Analisis Sentimen Publik Terhadap Isu RUU TNI pada Media Sosial X Menggunakan Algoritma Support Vector Machine dan Bidirectional Gated Recurrent Unit

Nurul Wijayanti Asyahiro
Fakultas Informatika
Universitas Telkom
Bandung
nurulwa@telkomuniversity.ac.id

Fitriyani
Fakultas Informatika
Universitas Telkom
Bandung
fitriyani@telkomuniversity.ac.id

***Abstrak* — Perkembangan teknologi digital telah mengubah pola komunikasi publik, dengan media sosial menjadi salah satu tempat dalam penyampaian opini masyarakat terhadap isu kebijakan publik, termasuk Rancangan Undang-Undang Tentara Nasional Indonesia (RUU TNI). Penyampaian opini dan pendapat yang berlangsung di *platform* X menunjukkan keberagaman sikap masyarakat, yang dapat diklasifikasikan ke dalam sentimen positif, negatif, dan netral. Namun, karakteristik bahasa yang tidak baku, penggunaan ragam informal, serta struktur kalimat yang variatif sering menjadi hambatan dalam proses klasifikasi sentimen. Untuk menjawab tantangan tersebut, penelitian ini menerapkan dua pendekatan pemodelan, yakni Bidirectional Gated Recurrent Unit (Bi-GRU) yang dipadukan dengan ekstraksi fitur FastText, serta Support Vector Machine (SVM) yang menggunakan TF-IDF sebagai representasi fitur. Data penelitian diperoleh dari *platform* X dengan total 7.198 cuitan yang dikumpulkan melalui API menggunakan kata kunci relevan. Selanjutnya, data dipraproses dan diberi label menggunakan model IndoRoBERTA untuk mengidentifikasi sentimen positif, negatif, dan netral. *Dataset* hasil pelabelan kemudian dibagi menjadi data latih dan data uji. Hasil eksperimen menunjukkan bahwa model Bi-GRU mencapai akurasi tertinggi sebesar 80,76%, sementara SVM dengan TF-IDF memperoleh performa terbaik pada skenario *baseline* dengan akurasi 79,93% dan *F1-score* tertinggi pada skenario *class weight* sebesar 0,6870.**

***Kata kunci*— Analisis Sentimen, RUU TNI, Support Vector Machine, Bidirectional Gated Recurrent Unit**

I. PENDAHULUAN

Perkembangan teknologi informasi pada era digital telah mengubah cara masyarakat menyampaikan opini, khususnya melalui media sosial yang berfungsi sebagai ruang publik untuk menyampaikan kritik, dukungan, dan pandangan terhadap kebijakan pemerintah serta isu politik secara cepat [1]. Salah satu isu yang banyak mendapat perhatian publik adalah Rancangan Undang-Undang Tentara Nasional Indonesia (RUU TNI), terutama terkait perluasan peran TNI di luar fungsi pertahanan sebagaimana diatur dalam Undang-Undang Nomor 34 Tahun 2004 tentang TNI, yang memunculkan kekhawatiran akan potensi kembalinya praktik dwifungsi ABRI seperti pada masa Orde Baru, ketika militer memiliki pengaruh sosial dan politik yang kuat dalam pemerintahan dan kehidupan publik [2]. Media sosial X menjadi salah satu platform utama yang dimanfaatkan masyarakat untuk mengekspresikan opini tersebut melalui cuitan, komentar, dan diskusi, sejalan dengan temuan Fauzi Syarief (2017) yang menyatakan bahwa media sosial berperan signifikan dalam pembentukan opini publik karena memungkinkan pengguna tidak hanya menerima informasi tetapi juga menanggapi dan mendiskusikan isu politik [3]. Lebih lanjut, Nurtikasari *et al.* (2021) menegaskan bahwa pemahaman opini publik secara efektif memerlukan metode analisis sentimen yang mampu mengolah data teks dalam jumlah besar secara akurat, sehingga media sosial tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai sumber data penting untuk menganalisis persepsi masyarakat terhadap isu sosial dan politik, termasuk RUU TNI [4]-[6].

Analisis sentimen merupakan salah satu teknik dalam pemrosesan bahasa alami (*Natural Language Processing*) yang digunakan untuk mengklasifikasikan opini masyarakat terhadap suatu topik berdasarkan data teks [7]-[9]. Chandra Dev *et al.* (2021) [10] menunjukkan bahwa pendekatan berbasis deep learning lebih efektif dalam memahami konteks opini, khususnya pada ulasan pelanggan. Temuan ini mengindikasikan bahwa analisis sentimen memiliki potensi yang luas untuk diterapkan pada berbagai isu yang berkembang di media digital.

Dalam penelitian ini digunakan dua pendekatan klasifikasi sentimen, yaitu Support Vector Machine (SVM) dan Bidirectional Gated Recurrent Unit (Bi-GRU). SVM dipilih karena kemampuannya dalam menangani fitur berdimensi tinggi dan data teks yang bersifat sparse. Penelitian oleh Zainuddin dan Selamat (2014) [11] serta Aulia *et al.* (2021) [12] menunjukkan bahwa SVM mampu menghasilkan klasifikasi yang akurat pada berbagai konteks data opini, termasuk data dengan distribusi kelas yang tidak seimbang. Di sisi lain, Bi-GRU sebagai bagian dari Recurrent Neural Network (RNN) memiliki keunggulan dalam memahami konteks sekuensial melalui pemrosesan dua arah, yaitu *forward* dan *backward*. Ali *et al.* (2021) [13] membuktikan bahwa Bi-GRU efektif dalam menangkap konteks sentimen pada level aspek, sehingga relevan untuk data media sosial yang tidak terstruktur. Kemampuan Bi-

GRU dalam memodelkan dependensi jangka pendek maupun panjang memungkinkan pemahaman nuansa makna dan polaritas sentimen secara lebih mendalam.

Melalui perbandingan performa SVM dan Bi-GRU dalam analisis sentimen terhadap isu RUU TNI, penelitian ini diharapkan dapat memberikan gambaran mengenai karakteristik opini publik serta mengevaluasi efektivitas kedua metode dalam mengolah data teks media sosial. Selain itu, hasil penelitian ini diharapkan dapat menjadi dasar rekomendasi pemilihan model klasifikasi sentimen yang sesuai untuk konteks opini masyarakat Indonesia, khususnya pada isu sosial yang bersifat sensitif dan dinamis, serta menjadi referensi bagi pengembangan sistem analisis sentimen otomatis yang lebih akurat.

