

Klasifikasi Stance Opini Publik Komentar TikTok Kasus Guru Menampar Murid memakai TF-IDF dan SVM

Salsabila Yusra*, Raissa Amanda Putri

Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Sumatera Utara, Medan, Indonesia

Email: ^{1,*}salsabilayusra086@gmail.com, ²raissa.ap@uinsu.ac.id

Email Penulis Korespondensi: salsabilayusra086@gmail.com*

Submitted: 29/01/2026; Accepted: 13/02/2026; Published: 31/03/2026

Abstrak– Media sosial TikTok menjadi salah satu platform utama dalam menyampaikan opini publik terhadap isu sosial yang bersifat viral, termasuk kasus guru yang menampar murid karena merokok di sekolah. Kasus ini memunculkan perbedaan pandangan yang tajam di masyarakat antara pihak yang mendukung tindakan guru sebagai bentuk pendisiplinan dan pihak yang menilai tindakan tersebut sebagai bentuk kekerasan dalam pendidikan. Penelitian ini bertujuan untuk menganalisis sentimen publik di TikTok terhadap kasus tersebut menggunakan metode Support Vector Machine (SVM). Data penelitian berupa 2.200 komentar publik dikumpulkan melalui teknik web scraping menggunakan Apify Web Scraper dan diolah menggunakan Google Colab. Tahapan penelitian meliputi preprocessing teks yang terdiri dari cleaning, case folding, tokenization, normalisasi, stopword removal, dan stemming. Selanjutnya, fitur diekstraksi menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) dan diklasifikasikan ke dalam empat kategori sentimen, yaitu Mendukung Guru, Menyalahkan Guru, Mendukung Murid, dan Menyalahkan Murid. Hasil penelitian menunjukkan bahwa metode SVM dengan kernel linear berhasil mengklasifikasikan sentimen publik dengan akurasi 94% (precision 0.94, recall 0.94, F1-score 0.94). Distribusi sentimen menunjukkan 69.2% komentar mendukung guru, 20.8% mendukung murid, 5.5% menyalahkan guru, dan 4.5% menyalahkan murid. Hasil ini memberikan gambaran komprehensif bahwa mayoritas masyarakat cenderung membenarkan tindakan disiplin guru, meskipun terdapat kesadaran yang berkembang terhadap perlindungan anak dan penolakan kekerasan dalam dunia pendidikan.

Kata Kunci: Analisis Sentimen; TikTok; Support Vector Machine; Text Mining; Media Sosial

Abstract– TikTok has become one of the primary social media platforms for expressing public opinion on viral social issues, including the case of a teacher slapping a student for smoking at school. This case triggered sharp divisions in public opinion between those who supported the teacher's action as a form of discipline and those who considered it an act of violence in education. This study aims to analyze public sentiment on TikTok regarding this case using the Support Vector Machine (SVM) method. The dataset consists of 2,200 public comments collected through web scraping using Apify Web Scraper and processed using Google Colaboratory. The research stages include text preprocessing, namely cleaning, case folding, tokenization, normalization, stopword removal, and stemming. Feature extraction was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method, and the data were classified into four sentiment categories: Supporting the Teacher, Blaming the Teacher, Supporting the Student, and Blaming the Student. The results indicate that the linear kernel SVM successfully classified public sentiment with an accuracy of 94% (precision 0.94, recall 0.94, and F1-score 0.94). The sentiment distribution shows that 69.2% of the comments support the teacher, 20.8% support the student, 5.5% blame the teacher, and 4.5% blame the student. These findings provide a comprehensive overview indicating that the majority of the public tends to justify the teacher's disciplinary action, while there is also a growing awareness of child protection and rejection of violence in the educational context.

Keywords: Sentiment Analysis; TikTok; Support Vector Machine; Text Mining; Social Media

1. PENDAHULUAN

Media sosial telah mengalami perkembangan yang sangat pesat dalam beberapa tahun terakhir dan menjadi bagian yang tidak terpisahkan dari kehidupan masyarakat modern. Platform media sosial tidak hanya berfungsi sebagai sarana hiburan, tetapi juga sebagai ruang publik digital tempat masyarakat menyampaikan pendapat, kritik, dan reaksi terhadap berbagai isu sosial [1], [2]. Perkembangan ini mendorong meningkatnya produksi data teks dalam jumlah besar yang merepresentasikan opini dan sikap (stance) publik secara alami. Oleh karena itu, pemanfaatan media sosial sebagai sumber data penelitian menjadi sangat relevan, khususnya dalam memahami kecenderungan opini masyarakat terhadap suatu peristiwa tertentu [3], [4].

Salah satu platform media sosial yang mengalami pertumbuhan signifikan di Indonesia adalah TikTok [5]. TikTok memiliki karakteristik berupa konten video singkat yang interaktif serta fitur komentar yang memungkinkan pengguna untuk memberikan tanggapan secara langsung [6]. Tingginya tingkat partisipasi pengguna menjadikan TikTok sebagai sumber data yang kaya untuk analisis opini publik. Komentar-komentar yang dihasilkan pengguna TikTok mencerminkan sikap dan posisi masyarakat terhadap isu-isu yang sedang viral, sehingga dapat dianalisis secara komputasional [7].

Isu yang sering menjadi perhatian publik di media sosial adalah isu yang berkaitan dengan dunia pendidikan. Pendidikan merupakan sektor yang sensitif karena menyangkut pembentukan karakter, moral, dan masa depan generasi muda [8]. Salah satu kasus yang viral di TikTok adalah peristiwa seorang guru yang

menampar murid karena merokok di lingkungan sekolah [9]. Kasus ini dengan cepat menyebar dan memicu perdebatan luas di masyarakat, baik di media sosial maupun media massa [10].

Peristiwa tersebut menimbulkan beragam reaksi dari masyarakat. Sebagian pihak menilai tindakan guru sebagai bentuk pendisiplinan yang wajar untuk menegakkan aturan sekolah, sementara pihak lain menganggap tindakan tersebut sebagai bentuk kekerasan fisik yang tidak dapat dibenarkan dalam dunia pendidikan modern [11]. Perbedaan pandangan ini menunjukkan adanya polarisasi opini publik yang bersifat kontras dan terarah, sehingga lebih tepat dianalisis sebagai klasifikasi opini atau stance daripada sekadar sentimen positif atau negatif [12].

