Progresif: Jurnal Ilmiah Komputer
https://ojs.stmik-banjarbaru.ac.id/index.php/progresif/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat - Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0877

Eksplorasi Sentimen Publik di Media Sosial terhadap Isu RUU TNI Menggunakan Pendekatan *Machine Learning*

DOI: http://dx.doi.org/10.35889/progresif.v21i2.2819

Creative Commons License 4.0 (CC BY – NC)

Sandy Aprilyanto[1*], Rudy Latuperissa[2]
Sistem Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia
*e-mail *Corresponding Author:* 682021002@student.uksw.edu

Abstract

The developments regarding the revision of the Indonesian National Armed Forces Law (RUU TNI) have sparked various reactions from the public, most of which have been expressed through social media platforms. This research aims to evaluate public opinion on the TNI Bill issue by utilizing machine learning technology and the Support Vector Machine (SVM) algorithm. Data were collected from three social networks, namely YouTube, Instagram, and X (formerly known as Twitter), with a total of 4,530 data points. The stages of data processing include web scraping, data cleaning, text preprocessing, automatic labelling using textblob, word weighting through TF-IDF, and data balancing using the SMOTE method. The sentiment classification results reveal that the majority of public opinions are positive and neutral, with the highest model accuracy achieved at parameter C = 1, namely 94.60% for YouTube, 95.80% for Instagram, and 97.33% for X. These findings demonstrate the effectiveness of the SVM approach in categorizing public opinions generated from social media, and imply that social media serves as an important source for understanding public views on national policy issues.

Keywords: *Machine Learning; Support Vector Machine; SMOTE; Text Mining; TF IDF*

Abstrak

Perkembangan mengenai revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI) telah menimbulkan berbagai reaksi dari masyarakat, yang sebagian besar disampaikan melalui platform media sosial. Penelitian ini bertujuan untuk mengevaluasi pandangan publik terhadap isu RUU TNI dengan memanfaatkan teknologi pembelajaran mesin dan algoritma *Support Vector Machine* (SVM). Data dikumpulkan dari tiga jejaring sosial, yaitu YouTube, Instagram, dan X (yang sebelumnya dikenal sebagai Twitter), dengan total sebanyak 4.530 data. Tahapan pengolahan data mencakup *web scraping*, pembersihan data, pra-pemrosesan teks, pelabelan otomatis menggunakan *TextBlob*, pemberian bobot kata melalui TF-IDF, serta penyeimbangan data dengan metode SMOTE. Hasil klasifikasi sentimen mengungkapkan bahwa sebagian besar pendapat masyarakat bersifat positif dan netral, dengan akurasi tertinggi dari model dicapai pada parameter C = 1, yaitu 94,60% untuk YouTube, 95,80% untuk Instagram, dan 97,33% untuk X. Temuan ini menunjukkan efektivitas pendekatan SVM dalam mengkategorikan pendapat publik yang dihasilkan dari media sosial, serta menyiratkan bahwa media sosial berfungsi sebagai sumber penting untuk memahami pandangan masyarakat terhadap masalah kebijakan nasional.

Kata kunci: *Machine Learning; Support Vector Machine; SMOTE; Text Mining; TF IDF*

1. Pendahuluan

Beberapa minggu ini, terjadi polemik yang menjadikan masyarakat Indonesia menjadi gusar. Salah satu diantaranya ialah masalah Rancangan Undang-Undang Tentara Nasional Indonesia, atau umumnya disingkat menjadi RUU TNI. Baru - baru ini Dewan Perwakilan Rakyat Republik Indonesia atau lebih sering dikenal dengan DPR RI mengeluarkan RUU mengenai Tentara Nasional Indonesia atau TNI. Tujuan diadakannya revisi terhadap Undang-Undang

Nomor 34 Tahun 2004 tentang Tentara Nasional Indonesia adalah membantu kinerja kementerian/lembaga negara terkait karena untuk memenuhi kebutuhan atas keahlian tertentu yang diperlukan oleh kementerian/lembaga terkait [1]. Kebijakan pemerintah dalam membuat aturan dalam penugasan terkait TNI, khususnya dalam hal peran TNI dalam mengambil peran di lingkungan masayarakat sipil. Oleh sebab itu, dalam memahami persepsi masyarakat terhadap kebijakan ini sangat penting dalam melakukan penilaian sentimen serta dampak sosial dari persepsi terkait RUU TNI.

Banyaknya pengguna media sosial terkhususnya di indonesia, dimana media sosial tidak hanya digunakan sebagai sarana hiburan semata, namun juga dipergunakan dalam penyampaian terkait opini publik dalam topik politik di Indonesia. Adanya berita pada berbagai kanal media sosial mengenai peran TNI yang masuk pada ranah sipil dan menimbulkan keresahan pada profesionalisme militer di Indonesia itu sendiri, sehingga memicu adanya berbagai tanggapan dari positif hingga negatif mengenai isu ini baik dari kalangan masyarakat baik dari kalangan akademisi maupun kalangan umum. Maka dari itu perlu dilakukan nya analisis sentimen terkait kebijakan dalam revisi terhadap Undang-Undang Nomor 34 Tahun 2004 tentang Tentara Nasional Indonesia.

Analisis sentimen merupakan suatu jenis pemrosesan bahasa alami yang menggunakan pembelajaran mesin dimana digunakan untuk mengekstrak informasi yang penting dari suatu data teks, termasuk perspektif, emosi, dan sikap [2]. Seperti pada penelitian Hayadi dan Maulita pada tahun 2025 yang berjudul "*Sentiment Analysis of Public Discourse on Education in Indonesia Using Support Vector Machine (SVM) and Natural Language Processing*", pada penelitian ini menggunakan algoritma SVM yang terbukti mampu dalam mengklasifikan teks dikarenakan kepresisisan dan adaptabilitasnya [3].

Berdasarkan latar belakang yang dibahas, penelitian ini bertujuan untuk menganalisis sentimen opini publik tentang RUU TNI di Indonesia dengan menggunakan Algoritma *Support Vector Machine* (SVM). Dibandingkan dengan beberapa metode klasifikasi pembelajaran mesin lainnya, algoritma SVM mampu mengenali pola dan menganalisis data dengan dimensi tinggi, yang menghasilkan tingkat akurasi yang lebih tinggi. Setelah itu, hasil analisis sentimen akan diperiksa dan dibahas secara menyeluruh, dan temuan akan disimpulkan. Dataset yang digunakan diperoleh melalui pencarian data di platform media sosial dengan menggunakan keyword atau kunci yang terkait dengan kata kunci RUU TNI di Indonesia.