II. KAJIAN TEORI

Penelitian analisis sentimen telah banyak dilakukan menggunakan pendekatan NLP dan *deep learning*, khususnya model Bi-GRU, melalui studi literatur yang mengumpulkan dan mengevaluasi penelitian relevan [14]. Wang *et al.* (2022) menunjukkan bahwa Bi-GRU efektif dalam menangani *dataset* kompleks dan memahami makna kata berdasarkan konteks kalimat, sehingga mampu menentukan ketepatan opini pada data teks sekuensial seperti media sosial [15]. Selain itu, Bi-GRU sering dikombinasikan dengan metode lain untuk meningkatkan performa, seperti Ziwen Gao *et al.* (2022) menggabungkannya dengan Convolutional Neural Network (CNN) dan memperoleh peningkatan akurasi klasifikasi sentimen aspek sebesar 12,12% dibandingkan CNN tunggal, meskipun belum menggunakan *embedding* pra-latih dan tidak membahas efisiensi komputasi [16]. Sementara itu, Setiawan *et al.* (2020) menerapkan Bi-GRU dengan character-enhanced token *embedding* dan *multi-level attention* pada ulasan layanan kesehatan daring dan mencapai *F1-score* tertinggi sebesar 88%, yang menunjukkan efektivitas Bi-GRU dalam menangani kata tidak dikenal dan kesalahan ejaan, meskipun penelitian tersebut masih terbatas pada domain kesehatan dan belum meninjau efisiensi pelatihan model [17],[18].

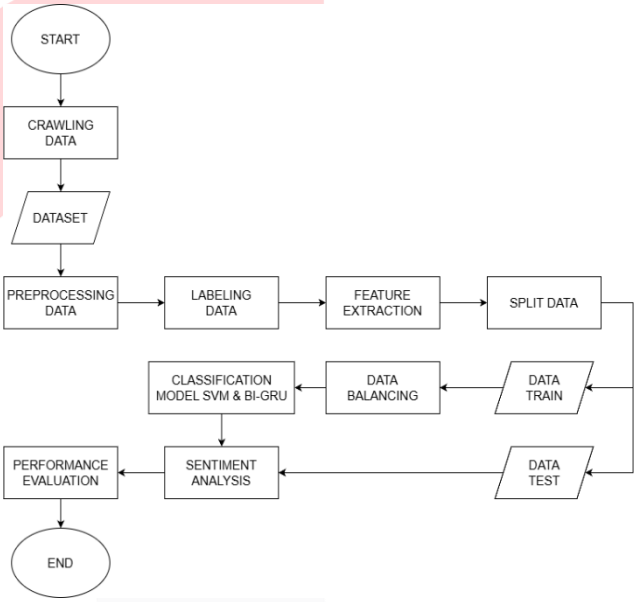
Sejumlah penelitian menunjukkan fleksibilitas Bi-GRU pada berbagai konteks data. Htet Myet Lynn *et al.* (2019) membuktikan efektivitas Bi-GRU dalam menangkap urutan waktu pada data sinyal ECG dan meningkatkan akurasi klasifikasi, meskipun memiliki kompleksitas model dan kebutuhan komputasi yang tinggi. Waqar Ali *et al.* (2021) mengusulkan Bi-GRU dengan embedding GloVe dan *pooling* tanpa *attention* yang lebih ringan dibandingkan LSTM dengan *attention* [13]. Jin Wang *et al.* (2022) memperkenalkan BiGRU-based Contextual Sentiment Embedding (CoSE) yang lebih akurat dibandingkan GloVe dan Word2Vec, tangguh terhadap kata *out-of-vocabulary*, serta lebih hemat parameter [15]. Selain itu, Bhuvaneswari *et al.* (2019) mengombinasikan Bi-GRU, LSTM, dan WordNet *embedding* untuk klasifikasi *tweet* bencana dan menunjukkan keunggulan Bi-GRU dibandingkan LSTM, meskipun penelitian tersebut terbatas pada data krisis tertentu dan menghadapi tantangan komputasi tinggi [19].

Selain pendekatan *deep learning*, metode *machine learning* juga banyak digunakan dalam analisis sentimen. Rahmat Syahputra *et al.* (2022) membandingkan SVM dan Naïve Bayes pada analisis sentimen pengguna Twitter dan menunjukkan bahwa SVM menghasilkan akurasi yang lebih tinggi, yaitu 78% dibandingkan 73% pada Naïve Bayes [20]. Hasil ini menunjukkan bahwa SVM lebih efektif dalam menangani data teks pendek dengan tingkat *noise* yang tinggi, meskipun memiliki kompleksitas model dan waktu pelatihan yang lebih besar. Temuan-temuan tersebut menjadi dasar pemilihan dan pembandingan metode SVM dan Bi-GRU pada penelitian ini.

III. METODE

A. Sistem yang Dibangun

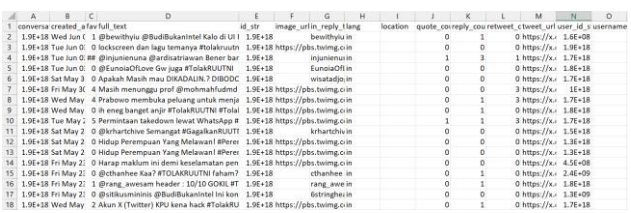
Sistem yang dibangun pada penelitian ini dirancang untuk melakukan analisis sentimen terhadap opini masyarakat mengenai isu RUU TNI yang diperoleh dari media sosial X. Alur sistem pada Gambar 1 mencakup tahapan pada penelitian ini.