Dalam konteks ini, analisis sentimen tradisional yang hanya membagi data ke dalam kategori positif, negatif, dan netral menjadi kurang memadai. Oleh karena itu, penelitian ini menggunakan pendekatan klasifikasi opini (stance classification) yang mengelompokkan komentar berdasarkan arah dukungan atau penolakan terhadap pihak yang terlibat dalam kasus. Pendekatan ini memungkinkan penangkapan makna opini publik secara lebih kontekstual dan sesuai dengan desain label penelitian [13].

Seiring dengan berkembangnya teknologi, analisis opini berbasis teks banyak dilakukan menggunakan pendekatan text mining dan machine learning [14]. Salah satu algoritma machine learning yang sering digunakan dalam klasifikasi data teks berdimensi tinggi adalah Support Vector Machine (SVM). Algoritma SVM dikenal memiliki kinerja yang baik, stabil, dan akurat dalam menangani data teks yang kompleks, sehingga dipilih sebagai metode klasifikasi utama dalam penelitian ini [15].

Beberapa penelitian sebelumnya telah menerapkan analisis sentimen pada berbagai platform media sosial seperti Twitter, Facebook, dan Instagram. Namun, penelitian yang secara khusus membahas klasifikasi opini publik pada platform TikTok, terutama pada isu pendidikan dengan pendekatan multikelas berbasis arah dukungan, masih sangat terbatas. Sebagian besar penelitian terdahulu masih menggunakan klasifikasi biner atau polaritas sentimen, sehingga belum mampu merepresentasikan kompleksitas opini publik yang bersifat multidimensi [16].

Berdasarkan kondisi tersebut, research gap penelitian ini terletak pada belum banyaknya kajian yang menguji efektivitas algoritma SVM dalam melakukan klasifikasi opini multikelas pada data komentar TikTok untuk isu pendidikan yang bersifat kontroversial. Gap ini bersifat terukur dan dapat diuji melalui evaluasi kinerja model menggunakan metrik klasifikasi dan confusion matrix [17].

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah: (1) bagaimana distribusi opini publik pada komentar TikTok terhadap kasus guru menampar murid, dan (2) seberapa baik kinerja algoritma SVM dalam mengklasifikasikan opini publik ke dalam empat kelas berbasis arah dukungan [18].

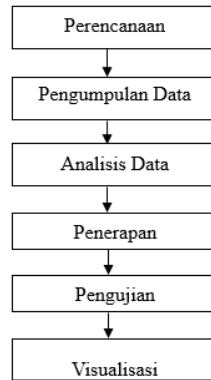
Oleh karena itu, penelitian ini bertujuan untuk menganalisis opini publik di TikTok terhadap kasus viral guru menampar murid menggunakan metode Support Vector Machine (SVM). Penelitian ini mengklasifikasikan komentar ke dalam empat kategori, yaitu mendukung guru, menyalahkan guru, mendukung murid, dan menyalahkan murid. Hasil penelitian diharapkan dapat memberikan kontribusi akademik dalam pengembangan klasifikasi opini berbasis media sosial serta menjadi bahan pertimbangan bagi pemangku kepentingan dalam dunia pendidikan [19].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan analisis sentimen untuk mengolah data teks secara objektif dan terukur. Kerangka penelitian disusun secara sistematis agar setiap tahapan penelitian berjalan terarah dan saling berkaitan. Metode Support Vector Machine (SVM) dipilih karena kemampuannya dalam mengklasifikasikan data teks berdimensi tinggi dengan tingkat akurasi yang baik, khususnya pada data opini di media sosial seperti TikTok.

Secara umum, tahapan penelitian meliputi perencanaan, pengumpulan data, analisis data, penerapan metode, pengujian, dan visualisasi hasil. Kerangka ini bertujuan untuk menghasilkan analisis sentimen yang akurat dan dapat menggambarkan persepsi publik secara komprehensif terhadap kasus viral guru menampar murid karena merokok di sekolah tahap penelitian dapat dilihat pada gambar 1.



Gambar 1. Kerangka Penelitian

2.2 Rencana Penelitian

2.2.1 Perencanaan

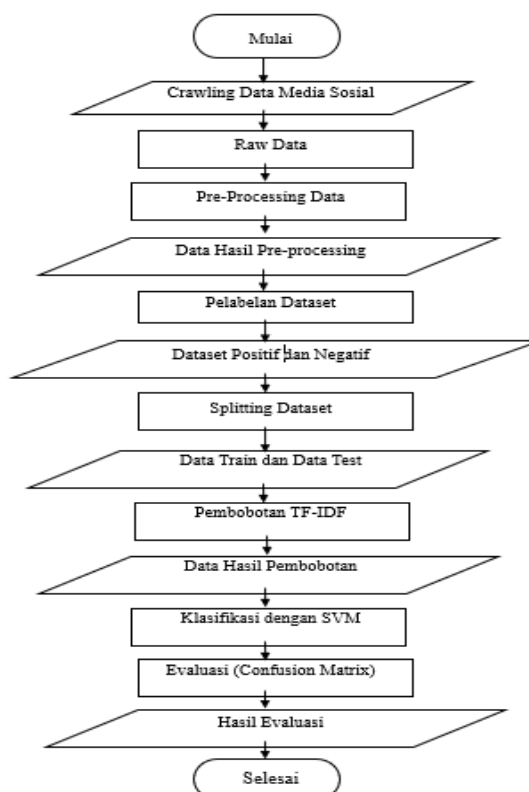
Tahap perencanaan diawali dengan penentuan topik dan ruang lingkup penelitian secara jelas, yaitu analisis sentimen publik di TikTok terhadap kasus viral guru menampar murid menggunakan metode SVM. Pada tahap ini juga ditentukan tujuan penelitian, metode yang digunakan, serta sumber data yang relevan untuk mendukung proses analisis.

2.2.2 Pengumpulan Data

Pengumpulan data dilakukan melalui dua pendekatan. Pertama, pengambilan data komentar publik di TikTok menggunakan Apify Web Scraper. Data yang diambil berupa komentar dari video-video yang membahas kasus guru menampar murid karena merokok di sekolah. Metode ini memungkinkan pengumpulan data secara otomatis dan terstruktur. Kedua, studi literatur dilakukan dengan menelaah jurnal ilmiah, buku, dan artikel terkait analisis sentimen, Support Vector Machine, serta penelitian sejenis pada media sosial untuk memperkuat landasan teoritis penelitian.

