

ANALISIS SENTIMEN OPINI WARGANET TERHADAP PENDIDIKAN MILITER BAGI PELAJAR MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Nuzul Karimah Fitrah¹, Novita Lestari Anggreini²

^{1,2,3}Politeknik TEDC Bandung

(Program Studi Teknik Informatika Politeknik TEDC Bandung)

(Jl. Politeknik - Pasantren Kota Cimahi, Jawa Barat telp. (022) 6645951)

e-mail: ¹nuzul5086@gmail.com, ²novitalestari@poltektedc.ac.id

Abstrak

Perkembangan media sosial, khususnya TikTok, telah menjadi ruang diskusi publik yang sarat opini terkait berbagai isu sosial dan kebijakan pemerintah. Salah satunya adalah program pendidikan militer bagi pelajar di Jawa Barat yang menimbulkan pro dan kontra di kalangan masyarakat. Penelitian ini bertujuan untuk menganalisis sentimen warganet terhadap program tersebut dengan mengklasifikasikan opini ke dalam kategori positif, negatif, dan netral. Data komentar diperoleh melalui crawling menggunakan APIfy TikTok Scraper dengan total 1.147 komentar. Tahapan penelitian meliputi preprocessing teks (case folding, cleansing, tokenisasi, stopword removal, stemming, dan normalisasi), ekstraksi fitur menggunakan TF-IDF, serta klasifikasi menggunakan algoritma Support Vector Machine (SVM) [1]. Hasil penelitian menunjukkan bahwa sentimen netral mendominasi sebesar 70,1%, disusul sentimen positif sebesar 28,6%, dan negatif hanya 1,3%. Evaluasi model menghasilkan tingkat akurasi yang baik, sehingga SVM terbukti efektif dalam menganalisis opini warganet berbasis teks di media sosial. Temuan ini diharapkan dapat menjadi masukan bagi pemerintah, khususnya dalam mengevaluasi kebijakan pendidikan militer, sekaligus memperkaya kajian penerapan analisis sentimen di bidang kebijakan publik.

Kata kunci: Analisis Sentimen, TikTok, Pendidikan Militer, Support Vector Machine, Opini Warganet

Abstract

The rapid growth of social media, particularly TikTok, has provided a public sphere for discussions filled with diverse opinions on social issues and government policies. One of the widely debated topics is the military education program for students in West Java, which has sparked both support and criticism among the public. This study aims to analyze netizens' sentiment toward the program by classifying opinions into positive, negative, and neutral categories. Data were collected through crawling using the APIfy TikTok Scraper, resulting in a total of 1,147 comments. The research process involved text preprocessing (case folding, cleansing, tokenization, stopword removal, stemming, and normalization), feature extraction with TF-IDF, and classification using the Support Vector Machine (SVM) algorithm. The results show that neutral sentiment dominates at 70.1%, followed by positive sentiment at 28.6%, and negative sentiment at only 1.3%. Model evaluation indicates a good level of accuracy, demonstrating that SVM is effective in analyzing netizen opinions from social media text. These findings are expected to provide valuable insights for the government in evaluating the military education policy while contributing to the broader study of sentiment analysis in public policy.

Keywords: Sentiment Analysis, TikTok, Military Education, Support Vector Machine, Netizen Opinion

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang pesat telah membawa perubahan signifikan dalam cara masyarakat mengakses, menyampaikan, dan merespons informasi di media

sosial [1]. Salah satu platform yang banyak digunakan adalah TikTok, yang awalnya hanya dikenal sebagai sarana hiburan, namun kini juga menjadi ruang diskusi publik melalui fitur komentar [2]. Fenomena ini menjadikan TikTok sebagai sumber data potensial untuk menggali opini masyarakat terkait isu sosial maupun kebijakan pemerintah [3]. Salah satu topik yang menimbulkan perdebatan publik adalah program pendidikan militer bagi pelajar di Jawa Barat yang digagas oleh Gubernur H. Dedi Mulyadi. Program ini ditujukan bagi pelajar yang memiliki permasalahan perilaku, seperti terlibat dalam tawuran, balap liar, maupun kenakalan remaja lainnya. Namun, dalam pelaksanaannya program ini juga melibatkan pelajar yang tidak bermasalah, berdasarkan permintaan sebagian orang tua yang menilai program tersebut mampu memberikan pembinaan mental dan pengembangan karakter [4]. Program yang dilaksanakan selama 12 hari ini menuai beragam tanggapan.

Di satu sisi, sebagian masyarakat mendukung program ini karena dianggap dapat menanamkan disiplin, tanggung jawab, dan jiwa kepemimpinan pada pelajar. Namun, di sisi lain terdapat pihak yang menolak dengan alasan metode pendidikan militer dianggap terlalu keras, bahkan lembaga negara seperti Komisi Perlindungan Anak Indonesia (KPAI), Komisi Nasional Hak Asasi Manusia (KOMNAS HAM), dan Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) turut menyampaikan kritik terhadap kebijakan tersebut [5]. Perbedaan opini masyarakat inilah yang melatarbelakangi perlunya analisis sentimen untuk memetakan persepsi publik secara lebih objektif [6]. Analisis sentimen memungkinkan identifikasi kecenderungan opini warganet ke dalam kategori positif, negatif, maupun netral. Mengingat jumlah komentar di media sosial sangat besar dan tidak mungkin dianalisis secara manual, diperlukan metode berbasis machine learning yang mampu mengklasifikasikan opini dengan cepat dan akurat [7].

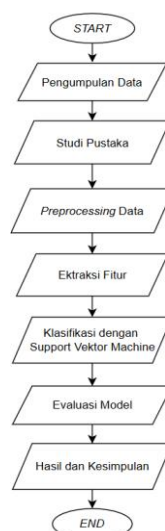
Dalam penelitian ini digunakan algoritma *Support Vector Machine* (SVM) yang dikenal memiliki performa baik dalam klasifikasi teks berdimensi tinggi, termasuk dalam analisis sentimen media sosial [8]. Dengan menggunakan data komentar warganet TikTok, penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai pandangan masyarakat terhadap program pendidikan militer bagi pelajar, sekaligus menjadi referensi bagi pemerintah dalam mengevaluasi kebijakan terkait [9].