GAMBAR 1
Diagram Analisis Sentimen SVM dan Bi-GRU

B. *Crawling Data*

Data yang digunakan dalam penelitian ini berupa opini masyarakat dan dikumpulkan melalui proses *crawling* menggunakan Tweet Harvest yang memanfaatkan Twitter *authentication token* untuk memperoleh data secara otomatis. *Dataset* berisi cuitan yang membahas isu RUU TNI, dikumpulkan menggunakan beberapa kata kunci pada rentang waktu Januari 2024 hingga September 2025. Total data yang berhasil dikumpulkan sebanyak 8.741 *tweets* dan disimpan dalam format CSV seperti pada Gambar 2.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	conversa	created_a	fav	full_text	id_str	image_url	in_reply_t	lang	location	quote_co	reply_cou	retweet_c	tweet_url	user_id_s	username
2	1.9E+18	Wed Jun (	1	@bewithyiu @BudiBukanIntel Kalo di UI l	1.9E+18		bewithyiu	in		0	1	0	https://x.-	1.6E+08	
3	1.9E+18	Tue Jun 0:	0	lockscreen dan lagu temanya #tolakruutn	1.9E+18	https://pbs.twimg.cin				0	0	0	https://x.-	1.9E+18	
4	1.9E+18	Tue Jun 0:	##	@injunienuna @ardisatriawan Bener bar	1.9E+18		injunienuna	in		1	3	1	https://x.-	1.7E+18	
5	1.9E+18	Tue Jun 0:	0	@EunoiaOfLove Gw juga #TolakRUUTNI	1.9E+18		EunoiaOfLo	in		0	0	0	https://x.-	1.8E+18	
6	1.9E+18	Sat May 3	0	Apakah Masih mau DIKADALIN.? DIBODC	1.9E+18		wisatadjo	in		0	0	0	https://x.-	1.7E+18	
7	1.9E+18	Fri May 3(	4	Masih menunggu prof @mohmahfudmd	1.9E+18	https://pbs.twimg.cin				0	0	3	https://x.-	1E+18	
8	1.9E+18	Wed May	4	Prabowo membuka peluang untuk menja	1.9E+18	https://pbs.twimg.cin				0	1	3	https://x.-	1.7E+18	
9	1.9E+18	Wed May	0	@ emeg.banget anjir #TolakRUUTNI #Tolal	1.9E+18	https://pbs.twimg.cin				0	1	0	https://x.-	1.8E+18	
10	1.9E+18	Tue May ;	5	Permintaan takedown lewat WhatsApp #	1.9E+18	https://pbs.twimg.cin				1	1	3	https://x.-	1.7E+18	
11	1.9E+18	Sat May 2	0	@krhartchive Semangat #GagalkanRUUTI	1.9E+18		krhartchiv	in		0	0	0	https://x.-	1.5E+18	
12	1.9E+18	Sat May 2	0	Hidup Perempuan Yang Melawan! #Perer	1.9E+18	https://pbs.twimg.cin				0	0	0	https://x.-	1.3E+18	
13	1.9E+18	Sat May 2	0	Hidup Perempuan Yang Melawan! #Perer	1.9E+18	https://pbs.twimg.cin				0	0	0	https://x.-	1.3E+18	
14	1.9E+18	Fri May 2(	0	Harap maklum ini demi keselamatan pen	1.9E+18	https://pbs.twimg.cin				0	0	0	https://x.-	4.5E+08	
15	1.9E+18	Fri May 2(	0	@cthanhee Kaa? #TOLAKRUUTNI faham?	1.9E+18		cthanhee	in		0	1	0	https://x.-	2.4E+09	
16	1.9E+18	Fri May 2(	1	@rang_awesam header : 10/10 GOKIL #T	1.9E+18		rang_awe	in		0	1	0	https://x.-	1.8E+18	
17	1.9E+18	Fri May 2(	0	@sitikusminiis @BudiBukanIntel Ini kon	1.9E+18		6stringhei	in		0	0	0	https://x.-	1.3E+09	
18	1.9E+18	Wed May	2	Akun X (Twitter) KPU kena hack #TolakRU	1.9E+18	https://pbs.twimg.cin				0	1	0	https://x.-	1.7E+18	

GAMBAR 2
Dataset RUU TNI dari Media Sosial X

C. *Preprocessing Data*

Tahap praproses data bertujuan untuk membersihkan dan menstandarkan data teks agar siap digunakan pada tahap

pemodelan. Praproses dilakukan secara berurutan yang bertujuan mengurangi *noise* pada data serta meningkatkan konsistensi representasi teks sehingga model dapat mempelajari pola sentimen secara lebih efektif. Seluruh tahapan praproses beserta contoh transformasi data dirangkum dalam Tabel 1.

TABEL 1
TAHAPAN PRAPROSES DATA

Tahapan	Deskripsi	*Full Text*	Hasil
Case Folding	Mengubah seluruh teks menjadi huruf kecil untuk menjaga konsistensi data	Barusan ngimpi masih sekolah tapi ada TNI masuk kelas bawa senpi anjir #TolakRUUTNI	barusan ngimpi masih sekolah tapi ada tni masuk kelas bawa senpi anjir #tolakruutni
Cleansing	Menghapus tanda baca, angka, URL, tagar, user mention, emotikon, dan karakter khusus	@kmeilyta udah puas gw ketawain pas dia baru nyesel di era demo RUU TNI wkwk	udah puas gw ketawain pas dia baru nyesel di era demo wkwk
Tokenizing	Memecah kalimat menjadi token kata [21]	gak ada untungnya semua mana ada ruu tni	[‘gak’, ’ada’, ’untungnya’, ’semua’, ’mana’, ’ada’, ’ruu’, ’tni’]
Normalisasi	Mengubah kata tidak baku atau bahasa gaul menjadi bentuk baku [22]	benar ruu tni ini utk siapa utk semua prajurit tni atau hanya utk mereka para petinggi	[‘benar’, ‘ruu’, ‘tn’, ‘ini’, ‘untuk’, ‘siapa’, ‘untuk’, ‘semua’, ‘prajurit’, ‘tni’, ‘atau’, ‘hanya’, ‘untuk’, ‘mereka’, ‘para’, ‘petinggi’]
Stopword Removal	Menghapus kata umum yang tidak memiliki kontribusi semantik [23]	benar ruu tni ini utk siapa utk semua prajurit tni atau hanya utk mereka para petinggi	[‘benar’, ‘ruu’, ‘tni’, ‘siapa’, ‘prajurit’, ‘tni’, ‘hanya’,’petinggi’’]
Stemming	Menghilangkan imbuhan untuk memperoleh kata dasar [24]	kalo udh sepi digass tu diem diemm jgn smpe kejadian kaya RUU TNI	['kalau', 'sudah', 'sepi', 'gas', 'itu', 'diam', 'diam', 'jangan', 'sampai', 'jadi', 'kaya', 'ruu', 'tni']

D. *LABELING DATA*

Setelah praproses, data dilabeli ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral secara semi-manual menggunakan IndoRoBERTA yang telah dilatih pada data berbahasa Indonesia dan sesuai untuk karakteristik teks media sosial X. Model ini digunakan untuk menghasilkan prediksi awal label sentimen, yang selanjutnya divalidasi secara manual guna memastikan kesesuaian label dengan konteks semantik kalimat, khususnya pada data yang ambigu atau menggunakan bahasa tidak baku. Jumlah data yang digunakan dalam penelitian adalah 7.198 *tweets*, dengan distribusi data 4.439 negatif, 2.210 netral, dan 549 positif.