2. Tinjauan Pustaka

Pada penelitian Wijaya dkk yang berjudul "Analisis sentimen seputar uu ite menggunakan menggunakan algoritma *Support Vector Machine*" yang mencapai tingkat akurasi sebesar 84%. Penelitian ini bertujuan untuk mengetahui bagaimana reaksi publik terhadap uu ite yang menggunakan SVM dan ekstraksi TF-IDF [4].

Selain itu, pada penelitian dalam Wardianto dkk dalam "Analisis Sentimen Publik Terhadap Program Makan Bergizi Gratis di Instagram menggunakan Algoritma *Support Vector Machine*". Penelitian ini menggunakan konsep penelitian berbasis pemodelan pembelajaran mesin yang didukung oleh algoritma *Support Vector Machine* yang melingkupi data sebanyak 10000 komentar terkait dengan Program makan bergizi yang memiliki akurasi sebesar 90% [5].

Pada penelitian Husein dkk yang berjudul "*Sentiment Analysis Od Face To Face School Policy On Twitter Social Media With Support Vector Machine (SVM)"*, Penelitian ini bertujuan untuk penganalisaan opini publik twitter tentang kebijakan pemerintah pasca pandemi Covid-19, dan penelitian ini menggunakan konsep pendekatan pembelajaran mesin yang menggunakan algoritma *Support Vector Machine* [6].

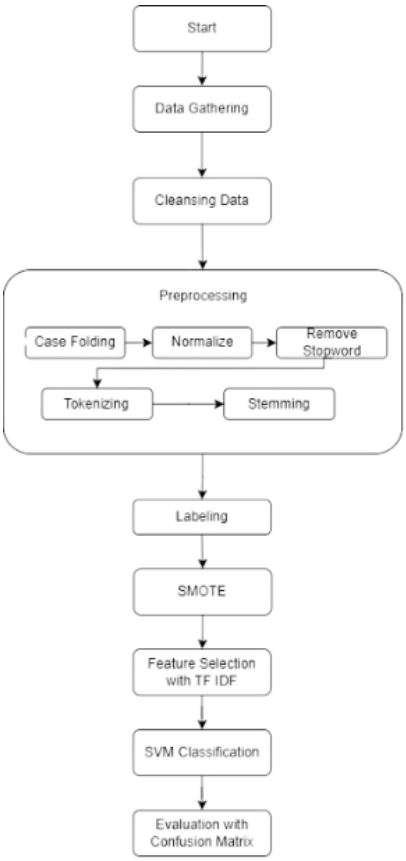
Lebih lanjut dalam penelitian dari Dharmapati dan merawati yang berjudul "Penerapan Algoritma *Support Vector Machine* Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan", dari 671 sampel penelitian prediksi dari algoritma *Support Vector Machine* menghasilkan akurasi sebesar 92%. Pada Penelitian ini membahas analisis sentimen masyarakat di Twitter terhadap kebijakan kenaikan tarif BPJS Kesehatan, Penelitian ini menggunakan metode *Support Vector Machine* (SVM) [7].

Pada penelitian dalam Amelia dkk berjudul "Analisis Sentimen Tanggapan Pengguna Media Sosial X Terhadap Program Beasiswa KIP-Kuliah dengan Menggunakan Algoritma *Support Vector Machine* (SVM)", tujuan dari penelitian ini berguna untuk melakukan analisis terkait beasiswa KIP-Kuliah menggunakan *Support Vector Machine* yang menghasilkan akurasi 86,27% yang hasil algoritma model lebih bias kepada sentimen negatif [8].

Pada penelitian kali ini, berbeda dengan penelitian terdahulu lainnya penelitian ini menggunakan metode berbasis *Support Vector Machine* dengan menambahkan pembobotan kata dan juga penggunaan *Synthetic Minority Over-sampling Technique (SMOTE)*, sehingga menjadikan penelitian ini berbeda dengan penelitian terdahulu sebelumnya.

3. Metodologi

Pada penelitian menggunakan penelitian analisis sentimen dengan menggunakan SVM sebagai algoritma dalam melakukan analisis sentimennya berikut metodologi penelitian yang dapat dilihat pada gambar 1.



Gambar 1. Metode Penelitian

3.1 Data Gathering

Pada tahapan pengumpulan data dilakukan dengan metode *web scraping* yakni metode yang mengekstrak informasi yang terjadi pada suatu halaman, hal ini bisa dilakukan pada python yakni menggunakan tools bernama *tweet harvest, ig comments export tool,* dan *Python.*

3.2 Cleansing Data

Tahapan ini membersihkan semua data hasil dari *gathering data* dengan membersihkan semua url, html, dan emoji dan karakter asing lainnya pada data sehingga dapat dilanjutkan pada proses *preprocessing data* sehingga data yang ingin diolah sudah dapat diolah dengan baik.

3.3 Preprocessing data

1) Case Folding

Case Folding merupakan tahapan dimana dilakukan penyamaan pada tulisan yang sebelumnya tidak beraturan menjadi *lower case*. Karakter seperti huruf dan angka akan dianggap sebagai *delimiter* sehingga dihilangkan.

2) Normalize

Pada tahapan ini penggunaan normalisasi berfungsi untuk mengembalikan kata yang bermakna lain sesuai dengan konteks dalam kamu KBBI, sehingga tidak menimbulkan kebingungan saat proses analisis sentimen.

3) Remove Stopwords

Stopwords berfungsi untuk menghilangkan kata-kata yang tidak penting dalam teks dan tidak ada hubungannya dengan isi teks yang dimaksudkan, dan juga tujuan dari pemakaian Stopwords adalah meningkatkan kinerja algoritma agar tepat dalam menganalisis sentimen [9].

4) Tokenizing

Proses *tokenizing* adalah memecah kata per kata, pada proses ini juga dilakukan juga penghilangan angka, tanda baca, dan tanda lain yang tidak memiliki pengaruh signifikan pada teks terkait.

5) Stemming

Stemming merupakan proses pengembalian kata ke bentuk aslinya sehingga kata yang sudah dilakukan proses *tokenizing* dikembalikan kepada bentuk kalimat yang sudah tidak ada tanda baca dan tanda lain yang mengganggu pemrosesan teks.