2.3 Analisis Data

Berikut ini adalah Proses analisis data yang dapat dilihat di gambar flowchart system yang terdapat pada gambar 2.



Gambar 2. Flowchart Sistem

Tahap analisis data dapat dilihat pada gambar 2 diawali dengan *crawling* data media sosial untuk memperoleh data mentah. Selanjutnya, data melalui proses preprocessing yang meliputi pembersihan teks, case folding, tokenisasi, penghapusan stopword, normalisasi, dan stemming sehingga menghasilkan data yang siap diolah. Data kemudian diberi label sentimen dan dibagi menjadi data latih dan data uji. Representasi teks dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengubah data teks menjadi bentuk numerik. Data hasil pembobotan selanjutnya diklasifikasikan menggunakan algoritma Support Vector Machine (SVM). Evaluasi kinerja model dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall, dan F1-score [20].

2.4 Penerapan Metode

Metode Support Vector Machine (SVM) diterapkan dalam penelitian ini untuk mengklasifikasikan sentimen dalam komentar publik di TikTok terkait kasus viral guru menampar murid karena merokok di sekolah. Tahapan utamanya meliputi preprocessing data (cleaning, case folding, tokenizing, stopword removal, dan stemming), ekstraksi fitur menggunakan TF-IDF untuk mengidentifikasi kata-kata kunci, serta klasifikasi dengan SVM yang memanfaatkan hyperplane optimal untuk memisahkan kategori sentimen (Mendukung Guru, Menyalahkan Guru, Mendukung Murid, dan Menyalahkan Murid) dengan margin terbesar. Pendekatan ini memungkinkan model tidak hanya melakukan klasifikasi secara akurat, tetapi juga menjaga konsistensi performa ketika menghadapi variasi data komentar baru, sehingga menghasilkan analisis sentimen yang andal dan komprehensif terhadap persepsi masyarakat mengenai kasus tersebut.

2.5 Pengujian

Pengujian dalam penelitian ini dilakukan untuk menyalakan keefektifan metode Support Vector Machine (SVM) dalam mengklasifikasikan sentimen publik di TikTok terkait kasus viral guru menampar murid karena merokok di sekolah. Proses pengujian meliputi pembagian data menjadi set pelatihan dan pengujian, serta penggunaan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengukur kinerja model. Hasil pengujian ini akan menyajikan sejauh mana SVM mampu membedakan sentimen Mendukung Guru, Menyalahkan Guru, Mendukung Murid, dan Menyalahkan Murid secara akurat berdasarkan komentar yang telah diproses. Pengujian ini penting untuk memastikan validitas model dalam merepresentasikan persepsi publik secara tepat sehingga hasil analisis sentimen yang diperoleh dapat dipercaya dan digunakan sebagai dasar interpretasi dan rekomendasi.

2.6 Visualisasi

Visualisasi data dalam penelitian analisis sentimen ini menggunakan beberapa teknik untuk menyajikan temuan secara komprehensif. Word cloud (awan kata) digunakan untuk menampilkan kata kunci yang paling sering muncul dalam komentar publik. Bar chart dan pie chart dimanfaatkan untuk memvisualisasikan distribusi empat kategori sentimen (Mendukung Guru, Menyalahkan Guru, Mendukung Murid, dan Menyalahkan Murid). Sementara itu, line chart dapat menunjukkan tren sentimen dari waktu ke waktu terkait perkembangan kasus viral guru menampar murid karena merokok di sekolah. Melalui visualisasi ini, pola sentimen masyarakat dapat diidentifikasi secara cepat dan efektif. Proporsi respons publik terhadap kasus tersebut serta tema-tema utama dalam diskusi digital dapat terlihat jelas tanpa perlu menelusuri ribuan data teks secara manual. Pendekatan visual ini memungkinkan pemahaman yang lebih mendalam tentang dinamika persepsi masyarakat mengenai kasus tersebut di platform TikTok.

2.7 Rencana Pembahasan

Rencana pembahasan dalam penelitian ini akan fokus pada analisis sentimen publik di TikTok terkait kasus viral guru menampar murid karena merokok di sekolah dengan menggunakan metode Support Vector Machine (SVM). Pembahasan dimulai dengan deskripsi proses pengumpulan dan pengumpulan data komentar publik, kemudian melanjutkan penerapan SVM untuk mengklasifikasi sentimen ke dalam kategori Mendukung Guru, Menyalahkan Guru, Mendukung Murid, dan Menyalahkan Murid. Selanjutnya, hasil klasifikasi akan divisualisasikan menggunakan berbagai grafik seperti word cloud untuk menggambarkan kata kunci dominan, serta bar chart, pie chart, dan line chart untuk menampilkan distribusi dan tren sentimen dari waktu ke waktu. Melalui pembahasan tersebut, penelitian berupaya mengungkap pola perilaku dan persepsi masyarakat di TikTok terhadap kasus tersebut, serta memahami dinamika komunikasi digital yang berkembang dalam konteks sosial dan pendidikan [21].

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan

3.1.1 Analisis Data

Penelitian ini melakukan analisis sentimen terhadap 2.200 komentar publik di TikTok terkait kasus guru menampar murid. Data dikumpulkan menggunakan Apify Web Scraper dengan Google Colaboratory sebagai lingkungan

pemrograman Python. Kata kunci yang digunakan meliputi "guru tampar murid", "guru memukul murid merokok", "kasus guru tampar viral", "murid merokok di sekolah", "guru menampar siswa", dan "guru hukum murid". Data disimpan dalam format CSV dan diberi label manual ke dalam empat kelas sentimen: mendukung guru, menyalahkan guru, mendukung murid, dan menyalahkan murid. Seluruh data melalui tahap pre-processing (cleaning, case folding, tokenization, normalisasi, stopword removal, dan stemming), pembobotan TF-IDF, serta pembagian data latih dan uji. Klasifikasi dilakukan menggunakan Support Vector Machine (SVM) dan dievaluasi dengan confusion matrix. Berikut pada tabel 1. contoh dataset yang digunakan.