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining dan algoritma Support Vector Machine (SVM) untuk melakukan analisis sentimen terhadap komentar warganet di platform TikTok. Tahapan penelitian secara umum terdiri atas pengumpulan data, preprocessing data, ekstraksi fitur, klasifikasi, evaluasi, serta penarikan kesimpulan [10].

2.1. Tahapan Penelitian

Tahapan penelitian ini mencakup proses dari pengumpulan data hingga interpretasi hasil analisis sentimen [11]. Alur penelitian disajikan pada Gambar 1.



Gambar 1 Tahapan Penelitian

1. Pengumpulan Data
Data diperoleh dari komentar warganet pada beberapa akun TikTok terkait isu pendidikan militer. Pengambilan data dilakukan menggunakan APIfy TikTok *Scraper*. Dari total 1.450 komentar yang terkumpul, hanya 1.147 komentar yang layak digunakan setelah melalui tahap penyaringan [12].
2. Studi Pustaka
Peneliti mempelajari berbagai literatur berupa jurnal, artikel, dan penelitian terdahulu sebagai referensi pendukung terkait analisis sentimen, text mining, serta penggunaan algoritma SVM [13].
3. *Preprocessing* Data
Data mentah hasil crawling masih mengandung noise seperti emoticon, tanda baca, angka, dan kata tidak baku. Tahap *preprocessing* dilakukan melalui beberapa langkah: *case folding*, *cleansing*, *tokenizing*, *stopword removal*, *stemming*, dan normalisasi [14].
4. Ekstraksi Fitur
5. Setelah data dibersihkan, dilakukan representasi teks menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) [15].
6. Klasifikasi dengan SVM
Algoritma *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan komentar ke dalam tiga kategori sentimen: positif, negatif, dan netral.
7. Evaluasi Model
Dataset dibagi menjadi 80% data latih dan 20% data uji. Performa model dievaluasi menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score*.
8. Hasil dan Kesimpulan
Hasil klasifikasi dianalisis lebih lanjut untuk mengetahui distribusi opini warganet. Informasi ini digunakan sebagai dasar dalam menarik kesimpulan mengenai respon masyarakat terhadap isu pendidikan militer.

2.2. Data Mining

Data mining merupakan teknik analisis yang digunakan untuk mengidentifikasi pola, tren, dan informasi berharga dari kumpulan *dataset* yang besar [16]. Teknik ini membantu mengubah data mentah menjadi informasi yang dapat digunakan untuk pengambilan keputusan strategis di berbagai bidang, seperti bisnis, kesehatan, pendidikan, dan teknologi. [17] Secara umum, penggunaan *data mining* dapat dikategorikan menjadi dua jenis, yaitu deskriptif dan prediktif. Dalam konteks deskriptif, *data mining* berfungsi untuk mengidentifikasi pola-pola yang menggambarkan karakteristik data. Sementara dalam konteks prediktif, *data mining* digunakan untuk membangun model-model yang memungkinkan untuk melakukan prediksi berdasarkan data yang ada

2.3. Dataset

Dataset merupakan kumpulan fakta dan angka (sering disebut sebagai data mentah) yang berkaitan dengan suatu permasalahan. Data memiliki dua elemen utama, yaitu objek dan atribut. Selain itu, dalam *data mining* dataset sangat penting sebagai objek untuk memperoleh pengetahuan. Secara statistik, dataset adalah kumpulan objek yang memiliki atribut atau variabel tertentu, di mana setiap objek tersebut mewakili individu dalam data yang memiliki sejumlah atribut atau variabel yang relevan [18].

2.4. Klasifikasi

Klasifikasi adalah jenis analisis data yang bertujuan untuk menentukan label kelas dari suatu sampel, termasuk dalam *supervised learning*, yang berfokus dalam mencari hubungan antara atribut input dan atribut target [19]. Tujuannya adalah untuk meningkatkan akurasi dan keandalan hasil yang dihasilkan dari proses analisis data. Sementara, [20] untuk menghasilkan hasil klasifikasi yang optimal, diperlukan algoritma klasifikasi yang cocok dengan karakteristik dataset, sehingga dapat memberikan performa terbaik dalam proses klasifikasi.

2.5. Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah algoritma yang digunakan untuk klasifikasi dan prediksi, bekerja dengan membagi data ke dalam kelas meskipun tidak terdapat garis atau bidang pemisah yang jelas antar kelas tersebut [21] metode ini telah terbukti memberikan hasil yang optimal dalam penerapannya, menjadikannya salah satu pendekatan yang efektif dalam pengolahan data yang kompleks.

2.6. Google Colab

Google Colab (singkatan dari Google Colaboratory) adalah platform berbasis cloud dari Google yang memungkinkan kamu menulis, menjalankan, dan membagikan kode Python langsung di browser, tanpa perlu menginstal Python atau paket tambahan di komputer sendiri.

2.7. Python

Python adalah bahasa pemrograman tingkat tinggi yang dirancang agar mudah dibaca, dipelajari, dan digunakan. Python bersifat serbaguna (*general-purpose*), artinya bisa digunakan untuk berbagai bidang seperti: pengembangan web, analisis data dan sains data, kecerdasan buatan (*artificial intelligence*), *machine learning*, automasi skripsi dan tugas, pengolahan teks (*natural language processing*), pembuatan aplikasi, game bahkan *internet of things (IoT)*. Python diciptakan oleh Guido van Rossum dan pertama kali dirilis tahun 1991. Nama "Python" diambil bukan dari ular, tapi dari acara komedi Inggris *Monty Python*.

