uji coba klasifikasi dilakukan menggunakan Naïve Bayes Classifier. Akurasi maksimum sebesar 83,74% diperoleh pada uji coba ke 2, yang menggunakan 80% data training dan 20% data uji. Secara empiris, banyak studi juga menunjukkan bahwa pembagian data 80:20 merupakan praktik umum yang menghasilkan performa optimal pada banyak kasus klasifikasi teks, seperti pada penelitian Wibowo dkk pada tahun 2022, hasil pengujian pada Algoritma Naïve Bayes Classifier mendapatkan nilai akurasi sebesar 88,5% [9].



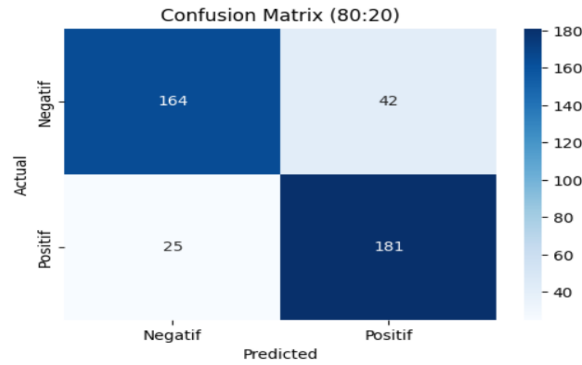
Gambar 6. Hasil Klasifikasi Naïve Bayes Pada Setiap Percobaan

3.6 Pengujian

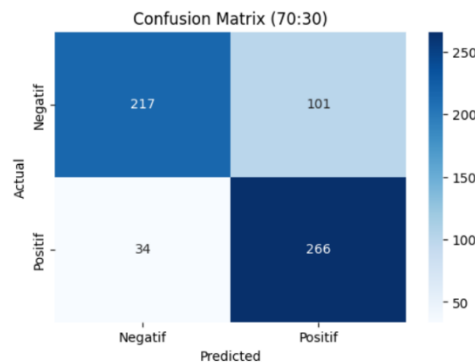
Untuk mengetahui seberapa efektif algoritma yang digunakan, 1547 titik data dibagi menjadi data latih dan data uji. Data dibagi menjadi tiga skenario. Pertama, 90% data latih dan 10% data uji digunakan; kedua, 80% data latih dan 20% data uji digunakan; dan ketiga, 70% data latih dan 30% data uji digunakan. Gambar 7, 8, dan 9 menunjukkan hasil pemeriksaan.



Gambar 7. Confusion Matrix Pada Data 90:10



Gambar 8. *Confusion Matrix* Pada Data 80:20



Gambar 9. *Confusion Matrix* Pada Data 70:30

Berdasarkan Gambar 8 hasil *Confusion Matrix* pada pembagian data latih 80% dan data uji 20%, didapatkan

a. *Accuracy*

TP = 89; FP = 30; FN = 12; TN = 75.

Menggunakan rumus (5), didapatkan nilai $(89 + 75 = 164) / (89 + 30 + 12 + 75 = 206) = 0,796$

b. *Precision*

TP = 89; FP = 30

Menggunakan rumus (6), didapatkan nilai $(89) / (89 + 30) = 0,747$

c. *Recall*

TP = 89; FN = 12

Menggunakan rumus (7), didapatkan nilai $(89) / (89 + 12) = 0,881$

d. *F1-Score*

Recall = 0,881; Precision = 0,747

Menggunakan rumus (8), didapatkan nilai $2 (0,881 * 0,747 = 0,658) / (0,881 + 0,747 = 1,628) = 0,808$.

Berdasarkan Gambar 8 hasil *Confusion Matrix* pada pembagian data latih 80% dan data uji 20%, didapatkan

a. *Accuracy*

TP = 181; FP = 42; FN = 25; TN = 164.

