



Sentiment Analysis of Motivational Content on Social Media Using Support Vector Machine

Nadiah Zakira¹, Salahuddin^{2*}

^{1,2}Jurusan Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe, Kota Lhokseumawe, 24301
INDONESIA

*Penulis Korespondensi : salahuddin@pnl.ac.id

INFORMASI ARTIKEL

Riwayat artikel:

Diajukan pada 01 November 2025

Direvisi pada 17 November 2025

Publikasi pada 20 Desember 2025

Kata kunci:

Analisis sentimen

Motivasi

Support Vector Machine

Machine Learning

Keywords:

Sentiment Analysis

Motivation Quote

Support Vector Machine

Machine Learning

ABSTRAK

Media sosial telah menjadi platform utama dalam berbagi informasi, termasuk kalimat-kalimat motivasi yang memiliki potensi besar dalam memengaruhi pola pikir dan perilaku pengguna. Motivasi, baik yang berasal dari dalam diri maupun dari lingkungan sekitar, berperan penting dalam mendorong aktivitas sehari-hari seseorang. Oleh karena itu, analisis sentimen terhadap konten motivasi di media sosial menjadi relevan untuk memahami dampaknya terhadap kondisi psikologis pengguna, seperti peningkatan semangat hidup atau produktivitas. Penelitian ini berfokus pada analisis sentimen terhadap kalimat motivasi di media sosial menggunakan metode Support Vector Machine (SVM). Metode SVM dipilih karena kemampuannya dalam mengklasifikasikan teks dengan akurasi tinggi, bahkan pada data yang kompleks dan berdimensi besar. Proses penelitian meliputi pengumpulan data (crawling), preprocessing teks menggunakan teknik NLP, pelabelan sentimen, serta pelatihan dan pengujian model klasifikasi menggunakan algoritma SVM. Evaluasi model dilakukan dengan menggunakan Confusion Matrix untuk mengukur nilai akurasi, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa sistem analisis sentimen yang dibangun mampu mencapai tingkat akurasi sebesar 96,62%. Selain itu, nilai F1-score yang diperoleh untuk label positif sebesar 97,19%, netral 96,84%, dan negatif 95,04%, menandakan bahwa metode SVM sangat efektif dalam melakukan klasifikasi sentimen terhadap kalimat motivasi. Dengan demikian, penelitian ini berhasil membuktikan bahwa SVM merupakan metode yang valid dan andal dalam menganalisis dampak emosional dari konten motivasi di media sosial.

ABSTRACT

Social media has become a primary platform for sharing information, including motivational content that holds significant potential to influence user mindset and behavior. Motivation, whether intrinsic or from one's environment, plays a crucial role in driving an individual's daily activities. Therefore, sentiment analysis of motivational content on social media is relevant for understanding its impact on users' psychological states, such as boosting life spirit or productivity. This research focuses on performing sentiment analysis on motivational sentences from social media using the Support Vector Machine (SVM) method. SVM was chosen for its proven capability in text classification with high accuracy, even on complex and high-dimensional data. The research process included data collection (crawling), text preprocessing using NLP techniques, sentiment labeling, and the training and testing of the classification model using the SVM algorithm. The model was evaluated using a Confusion Matrix to measure accuracy, precision, recall, and F1-score. The test results demonstrate that the built sentiment analysis system achieved an accuracy of 96.62%. Furthermore, the obtained F1-scores were 97.19% for the positive label, 96.84% for neutral, and 95.04% for negative, indicating that the SVM method is

highly effective in classifying the sentiment of motivational content. Thus, this study successfully proves that SVM is a valid and reliable method for analyzing the emotional impact of motivational content on social media.