2.8. Application Programming Interface Factory (APIfy)

Application Programming Interface Factory (APIfy) adalah platform cloud untuk mengotomatiskan pengambilan data dari web (*web scraping*), pemrosesan data, dan automasi berbasis web. APIfy menyediakan alat siap pakai untuk menarik data dari web dalam format terstruktur seperti JSON, yang kemudian bisa Anda gunakan untuk analisis, termasuk dalam proyek seperti analisis sentimen.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian mengenai analisis sentimen komentar warganet terkait program pendidikan militer bagi pelajar di Jawa Barat. Data komentar dikumpulkan dari platform TikTok dan dianalisis menggunakan algoritma *Support Vector Machine (SVM)*. Hasil analisis tidak hanya menampilkan distribusi sentimen, tetapi juga mengevaluasi performa model klasifikasi dan membahas temuan dalam konteks literatur terkait.

3.1. Data Penelitian

Data penelitian berupa komentar warganet pada platform TikTok terkait program pendidikan militer bagi pelajar di Jawa Barat. Data dikumpulkan menggunakan APIfy TikTok Scraper dari 12 tautan video. Dari total 1.450 komentar, sebanyak 1.147 komentar layak digunakan setelah melalui proses pembersihan data.

3.2. Hasil Preprocessing

Data komentar yang telah dikumpulkan masih mentah karena masih mengandung simbol, angka, *emoticon*, serta kata-kata tidak baku. Oleh karena itu, dilakukan tahapan preprocessing yang meliputi:

a. Case folding

Tabel 1. Hasil Case Folding

Sebelum	Sesudah
HARUSNYA PRESIDEN LEBIH PEKA DARI DULU. malah pak Gubernur yang Lebih Peka terhadap Keadaan Anak nakalnya.	harusnya presiden lebih peka dari dulu. malah pak gubernur yang lebih peka terhadap keadaan anak nakalnya.

Dari Tabel 1. Hasil *Case Folding* dapat dilihat bahwa proses *case folding* berhasil mengubah seluruh huruf kapital menjadi huruf kecil. Hal ini penting untuk menyeragamkan teks sehingga kata-kata seperti " HARUSNYA PRESIDEN LEBIH PEKA DARI DULU...." dan " harusnya presiden lebih peka dari dulu....." dianggap sama oleh sistem, yang pada akhirnya membantu meningkatkan akurasi dalam proses analisis dan klasifikasi teks.

b. Cleaning

Tabel 2. Hasil *Cleaning*

Sebelum	Sesudah
orang tuanya aja kaga masalah, tapi ada aja yang berusaha untuk menghalangi program ini ya 😊	orang tuanya aja kaga masalah tapi ada aja yang berusaha untuk menghalangi program ini ya.

Dari Tabel 2. Hasil *Cleaning* dapat dilihat bahwa proses *cleaning* berhasil menghapus tanda baca yang tidak diperlukan seperti tanda seru dan titik, serta menghilangkan karakter huruf yang berulang secara berlebihan. Proses ini bertujuan untuk menyederhanakan teks agar lebih mudah diproses pada tahap selanjutnya dalam analisis teks.

c. *Tokenizing dan Stopword removal*

Tabel 4. Hasil *Tokenizing dan Stopword removal*

Sebelum	Sesudah
waduhhh ini sih udah kanyak pendidikan militer saya yakin pulang dari situ jiwa pemimpin nya keluar rasa takutnya berkurang tapi sopan nya pasti di utamakan percaya sama saya	['waduhhh', 'sih', 'udah', 'kanyak', 'pendidikan', 'militer', 'yakin', 'pulang', 'situ', 'jiwa', 'pemimpin', 'keluar', 'rasa', 'takutnya', 'berkurang', 'tapi', 'sopan', 'pasti', 'utamakan', 'percaya', 'sama']

Dari tabel 4. Hasil *Tokenizing dan Stopword removal* di atas, terlihat bahwa teks asli telah dipecah menjadi token-token individual. Beberapa kata yang bersifat pengisi atau umum dapat dihapus pada tahap selanjutnya untuk meningkatkan kualitas fitur yang digunakan dalam analisis teks.

d. *Stemming*

Tabel 4. Hasil *Stemming*

Sebelum	Sesudah
komennya kenapa dah gue juga pernah pendidikan militer malah lebih berkarakter dan bagus daripada ke pesantren gue akui malah lebih dalem pendidikan bela negaranya juga	['komennya', 'kenapa', 'dah', 'gue', 'pernah', 'didik', 'militer', 'malah', 'lebih', 'karakter', 'bagus', 'daripada', 'pesantren', 'gue', 'aku', 'malah', 'lebih', 'dalem', 'didik', 'bela', 'negara']

Dari hasil ini terlihat bahwa kata “pendidikan” berubah menjadi “didik”, “berkarakter” menjadi “karakter”, “mengakui” menjadi “aku”, dan “bela negaranya” menjadi “bela” dan “negara”. Tahap stemming ini penting karena mengurangi variasi bentuk kata sehingga analisis teks lebih fokus pada makna inti, bukan perbedaan bentuk kata.

e. *Normalisasi Teks*

Tabel 5. Hasil *Normalisasi Teks*

Sebelum	Sesudah
dikasih solusi malah banyak yang ragu giliran dibiarin malah nyalahin pemerintah karna tak bisa cari solusi	kasih solusi banyak ragu giliran biar salah pemerintah karena cari solusi"

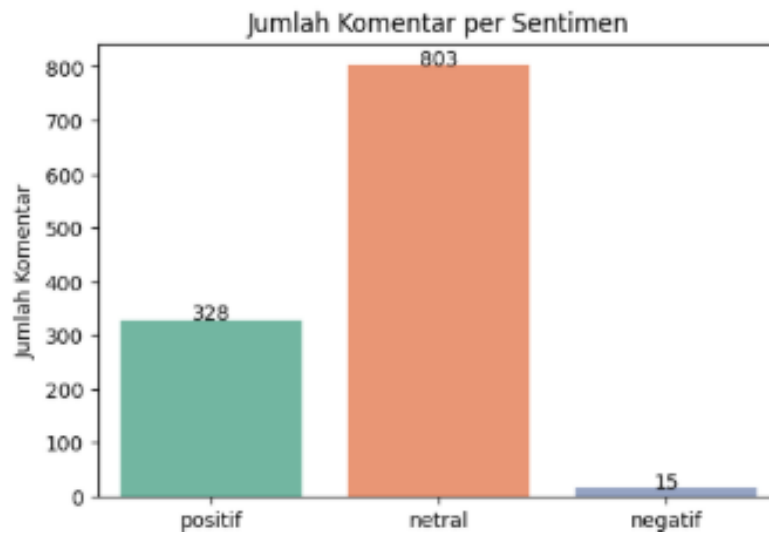
Dari tabel diatas terlihat bahwa beberapa kata informal dan tidak baku telah diubah:

1. 'dikasih' → 'kasih'
2. 'dibiarin' → 'biar'
3. 'nyalahin' → 'salah'
4. 'karna' → 'karena'

Selain itu, kata-kata umum yang tidak terlalu penting seperti *malah* atau *yang* bisa dihapus jika diperlukan untuk analisis lebih lanjut. Hasil normalisasi ini membuat teks menjadi lebih bersih, konsisten, dan siap untuk tahap ekstraksi fitur atau analisis teks berikutnya.

3.3. Distribusi Sentimen

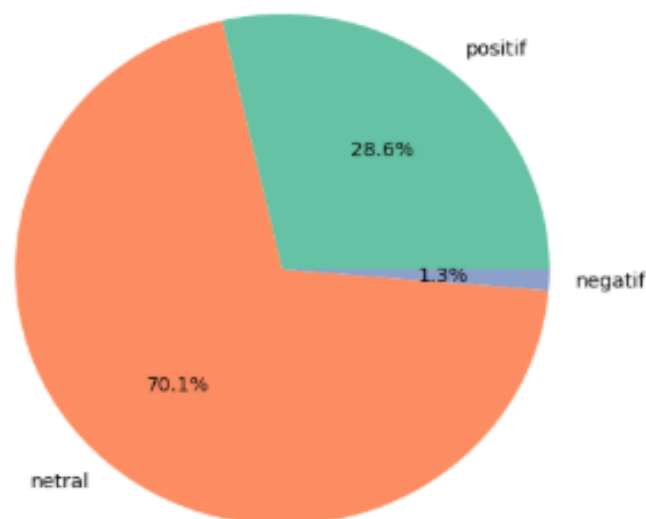
Setelah dilakukan labeling dan klasifikasi, distribusi sentimen komentar warganet, perhatikan gambar berikut ini:



Gambar 2. Visualisasi Distribusi Sentimen

Berdasarkan grafik jumlah komentar per sentimen, terlihat bahwa mayoritas komentar warganet bersifat netral, dengan total 803 komentar. Komentar positif menempati posisi kedua dengan 328 komentar, sementara komentar negatif paling sedikit, hanya 15 komentar. Hal ini menunjukkan bahwa sebagian besar warganet cenderung memberikan tanggapan yang netral terhadap topik yang dibahas, dengan tingkat sentimen negatif yang sangat rendah. Temuan ini dapat menjadi indikasi bahwa isu yang dianalisis tidak menimbulkan kontroversi yang signifikan, dan persepsi publik cenderung stabil atau moderat.

Perhatikan *Bar Chart* berikut ini untuk melihat hasil presentasi dari distribusi sentimen



Gambar 3. Visualisasi Persentasi Distribusi Sentimen

Keterangan:

- a. Netral: 803 komentar (70,1%).
- b. Positif: 328 komentar (28,6%).
- c. Negatif: 15 komentar (1,3%).

Dari gambar di atas menunjukkan distribusi sentimen komentar warganet terhadap program pendidikan militer bagi pelajar di Jawa Barat. Dari total komentar yang dianalisis, mayoritas berada pada kategori netral, yaitu sebesar 70,1%. Hal ini menunjukkan bahwa sebagian besar warganet memberikan tanggapan yang bersifat deskriptif atau informatif, tanpa menunjukkan dukungan atau penolakan yang jelas terhadap program tersebut. Selanjutnya, komentar dengan sentimen positif mencapai 28,6%, menandakan adanya kelompok warganet yang mendukung program ini, kemungkinan karena melihat manfaat dalam pembinaan disiplin, tanggung jawab, dan pengembangan karakter pelajar. Sementara itu, komentar dengan sentimen negatif hanya 1,3%, yang menunjukkan bahwa penolakan terhadap kebijakan ini relatif minor di kalangan pengguna TikTok.

Distribusi ini memperlihatkan bahwa opini warganet cenderung condong ke arah netral, dengan dukungan positif yang cukup signifikan dan penolakan yang minimal. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa pengguna media sosial cenderung lebih sering mengekspresikan sentimen netral dalam isu kebijakan publik, sementara sentimen positif dan negatif muncul dalam proporsi yang lebih rendah. Berdasarkan hasil klasifikasi sentimen terhadap komentar warganet pada *platform* TikTok ditunjukkan pada Gambar 2. Grafik tersebut memperlihatkan bahwa kategori netral mendominasi dengan jumlah 803 komentar atau sekitar 70,1% dari total data. Selanjutnya, komentar dengan kategori positif berjumlah 328 komentar atau 28,6%. Sementara itu, kategori negatif hanya ditemukan pada 15 komentar atau sebesar 1,3%..

Distribusi ini menunjukkan bahwa mayoritas warganet lebih banyak memberikan komentar yang bersifat deskriptif dan tidak berpihak, sehingga digolongkan sebagai sentimen netral. Persentase sentimen positif yang cukup tinggi menggambarkan adanya dukungan warganet terhadap isu yang dibahas, sedangkan persentase negatif yang relatif kecil mengindikasikan bahwa penolakan terhadap kebijakan tersebut tidak dominan. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa warganet cenderung lebih sering mengekspresikan sentimen netral ketika membahas isu kebijakan publik di media sosial, sementara sentimen positif dan negatif muncul dalam proporsi yang lebih rendah. Dengan demikian, hasil analisis ini dapat menjadi dasar bagi pemangku kebijakan untuk memahami kecenderungan opini publik dari perspektif pengguna TikTok.