3.4 Labeling Data

Setelah proses pengumpulan data selesai dilakukan proses *labelling data* yakni melakukan pelabelan data jika data n>1 maka dikatakan positif sedangkan apabila data n=0 maka dikatakan netral sedangkan n<0 maka data dikatakan negatif.

3.5 Feature Selection

Term Frequency - Inverse Document Frequency (TF-IDF) merupakan teknik yang menilai hubungan antara kata dan dokumen dengan menggunakan dua konsep penghitungan bobot: frekuensi kata dalam dokumen tertentu dan frekuensi inverse dari dokumen yang mengandung kata tersebut [10]. *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) membentuk TF-IDF. Frekuensi kemunculan sebuah kata dalam manuskrip dikenal sebagai TF-nya. Sebaliknya, IDF adalah kebalikan dari frekuensi dokumen [10].

3.6 SMOTE: Synthetic Minority Over-sampling Technique

Synthetic Minority Over-sampling Technique atau biasa disingkat dengan *SMOTE* adalah teknik yang kelebihan beban untuk mengkompensasi jumlah distribusi catatan data di kelas minoritas dengan menyelaraskan catatan data minoritas dengan jumlah catatan data di kelas mayoritas [11].

3.7 SVM Classification

Support Vector Machine adalah pertama kali diperkenalkan oleh Cortes dan Vapnik (1995) sebagai cara untuk menjaga agar nilai dari bahaya eksperimental tetap stabil, meminimalkan ketidakpastian sementara [12]. Strategi ini digunakan untuk menemukan jalur hiper terbaik dengan meningkatkan jarak antar lapisan [13]. Berikut merupakan sistematika kinerja dari Algoritma SVM.

$$w \cdot x_i + b = 0 \quad (1)$$

Keterangan:

- w = vektor bobot
- xi = atribut suku ke-i pada data
- b = suku bias

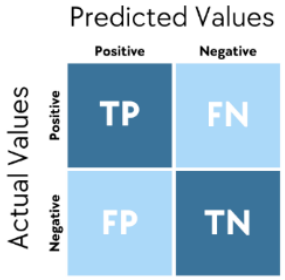
Dalam konteks perumusan sentimen negatif dan positif pada rumus 2 untuk positif dan negatif pada rumus 3.

$$w \cdot x_i + b \geq 1 \quad (2)$$

$$w \cdot x_i + b \leq -1 \quad (3)$$

3.8 Evaluation with Confusion Matrix

Pada penelitian ini menggunakan evaluasi *confusion matrix*, pada dasarnya *confusion matrix* digunakan sebagai tolak ukur kinerja setelah data diolah dengan model yang sudah ditentukan dengan sedemikian rupa [14]. Pengukuran menggunakan metode ini berguna dalam memberikan informasi perbandingan hasil algoritma yang digunakan dalam hasil klasifikasi yang sudah digunakan sebelumnya [14].



Gambar 2. *Confusion Matrix*

Keterangan:
1. TP *(True Positive)* = Data yang bernilai 1 diklasifikasikan sebagai nilai 1
2. FP *(False Positive)* = Data yang bernilai 0 diklasifikasikan sebagai nilai 0
3. TN *(True Negative)* = Data yang bernilai 0 diklasifikasikan sebagai nilai 1
4. FN *(False Negative)* = Data yang bernilai 1 diklasifikasikan sebagai nilai 0

Pada penggunaannya *confusion matrix* sering digunakan dalam 666able666n data untuk mencari nilai *accuracy, precision, recall,* dan *f1-score*. Berikut beberapa rumus untuk menghitungnya.

1. *Accuracy*

Nilai ini menggambarkan seberapa akurat suatu algoritma dalam melakukan pengklasifikasian suatu data tertentu.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

2. *Precision*

Precision disini digunakan dalam penggambaran akurasi antara data yang diminta dengan hasil prediksi data yang sudah dilakukan oleh model algoritma.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

3. *Recall*

Recall merupakan penggambaran hasil evaluasi model dalam menghasilkan pengklasifikasian kelas positif dengan benar.

$$Recall = \frac{TP}{TP+FN} \times 100\ \% \quad (5)$$

4. *f1-score*

f1-score merupakan pencerminan keseimbangan antara *precision* dan *recall* yang telah di bobotkan.

$$f1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \times 100\% \quad (6)$$

4. Hasil dan Pembahasan

4.1 Data Gathering

Proses *data gathering* dilaksanakan dari awal pembahasan tentang revisi RUU TNI yang dimulai pada tanggal 24 maret 2025 sejak dimulainya tagar RUU TNI dari berbagai kanal media sosial dimulai dari *youtube, Instagram* dan *x*. Dengan hasil data yang ditunjukkan pada gambar 3, 4 dan 5.

	publishedAt	authorDisplayName	textDisplay	likeCount
0	2025-04-25T00:31:25Z	@Soppo-kp1sp	TOLONG YA PAK PRABOWO.... DPR..... YANG BENER ...	0
1	2025-04-23T13:04:27Z	@__-wt8qk	Revisi terbaru UU TNI menuai kritik tajam kare...	1
2	2025-04-19T16:00:05Z	@irmanamonk6085	Materinya sangat dapat dipahami, tapi kenapa m...	0
3	2025-04-16T09:20:49Z	@setyawandanurwendha5383	TNI hars berinovasai bikin super soldier seper...	0
4	2025-04-12T02:23:06Z	@Baharichanel121	Tidak ada Dwi Pungsi kembali Pemerintahan sipi...	0
...	...	...	...	...
1396	2025-03-23T08:32:50Z	@cillockcube	@@Payday5sebutkan apa bagusnya anies? Klau ngo...	0
1397	2025-03-23T08:37:09Z	@Payday5	@@cillockcube gua tanya balik bagusnya prabowo...	1
1398	2025-03-23T08:45:31Z	@cillockcube	@ milih pemimpin ga cuma dri janji2 kampanye, ...	0
1399	2025-03-23T07:05:36Z	@farizandrojuventino24	mampir	0
1400	2025-03-23T07:02:17Z	@Random_Netizen_	Hmm...	1