Menggunakan rumus (5), didapatkan nilai $(181 + 164 = 345) / (181 + 25 + 42 + 164 = 412) = 0,837$

b. *Precision*

TP = 181; FP = 42

Menggunakan rumus (6), didapatkan nilai $(181) / (181 + 42) = 0,811$

c. *Recall*

TP = 181; FN = 25

Menggunakan rumus (7), didapatkan nilai $(181) / (181 + 25) = 0,878$

d. *F1-Score*

Recall = 0,878; Precision = 0,811

Menggunakan rumus (7), didapatkan nilai $2 (0,811 * 0,878 = 0,712) / (0,811 + 0,878 = 1,689) = 0,843$

Berdasarkan Gambar 9 hasil *Confusion Matrix* pada pembagian data latih 70% dan data uji 30%, didapatkan

a. *Accuracy*

TP = 266; FP = 101; FN = 34; TN = 217.

Menggunakan rumus (5), didapatkan nilai $(266 + 217 = 483) / (266 + 34 + 101 + 217 = 618) = 0,781$

b. *Precision*

TP = 266; FP = 101



Menggunakan rumus (6), didapatkan nilai $(266) / (266 + 101) = 0,724$

c. Recall

TP = 266; FN = 34

Menggunakan rumus (7), didapatkan nilai $(266) / (266+34) = 0,866$

d. F1-Score

Recall = 0,866; Precision = 0,724

Menggunakan rumus (7), didapatkan nilai $2 (0,724 * 0,866 = 0,626) / (0,724 + 0,866 = 1,590) = 0,787$

Tabel menampilkan hasil uji coba pada seluruh skema pembagian data, yang diperoleh dari item dalam Confusion Matriks (Gambar 7, 8, dan 9).

Tabel 3. Hasil Uji Coba

Pembagian Data	Akurasi	Precision	Recall	F1-Score
90%:10%	79,61%	74,79%	88,12%	80,91%
80%:20%	83, 74%	81,17%	87,86%	84,38%
70%:30%	78,16%	72,48%	88,67%	79,76%

Berdasarkan Tabel 3, skenario pembagian 80%:20% memiliki rasio yang paling seimbang. Nilai *recall* lebih tinggi daripada nilai *precision*. Model berhasil menangkap semua komentar yang diprediksi negatif. Namun, masih salah untuk menganggap beberapa komentar yang positif sebagai negatif. Ini disebabkan oleh fakta bahwa model Naive Bayes menghitung kemungkinan berdasarkan frekuensi kata yang ada di setiap kelas. Model mungkin salah mendeteksi kata-kata yang sering muncul di kelas negatif dan juga muncul di komentar positif, terutama jika data negatif lebih dominan atau kosakatanya lebih variatif.

4. KESIMPULAN

Penelitian ini mengklasifikasikan 1.547 komentar Instagram dari jumlah total 1755 komentar yang dikumpulkan selama proses *scrapping* terkait Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI), terdiri dari 517 komentar positif dan 1030 komentar negatif. Data dikumpulkan dari 2 akun media sosial berita. Pengujian ini mengimplementasikan metode NBC (Naïve Bayes Classifier) menggunakan rancangan pembagian data 80% untuk data latih dan 20% untuk data uji. Rancangan pembagian 80%:20% memiliki rasio yang paling seimbang. Model berhasil menangkap semua komentar yang diprediksi negatif. Namun, masih salah untuk menganggap beberapa komentar yang positif sebagai negatif. Ini disebabkan oleh fakta bahwa model *Naive Bayes* menghitung kemungkinan berdasarkan frekuensi kata yang ada di setiap kelas. Model mungkin salah mendeteksi kata-kata yang sering muncul di kelas negatif dan juga muncul di komentar positif, terutama jika data negatif lebih dominan atau kosakatanya lebih variatif. Hasil pengujian menunjukkan bahwa model ini mencapai akurasi sebesar 83,74%, precision 81,17%, recall 87,86%, dan F1-score 84,38%. Penelitian ini memiliki keterbatasan, antara lain data yang dikumpulkan masih terbatas. ketidakseimbangan jumlah data antar kategori sentimen, data hanya dari satu platform media sosial, serta belum optimalnya identifikasi terhadap sentimen positif dan negatif dan disarankan untuk penelitian selanjutnya supaya membandingkan dengan metode klasifikasi lainnya, memperluas dataset, memperluas ruang lingkup pengumpulan data dengan melibatkan berbagai platform media sosial dengan rentang waktu yang lebih luas, sehingga dapat memberikan gambaran pandangan masyarakat yang lebih menyeluruh, serta menguji kombinasi algoritma yang lebih luas. Studi ini dapat dijadikan sebagai indikator awal untuk evaluasi kebijakan secara cepat, dimana komentar positif atau negatif dari masyarakat di media sosial bisa menjadi masukan penting dalam meninjau efektivitas suatu kebijakan.