3.4. Evaluasi

Berikut beberapa hasil evaluasi dalam penelitian ini;

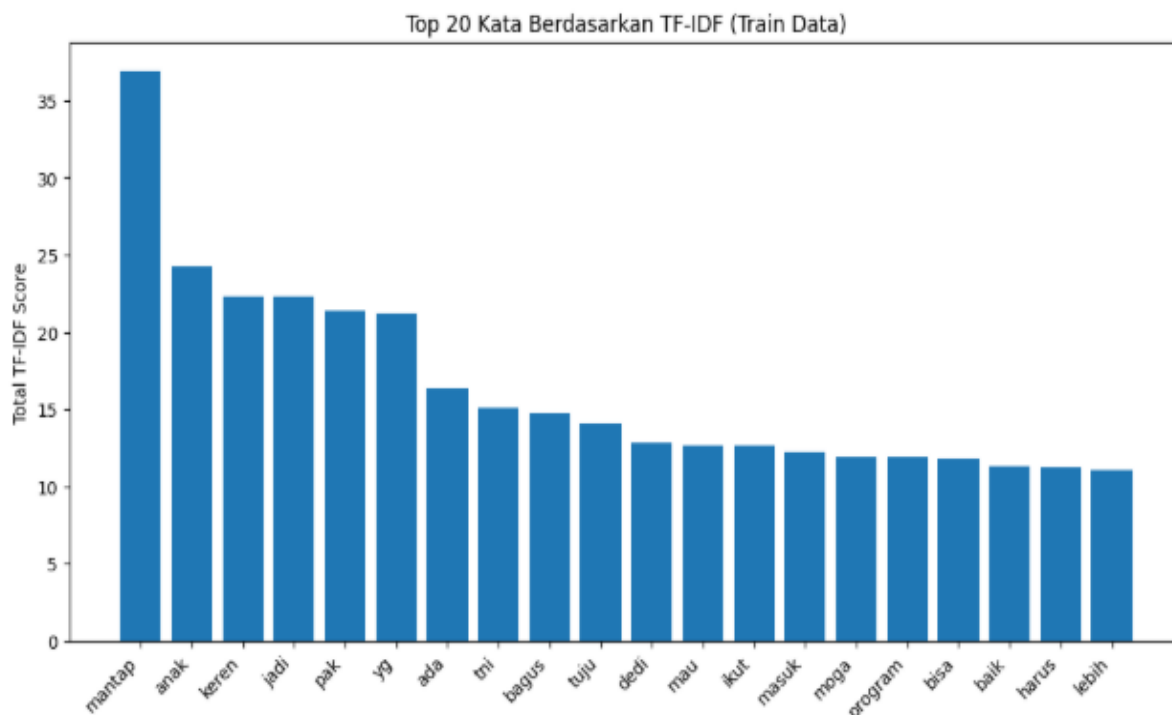
- a. Vektorisasi

```
Jumlah data setelah dibersihkan: 1144
Cek jumlah NaN: 0
Jumlah data train: 915
Jumlah data test : 229
Ukuran data train (vectorized): (915, 1902)
Ukuran data test (vectorized): (229, 1902)
```

Gambar 4. Pembagian Data dan Proses Vektorisasi

Berdasarkan hasil pengolahan data, jumlah data yang digunakan setelah melalui tahap pembersihan adalah sebanyak 1.144 data, tanpa ada nilai kosong (NaN). Data tersebut kemudian dibagi menjadi dua bagian, yaitu data latih (*train*) sebanyak 915 data dan data uji (*test*) sebanyak 229 data. Proses vektorisasi menghasilkan representasi numerik dengan dimensi fitur sebanyak 1.902, sehingga ukuran data latih menjadi (915, 1902) dan ukuran data uji menjadi (229, 1902). Hal ini menunjukkan bahwa setiap data direpresentasikan dalam 1.902 fitur, yang siap digunakan dalam proses pelatihan dan pengujian model *machine learning*.

b. TF-IDF



Gambar 5. Hasil Visualisasi TF-IDF

Berdasarkan grafik batang vertikal pada gambar 4. Hasil Visualisasi TF-IDF ini menampilkan 20 kata teratas berdasarkan skor TF-IDF (*Term Frequency-Inverse Document Frequency*), yang menunjukkan kata-kata paling penting atau sering muncul dalam kumpulan data yang dianalisis. Kata dengan skor tertinggi adalah “mantap”, “anak”, “keren”, “jadi”, “pak”, “yg”, “ada”, “tni”, “bagus”, “dedi”, “mau”, “masuk”, “moga”, “program”, “bisa”, “baik”, “harus”, “lebih”. Grafik ini membantu mengidentifikasi kata kunci utama yang sering muncul dan dianggap relevan dalam analisis teks yang dilakukan.

c. Hasil Evaluasi *Performance Support Vector Machine*

Evaluasi dilakukan menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score*. Hasil menunjukkan bahwa model SVM mampu melakukan klasifikasi komentar dengan tingkat akurasi yang baik serta nilai *presisi*, *recall*, dan *F1-score* yang konsisten. Hal ini membuktikan bahwa SVM efektif digunakan dalam analisis sentimen berbasis teks di media sosial.