E. *FEATURE EXCTRACTION*

Metode ekstraksi fitur disesuaikan dengan model klasifikasi yang digunakan. Untuk model SVM, digunakan TF-IDF (Term Frequency-Inverse Document Frequency) sebagai metode pembobotan kata dengan memanfaatkan *library* scikit-learn. TF-IDF menghasilkan representasi vektor bernilai antara 0 hingga 1 berdasarkan frekuensi kemunculan kata dan tingkat kepentingannya dalam dokumen [9]. Frekuensi kata dalam dokumen dihitung dengan membagi jumlah kemunculan sebuah kata dengan jumlah total kata dalam dokumen. Berikut adalah rumus untuk pembobotan TF- IDF [9]:

$$TF - IDF = TF(t) * IDF(t) \quad (1)$$

Untuk model Bi-GRU, digunakan *embedding* FastText yang merepresentasikan kata dalam bentuk vektor berdimensi tetap dengan mempertimbangkan *subword information* [25]. FastText dipilih karena kemampuannya menangani variasi penulisan yang tidak baku pada data media sosial, serta lebih efektif dalam merepresentasikan makna dan konteks kata dibandingkan pendekatan berbasis frekuensi seperti TF-IDF.

F. *SPLIT DATA*

Split data dilakukan dimana pembagian *dataset* menjadi dua bagian, yaitu data latih dan data uji. Data latih mengambil sekitar 80% dari *dataset*, sementara 20% sebagai data uji. Pembagian ini bertujuan untuk memastikan bahwa model dapat belajar dari data latih dan diuji pada data baru untuk mengevaluasi performanya.

G. *DATA BALANCING*

Distribusi kelas pada *dataset* bersifat tidak seimbang, sehingga diperlukan teknik penyeimbangan data agar model tidak bias terhadap kelas mayoritas. Teknik pertama yang digunakan adalah SMOTE (Synthetic Minority Over-sampling Technique), yaitu metode *oversampling* yang menghasilkan data sintetis pada kelas minoritas melalui interpolasi antar tetangga terdekat dalam ruang fitur [26], [27]. Selain itu, diterapkan *class weight* dengan memberikan bobot kesalahan yang lebih besar pada kelas minoritas selama proses pelatihan, sehingga model lebih sensitif terhadap kesalahan prediksi pada kelas dengan jumlah data lebih sedikit [28].

H. MODEL KLASIFIKASI SENTIMEN

1) Support Vector Machine

Model pertama yang digunakan adalah SVM, yaitu algoritma *supervised learning* yang banyak digunakan pada tugas klasifikasi teks karena stabilitas dan akurasinya yang tinggi [5], [6]. SVM bekerja dengan memetakan data ke ruang fitur berdimensi lebih tinggi dan mencari sebuah *hyperplane* optimal yang memisahkan data berbagai kelas dengan margin maksimum. Prinsip ini memungkinkan SVM mempertahankan performa yang baik meskipun data memiliki dimensi tinggi dan mengandung *noise*, seperti lazim ditemukan pada data teks media sosial [6],[9].

$$w . x + b = 0 \quad (2)$$

$$w . x_i + b \geq -1 \text{ Kelas Positif} \quad (3)$$

$$w . x_i + b \geq 1 \text{ Kelas Negatif} \quad (4)$$

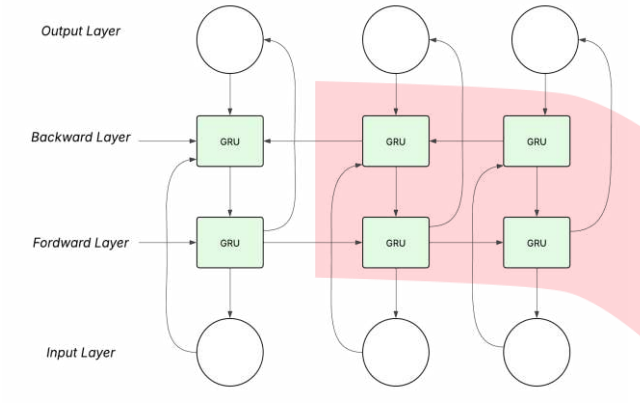
$$-1 < w . x_i + b < +1 \text{ Kelas Netral} \quad (5)$$

Secara matematis, SVM direpresentasikan melalui fungsi pemisah *linier* pada Persamaan (2), di mana w adalah vektor bobot, x adalah vektor fitur *input*, dan b adalah bias [25]. Batas kelas positif dan negatif ditentukan masing-masing oleh Persamaan (3) dan (4), sementara kelas netral diasumsikan berada di antara kedua batas tersebut dengan nilai fungsi berada pada rentang -1 hingga +1 sebagaimana ditunjukkan pada Persamaan (5). Pendekatan ini memungkinkan pemisahan tiga kelas sentimen dengan

mempertimbangkan jarak data terhadap *hyperplane*, sehingga klasifikasi tidak hanya bergantung pada label, juga pada tingkat keyakinan posisi data terhadap batas Keputusan [29],[30].

2) Bidirectional Gated Recurrent Unit

Model kedua adalah Bi-GRU, yang memiliki kemiripan dengan LSTM, namun dengan struktur yang lebih sederhana dan efisien [31]. Bi-GRU memproses teks secara dua arah, yaitu *forward* dan *backward*, sehingga mampu menangkap konteks kata berdasarkan informasi masa lalu dan masa depan secara simultan [11].



GAMBAR 3
Arsitektur Bi-GRU [7]

Arsitektur Bi-GRU pada Gambar 3 terdiri atas dua jaringan GRU independen yang memproses urutan *input* yang sama dari arah berlawanan [12]. Masing-masing jaringan menghasilkan *hidden state* pada setiap *timestep* yang kemudian digabungkan. Proses ini dikendalikan *oleh update gate* untuk mempertahankan informasi lama dan *reset gate* untuk mengontrol pengabaian informasi masa lalu.

I. EVALUASI PERFORMA

Evaluasi performa model dilakukan untuk menilai kemampuan klasifikasi sentimen pada setiap kelas. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score*, yang memungkinkan pengukuran performa model secara menyeluruh, termasuk pada kelas minoritas. Selain itu, digunakan *confusion matrix* untuk memvisualisasikan hasil prediksi benar dan salah pada setiap kelas sentimen.