Tabel 1. Contoh dataset mentah

No	Text	diggCount	replyCommentTotal	createTimeISO
1	"Butterfly Effect" 🦋	1	0	2025-10-18T00:54:46.000Z
2	"di sekolah tempat belajar bukan tempat kekerasan" lantas? ngapa lu ngerokok di sekolah? emang sekolah tempat ngerokok? 😂😂😂	1	0	2025-10-20T06:35:59.000Z
3	"DI SEKOLAH TEMPAT BELAJAR, BUKAN MEROKOK"	2	0	2025-10-17T15:50:45.000Z
4	"di sekolah tempat belajar, bukan tempat MEROKOK"	2	0	2025-10-15T13:30:05.000Z
5	"di sikilih timpit bilijir bikin timpit kikirisin"	2	4	2025-10-15T07:43:14.000Z
6	"gak mungkin lah"	0	0	2025-10-19T01:07:15.000Z
7	"gak mungkin" jir lah 🤔	1	0	2025-10-17T09:42:29.000Z
8	"JELAS JELAS ORANG TUA DI SEKOLAH ADALAH GURU" ITU MENURUT KU HANYA BERLAKU ANAK JAMAN DULU YANG MENGANGGAP GURU ITU ORANG TUA SAAT DI SEKOLAH	4	0	2025-10-15T19:36:03.000Z
9	"kalo mau nakal jangan sampe org rumah tau" ahh sudah di lupakan ya 🤔	2	0	2025-10-18T09:42:11.000Z
10	"mamih, aku di tampar" 🐱🐱	2	0	2025-10-17T08:13:21.000Z

3.1.2 Pre-processing Data

Data komentar melalui tahapan pre-processing untuk mengubah teks tidak terstruktur menjadi data bersih:

a. Cleaning & Case Folding

Menghilangkan URL, hashtag, mention, angka, simbol, emoji, dan tanda baca, serta mengubah teks menjadi huruf kecil. Hasil *proses cleaning* bisa dilihat pada tabel 2.

Tabel 2. Sampel Proses *Cleaning*

text	clean&casefolding
"Butterfly Effect" 🦋	butterfly effect

"di sekolah tempat belajar bukan tempat kekerasan" lantas? ngapa lu ngerokok di sekolah? emang sekolah tempat ngerokok? 🤔🤔🤔	di sekolah tempat belajar bukan tempat kekerasan lantas ngapa lu ngerokok di sekolah emang sekolah tempat ngerokok
"DI SEKOLAH TEMPAT BELAJAR, BUKAN MEROKOK"	di sekolah tempat belajarbukan merokok
"di sekolah tempat belajar, bukan tempat MEROKOK"	di sekolah tempat belajarbukan tempat merokok
"di sikilih timpit bilijir bikin timpit kikirisin"	di sikilih timpit bilijir bikin timpit kikirisin
"gak mungkin lah"	gak mungkin lah
"gak mungkin" jir lah 🤔	gak mungkin jir lah
"JELAS JELAS ORANG TUA DI SEKOLAH ADALAH GURU" ITU MENURUT KU HANYA BERLAKU ANAK JAMAN DULU YANG MENGANGGAP GURU ITU ORANG TUA SAAT DI SEKOLAH	jelas jelas orang tua di sekolah adalah guru itu menurut ku hanya berlaku anak jaman dulu yang menganggap guru itu orang tua saat di sekolah
"kalo mau nakal jangan sampe org rumah tau" ahh sudah di lupakan ya 🤔	kalo mau nakal jangan sampe org rumah tau ahh sudah di lupakan ya
"mamih, aku di tampar" 🐱🐱	mamih aku di tampar

b. Tokenization

Memecah teks menjadi token/kata berdasarkan spasi dapat dilihat pada tabel 3 berikut ini.

Tabel 3. Sampel proses tokenization

clean&casefolding	Tokenization
butterfly effect	['butterfly', 'effect']
di sekolah tempat belajar bukan tempat kekerasan lantas ngapa lu ngerokok di sekolah emang sekolah tempat ngerokok	['di', 'sekolah', 'tempat', 'belajar', 'bukan', 'tempat', 'kekerasan', 'lantas', 'ngapa', 'lu', 'ngerokok', 'di', 'sekolah', 'emang', 'sekolah', 'tempat', 'ngerokok']
di sekolah tempat belajarbukan merokok	['di', 'sekolah', 'tempat', 'belajarbukan', 'merokok']
di sekolah tempat belajarbukan tempat merokok	['di', 'sekolah', 'tempat', 'belajarbukan', 'tempat', 'merokok']
di sikilih timpit bilijir bikin timpit kikirisin	['di', 'sikilih', 'timpit', 'bilijir', 'bikin', 'timpit', 'kikirisin']
gak mungkin lah	['gak', 'mungkin', 'lah']
gak mungkin jir lah	['gak', 'mungkin', 'jir', 'lah']
jelas jelas orang tua di sekolah adalah guru itu menurut ku hanya berlaku anak jaman dulu yang menganggap guru itu orang tua saat di sekolah	['jelas', 'jelas', 'orang', 'tua', 'di', 'sekolah', 'adalah', 'guru', 'itu', 'menurut', 'ku', 'hanya', 'berlaku', 'anak', 'jaman', 'dulu', 'yang', 'menganggap', 'guru', 'itu', 'orang', 'tua', 'saat', 'di', 'sekolah']
kalo mau nakal jangan sampe org rumah tau ahh sudah di lupakan ya	['kalo', 'mau', 'nakal', 'jangan', 'sampe', 'org', 'rumah', 'tau', 'ahh', 'sudah', 'di', 'lupakan', 'ya']
mamih aku di tampar	['mamih', 'aku', 'di', 'tampar']

c. Tahap Normalisasi

Mengubah kata tidak baku menjadi kata baku menggunakan file kamusGaul.txt (~10.000 kata dari GitHub). File hasil normalisasi dapat di lihat pada table 4 berikut ini.

Tabel 4. File normalisasi github

Sebelum	Sesudah
woww	wow
hallo	halo
eeeehhhh	eh
kata2nyaaa	kata-katanya
smga	semoga
slalu	selalu
kk	kakak
bgt	banget
gitu	begitu
aja	saja
lg	lagi
mantep	mantap
daah	dah
akuuuuu	aku
skarang	sekarang
sampe	sampai
sma	sama
bener2	benar-benar
sdikit	sedikit
jdi	jadi
kau	kamu

Sampel proses normalisasi dapat dilihat pada tabel 5 berikut ini.