1401 rows × 4 columns

Gambar 3. Data Youtube

	user_id	profile_pic_url	text	username
0	44360008603	https://instagram.fsoc1-2.fna.fbcdn.net/v/t51....	Heeemm miris btl	sukmayaaris
1	65390624368	https://instagram.fsoc1-1.fna.fbcdn.net/v/t51....	Aneh dr pasal2 itu saja sdh kelihatan ga ada y...	akhtarmalik4887
2	13972150218	https://instagram.fsoc1-2.fna.fbcdn.net/v/t51....	Yang harusnya berada pada zona netral kini sud...	putriandayanimuslim
3	1506741981	https://instagram.fsoc1-2.fna.fbcdn.net/v/t51....	@rsyafiims mungkin disini yg di maksut kerjaa...	dodikbintoroe
4	2070272562	https://instagram.fsoc1-2.fna.fbcdn.net/v/t51....	@ptrdr 19 juta lapangan pekerjaan 😂	fadhil.shl
...	...	...	...	...
2187	7044990267	https://instagram.fsoc1-1.fna.fbcdn.net/v/t51....	@ru.unplugged	usamthings
2188	64253264668	https://instagram.fsoc1-1.fna.fbcdn.net/v/t51....	@mfrtzky94 blokk	rumah_puisi03
2189	4580994133	https://instagram.fsoc1-1.fna.fbcdn.net/v/t51....	Yaaa masuk kejaksaan 😂	krismukk
2190	4180703319	https://instagram.fsoc1-2.fna.fbcdn.net/v/t51....	Terima kasih kaum 56% kalian membuat kami semu...	_aafadli
2191	42580144839	https://instagram.fsoc1-2.fna.fbcdn.net/v/t51....	Bisaaa biar mrk kaya... korup atau perbuatan s...	boeseey

2192 rows × 4 columns

Gambar 4. Data Instagram

	created_at	full_text	username
0	Mon Mar 24 23:54:40 +0000 2025	@puanmaharani_ri Di luar sana lagi banyak yg d...	canopusky
1	Mon Mar 24 23:54:04 +0000 2025	Di luar sana lagi banyak yg demo minta batalka...	canopusky
2	Mon Mar 24 23:43:42 +0000 2025	Di Traffic Light simpang tiga Ngangkruk Ds Nga...	wartabengawan
3	Mon Mar 24 23:35:44 +0000 2025	Masyarakat yang turun ke jalanan untuk aksi de...	Asep727446
4	Mon Mar 24 23:15:27 +0000 2025	@woonhakpoca kiyowookkk	sanhiraeth10
...	...	...	...
1467	Sun Apr 20 13:13:44 +0000 2025	Apakah langka ya Wakil Rakyat yg benar2 menyua...	mikuroQ
1468	Sun Apr 20 12:24:35 +0000 2025	@fairylixee baca dikit-dikit aja kan udah usaha	hynciwww
1469	Sun Apr 20 08:05:48 +0000 2025	@panggilakunorma Coretax mbg danantara korupsi...	567O23
1470	Sun Apr 20 01:53:34 +0000 2025	Revisi UU TNI baik untuk pertahanan dan keaman...	mediamediasi
1471	Sun Apr 20 00:54:08 +0000 2025	@creeping5ilence Pengesahan UU TNI bukan memun...	NabilaKeisha7

1472 rows × 3 columns

Gambar 5. Data X

4.2 Cleansing Data

Tahapan ini membersihkan semua data hasil dari *gathering data* dengan membersihkan semua url, html, dan emoji dan karakter asing lainnya pada data sehingga dapat dilanjutkan pada proses *preprocessing data.* Pada tahapan ini dapat mempengaruhi analisis teks pada proses selanjutnya, apabila tidak dilakukan pada tahapan ini maka pada saat praproses data maka akan terjadi kerancuan saat algoritma dalam mengolah data yang tidak dibersihkan secara benar. Data hasil *cleansing* ditampilkan pada gambar 6 ,7, dan 8.

	authorDisplayName	textDisplay	likeCount
0	@Soppo-kp1sp	TOLONG YA PAK PRABOWO DPR YANG BENER donkkkk J...	0
1	@__-wt8qk	Revisi terbaru UU TNI menuai kritik tajam kare...	1
2	@irmanamonk6085	Materinya sangat dapat dipahami tapi kenapa mb...	0
3	@setyawandanurwendha5383	TNI hars berinovasai bikin super soldier seper...	0
4	@Baharichanel121	Tidak ada Dwi Pungsi kembali Pemerintahan sipi...	0
...	...	...	...
1396	@cillockcube	apa bagusnya anies Klau ngomong ga pernah tu t...	0
1397	@Payday5	gua tanya balik bagusnya prabowo apa Orba bali...	1
1398	@cillockcube	milih pemimpin ga cuma dri janji kampanye tpi ...	0
1399	@farizandrojuventino24	mampir	0
1400	@Random_Netizen_	Hmm	1

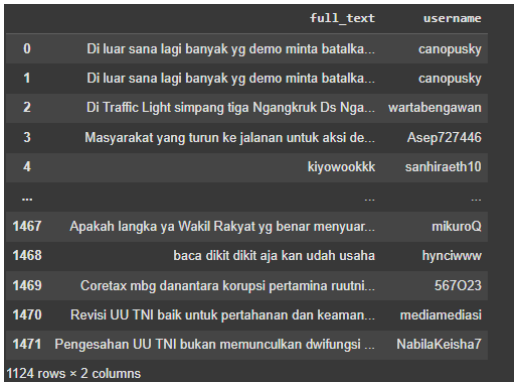
1388 rows × 3 columns

Gambar 6. Cleansing youtube

	user_id	text	username
0	44360008603	Heeemm miris btl	sukmayaaris
1	65390624368	Aneh dr pasal itu saja sdh kelihatan ga ada yg...	akhtarmalik4887
2	13972150218	Yang harusnya berada pada zona netral kini sud...	putriandayanimuslim
3	1506741981	mungkin disini yg di maksut kerjaaan tni namba...	dodikbintoroe
4	2070272562	juta lapangan pekerjaan	fadhil.shl
...	...	...	...
2187	7044990267	unplugged	usamthings
2188	64253264668	blokk	rumah_puisi03
2189	4580994133	Yaaa masuk kejaksaan	krismukk
2190	4180703319	Terima kasih kaum kalian membuat kami semua ke...	_aafadli
2191	42580144839	Bisaaa biar mrk kaya korup atau perbuatan sala...	boeseey