Accuracy : 0.9475982532751092
 Precision: 0.9498215187294614
 Recall : 0.9475982532751092
 F1-Score : 0.9448001639155323

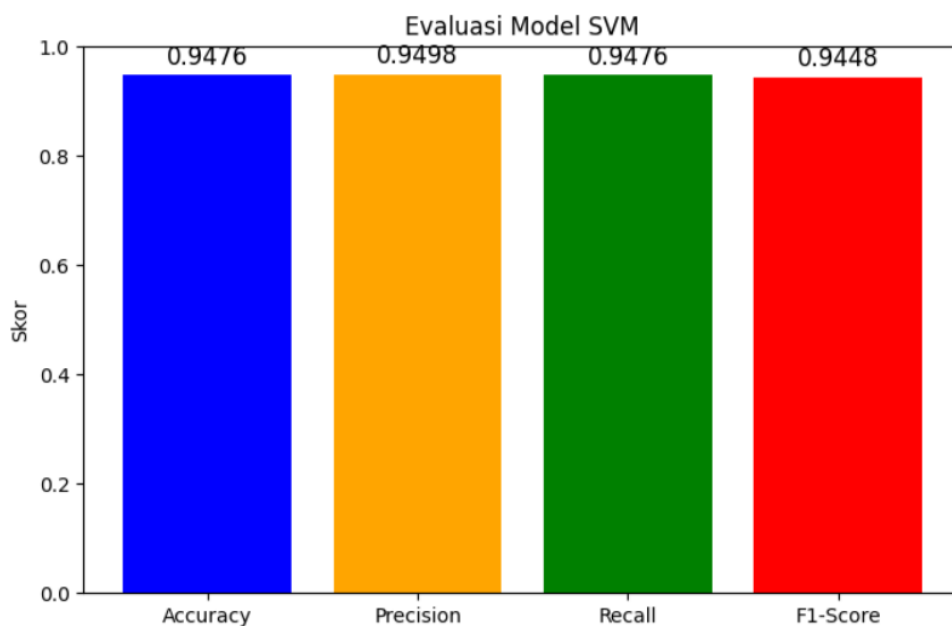
Classification Report:				
	precision	recall	f1-score	support
negatif	1.00	0.33	0.50	3
netral	0.94	0.99	0.96	160
positif	0.98	0.86	0.92	66
accuracy			0.95	229
macro avg	0.97	0.73	0.79	229
weighted avg	0.95	0.95	0.94	229

Gambar 6. Hasil Akurasi SVM

Berdasarkan hasil evaluasi performa model klasifikasi sentimen, diperoleh nilai akurasi sebesar 0.9476 (94,76%), yang menunjukkan bahwa model mampu memprediksi data dengan tingkat ketepatan yang cukup tinggi. Nilai *precision* (0.9498), *recall* (0.9476), dan *F1-score* (0.9448) juga menunjukkan keseimbangan yang baik antara ketepatan prediksi dan kemampuan model dalam mengidentifikasi kelas dengan benar.

Pada hasil *classification report*, kelas *netral* memiliki performa terbaik dengan *precision* 0.94, *recall* 0.99, dan *F1-score* 0.96. Hal ini menunjukkan bahwa model sangat efektif dalam mengklasifikasikan data dengan sentimen netral, yang juga didukung oleh jumlah data uji yang lebih besar (160 data). Untuk kelas *positif*, diperoleh nilai *precision* 0.98, *recall* 0.86, dan *F1-score* 0.92, menandakan model cukup akurat meskipun masih terdapat beberapa kesalahan dalam mengidentifikasi sentimen positif. Sementara itu, performa model pada kelas *negatif* masih relatif rendah dengan *precision* 1.00 tetapi *recall* hanya 0.33, yang berarti meskipun prediksi negatif yang dihasilkan model benar, sebagian besar data negatif tidak berhasil teridentifikasi dengan baik. Hal ini kemungkinan dipengaruhi oleh jumlah data uji negatif yang jauh lebih sedikit (hanya 3 data), sehingga distribusi kelas yang tidak seimbang (*imbalanced dataset*) memengaruhi hasil klasifikasi.

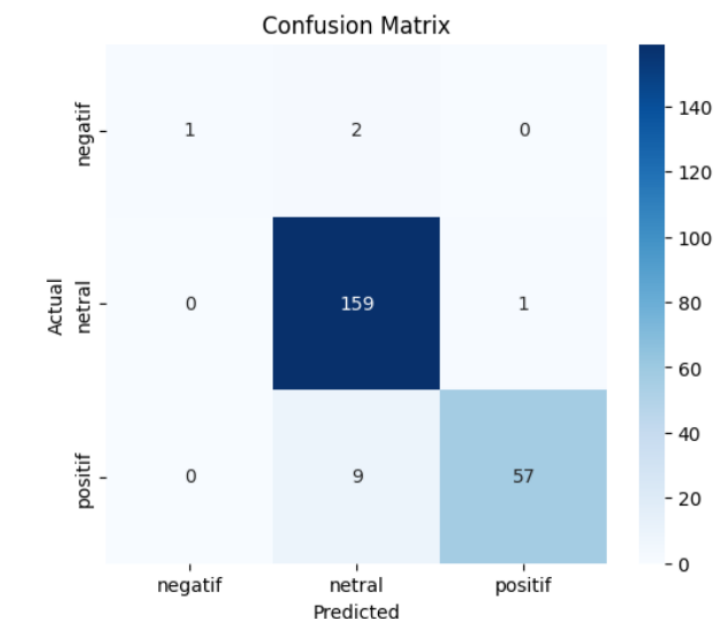
Secara keseluruhan, hasil pengujian menunjukkan bahwa model memiliki kinerja yang sangat baik dalam mendeteksi sentimen *netral* dan *positif*, namun masih perlu dilakukan perbaikan dalam mendeteksi sentimen *negatif*. Upaya yang dapat dilakukan antara lain adalah menyeimbangkan jumlah data latih pada tiap kelas, serta menerapkan teknik *resampling* atau *class weighting* untuk meningkatkan performa model pada kelas minoritas.



Gambar 7. Visualisasi Model SVM

Gambar di atas menunjukkan hasil evaluasi performa model klasifikasi *Support Vector Machine* (SVM) menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Secara umum, model SVM memiliki kinerja yang sangat baik dengan nilai evaluasi di atas 0.94 untuk semua metrik. Nilai akurasi yang dicapai adalah 0.9476, menunjukkan bahwa hampir 95% data berhasil diklasifikasikan dengan benar oleh model. *Precision* sebesar 0.9498 menandakan bahwa sebagian besar prediksi positif yang dihasilkan oleh model sesuai dengan label sebenarnya. Sementara itu, *recall* sebesar 0.9476 menunjukkan kemampuan model dalam mendeteksi data sesuai dengan kelasnya secara konsisten. Terakhir, nilai *F1-Score* sebesar 0.9448 mencerminkan keseimbangan yang baik antara *precision* dan *recall*.