REFERENCES

- [1] Y. Nur Fathoni, A. Daffa Ahmad, J. Benrivo Sipayung, R. Rau, atur Rushufah, and J. Indrawan, "Analisis Implikasi Pengesahan RUU TNI Dalam Sinergi Sipil-Militer dan Konsolidasi Demokrasi di Indonesia," *Global Komunika*, vol. 8, no. 1, pp. 3–15, May 2025.
- [2] M. Z. Habibie *et al.*, "Yayasan Lembaga Bantuan Hukum Indonesia (YLBHI) Menolak Upaya Menghidupkan Kembali Dwi Fungsi Melalui Revisi Undang-Undang TNI (Tentara Negara Indonesia)," *Quantum Juris: Jurnal Hukum Modern*, vol. 07, no. 2, pp. 157–168, Apr. 2025.
- [3] M. Tomz, J. L.P. Weeks, and Y.-M. Keren, "Public Opinion and Decisions About Military Force in Democracies," *Int Organ*, vol. 74, no. 1, pp. 1–47, Jan. 2020.
- [4] T. B. Silitonga, "Tantangan globalisasi, peran negara, dan implikasinya terhadap aktualisasi nilai-nilai ideologi negara," *Jurnal Civics: Media Kajian Kewarganegaraan*, vol. 17, no. 1, pp. 15–28, May 2020, doi: 10.21831/jc.v17i1.29271.
- [5] F. Fathoni, A. Ibrahim, F. Rizka Mumtaz, M. Azmi Zaky, M. Jodi Pratama, and I. Akbar Kurniawan, "Analisis Sentimen Public Twitter Terhadap Kebijakan Pemerintah Menggunakan Metode SVM (Studi Kasus: RUU TNI)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6322–6329, May 2025, doi: 10.36040/jati.v9i4.14036.
- [6] E. Sjarief, "Sikap Politik Anggota DPR Periode 1999-2004 Dalam Proses Pembentukan Undang-Undang (Studi Analisis Isi Proses Pembentukan Undang-Undang Nomor 20 Tahun 2002 Tentang Ketenagalistrikan)," *Jurnal Mitra Manajemen*, vol. 5, no. 2, pp. 1–18, Oct. 2020.
- [7] W. P. Anggraini and M. S. Utami, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Pekerja Di Indonesia," *Faktor Exacta*, vol. 13, no. 4, p. 255, Feb. 2021, doi: 10.30998/faktorexacta.v13i4.7964.