Tabel 5. Sampel proses *normalization*

Tokenization	Normalisasi
['butterfly', 'effect']	['butterfly', 'effect']
['di', 'sekolah', 'tempat', 'belajar', 'bukan', 'tempat', 'kekerasan', 'lantas', 'ngapa', 'lu', 'ngerokok', 'di', 'sekolah', 'emang', 'sekolah', 'tempat', 'ngerokok']	['di', 'sekolah', 'tempat', 'belajar', 'bukan', 'tempat', 'kekerasan', 'lantas', 'ngapa', 'lu', 'ngerokok', 'di', 'sekolah', 'memang', 'sekolah', 'tempat', 'ngerokok']
['di', 'sekolah', 'tempat', 'belajarbukan', 'merokok']	['di', 'sekolah', 'tempat', 'belajarbukan', 'merokok']
['di', 'sekolah', 'tempat', 'belajarbukan', 'tempat', 'merokok']	['di', 'sekolah', 'tempat', 'belajarbukan', 'tempat', 'merokok']
['di', 'sikilih', 'timpit', 'bilijir', 'bikin', 'timpit', 'kikirisin']	['di', 'sikilih', 'timpit', 'bilijir', 'bikin', 'timpit', 'kikirisin']
['gak', 'mungkin', 'lah']	['enggak', 'mungkin', 'lah']
['gak', 'mungkin', 'jir', 'lah']	['enggak', 'mungkin', 'jir', 'lah']
['jelas', 'jelas', 'orang', 'tua', 'di', 'sekolah', 'adalah', 'guru', 'itu', 'menurut', 'ku', 'hanya', 'berlaku', 'anak', 'jaman', 'dulu', 'yang', 'menganggap', 'guru', 'itu', 'orang', 'tua', 'saat', 'di', 'sekolah']	['jelas', 'jelas', 'orang', 'tua', 'di', 'sekolah', 'adalah', 'guru', 'itu', 'menurut', 'ku', 'hanya', 'berlaku', 'anak', 'jaman', 'dulu', 'yang', 'menganggap', 'guru', 'itu', 'orang', 'tua', 'saat', 'di', 'sekolah']
['kalo', 'mau', 'nakal', 'jangan', 'sampe', 'org', 'rumah', 'tau', 'ahh', 'sudah', 'di', 'lupakan', 'ya']	['kalo', 'mau', 'nakal', 'jangan', 'sampai', 'orang', 'rumah', 'tau', 'ah', 'sudah', 'di', 'lupakan', 'ya']
['mamih', 'aku', 'di', 'tampar']	['mamih', 'aku', 'di', 'tampar']

d. Stopwords

Menghapus kata umum yang tidak berpengaruh signifikan seperti "dan", "yang", "di", "ke". Hasilnya dapat dilihat pada tabel 6 berikut ini.

Tabel 6. Text stopwords

Normalisasi	Stopwords
['butterfly', 'effect']	['butterfly', 'effect']
['di', 'sekolah', 'tempat', 'belajar', 'bukan', 'tempat', 'kekerasan', 'lantas', 'ngapa', 'lu', 'ngerokok', 'di', 'sekolah', 'memang', 'sekolah', 'tempat', 'ngerokok']	['sekolah', 'belajar', 'kekerasan', 'lantas', 'ngapa', 'lu', 'ngerokok', 'sekolah', 'sekolah', 'ngerokok']
['di', 'sekolah', 'tempat', 'belajarbukan', 'merokok']	['sekolah', 'belajarbukan', 'merokok']

['di', 'sekolah', 'tempat', 'belajarbukan', 'tempat', 'merokok']	['sekolah', 'belajarbukan', 'merokok']
['di', 'sikilih', 'timpit', 'bilijir', 'bikin', 'timpit', 'kikirisin']	['sikilih', 'timpit', 'bilijir', 'bikin', 'timpit', 'kikirisin']
['enggak', 'mungkin', 'lah']	[]
['enggak', 'mungkin', 'jir', 'lah']	['jir']
['jelas', 'jelas', 'orang', 'tua', 'di', 'sekolah', 'adalah', 'guru', 'itu', 'menurut', 'ku', 'hanya', 'berlaku', 'anak', 'jaman', 'dulu', 'yang', 'menganggap', 'guru', 'itu', 'orang', 'tua', 'saat', 'di', 'sekolah']	['orang', 'tua', 'sekolah', 'guru', 'ku', 'berlaku', 'anak', 'jaman', 'menganggap', 'guru', 'orang', 'tua', 'sekolah']
['kalo', 'mau', 'nakal', 'jangan', 'sampai', 'orang', 'rumah', 'tau', 'ah', 'sudah', 'di', 'lupakan', 'ya']	['kalo', 'nakal', 'orang', 'rumah', 'tau', 'ah', 'lupakan', 'ya']
['mamih', 'aku', 'di', 'tampar']	['mamih', 'tampar']

e. *Steaming*

Mengubah kata berimbuhan menjadi kata dasar, hasilnya dapat kita lihat pada tabel 7 berikut ini.

Tabel 7. Sampel proses *steaming*

Stopwords	stemming_data
['butterfly', 'effect']	butterfly effect
['sekolah', 'belajar', 'kekerasan', 'lantas', 'ngapa', 'lu', 'ngerokok', 'sekolah', 'sekolah', 'ngerokok']	sekolah ajar keras lantas ngapa lu ngerokok
['sekolah', 'belajarbukan', 'merokok']	sekolah sekolah ngerokok
['sekolah', 'belajarbukan', 'merokok']	sekolah belajarbukan rokok
['sekolah', 'belajarbukan', 'merokok']	sekolah belajarbukan rokok
['sikilih', 'timpit', 'bilijir', 'bikin', 'timpit', 'kikirisin']	sikilih timpit bilijir bikin timpit kikirisin