1) *Accuracy*

Akurasi digunakan untuk mengukur proporsi prediksi yang benar dari keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

2) *Precision*

Presisi digunakan sebagai rasio prediksi positif yang benar terhadap total prediksi positif.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

3) *Recall*

Recall digunakan sebagai rasio prediksi positif yang benar terhadap total positif yang sebenarnya.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

4) *F1-Score*

F1-Score menggabungkan presisi dan *recall* ke dalam satu metrik dengan menghitung rata-rata harmoniknya, sehingga memberikan keseimbangan di antara keduanya.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

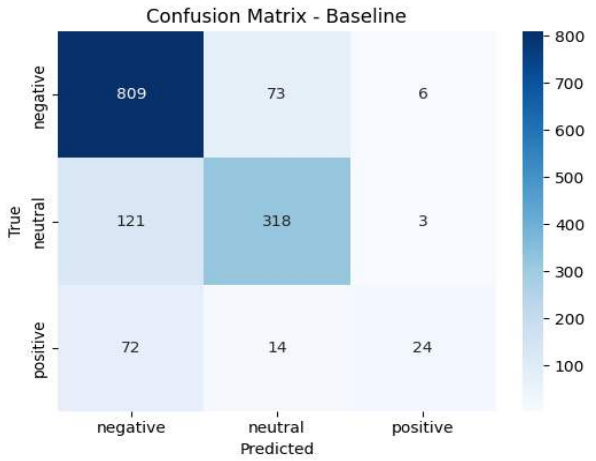
Data penelitian dikumpulkan melalui API media sosial X menggunakan *auth token* dengan kata kunci pada periode Januari 2024 hingga September 2025, menghasilkan 7.198 *tweet* dan disimpan dalam format CSV. Perbedaan sentimen tercermin dari kata kontekstual seperti “tolak” pada kelas negatif dan “dukung” pada kelas positif yang merepresentasikan sikap publik terhadap isu yang sama. Data kemudian dilabeli ke dalam tiga kelas sentimen, dengan distribusi netral 2.210 data, negatif 4.439 data, dan positif 549 data, yang menunjukkan ketidakseimbangan kelas signifikan dengan dominasi kelas negatif sehingga diperlukan strategi khusus untuk mencegah bias model terhadap kelas mayoritas.

B. Hasil Pengujian Klasifikasi SVM

Bagian ini menyajikan hasil pengujian kinerja model klasifikasi SVM melalui beberapa skenario percobaan yang dirancang untuk mengevaluasi pengaruh strategi penanganan ketidakseimbangan data serta kestabilan model. Pengujian pertama dengan skenario *baseline* menggunakan ekstraksi fitur TF-IDF tanpa penyesuaian distribusi kelas. Selanjutnya, dilakukan dua skenario penanganan ketidakseimbangan data menggunakan SMOTE dan *class weight*. Pada skenario keempat, dilakukan skema *k-fold cross validation* untuk menilai konsistensi dan stabilitas performa model pada berbagai *subset* data.

1) Hasil Percobaan 1

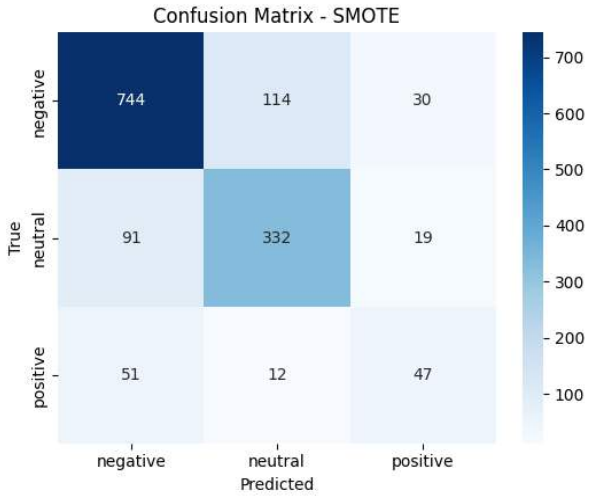
Model *baseline* pada penelitian ini menggunakan Support Vector Classifier (SVC) dengan ekstraksi fitur TF-IDF, di mana penentuan parameter optimal dilakukan menggunakan GridSearchCV dengan 3-*Fold Cross Validation*. Hasil pencarian menunjukkan bahwa konfigurasi terbaik diperoleh pada kernel *linear* dengan nilai C = 1 dan gamma = *scale*. Evaluasi kinerja model *baseline* menghasilkan akurasi sebesar 79,93%. Pada kelas negatif, model memperoleh *precision* 0,77 dan *recall* 0,62, sedangkan pada kelas netral *precision* dan *recall* masing-masing sebesar 0,71 dan 0,65. Kelas positif menunjukkan performa terendah dengan *precision* 0,65 dan *recall* 0,62. *Confusion matrix* pada Gambar 4 menunjukkan bahwa model paling efektif dalam mengklasifikasikan kelas negatif dengan 809 prediksi benar. Pada kelas netral, terdapat 318 prediksi benar, namun masih ditemukan kesalahan sebanyak 121 *tweet* yang diprediksi sebagai negatif. Performa terendah terlihat pada kelas positif dengan 24 prediksi benar. Pola kesalahan ini menegaskan bahwa ketidakseimbangan distribusi label berdampak signifikan terhadap sensitivitas model terhadap kelas minoritas.



GAMBAR 4
Confusion Matrix Baseline

2) Hasil Percobaan 2

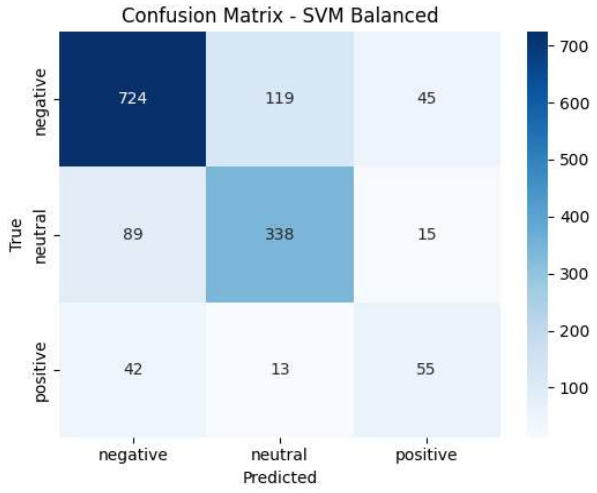
Penerapan SMOTE dilakukan untuk menyeimbangkan distribusi kelas pada data latih yang semula didominasi kelas negatif, kemudian setelah oversampling seluruh kelas diseimbangkan menjadi masing-masing 3.551 *tweet*. Hasil evaluasi setelah penerapan SMOTE menunjukkan akurasi sebesar 77,99%, dengan nilai *precision* dan *recall* yang relatif seimbang pada ketiga kelas, yakni di kisaran 0,66-0,68, menandakan peningkatan sensitivitas model terhadap kelas positif dibandingkan skenario *baseline*. *Confusion matrix* pada Gambar 5 memperlihatkan bahwa meskipun performa antar kelas menjadi lebih seimbang, model masih paling konsisten dalam mengenali kelas negatif.