2021 rows × 3 columns

Gambar 7. Cleansing Instagram



	full_text	username
0	Di luar sana lagi banyak yg demo minta batalka...	canopusky
1	Di luar sana lagi banyak yg demo minta batalka...	canopusky
2	Di Traffic Light simpang tiga Ngangkruk Ds Nga...	wartabengawan
3	Masyarakat yang turun ke jalanan untuk aksi de...	Asep727446
4	kiyowookkk	sanhiraeth10
...	...	...
1467	Apakah langka ya Wakil Rakyat yg benar menyuar...	mikuroQ
1468	baca dikit dikit aja kan udah usaha	hynciwww
1469	Coretax mbg danantara korupsi pertamina ruutni...	567O23
1470	Revisi UU TNI baik untuk pertahanan dan keaman...	mediamediasi
1471	Pengesahan UU TNI bukan memunculkan dwifungsi ...	NabilaKeisha7

1124 rows × 2 columns

Gambar 8. Cleansing x

4.3 Preprocessing Data

Case folding, normalize, stopwords, tokenizing, dan *stemming* dilakukan pada proses ini. Proses yang terjadi pada tahapan ini ialah mengubah semua karakter yang kapital menjadi huruf kecil. Selanjutnya melakukan penghapusan tanda baca dan karakter asing seperti titik, tanda tanya, tanda seru, koma, dan penghapusan simbol asing, kata henti, mengubah kata majemuk menjadi kata dasar kemudian pemisahan teks.

Tabel 1 Contoh Hasil *preprocessing data*

Sebelum	Sesudah
[‘pemerintah’, ‘indonesia’, ‘memastikan’, ‘revisi’, ‘undang’, ‘undang’, ‘tentara’, ‘nasional’, ‘indonesia’, ‘tentara’, ‘nasional’, ‘indonesia’, ‘baru’, ‘disahkan’, ‘bertujuan’, ‘mengembalikan’, ‘dwifungsi’, ‘tentara’, ‘nasional’, ‘indonesia’, ‘lebih’, ‘difokuskan’, ‘penguatan’, ‘peran’, ‘tentara’, ‘nasional’, ‘indonesia’, ‘menjaga’, ‘stabilitas’, ‘negara’]	perintah indonesia pasti revisi undang undang tentara nasional indonesia tentara nasional indonesia baru sah tuju kembali dwifungsi tentara nasional indonesia lebih fokus kuat peran tentara nasional indonesia jaga stabilitas negara

4.4 Labeling Data

Pada tahapan pelabelan data kali ini menggunakan library dari python yang bernama Textblob sebelum melakukan pelabelan data menggunakan Textblob secara otomatis, library ini mewajibkan untuk melakukan penerjemahan kedalam bahasa inggris terlebih dahulu sebelum melakukan pelabelan data yang terkait. Hasil dari terjemahan dan sentimen dapat dilihat pada tabel 2.

Tabel 2. Contoh hasil *labelling*

Sebelum	Sesudah	Sentimen
perintah indonesia pasti revisi undang undang tentara nasional indonesia tentara nasional indonesia baru sah tuju kembali dwifungsi tentara nasional indonesia lebih fokus kuat peran tentara nasional indonesia jaga stabilitas negara	the indonesian government ensured that the revision of the indonesian national army national army law was approved aimed at restoring the dual function of the indonesian national army more focused on strengthening the role of the indonesian national army to maintain the stability of the country	**Positif**

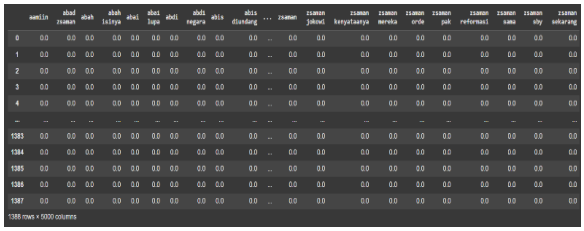
Dari hasil pelabelan data menggunakan *textblob* dari data dari *youtube* mendapatkan data sebanyak 1.385 data dengan jumlah sentimen positif sebanyak 580, netral sebanyak 548, dan negatif sebanyak 257.Sedangkan untuk data dari *instagram* mendapatkan data sebanyak 2.021

data dengan jumlah sentimen positif sebanyak 635, netral sebanyak 1.096, dan negatif sebanyak 290. Terakhir untuk data dari *X* mendapatkan data sebanyak 1.124 data dengan jumlah sentimen positif sebanyak 398, netral sebanyak 517, dan negatif sebanyak 296. Dengan total data dari 3 media sosial berjumlah 4.530 data dengan jumlah sentimen positif sebanyak 1.618, sentimen netral sebanyak 2.161 dan sentimen negatif sebanyak 843.

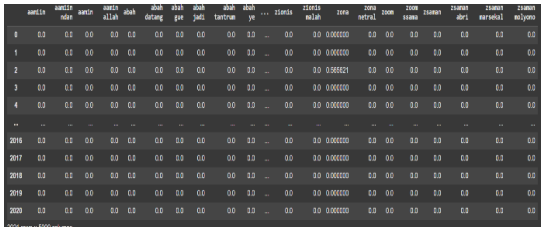
Disini dapat dilihat hasil analisis sentiment dari *textblob* menunjukkan secara garis besar bahwa mayoritas sentimen tertuju pada arah netral namun dapat dilihat bahwa ada beberapa data yang menunjukkan ketimpangan seperti pada data yang ada pada media *instagram* dengan data negatif yang cukup minim, maka dari itu kita akan melakukan proses selanjutnya yakni *feature selection* dan *smote* guna menyeimbangkan data yang sudah dilakukan pelabelan tersebut.

4.5 Feature Selection

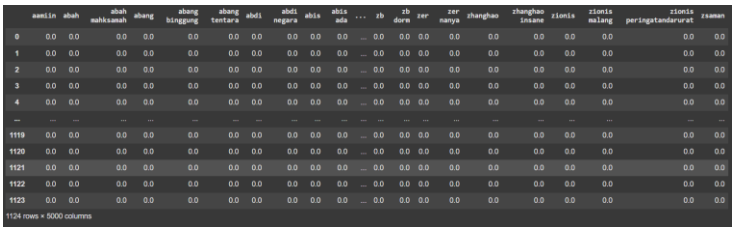
Sebelum proses *smote* dilakukan, terlebih dahulu akan dilakukan proses pembobotan kata atau tf idf terlebih dahulu dikarenakan agar kata yang sering muncul dalam dokumen dapat digunakan saat *balancing* data saat dilakukannya proses *smote* yang akan meningkatkan akurasi algoritma mesin. Hasil dari *tf idf* dapat dilihat pada gambar 9, 10, dan 11.