Secara keseluruhan, grafik ini memperlihatkan bahwa model SVM mampu bekerja dengan stabil dan konsisten dalam mengklasifikasikan data sentimen. Dengan performa yang merata pada seluruh metrik evaluasi, SVM dapat dikatakan cukup handal untuk digunakan pada tahap analisis selanjutnya.

Gambar 7. Hasil *Matrix Confusion*

Berdasarkan *confusion matrix* di atas, dapat dijelaskan bahwa sistem klasifikasi sentimen bekerja cukup baik, terutama pada kelas netral. Dari total data uji, terdapat 159 data netral yang berhasil diprediksi dengan benar, hanya terjadi sedikit kesalahan di mana 1 data netral diprediksi sebagai positif. Untuk kelas positif, sebanyak 57 data positif berhasil diprediksi benar, namun terdapat 9 data positif yang salah diprediksi sebagai netral. Sementara itu, pada kelas negatif, terdapat 1 data negatif diprediksi benar, namun ada 2 data negatif yang salah diklasifikasikan sebagai netral. Secara umum, performa model menunjukkan kecenderungan yang lebih tinggi dalam mengklasifikasikan teks sebagai netral, dibandingkan dengan positif atau negatif. Hal ini terlihat dari jumlah prediksi benar pada kelas netral yang jauh lebih tinggi dibandingkan kelas lainnya.

3.5. Pembahasan

Hasil penelitian menunjukkan bahwa model *Support Vector Machine* (SVM) mampu memberikan performa yang baik dalam mengklasifikasikan komentar warganet TikTok terkait program pendidikan militer. Dengan nilai akurasi sebesar 94,76%, model ini dapat dikatakan cukup andal dalam mendeteksi sentimen. Hal ini sejalan dengan karakteristik SVM yang dikenal efektif dalam menangani data berdimensi tinggi seperti teks hasil vektorisasi TF-IDF. Dari sisi distribusi sentimen, mayoritas komentar warganet tergolong netral (70,1%), diikuti komentar positif (28,6%), dan negatif hanya 1,3%. Pola ini menunjukkan bahwa isu pendidikan militer tidak memicu kontroversi besar di masyarakat digital, melainkan lebih banyak direspon dengan nada deskriptif atau informatif. Tingginya persentase komentar netral juga berkontribusi pada performa model yang lebih baik pada kelas tersebut.

Evaluasi lebih lanjut melalui *classification report* dan *confusion matrix* memperlihatkan bahwa model sangat baik dalam mengklasifikasikan sentimen netral (*precision* 0.94, *recall* 0.99, *F1-score* 0.96). Namun, performa pada kelas negatif masih lemah, ditunjukkan dengan *recall* hanya 0.33, yang berarti sebagian besar data negatif tidak berhasil terdeteksi. Penyebab utamanya adalah distribusi data yang tidak seimbang (*imbalanced dataset*), di mana jumlah komentar negatif sangat sedikit dibanding netral maupun positif. Hal ini membuat model cenderung bias terhadap kelas dengan jumlah data yang lebih besar.

Kinerja pada kelas positif cukup baik (*precision* 0.98, *recall* 0.86, *F1-score* 0.92), meskipun masih terdapat 9 komentar positif yang salah diklasifikasikan sebagai netral. Temuan ini mengindikasikan bahwa beberapa komentar positif memiliki gaya bahasa yang dekat dengan komentar netral, sehingga menyulitkan model dalam membedakan keduanya. Secara keseluruhan, hasil penelitian ini mengonfirmasi bahwa SVM dapat digunakan secara efektif untuk analisis sentimen pada media sosial, khususnya dengan *representasi* TF-IDF. Namun, tantangan utama terletak pada ketidakseimbangan

data antar kelas, terutama pada kelas negatif. Perbaikan dapat dilakukan melalui pendekatan *oversampling*, *undersampling*, atau penerapan *class weighting*, sehingga distribusi data lebih seimbang dan performa model meningkat pada kelas minoritas.

Temuan ini sejalan dengan penelitian terdahulu yang menunjukkan bahwa SVM memiliki performa stabil pada data teks dengan jumlah fitur besar, namun akurasi dapat menurun apabila distribusi kelas tidak merata. Dengan demikian, penelitian ini tidak hanya memberikan gambaran tentang sentimen publik terhadap isu pendidikan militer, tetapi juga menegaskan pentingnya pengolahan dataset yang seimbang dalam membangun model klasifikasi sentimen yang lebih adil dan akurat.

4. KESIMPULAN

Penelitian ini berhasil mengumpulkan 1.144 komentar warganet TikTok terkait program pendidikan militer bagi pelajar, setelah melalui *preprocessing* dan vektorisasi menghasilkan 1.902 fitur. Analisis sentimen menunjukkan mayoritas komentar netral (70, 1%), diikuti positif (28, 6%), dan negatif (1, 3%). Berdasarkan hasil evaluasi model SVM menghasilkan performa yang sangat baik. Terutama setelah penyeimbangan data menggunakan SMOTE dengan nilai *accuracy* 94, 76%, *precision* 94, 98%, *recall* 94, 76%, dan *f1-score* 94,48%. Model SVM mampu mengenali komentar netral dengan sangat akurat, cukup baik dalam mengklasifikasikan komentar positif, namun masih memiliki keterbatasan dalam mengenali komentar negatif karena jumlah data yang sangat sedikit (*imbalanced*).

Secara keseluruhan, penelitian ini menunjukkan bahwa SVM efektif dalam menganalisis sentimen warganet TikTok dengan akurasi yang tinggi. Hasil analisis dapat menjadi gambaran objektif mengenai opini publik terhadap program pendidikan militer bagi pelajar. Temuan ini juga dapat dimanfaatkan oleh pihak pemerintah sebagai bahan evaluasi kebijakan, khususnya dalam melihat bagaimana respon masyarakat di media sosial.