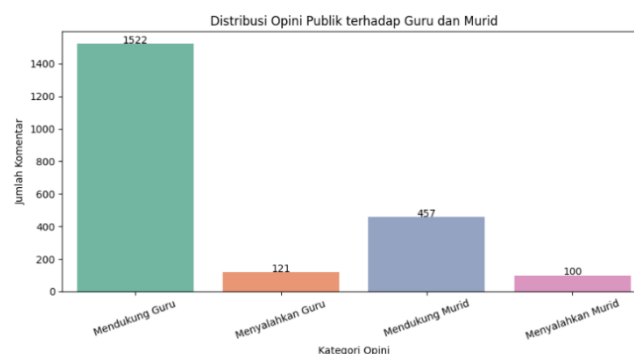
3.1.3 Pemberian Label Sentiment Berbasis Kata Kunci

Pelabelan dilakukan secara otomatis menggunakan pendekatan berbasis kata kunci dengan empat kelompok kamus yang merepresentasikan masing-masing kelas sentimen. Label ditentukan berdasarkan skor tertinggi dari keempat kelas. Contoh hasil pelabelan data komentar, yang menampilkan teks hasil stemming beserta label sentimen yang diperoleh, disajikan pada Tabel 8.

Tabel 8. Pelabelan dataset

stemming_data	Labelisasi
butterfly effect	Mendukung Guru
sekolah ajar keras lantas ngapa lu ngerokok sekolah	Mendukung Guru
di sekolah guru itu seperti orang tua mendidik anak anak zaman sekarang	Mendukung Guru
kalau anak nakal orang rumah tahu atau purapura lupa	Menyalahkan Murid
masa iya guru sampai nampang murid	Menyalahkan Guru
muridnya masih muda kasihan juga kalau diperlakukan begitu.	Mendukung Murid
pria itu cuma menegur demi kedisiplinan	Mendukung Guru
pria itu tegas demimendidik	Mendukung Guru
dia itu cuma menjalankan tugasnya sebagai guru	Mendukung Guru
sekolah ngajarin disiplin bukan tempat buat ngerokok lucu aja	Mendukung Guru

Jumlah sentimen terdapat pada gambar 3 berikut, dimana proses dilakukan menggunakan pemrograman python di google collabs.



Gambar 3. Visualisasi jumlah analisis sentiment

Gambar tersebut menunjukkan distribusi opini publik terhadap kasus guru menampar murid berdasarkan hasil pelabelan sentimen menggunakan bahasa pemrograman Python pada lingkungan Google Colaboratory. Grafik memperlihatkan pembagian komentar ke dalam empat kategori, yaitu mendukung guru, menyalahkan guru, mendukung murid, dan menyalahkan murid.

Berdasarkan hasil tersebut, kategori mendukung guru merupakan yang paling dominan dengan jumlah 1.522 komentar, yang menunjukkan bahwa mayoritas masyarakat cenderung membenarkan tindakan guru sebagai bentuk pendisiplinan di lingkungan sekolah. Kategori mendukung murid berada pada urutan kedua dengan 457 komentar, mencerminkan adanya perhatian terhadap aspek perlindungan anak dan dampak psikologis akibat kekerasan. Sementara itu, kategori menyalahkan guru dan menyalahkan murid memiliki jumlah yang relatif lebih kecil, masing-masing sebanyak 121 dan 100 komentar. Secara keseluruhan, distribusi ini menunjukkan bahwa opini publik lebih banyak berpihak kepada guru, meskipun masih terdapat keberagaman pandangan masyarakat terhadap kasus tersebut.

3.1.4 Pembobotan TF-IDF

Metode TF-IDF mengukur tingkat kepentingan kata dalam dokumen. Semakin sering kata muncul dalam satu dokumen dan semakin jarang di dokumen lain, semakin tinggi bobotnya.

Rumus:

a. Menghitung Term Frequency (TF): Frekuensi kemunculan kata dalam dokumen.

b. Menghitung IDF: $idf_{(t,D)} = \log \frac{\text{Total Jumlah Dokumen}}{\text{Jumlah Dokumen dari TF}}$ (1)

c. Menghitung TF-IDF: $tf - idf_{t,d} = tf_{td} * idf_t$ (2)

Hasil dari perhitungan TF-IDF dapat dilihat pada tabel 9.

Tabel 9. Tabel TF-IDF

Kata	Dokumen	TF-IDF
tegas	1	0.06
mendidik	1	0.06
disiplin	1	0.06
guru	1	0
murid	1	0
kasar	2	0.06
tidak	2	0.06
pantas	2	0.06
menyakiti	2	0.06
guru	2	0
murid	2	0

Berdasarkan hasil perhitungan TF-IDF tersebut, dapat diketahui bahwa kata-kata yang memiliki nilai TF-IDF tinggi merupakan kata yang lebih berpengaruh dalam menentukan sentimen suatu komentar. Nilai-nilai TF-IDF ini selanjutnya digunakan sebagai fitur numerik dalam proses klasifikasi sentimen menggunakan metode Support Vector Machine (SVM).

3.1.5 Klasifikasi Support Vector Machine

Dataset (2.200 komentar) dibagi dengan rasio 80:20 (data latih:data uji). SVM menggunakan kernel linear untuk memisahkan data berdasarkan fitur TF-IDF.

Parameter yang digunakan:

- $\alpha_i = 0$
- $C = 1$
- γ (learning rate) = 0.1
- $\lambda = 0.5$
- Iterasi = 2

Proses meliputi:

- Perhitungan kernel linear: $K(x,y) = x \cdot y$
- Perhitungan matriks Hessian
- Sequential training untuk nilai α
- Penentuan support vector
- Perhitungan bias dan fungsi keputusan $f(x)$

Berikut ini gambar *Wordcloud* untuk sentimen yang dapat dilihat pada gambar 4 dan 5 dibawah ini.



Gambar 4. Wordcloud untuk sentimen mendukung guru dan menyalahkan guru



Gambar 5. Wordcloud untuk sentimen mendukung murid dan menyalahkan murid

3.2 Hasil Pengujian

Bagian ini membahas hasil penelitian analisis sentimen terhadap komentar pengguna TikTok terkait kasus guru menampar murid menggunakan metode Support Vector Machine (SVM). Tahapan pada bagian ini meliputi implementasi sistem analisis sentimen, pengujian sistem, serta evaluasi hasil klasifikasi berdasarkan performa model yang dibangun.

3.2.1 Implementasi Sistem Analisis Sentimen

Implementasi menggunakan Google Colab dan Python dengan tahapan lengkap preprocessing, pelabelan, splitting data, dan klasifikasi SVM.