Gambar 9. Hasil TF IDF Youtube



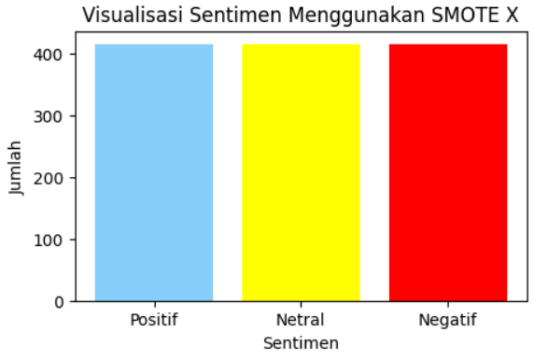
Gambar 10. Hasil TF IDF Instagram



Gambar 11. Hasil TF IDF X

4.6 SMOTE (Synthetic Minority Over-sampling Technique)

Setelah dilakukan proses pembobotan kata atau *tf idf* selanjutnya akan dilakukan proses *balancing data* atau *smote* yakni menjadikan data kelas minoritas diselaraskan dengan jumlah data minoritas dengan data mayoritas yang ada pada suatu data. *SMOTE* meningkatkan kinerja prediksi model pembelajaran mesin dengan menyeimbangkan dataset dengan cara ini, mencegah bias terhadap kelas mayoritas dan memungkinkan generalisasi yang lebih besar pada data yang belum terlihat.



Gambar 12. Contoh hasil Visualisasi SMOTE

4.7 SVM Classification

Tahap selanjutnya dalam setelah dilakukan penyelarasan pada data dari 3 media sosial yang ada yakni *youtube, Instagram,* dan *x* selanjutnya yakni pengklasifikasian menggunakan algoritma mesin menggunakan *support vector machine*, dengan menggunakan parameter c yakni dari 0.01, 0.05, 0.25, 0.5, 0.75, dan 1. Berikut hasil uji dari data dari 3 media sosial yang sudah diolah yang dapat dilihat pada tabel 3.

Tabel 3. Hasil Klasifikasi Data Media Sosial

Complexity	**Accuracy**		
	Youtube	*Instagram*	*X*
0,01	42.81%	60.74%	45.33%
0,05	68.35%	81.98%	81.78%
0,25	83.09%	88.89%	93.33%
0,5	89.93%	93.33%	96.00%
0,75	92.81%	94.32%	97.33%
1	94.60%	95.80%	97.33%

Dapat dilihat dari ketiga tabel diatas parameter C = 1 membuktikan bahwa memiliki rata-rata akurasi yang tinggi dari ketiga data berdasarkan hasil olah data dari pembobotan kata dan penyelarasan data terkait. Hasil akurasi yang didapat sangat bagus yakni 94,60% untuk *youtube*, 95,80% untuk *instagram*, dan 97,33% untuk *x*. Berikut adalah hasil laporan lengkap dari hasil pengolahan data terkait yang dapat dilihat pada gambar 13, 14, dan 15.

	precision	recall	f1-score	support
Negatif	0.94	1.00	0.97	49
Netral	0.98	0.91	0.94	118
Positif	0.91	0.96	0.94	111
accuracy			0.95	278
macro avg	0.95	0.96	0.95	278
weighted avg	0.95	0.95	0.95	278

Akurasi nya adalah : 94.60%
Precision nya adalah : 94.79%
Recall nya adalah : 94.60%
F1 Score nya adalah : 94.59%

Gambar 13. Hasil laporan klasifikasi youtube

	precision	recall	f1-score	support
Negatif	0.93	0.97	0.95	66
Netral	0.98	0.95	0.96	212
Positif	0.95	0.97	0.96	127
accuracy			0.96	405
macro avg	0.95	0.96	0.96	405
weighted avg	0.96	0.96	0.96	405

Akurasi nya adalah : 95.80%
Precision nya adalah : 95.86%
Recall nya adalah : 95.80%
F1 Score nya adalah : 95.81%

Gambar 14. Hasil laporan klasifikasi instagram

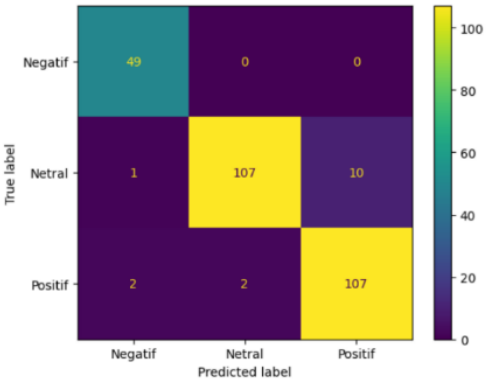
	precision	recall	f1-score	support
Negatif	0.98	1.00	0.99	53
Netral	0.99	0.95	0.97	102
Positif	0.95	0.99	0.97	70
accuracy			0.97	225
macro avg	0.97	0.98	0.98	225
weighted avg	0.97	0.97	0.97	225

Akurasi nya adalah : 97.33%
Precision nya adalah : 97.40%
Recall nya adalah : 97.33%
F1 Score nya adalah : 97.33%

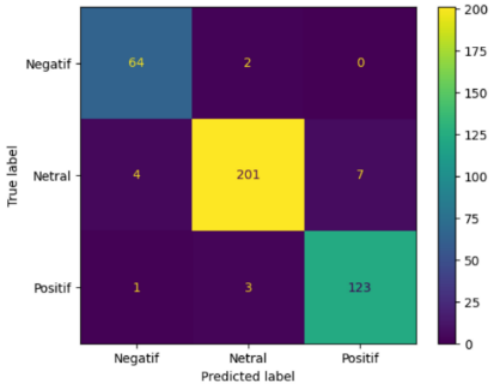
Gambar 15. Hasil laporan klasifikasi X

4.8 Evaluation with confusion matrix

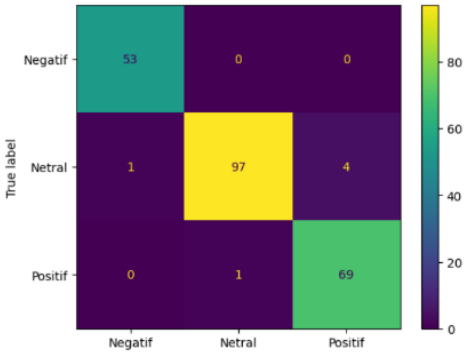
Selanjutnya setelah melakukan klasifikasi dengan algoritma *support vector machine*, pada tahap ini kita akan melakukan evaluasi dengan bantuan *confusion matrix* untuk mengevaluasi hasil kinerja dari algoritma yang sudah kita bangun untuk menganalisa sentimen RUU TNI pada ketiga media sosial.