Saran, untuk pengembangan penelitian selanjutnya, beberapa saran yang dapat diberikan adalah sebagai berikut:

1. Menambah jumlah data latih, terutama pada kelas negatif, agar distribusi data lebih seimbang dan model dapat belajar dengan lebih baik.
2. Menggunakan teknik penanganan data *imbalanced*, seperti SMOTE (*Synthetic Minority Oversampling Technique*) atau *class weight adjustment*, untuk membantu meningkatkan performa model pada kelas minoritas.
3. Membandingkan SVM dengan algoritma lain, misalnya *Naïve Bayes*, *Random Forest*, *Logistic Regression*, atau model *deep learning* seperti LSTM dan BERT, untuk mengetahui algoritma mana yang memiliki kinerja terbaik pada data komentar yang digunakan.
4. Mengeksplorasi representasi teks lain selain TF-IDF, seperti *Word2Vec*, *FastText*, atau *BERT embeddings*, yang dapat menangkap konteks kata dengan lebih baik dibandingkan pendekatan berbasis frekuensi.
5. Mengembangkan sistem aplikasi berbasis web atau *mobile*, sehingga model klasifikasi sentimen dapat dimanfaatkan secara langsung dalam analisis opini publik, misalnya untuk memantau komentar media sosial, forum diskusi, maupun ulasan produk.

Daftar Pustaka

- [1] D. Andriyani, M. A. Ramdhani, dan E. Kurniawan, "Analisis sentimen menggunakan TF-IDF untuk klasifikasi teks," *Jurnal Informatika*, vol. 10, no. 2, hlm. 45–52, 2023.
- [2] A. P. Astuti, S. Alam, dan I. Jaelani, "Komparasi algoritma Support Vector Machine dengan Naive Bayes untuk analisis sentimen pada aplikasi BRImo," *Jurnal Bangkit Indonesia*, vol. 11, no. 2, hlm. 1–7, 2022.
- [3] D. Wahyuni dan T. Ramadhan, "Studi komparatif preprocessing teks untuk analisis sentimen pada media sosial," *Jurnal Penelitian Teknologi Informasi*, vol. 12, no. 3, hlm. 150–159, 2021.
- [4] S. A. Putri dan A. Nugroho, "Evaluasi performa algoritma klasifikasi menggunakan metrik akurasi, presisi, recall, dan F1-score," *Jurnal Sistem Komputer*, vol. 10, no. 4, hlm. 55–63, 2020.

- [5] F. Baharuddin dan A. Tjahyanto, "Peningkatan performa klasifikasi machine learning melalui perbandingan metode machine learning dan peningkatan dataset," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 1, hlm. 25–31, 2022.
- [6] N. F. Budiarti, A. Riandhita, N. Audra, dan S. Tally, "Preferensi Gen Z terhadap kemajuan teknologi pada aplikasi TikTok dan YouTube," *Proceeding Universitas Negeri Surabaya*, vol. 1, no. 1, hlm. 501–508, 2022.
- [7] L. Rahmawati dan A. Hasan, "Analisis data untuk pengambilan keputusan menggunakan algoritma machine learning," *Jurnal Sistem Informasi*, vol. 14, no. 2, hlm. 70–78, 2023.
- [8] R. Hidayat dan N. Yuliani, "Penerapan analisis sentimen untuk media sosial TikTok menggunakan pendekatan machine learning," *Jurnal Ilmu Komputer dan Aplikasi*, vol. 15, no. 2, hlm. 89–97, 2023.
- [9] A. A. Devi, "Pemanfaatan aplikasi TikTok sebagai media pembelajaran," *Jurnal Epistema*, vol. 3, no. 1, hlm. 1–17, 2022.
- [10] M. Sahir, "Pendekatan penarikan kesimpulan berdasarkan bukti penelitian," *Jurnal Penelitian dan Evaluasi*, vol. 9, no. 1, hlm. 22–29, 2021.
- [11] W. Ningsih, B. Alfianda, R. Rahmaddeni, dan D. Wulandari, "Perbandingan algoritma SVM dan Naive Bayes dalam analisis sentimen Twitter pada penggunaan mobil listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, hlm. 57–562, 2024.
- [12] F. Sari dan A. Prasetyo, "Implementasi klasifikasi teks pada media sosial menggunakan algoritma SVM," *Jurnal Informatika Terapan*, vol. 7, no. 1, hlm. 35–44, 2021.
- [13] R. Putra dan T. Hartono, "Analisis sentimen opini publik terhadap kebijakan pemerintah," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 3, hlm. 77–85, 2020.
- [14] D. Lestari dan P. Nugraha, "Evaluasi performa machine learning untuk klasifikasi teks berita," *Jurnal Sistem Informasi dan Komputer*, vol. 12, no. 2, hlm. 66–74, 2022.
- [15] M. Santoso dan R. Handoko, "TF-IDF dan teknik ekstraksi fitur pada analisis teks," *Jurnal Ilmu Komputer Indonesia*, vol. 5, no. 3, hlm. 21–30, 2020.
- [16] R. P. Sari, "Apa itu Data Mining? Pengertian, Metode dan Penerapannya," *Cloud Computing Indonesia*. [Online]. Available: <https://www.cloudcomputing.id/pengetahuan-dasar/apa-itu-data-mining>
- [17] N. L. W. S. R. Ginantra et al., *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis, 2021.
- [18] M. A. Muslim et al., *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program komputer*, vol. 11, no. 1. Semarang, 2019.
- [19] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [20] F. Baharuddin and A. Tjahyanto, "Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Peningkatan Dataset," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 1, pp. 25–31, 2022, doi: 10.32736/sisfokom.v11i1.1337
- [21] W. Ningsih, B. Alfianda, Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 557–562, 2024, doi: 10.57152/malcom.v4i2.1253.



ZONAsi: Jurnal Sistem Informasi

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)