3.2.2 Pengujian

Tahap pengujian bertujuan untuk mengevaluasi kemampuan sistem dalam mengklasifikasikan opini masyarakat terhadap kasus guru menampar murid berdasarkan komentar pengguna TikTok. Pengujian dilakukan menggunakan Google Colaboratory dengan bahasa pemrograman Python, serta menerapkan model Support Vector Machine (SVM) berbasis fitur TF-IDF. Data uji diperoleh dari pemisahan dataset dengan rasio 80% data latih dan 20% data uji, sehingga dari total 2.200 komentar digunakan 440 data sebagai data pengujian. Kinerja model dievaluasi menggunakan confusion matrix dan classification report untuk mengukur nilai accuracy, precision, recall, dan F1-score pada setiap kelas opini, yaitu Mendukung Guru, Mendukung Murid, Menyalahkan Guru, dan Menyalahkan Murid.

3.2.3 Evaluasi Hasil

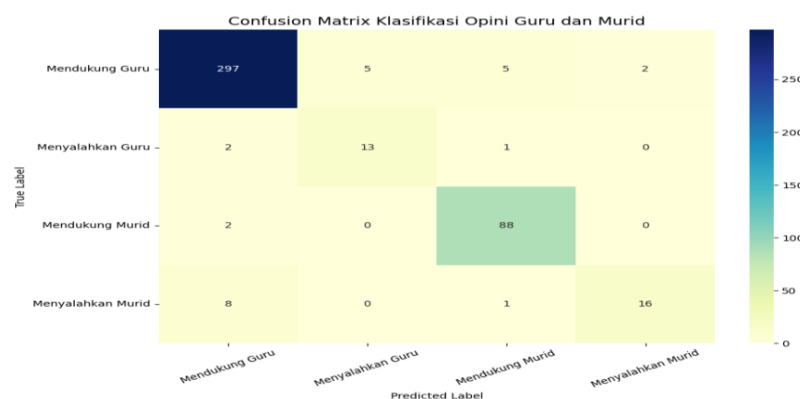
Setelah proses pelatihan dan pengujian model Support Vector Machine (SVM) selesai dilakukan, sistem menghasilkan prediksi label opini pada data uji berdasarkan pola yang telah dipelajari dari data latih. Evaluasi performa model dilakukan menggunakan classification report untuk melihat nilai precision, recall, dan F1-score pada masing-masing kelas opini, serta confusion matrix untuk mengetahui distribusi hasil prediksi model terhadap label sebenarnya dapat dilihat pada gambar 8.

Accuracy: 0.9409090909090909				
	precision	recall	f1-score	support
Mendukung Guru	0.96	0.96	0.96	309
Mendukung Murid	0.93	0.98	0.95	90
Menyalahkan Guru	0.72	0.81	0.76	16
Menyalahkan Murid	0.89	0.64	0.74	25
accuracy			0.94	440
macro avg	0.87	0.85	0.86	440
weighted avg	0.94	0.94	0.94	440

Gambar 6. Classification report

Gambar diatas menunjukkan nilai ketepatan, ketepatan, recall, dan skor F1. Ketepatan adalah metrik yang

mengukur seberapa baik model klasifikasi memprediksi kelas data yang benar, sementara ketepatan adalah metrik yang mengukur seberapa baik model klasifikasi memprediksi kelas data yang benar. Skor F1 adalah metrik yang mengimbangi ketepatan dan recall. Untuk melihat proses Confusion Matrix pada gambar 18 sebagai berikut.



Gambar 7. Proses Confusion Matrix

Confusion matrix digunakan untuk melihat secara rinci jumlah prediksi yang benar dan salah pada masing-masing kelas opini. Berdasarkan confusion matrix yang dihasilkan, sebagian besar data uji berhasil diklasifikasikan dengan benar, terutama pada kelas Mendukung Guru dan Mendukung Murid, yang memiliki jumlah data relatif lebih besar dibandingkan kelas lainnya. Dari gambar confusion matrix diatas maka dapat dihitung nilai dari *accuracy*, *precision*, *recall* dan *f1-score* dengan menggunakan persamaan di bawah ini:

$$\text{Accuracy} = \frac{297+13+88+16}{440} = \frac{414}{440} = 0.9409 = 94\%$$

$$\text{Precision} = \frac{297}{297+12} = 0.96$$

$$\text{Recall} = \frac{297}{297+12} = 0.96$$

$$\text{F1-Score} = \frac{2 \times 0.96 \times 0.96}{0.96+0.96} = 0.96$$

3.2.4 Penerapan

Penelitian ini bertujuan untuk mengembangkan sebuah sistem yang mampu melakukan klasifikasi opini masyarakat terhadap kasus guru menampar murid berdasarkan komentar pengguna TikTok. Dengan menerapkan metode Support Vector Machine (SVM), sistem mampu mengelompokkan komentar ke dalam empat kelas opini, yaitu Mendukung Guru, Mendukung Murid, Menyalahkan Guru, dan Menyalahkan Murid. Hasil pengujian menunjukkan bahwa sistem memiliki tingkat akurasi yang tinggi dan mampu menangkap kecenderungan sikap publik terhadap pihak-pihak yang terlibat dalam kasus tersebut. Dengan demikian, sistem yang dibangun dapat digunakan sebagai alat bantu analisis opini publik pada kasus sosial serupa di media sosial.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, metode Support Vector Machine (SVM) dengan pembobotan Term Frequency–Inverse Document Frequency (TF-IDF) berhasil diterapkan untuk mengklasifikasikan opini publik terhadap kasus guru menampar murid berdasarkan 2.200 komentar TikTok yang dibagi menjadi 80% data latih (1.760 komentar) dan 20% data uji (440 komentar) ke dalam empat kategori: Mendukung Guru, Mendukung Murid, Menyalahkan Guru, dan Menyalahkan Murid. Setelah melalui tahapan preprocessing (cleaning, case folding, tokenisasi, normalisasi menggunakan kamus ~10.000 kata, stopword removal, dan stemming) serta ekstraksi fitur TF-IDF, model SVM dengan kernel linear mencapai akurasi 94% dengan weighted average precision, recall, dan F1-score sebesar 0.94. Lebih penting dari metrik akurasi, model berhasil mengidentifikasi distribusi opini publik yang menunjukkan 69.2% komentar mendukung tindakan guru, 20.8% mendukung murid, 5.5% menyalahkan guru, dan 4.5% menyalahkan murid, memberikan wawasan bahwa mayoritas masyarakat cenderung membenarkan tindakan disiplin guru meskipun kesadaran terhadap perlindungan anak mulai berkembang. Performa model menunjukkan variasi antar kelas dengan F1-score tertinggi pada kelas Mendukung Guru (0.96) dan Mendukung Murid (0.94), sedangkan kelas minoritas memiliki F1-score lebih rendah (Menyalahkan Guru 0.53 dan Menyalahkan Murid 0.78), mengindikasikan tantangan klasifikasi pada data tidak seimbang yang umum terjadi pada opini publik di media sosial. Untuk penelitian selanjutnya, disarankan penanganan imbalanced dataset menggunakan teknik SMOTE atau class weighting, optimasi hyperparameter