Gambar 16. Hasil confusion matrix youtube



Gambar 17. Hasil confusion matrix



Gambar 18. Hasil confusion matrix x

Dapat dilihat pada gambar 16, bahwa untuk label negatif dari 49 data uji yang diujikan pada algoritma terkait menyatakan bahwa 49 label negatif yang terkait diprediksikan benar sebagai negatif, untuk label netral dari 118 label netral untuk data uji 107 diantaranya diprediksikan bahwa benar merupakan label netral dan 10 diantaranya yang sebenarnya netral diprediksikan sebagai positif dan 1 sentimen diantaranya yang sebenarnya netral diprediksikan sebagai negatif. Selanjutnya untuk sentimen positif dari 111 data uji diprediksikan bahwa 107 data diprediksikan sebagai benar bahwa merupakan sentimen positif dan 2 diantaranya diantaranya yang sebenarnya positif diprediksikan sebagai netral dan 2 lainnya yang sebenarnya positif diprediksikan sebagai negatif.

Dapat dilihat pada gambar 17, bahwa untuk label negatif dari 64 data uji yang diujikan pada algoritma terkait menyatakan bahwa 64 label negatif yang terkait diprediksikan benar sebagai negatif, dan 2 diantaranya yang sebenarnya negatif diprediksikan sebagai netral, untuk label netral dari 212 label netral untuk data uji 201 diantaranya diprediksikan bahwa benar merupakan label netral dan 7 diantaranya yang sebenarnya netral diprediksikan sebagai positif dan 4 sentimen yang sebenarnya netral diprediksikan sebagai negatif. Selanjutnya untuk sentimen positif dari 126 data uji diprediksikan bahwa 123 data diprediksikan sebagai benar bahwa merupakan sentimen positif dan 3 yang sebenarnya positif diantaranya diprediksikan sebagai netral dan 1 lainnya yang sebenarnya positif diprediksikan sebagai negatif.

Dapat dilihat pada gambar 18, bahwa untuk label negatif dari 53 data uji yang diujikan pada algoritma terkait menyatakan bahwa 53 label negatif yang terkait diprediksikan benar sebagai negatif, untuk label netral dari 102 label netral untuk data uji 97 diantaranya diprediksikan bahwa benar merupakan label netral dan 4 diantaranya yang sebenarnya netral diprediksikan sebagai positif dan 1 sentimen yang sebenarnya netral diprediksikan sebagai negatif. Selanjutnya untuk sentimen positif dari 70 data uji diprediksikan bahwa 69 data diprediksikan sebagai benar bahwa merupakan sentimen positif dan 1 diantaranya yang sebenarnya positif diprediksikan sebagai netral.

Tabel 4. Hasil Sentimen dari 3 Media Sosial

Sentimen	*Youtube*	*Instagram*	*X*
Positif	41.6%	32.6%	41.6%
Netral	38.9%	52.7%	38.9%
Negatif	19.5%	14.7%	19.5%

Dapat dilihat pada tabel 4 diatas hasil dari ketiga media sosial memiliki perbedaan, untuk *youtube* sendiri memiliki kecenderungan komentar yang lebih kepada arah sentimen positif sebanyak 41.6%, sedangkan untuk *instagram* cenderung lebih kepada komentar netral sebanyak 52.7% sedangkan pada aplikasi *x* lebih cenderung kepada sentimen netral sebanyak 46%. Dari hasil klasifikasi menggunakan algoritma *SVM* secara garis besar dari ketiga sentimen berdasarkan data dari ketiga media sosial lebih cenderung kepada sentimen positif dan netral.

4.9. Pembahasan

Dari pembelajaran mesin yang telah dilakukan dari ketiga dataset yang sudah dikumpulkan selama beberapa hari performa terbaik dihasilkan oleh performa dengan dataset dari x dengan Tingkat keakuratan mencapai 97.33%, sementara pada data set akurasi tertinggi 94.60 % untuk youtube dan 95.80% untuk Instagram, meskipun terlihat tinggi akurasi dan sudah dilakukan penyelarasan dalam jumlah data yang tersebar dapat dilihat masih terdapat bias sentimen positif dan netral, dapat dilihat pada hasil sentimen pada instagram yang memiliki bias netral yang tinggi yakni 52.7%. Bias netral ini dapat disebabkan oleh beberapa komentar yang tidak sesuai dengan konteks pembicaraan yang sedang berlangsung sehingga terjadi kebiasan terhadap suatu topik tertentu.

Seperti yang dikemukakan oleh Nurhasananda dan Akbar pada “Analisis Sentimen Masyarakat Terhadap Kebijakan Undang-Undang Tentara Nasional Indonesia (UU TNI) Menggunakan *Support Vector Machine*”, Penelitian mereka tidak mencakup beberapa media sosial melainkan hanya 1 media sosial saja yakni Twitter dan juga penggunaan metode lain dalam meningkatkan akurasi hasil klasifikasi [15]. Selain itu juga pada penelitian Hasibuan dkk dalam “Analisis Sentimen Kebijakan Ekspor Pasir Laut Pada Sosial Media Twitter Menggunakan Algoritma *Support Vector Machine*”, menyarankan bahwasanya harus ada dataset yang lebih banyak dari platform media sosial lain [16]. Dengan demikian, kontribusi utama dari penelitian ini adalah penggabungan antar platform, kualitas data, dan teknik pengolahan seperti pembobotan kata dan SMOTE, yang secara kolektif meningkatkan reliabilitas hasil klasifikasi sentimen kebijakan TNI.

Dengan integrasi kontribusi ini, pemahaman kita tentang bagaimana mengoptimalkan SVM dalam konteks opini kebijakan nasional semakin kuat. Oleh karena itu, penelitian ini berkontribusi pada pengembangan metode klasifikasi sentimen berbasis machine learning yang multikanal dan fleksibel. Metode ini relevan untuk studi kebijakan, sosial-politik, dan media digital.