GAMBAR 5
Confusion Matrix SMOTE

3) Hasil Percobaan 3

Penerapan *class weight* menghasilkan akurasi sebesar 77,57% dengan nilai *precision* dan *recall* yang relatif seimbang pada seluruh kelas, yaitu sekitar 0,68-0,69. Keseimbangan ini menunjukkan bahwa penyesuaian bobot kelas mampu meningkatkan sensitivitas model terhadap kelas positif yang sebelumnya sulit terdeteksi akibat ketidakseimbangan data. *Confusion matrix* pada Gambar 6 memperlihatkan bahwa distribusi prediksi antar kelas menjadi lebih merata. Namun, masih terlihat kecenderungan sebagian data diprediksi ke kelas negatif. Hal ini mengindikasikan bahwa meskipun bobot kelas telah diatur, pola linguistik dari kelas mayoritas tetap memiliki pengaruh kuat terhadap keputusan model. Dengan kata lain, *class weight* membantu mereduksi bias, tetapi belum sepenuhnya menghilangkannya.



GAMBAR 6
Confusion Matrix Class Weight

4) Hasil Percobaan 4

Skenario 5 *Fold Cross Validation* digunakan untuk menilai kestabilan performa pada data uji yang berbeda. Hasil pengujian pada Gambar 7 menunjukkan akurasi rata-rata sebesar 79,58% dengan standar deviasi 0,0117, yang mengindikasikan performa model relatif konsisten di setiap *fold*. Nilai *precision* rata-rata sebesar 0,7566 menandakan kemampuan model yang baik dalam mengenali kelas mayoritas. Sementara itu, *recall* rata-rata 0,6220 menunjukkan bahwa sebagian data pada kelas minoritas masih belum teridentifikasi secara optimal. *F1-Score* rata-rata sebesar 0,6535. Standar deviasi yang kecil pada seluruh metrik menegaskan bahwa model memiliki stabilitas performa yang baik ketika diuji pada *subset* data yang berbeda.

```
5-fold CV, SVC dengan best params dari Baseline:
    accuracy: mean=0.7958 | std=0.0117
          f1: mean=0.6535 | std=0.0245
   precision: mean=0.7566 | std=0.0297
      recall: mean=0.6220 | std=0.0184
```

GAMBAR 7
Hasil SVM 5 *Fold Cross Validation*

5) *Error Analysis*

Berdasarkan Tabel 2, sebagian besar kesalahan prediksi yang dilakukan oleh model SVM terjadi pada *tweet* dengan struktur kalimat ambigu, serta mengandung campuran sikap dukungan dan kritik dalam satu konteks. Model SVM cenderung mengklasifikasikan *tweet* tersebut sebagai sentimen netral karena TF-IDF lebih menekankan frekuensi dan keberadaan kata, tanpa mampu memahami relasi semantik dan arah sentimen secara kontekstual. Selain itu, penggunaan bahasa informal, sarkasme, serta pergeseran sudut pandang dalam satu *tweet* menyebabkan model kesulitan menentukan polaritas dominan. Temuan ini menunjukkan bahwa meskipun SVM memiliki performa yang cukup, namun memiliki keterbatasan dalam menangkap nuansa makna dan konteks, sehingga berpotensi kesalahan klasifikasi pada opini yang kompleks dan tidak eksplisit.

TABEL 2
CONTOH KESALAHAN PREDIKSI SVM

No	*Tweet*	Label Asli	Label Prediksi
1	dia cuma caper ew gua tolak ruu tni tapi untuk giat yang bener ya gua dukung ini orang sebar benci aja sih	Negatif	Netral
2	gua benci gibran tp tanpa gibran wowo versi buruk bakal muncul solusi yang bener adalah turun 22nya tp ga mungkin lebih mudah lawan wiwi dan bikin wiwi tobat krn sama sipil dan dia msh punya malu wowo tuh udh batu enak jidat dan militer yeay bawa senjata	Negatif	Netral
3	banyak pihak memang khawatir luas wewenang militer ke ranah sipil di indonesia justru bentuk patuh masyarakat karena rasa takut bukan karena ubah karakter yang sehat atau sukarela lapor dari bagai organisasi ham dan media tunjuk ada tingkat	Negatif	Netral

C. Hasil Pengujian Model Bi-GRU

Model Bi-GRU pada penelitian ini menggunakan representasi kata berbasis FastText berbahasa Indonesia. Proses evaluasi dilakukan dengan skema 5-*Fold Cross Validation*, sehingga setiap *fold* memiliki data uji yang berbeda dan mampu memberikan gambaran performa yang lebih stabil. Hasil evaluasi menunjukkan bahwa Bi-GRU memperoleh akurasi tertinggi sebesar 80,76%, melampaui seluruh skenario yang diuji pada model SVM. Pada kelas negatif, nilai rata-rata *precision* sebesar 0,73 dan *recall* 0,61 menunjukkan bahwa mayoritas *tweet* negatif berhasil diidentifikasi. Untuk kelas netral, model mencapai *precision* 0,70 dan *recall* 0,60, yang menandakan bahwa model cukup efektif mengenali sentimen netral. Pada kelas positif, *precision* sebesar 0,62 dan *recall* 0,61 mengindikasikan bahwa meskipun masih terdapat sejumlah *tweet* positif yang tidak terdeteksi, performa model tetap lebih baik dibandingkan dengan SVM. Nilai rata-rata *F1-Score* sebesar 0,6235 serta standar deviasi yang relatif kecil menunjukkan bahwa performa Bi-GRU konsisten pada setiap *fold* dan mampu menjaga keseimbangan antara *precision* dan *recall*.

```
HASIL AKHIR:
ACCURACY: 0.807587 ± 0.015042
PRECISION: 0.730186 ± 0.123988
RECALL: 0.608648 ± 0.041729
F1_SCORE: 0.623472 ± 0.058519
```

GAMBAR 8
Hasil Bi-GRU 5 *Fold Cross Validation*

Tabel 3 memperlihatkan keterbatasan model dalam menangkap nuansa semantik dan konteks implisit. Beberapa *tweet* dengan sentimen positif diprediksi sebagai netral karena penggunaan kosakata yang bersifat informatif, seperti dukungan terhadap RUU TNI yang disampaikan tanpa ekspresi emosional, sehingga cenderung diasosiasikan sebagai pernyataan netral. Sebaliknya, *tweet* yang bersifat netral namun memuat isu sensitif diprediksi sebagai negatif, yang mengindikasikan bahwa model masih terpengaruh oleh keberadaan kata kunci bernuansa kritis meskipun tidak secara langsung mengekspresikan sentimen negatif.