melalui grid search, perbandingan dengan baseline model seperti Naïve Bayes dan Random Forest, eksplorasi metode ensemble, perluasan dataset dengan periode pengambilan lebih lama, serta evaluasi menggunakan metrik tambahan seperti macro F1-score dan Cohen's Kappa yang lebih sensitif terhadap data tidak seimbang, sehingga penelitian ini memberikan kontribusi dalam memahami dinamika opini publik di media sosial dan menunjukkan bahwa klasifikasi opini multikelas dapat menjadi alternatif yang lebih informatif dibandingkan analisis sentimen biner tradisional dalam menganalisis isu kompleks yang melibatkan berbagai perspektif.

REFERENCES

- [1] H. Setiawan, H. Oktaviana, F. Di, D. Andawas, M. Noval, and W. Saripah, "Pengaruh Media Sosial TikTok Terhadap Pola Belajar Mahasiswa," vol. 2, no. 1, pp. 28–34, 2022.
- [2] S. Heristian, M. Napiah, and W. Erawati, "Analisis Sentimen Ulasan Pelanggan Menggunakan Algoritma Naive Bayes pada Aplikasi Gojek," vol. 5, no. 1, pp. 35–41, 2025.
- [3] U. A. Purwokerto and K. Kunci, "PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN DECISION TREE UNTUK ANALISIS SENTIMEN REVIEW KOMENTAR PADA APLIKASI TRANSPORTASI ONLINE," vol. 2, no. 2, 2021.
- [4] R. R. Wahid, F. T. Anggraeni, and B. Nugroho, "Implementasi Metode Extreme Learning Machine untuk Klasifikasi Tumor Otak pada Citra Magnetic Resonance Imaging," vol. 1, pp. 16–20, 2020.
- [5] J. A. Zulqornain and P. P. Adikara, "Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD)," vol. 5, no. 7, pp. 2886–2890, 2021.
- [6] H. Hassani, C. Beneki, S. Unger, and M. T. Mazinani, "Text Mining in Big Data Analytics," pp. 1–34, 2021, doi: 10.3390/bdcc4010001.
- [7] H. Judul, F. T. Industri, and U. I. Indonesia, "ANALISIS SENTIMEN ULASAN PRODUK SKINCARE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (STUDI KASUS: FORUM FEMALE DAILY) MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (STUDI KASUS: FORUM FEMALE DAILY)," 2021.
- [8] N. H. Ovirianti, M. Zarlis, and H. Mawengkang, "Support Vector Machine Using A Classification Algorithm," vol. 6, no. 3, pp. 2103–2107, 2022.
- [9] L. Setiyani, "ANALISIS PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE DATA MINING NAÏVE BAYES: SYSTEMATIC REVIEW," vol. 13, no. 1, pp. 35–43, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [10] D. P. Utomo, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," vol. 4, no. April, pp. 437–444, 2020, doi: 10.30865/mib.v4i2.2080.
- [11] E. Apriani, I. F. Hanif, and F. Oktavianalisti, "Sentiment Analysis of Using TikTok as a Learning Media Using the Naïve Bayes Classifiers Algorithm Analisis Sentimen Penggunaan TikTok Sebagai Media Pembelajaran Menggunakan Algoritma Naïve Bayes Classifier," vol. 4, no. July, pp. 1160–1168, 2024.
- [12] C. Chazar and B. E. Widhiaputra, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," vol. 12, no. 1, pp. 67–80, 2020.
- [13] T. Buana and D. Maharani, "Penggunaan aplikasi tik tok (versi terbaru) dan kreativitas anak," vol. 14, no. 1, pp. 1–10, 2020.
- [14] S. Barta, D. Belanche, A. Fern, and M. Flavi, "Influencer marketing on TikTok: The effectiveness of humor and followers' hedonic experience," vol. 70, no. July 2022, 2023, doi: 10.1016/j.jretconser.2022.103149.
- [15] S. I. Reprint, X. Kong, W. Wang, and H. Liu, *AI Empowered Sentiment Analysis*. 2022.
- [16] C. A. Iglesias and A. Moreno, *Sentiment Analysis for Social Media*. 2022.
- [17] S. N. Hakim, "ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES CLASSIFIER (NBC)," 2021.
- [18] R. Andarsyah and A. Yanuar, "SENTIMEN ANALISIS APLIKASI POSAJA PADA GOOGLE PLAYSTORE UNTUK PENINGKATAN POSPAY SUPERAPP MENGGUNAKAN SUPPORT VECTOR MEACHINE Roni," vol. 16, no. 2, pp. 1–7, 2024.
- [19] R. MAHERI, "ANALISIS SENTIMEN ULASAN APLIKASI M-PASPOR MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE," 2024.
- [20] T. Bahtiar, "SISTEM ANALISIS SENTIMEN TERHADAP PRODUK SUNSCREEN PADA MARKETPLACE SHOPEE MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) LAPORAN," 2024.
- [21] I. Busrayan and Andrianingsih, "ANALISIS SENTIMEN PELANGGAN TERHADAP APLIKASI WONDR BY BNI MENGGUNAKAN NAIVE BAYES, SUPPORT VECTOR MACHINE (SVM), DAN K- NEAREST NEIGHBOR (KNN) Irzan," vol. 2, no. 2, pp. 263–274, 2025.