5. Simpulan

Berdasarkan hasil dari penelitian yang dilakukan dari ketiga data dari media sosial sebanyak 4.530 data yang diolah dapat disimpulkan bahwa algoritma *Support Vector Machine* dapat membantu dalam melakukan penganalisaan terhadap sentimen masyarakat yang dibantu dengan TF-IDF dan juga SMOTE dalam suatu topik tertentu, yang pada penelitian kali ini menitikberatkan pada topik ruu tni. Dapat dilihat hasil tuning dari parameter dengan menggunakan parameter C yang mendapatkan akurasi tertinggi ialah menggunakan parameter 1 dengan youtube sebesar 94.60%, instagram sebesar 95,80% dan x sebesar 97,33%. Namun mengingat menggunakan textblob sebagai alat untuk melakukan pelabelan otomatis membawa keterbatasan karena tidak sepenuhnya mampu menangkap makna sebenarnya dari nuansa politik yang bergejolak di Indonesia.

Dapat diambil benang merah bahwa ketiga media sosial yang terkait memiliki sarana dalam menyebarkan sentimen masyarakat terhadap suatu isu tertentu baik dalam skala nasional maupun global. Penelitian ini merupakan bagian dasar untuk pemerintah Indonesia dalam analisis sentimen masyarakat terkhususnya dalam isu RUU TNI, yang akan terus mengalami perkembangan seiring dengan perubahan kebijakan dan waktu. Budaya digital di kalangan masyarakat juga mempengaruhi cara mereka berkomunikasi, termasuk dalam menyampaikan pendapat lewat media sosial. Memahami opini publik dalam penelitian ini sangat bermanfaat untuk menggali opini masyarakat di dunia digital.

Namun, pada penelitian ini masih memiliki kekurangan yakni masih menggunakan 1 kernel yang tersedia pada algoritma SVM diharapkan pula pada penelitian yang akan datang dapat memakai kernel yang lain selain itu juga menggunakan optimasi parameter tuning yang lebih kompleks dan juga melakukan pelabelan secara manual dan menyeluruh sehingga menghasilkan analisis sentimen yang kuat terhadap suatu topik tertentu baik isu global maupun nasional.

Daftar Referensi

[1] Dewan Perwakilan Rakyat Republik Indonesia, "Rancangan Undang Undang Perubahan Atas Undang-UNDANG NOMOR 34 Tahun 2004 Tentang Tentara Nasional Indonesia," in *Sidang Paripurna*, Jakarta, 2024.

[2] P. Durga and D. Godavarthi, "Deep-Sentiment: An Effective Deep Sentiment Analysis using a Decision-based Recurrent Neural Network (D-RNN)," *IEEE Access,* vol. 11, no. 2, pp. 108433-108447, 2023.

[3] B. H. Hayadi and I. Maulita, "Sentiment Analysis of Public Discourse on Education in Indonesia Using Support Vector Machine (SVM) and Natural Language Processing," *Journal of Digital Society,* vol. 1, no. 1, pp. 68-90, 2025.

[4] Y. V. Wijaya, A. Erfina and C. Warman, "Analisis Sentimen Seputar UU ITE Menggunakan Algoritma Support Vector Machine," *Progresif: Jurnal Ilmiah Komputer,* vol. 17, no. 2, pp. 1-14, 2021.

[5] Wardianto, P. M. Jakak and M. Rohman, "Analisis Sentimen Public Program Makan Bergizi Gratis Platform Instagram dengan Algoritma Support Vector Machine," *SMARTICS,* vol. 11, no. 1, pp. 14-20, 2025.

[6] A. M. Husein, B. Sipahutar, R. Dashuah and E. Hutauruk, "Sentiment Analysis Od Face To Face School Policy On Twitter Social Media With Support Vector Machine(SVM)," *Sinkron : Jurnal Dan Penelitian Teknik Informatika,* vol. 7, no. 1, pp. 480-486, 2023.

[7] P. M. N. Dharmapatni and N. L. P. Merawati, "Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan," *Jurnal Bumigora Information Technology (BITe),* vol. 2, no. 2, pp. 105-112, 2020.

[8] I. Amelia, Sugiyono, F. M. Sarimole and Tundo, "Analisis Sentimen Tanggapan Pengguna Media Sosial X Terhadap Program Beasiswa KIP-Kuliah dengan Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi,* vol. 5, no. 3, pp. 2994-3003, 2024.

[9] C. B. Vista, O. M. Sihono and A. T. Firdausi, "Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Selama Pandemi Covid-19 Menggunakan Metode Support Vector Machine," *Jurnal Informatika Polinema,* vol. 9, no. 3, pp. 259-264, 2023.

[10] O. I. Gifari, M. Adha, I. R. Hendrawan and F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *Journal of Information Technology,* vol. 2, no. 1, pp. 36-40, 2022.

[11] H. Leidiyana, T. Misriati and R. Aryanti, "Klasifikasi Sentimen Terhadap Kebijakan Tapera Menggunakan Komparasi Machine Learning dan SMOTE," *Jurnal Komtika (Komputasi dan Informatika),* vol. 8, no. 2, pp. 125-135, 2024.

[12] International Business Machines, "Apa itu Support Vector Machines?," International Business Machines, 27 Desember 2023. [Online]. Available: https://www.ibm.com/id-id/think/topics/support-vector-machine. [Accessed 28 Juli 2025].

[13] F. Rahmadayana and Y. Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi),* vol. 5, no. 5, pp. 936-942, 2021.

[14] P. L. Romadloni, B. A. Kusuma and W. M. Baihaqi, "Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan," *JATI(Jurnal Mahasiswa Teknik Informatika),* vol. 6, no. 2, pp. 622-628, 2022.

[15] A. A. Nurhasananda and M. Akbar, "Analisis Sentimen Masyarakat Terhadap Kebijakan Undang-Undang Tentara Nasional Indonesia (UU TNI) Menggunakan Support Vector Machine," *Jurnal Komputer, Informasi Dan Teknologi,* vol. 5, no. 2, pp. 1-14, 2025.

[16] R. A. Hasibuan, D. E. Ratnawati and R. S. Perdana, "Analisis Sentimen Kebijakan Ekspor Pasir Laut pada Sosial Media Twitter Menggunakan Algoritma Support Vector Machine," *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi,* vol. 5, no. 1, pp. 24-33, 2024.