TABEL 3
CONTOH KESALAHAN PREDIKSI BIGRU

No	*Tweet*	Label Asli	Label Prediksi
1	ruu tni hadir dukung aman nasional dan menteri lebih siap dengan tenaga profesional tni tni kuat tahan	Positif	Netral
2	nanya ya tapi kamu jawab yang jujur andai kamu itu rakyat indonesia yang usia 20 tahun apa dapat tentang uu tni ruu jaksa ruu polri dan ruu hap kalau kamu jadi mahasiswa apa kamu akan ikut demo	Netral	Negatif
3	dukung ruu tni sesuai bagai tanda cinta buat tni dan bangsa	Positif	Netral

D. Analisis Hasil Pengujian

Bagian ini membandingkan performa model SVM dan Bi-GRU dalam klasifikasi sentimen. Berdasarkan Tabel 4, SVM *baseline* menghasilkan akurasi tertinggi di antara skenario SVM sebesar 79,93%, diikuti SVM 5-*Fold Cross Validation* dengan 79,58%, sementara pendekatan SMOTE dan *class weight* justru menurunkan akurasi. Hal ini menunjukkan bahwa pada karakteristik data penelitian ini, penanganan *imbalance* tidak otomatis meningkatkan performa keseluruhan. Model Bi-GRU dengan 5-*Fold Cross Validation* mencapai akurasi tertinggi secara keseluruhan yaitu 80,76%, mengindikasikan keunggulan model berbasis konteks dalam menangkap variasi bahasa media sosial. Namun, akurasi tinggi tidak selalu berarti performa merata di semua kelas.

TABEL 4
PERFORMA MASING-MASING MODEL

Model	Akurasi	*Precision*	*Recall*	*F1 Score*
SVM *Baseline*	79,93%	0,7733	0,6162	0,6475
SVM SMOTE	77,99%	0,6847	0,6721	0,6776
SVM *Class Weight*	77,57%	0,6814	0,6933	0,6870
SVM 5-Fold CV	79,58%	0,7566	0,6220	0,6535
Bi-GRU 5-Fold CV	80,76%	0,7302	0,6086	0,6235

Jika ditinjau lebih dalam melalui Tabel 4, terlihat bahwa model dengan akurasi tertinggi tidak selalu memiliki keseimbangan metrik terbaik. SVM *Class Weight* mencatat *F1-Score* tertinggi sebesar 0,6870, menandakan keseimbangan terbaik antara *precision* dan *recall* pada seluruh kelas sentimen. Sebaliknya, Bi-GRU meskipun unggul dari sisi akurasi memiliki *F1-Score* lebih rendah (0,6235), yang menunjukkan kecenderungan model dalam memprediksi kelas mayoritas dengan lebih baik. Analisis ini menegaskan bahwa evaluasi model pada data sentimen yang tidak seimbang tidak dapat bergantung pada akurasi semata. *F1-Score* menjadi metrik yang lebih representatif untuk menilai kualitas model secara menyeluruh.

V. KESIMPULAN

Berdasarkan percobaan lima skenario (SVM *Baseline*, SVM SMOTE, SVM *Class Weight*, SVM 5 Fold CV, dan Bi-GRU 5 Fold CV), seluruh model mampu mengolah data teks berbahasa Indonesia meskipun dihadapkan pada tantangan *imbalance data*, dengan hasil yang menunjukkan perbedaan dampak strategi penanganannya terhadap performa model. Pada model SVM, penerapan *class weight* terbukti paling efektif dalam menyeimbangkan performa antar kelas dengan menghasilkan *F1-score* tertinggi sebesar 0,6870 serta meningkatkan kemampuan prediksi pada kelas minoritas, sementara penggunaan SMOTE lebih berkontribusi pada peningkatan *recall* kelas minoritas. Pada model Bi-GRU, penanganan *imbalance* dilakukan melalui pembelajaran sekuensial dan validasi silang sehingga menghasilkan

performa yang relatif stabil tanpa memerlukan *oversampling*. Secara keseluruhan, Bi-GRU menunjukkan kinerja terbaik dengan akurasi tertinggi sebesar 80,76% serta nilai *precision*, *recall*, dan *F1-score* yang konsisten pada seluruh *fold*, menegaskan keunggulan pendekatan *deep learning* dalam menangkap konteks dan dependensi urutan kata dibandingkan *machine learning*, meskipun SVM tetap kompetitif dengan akurasi tertinggi sebesar 79,93% serta kelebihan pada efisiensi pelatihan dan stabilitas performa ketika dikombinasikan dengan *class weight*.

Penelitian berhasil mencapai tujuan, memberikan kontribusi pada performa *machine learning* dan *deep learning* untuk sistem analisis sentimen terkait isu RUU TNI. Meski demikian, dalam menangani data dihadapkan dengan keterbatasan menangani data yang tidak seimbang. Hal ini menunjukan bahwa Model Bi-GRU direkomendasikan sebagai implementasi pada sistem analisis sentimen, sementara model SVM tetap relevan pada skenario dengan waktu pelatihan yang lebih efisien. Penelitian selanjutnya disarankan untuk mengeksplorasi *pre-trained language models* guna meningkatkan pemahaman konteks dan mengatasi ketidakseimbangan kelas lebih efektif. Selain itu, perlu dilakukan perluasan *dataset* serta pengujian periode isu yang berbeda agar model yang dihasilkan memiliki kemampuan generalisasi lebih kuat.

REFERENSI

[1] M. A. Hayat, S. M. Soraidah, M. N. Rofif, A. R. Asriani, and Parihin, "Analyzing political trends and discourse on Twitter of influential Indonesian accounts," *Nyimak Journal of Communication*, vol. 8, no. 1, pp. 141–156, Mar. 2024.

[2] K. Rikan, "Konsep Dwifungsi ABRI dan perannya di masa pemerintahan Orde Baru tahun 1965-1998," *Program Studi Pendidikan Sejarah, Fakultas Keguruan dan Ilmu Pendidikan, Universitas PGRI Yogyakarta*, 2014.

[3] F. Syarief, "Pemanfaatan Media Sosial dalam Proses Pembentukan Opini Publik (Analisa Wacana Twitter SBY)," *Jurnal Ilmu Komunikasi*, 2018.

[4] Y. Nurtikasari, S. Alam, dan T. I. Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *Jurnal Informatika, Sekolah Tinggi Teknologi Wastukancana*, 2021.

[5] F. Z. Emeraldien, R. J. Sunarsono, and R. Alit, "Twitter sebagai platform komunikasi politik di Indonesia," *SCAN: Jurnal Komunikasi*, vol. 14, no. 1, pp. 21–30, Feb. 2019.