- [8] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.
- [9] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, "Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier," *Jurnal Asimetrik: Jurnal Ilmiah Rekayasa & Inovasi*, pp. 239–248, Jul. 2022, doi: 10.35814/asiimetrik.v4i1.3577.
- [10] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *SISTEMASI*, vol. 10, no. 1, p. 222, Jan. 2021, doi: 10.32520/stmsi.v10i1.1179.
- [11] N. Nopan and E. Mailoa, "Perbandingan Beberapa Algoritma Machine Learning Dalam Analisis Sentimen Terkait Pemilihan Presiden RI 2024," *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 13, no. 2, p. 1096, Aug. 2024, doi: 10.35889/jutisi.v13i2.1980.
- [12] A. Mondaref Jon and I. Vitra Paputungan, "Analisis Sentimen Pada Media Sosial Instagram Klub Persija Jakarta Menggunakan Metode Naive Bayes," *AUTOMATA*, vol. 4, no. 1, pp. 1–8, Feb. 2023.
- [13] H. Siagian *et al.*, "Strategi Peran TNI-POLRI Guna Mendukung Transformasi Ekonomi Yang Inklusif Dalam Rangka Percepatan Pemulihan Ekonomi Nasional," vol. 13, no. 1, pp. 81–88, Jan. 2025, doi: 10.37081/ed.v13i1.6473.
- [14] Z. Annisa and B. S. S. Ulama, "Analisis Sentimen Data Ulasan Pengguna Aplikasi 'PeduliLindungi' pada Google Play Store Menggunakan Metode Naïve Bayes Classifier Model Multinomial," *Jurnal Sains dan Seni ITS*, vol. 11, no. 6, May 2023, doi: 10.12962/j23373520.v11i6.94064.
- [15] I. Ghozali and W. H. Sugiharto, "Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1340–1348, Jun. 2023, doi: 10.30865/klik.v3i6.936.
- [16] E. F. Saraswita, D. P. Rini, and A. Abdiansah, "Analisis Sentimen E-Wallet di Twitter Menggunakan Support Vector Machine dan Recursive Feature Elimination," *Jurnal Media Informatika Budidarma*, vol. 5, no. 4, p. 1195, Oct. 2021, doi: 10.30865/mib.v5i4.3118.
- [17] S. W. Ritonga, . Y., M. Fikry, and E. P. Cynthia, "Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naïve Bayes Classifier," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3535.
- [18] P. Pandunata, C. K. Ananta, and Y. Nurdiansyah, "Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naïve Bayes Classifier," *INFORMAL: Informatics Journal*, vol. 7, no. 2, p. 146, Aug. 2022, doi: 10.19184/isj.v7i2.33928.
- [19] O. N. Julianti, N. Suarna, and W. Prihartono, "Penerapan Natural Language Processing Pada Analisis Sentimen Judi Online Di Media Sosial Twitter," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2936–2941, May 2024, doi: 10.36040/jati.v8i3.9613.
- [20] A. Atikah Putri, S. Agustian, and R. Abdillah, "Penerapan Metode Logistic Regression Untuk Klasifikasi Sentimen Pada Dataset Twitter Terbatas," *Jurnal Sistem Informasi*, vol. 7, no. 1, pp. 95–107, Jan. 2025.
- [21] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, Dec. 2023, doi: 10.57152/malcom.v4i1.983.
- [22] M. F. Lazuardi, R. Hiunarto, K. F. Ramadhani, N. Noviandi, R. Widayanti, and M. H. Arfian, "Hoax News Detection Using Passive Aggressive Classifier And TfIdfVectorizer," *JURNAL TEKNIK INFORMATIKA*, vol. 16, no. 2, pp. 185–193, Dec. 2023, doi: 10.15408/jti.v16i2.34084.
- [23] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, Sep. 2021.
- [24] E. M. Sipayung, H. Maharani, and I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *Jurnal Sistem Informasi (JSI)*, vol. 8, no. 1, pp. 2355–4614, Apr. 2020.
- [25] H. P. Doloksaribu and Yusran Timur Samuel, "Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi," *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 16, no. 1, pp. 1–11, Jan. 2022, doi: 10.47111/jti.v16i1.3747.
- [26] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *Jurnal Ilmiah Computer Science*, vol. 2, no. 2, 2024, doi: 10.58602/jics.v2i2.24.
- [27] M. Murni, I. Riadi, and A. Fadlil, "Analisis Sentimen HateSpeech pada Pengguna Layanan Twitter dengan Metode Naïve Bayes Classifier (NBC)," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 566, Apr. 2023, doi: 10.30865/jurikom.v10i2.5984.