[6] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis sentimen kinerja Dewan Perwakilan Rakyat (DPR) pada Twitter menggunakan metode Naive Bayes Classifier," *Jurnal Informatika Dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.

[7] H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, Jan. 2016, doi: 10.1109/icatcct.2016.7912034.

[8] R. Kumar and S. Garg, "Aspect-Based sentiment analysis using deep Learning convolutional neural network," *in Advances in intelligent systems and computing*, 2019, pp. 43–52. doi: 10.1007/978-981-13-7166-0_5.

[9] V. R. Prasetyo, N. Benarkah, and V. J. Chrisintha, "Implementasi natural language processing dalam pembuatan chatbot pada Program Information Technology Universitas Surabaya," *Teknika*, vol. 10, no. 2, pp. 114–121, Jul. 2021, doi: 10.34148/teknika.v10i2.370.

[10] V. Chandradev, I. M. A. D. Suarjaya, dan I. P. A. Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *Jurnal Ilmu Komputer dan Teknologi Informasi, Universitas Udayana*, 2021.

[11] N. Zainuddin and A. Selamat, "Sentiment analysis using Support Vector Machine," *2014 International Conference on Computer, Communications, and Control Technology (I4CT)*, pp. 333–337, Sep. 2014, doi: 10.1109/i4ct.2014.6914200.

[12] T. M. P. Aulia, N. Arifin, dan R. Mayasari, "Perbandingan Kernel Support Vector Machine (SVM) dalam Penerapan Analisis Sentimen Vaksinasi COVID-19," *Jurnal Informatika, Universitas Singaperbangsa Karawang*, 2021.

[13] W. Ali, Y. Yang, X. Qiu, Y. Ke, dan Y. Wang, "Aspect-Level Sentiment Analysis Based on Bidirectional-GRU in SIoT," *IEEE Access*, vol. 9, pp. 68445–68454, 2021, doi: 10.1109/ACCESS.2021.3077377.

[14] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, pp. 333–339, 2019.

[15] Jin Wang, You Zhang, Liang-Chih Yu, Xuejie Zhang. (2022). "Contextual sentiment embeddings via bi-directional GRU language model". *Knowledge-Based Systems*, 235, 107663. https://doi.org/10.1016/j .knosys.2021.107663.

[16] Z. Gao, Z. Li, J. Luo, and X. Li, "Short text aspect-based sentiment analysis based on CNN + BiGRU," *Applied Sciences*, vol. 12, no. 2707, 2022. doi: 10.3390/app12052707.

[17] E. I. Setiawan, F. Ferry, J. Santoso, S. Sumpeno, K. Fujisawa, and M. H. Purnomo, "Bidirectional GRU for targeted aspect-based sentiment analysis based on character-enhanced token-embedding and multi-level attention," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 4, pp. 392–402, 2020.

[18] M. A. Rohman, Suhartono, dan T. Chamidy, "Bidirectional GRU dengan Attention Mechanism pada Analisis Sentimen PLN Mobile," *Techno.COM*, vol. 22, no. 2, pp. 358–372, Mei 2023.

[19] A. Bhuvaneswari, J. T. Jones Thomas, and P. Kesavan, "Embedded Bi-directional GRU and LSTMLearning Models to Predict Disasterson Twitter Data," *Procedia Comput. Sci.*, vol. 165, pp. 511–516, 2019, doi: 10.1016/j.procs.2020.01.020.

[20] R. Syahputra, G. J. Yanris, and D. Irmayani, "SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter," *SinkrOn*, vol. 7, no. 2,

pp. 671–678, May 2022, doi: 10.33395/sinkron.v7i2.11430.

[21] M. A. Rofiqi, A. C. Fauzan, A. P. Agustin, A. A. Saputra, and H. D. Fahma, "Implementasi term-frequency inverse document frequency (TF-IDF) untuk mencari relevansi dokumen berdasarkan query," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 1, no. 2, pp. 58–64, Dec. 2019.

[22] C. Firalya, D. E. Ratnawati, and N. Y. Setiawan, "Analisis sentimen data ulasan aplikasi Pegadaian Digital menggunakan metode Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, Sep. 2024.

[23] A. P. Wibawa, F. Miftahuddin, and Suyono, "K-Medoids clustering untuk pembentukan database stopword Bahasa Jawa," *Ranah: Jurnal Kajian Bahasa*, vol. 10, no. 2, pp. 261–269, 2021, doi: 10.26499/rnh/v9i2.1490.

[24] A. Guterres, Gunawan, and J. Santoso, "Stemming Bahasa Tetun menggunakan pendekatan rule based," *Magister Teknologi Informasi, Sekolah Tinggi Teknik Surabaya*, pp. 142–150.

[25] A. F. Pangestu, B. Rahmat, and A. N. Sihananto, "Analisis sentimen pada media sosial X terhadap implementasi Kurikulum Merdeka menggunakan metode FastText dan Long Short-Term Memory (LSTM)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2271–2280, Dec. 2024, doi: 10.29100/jipi.v9i4.5665.

[26] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk mengatasi imbalance class dalam klasifikasi television advertisement performance rating menggunakan Artificial Neural Network," *JEPIN: Jurnal Edukasi dan Penelitian Informatika*, vol. 6, no. 3, pp. 379–388, Dec. 2020.

[27] A. A. Bhirawa and U. P. Sanjaya, "From data imbalance to precision: SMOTE-Driven Machine learning for early detection of kidney disease," *INOVTEK Polbeng - Seri Informatika*, vol. 10, no. 1, pp. 514–525, Mar. 2025, doi: 10.35314/7jgjmg64.

[28] H. Akbar and W. K. Sanjaya, "Kajian Performa Metode Class Weight Random Forest pada Klasifikasi Imbalance Data Kelas Curah Hujan*," Jurnal Sains Nalar Dan Aplikasi Teknologi Informasi*, vol. 3, no. 1, Dec. 2023, doi: 10.20885/snati.v3i1.30.

[29] A. M. Rahat, A. Kahir, and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," *Th International Conference on System Modeling & Advancement in Research Trends*, pp. 266–270, Nov. 2019, doi: 10.1109/smart46866.2019.9117512.

[30] Kowalczyk, A. (2017). Support Vector Machines Succinctly. USA: Syncfusion.

[31] S. S. M. M. Rahman, K. B. Md. B. Biplob, Md. H. Rahman, K. Sarker, and T. Islam, "An investigation and evaluation of N-Gram, TF-IDF and ensemble methods in sentiment classification," *in Springer eBooks*, 2020, pp. 391–402. doi: 10.1007/978-3-030-52856-0_31.