1. Pendahuluan

Pada era digital saat ini, media sosial telah menjadi salah satu platform utama bagi masyarakat untuk berbagi informasi, mengekspresikan diri, dan berinteraksi [1]. Kemajuan teknologi informasi telah membuka peluang untuk melakukan analisis mendalam terhadap dampak pada konten digital, termasuk kalimat motivasi yang tersebar di media sosial. Proses analisis ini mencakup ekstraksi, identifikasi, dan klasifikasi sentimen yang terdapat dalam teks, sehingga memberikan pemahaman yang lebih baik mengenai pandangan dan opini pengguna [2]. Setiap individu memiliki kondisi internal yang berpengaruh signifikan terhadap aktivitas sehari-harinya. Salah satu kondisi tersebut adalah motivasi, yang merupakan dorongan dasar yang mendorong seseorang untuk bertindak [3].

Motivasi dapat dipahami sebagai serangkaian sikap dan nilai yang memengaruhi individu dalam mencapai tujuan tertentu, baik yang berasal dari dalam diri maupun dari lingkungan sekitar [4]. Kalimat motivasi pada media sosial menjadi salah satu bentuk konten yang banyak dibagikan, namun tidak selalu diterima dengan respon yang sama oleh setiap pengguna. Ada kalimat yang menumbuhkan semangat, tetapi ada juga yang dianggap netral bahkan memicu respon negatif. Melalui analisis sentimen, peneliti dapat memahami lebih dalam bagaimana pesan-pesan tersebut diterima oleh pengguna serta dampaknya terhadap perilaku dan pola pikir mereka [5]. Salah satu platform media sosial yang populer saat ini adalah X, yang banyak digunakan masyarakat untuk mengekspresikan diri dan membagikan berbagai bentuk konten, seperti video, gambar, maupun kalimat. Platform ini juga menjadi ruang bagi para pengguna, khususnya generasi muda, untuk menyebarkan dan menemukan informasi dengan cepat. Salah satu jenis konten yang cukup sering muncul di X adalah kalimat motivasi, karena sifat platform ini yang memungkinkan pesan singkat menjangkau audiens secara luas. Karakteristik teks pendek pada X membuatnya relevan sebagai sumber data dalam penelitian analisis sentimen, dengan demikian platform X dipilih sebagai media pengumpulan data dalam penelitian ini untuk mendukung analisis sentimen [6].

Dampak dari konten motivasi di media sosial dapat ditelusuri melalui respon emosional yang ditunjukkan pengguna [7]. Salah satu cara untuk memahami respon tersebut adalah dengan melakukan analisis sentimen terhadap kalimat-kalimat motivasi yang dibagikan. Analisis ini memungkinkan pengelompokan konten ke dalam kategori sentimen positif, negatif, maupun netral, sehingga memberikan gambaran umum mengenai persepsi pengguna terhadap pesan motivasi di media sosial [7][8]. Dengan demikian analisis sentimen berperan penting dalam menilai sejauh mana kalimat motivasi benar-benar memberikan pengaruh yang konstruktif atau sebaliknya [8]. Oleh karena itu, diperlukan metode klasifikasi yang mampu menghasilkan tingkat akurasi tinggi untuk memastikan hasil analisis dapat diandalkan.

Dalam melakukan analisis sentimen terhadap konten motivasi, salah satu pendekatan yang efektif adalah machine learning [9][10]. Melalui teknik pengolahan bahasa alami atau NLP (Natural Language Processing), data dari media sosial dapat diolah untuk mengidentifikasi emosi, opini, maupun sikap pengguna terhadap suatu topik [11]. Salah satu algoritma machine learning yang populer adalah SVM (Support Vector Machine), yang terbukti efektif dalam klasifikasi teks karena kemampuannya mengolah

data berdimensi tinggi dengan tingkat akurasi yang baik [12][13]. Penelitian sebelumnya menunjukkan bahwa SVM mampu mencapai akurasi hingga 82% dalam klasifikasi opini di media sosial, dengan proporsi sentimen negatif 77%, positif 8%, dan netral 15%, sehingga menjadikannya metode yang banyak digunakan dalam analisis sentimen [10][14]. Berdasarkan keunggulan tersebut, penelitian ini dilakukan untuk mengukur tingkat akurasi metode SVM dalam menganalisis sentimen terhadap kalimat motivasi di media sosial [15].

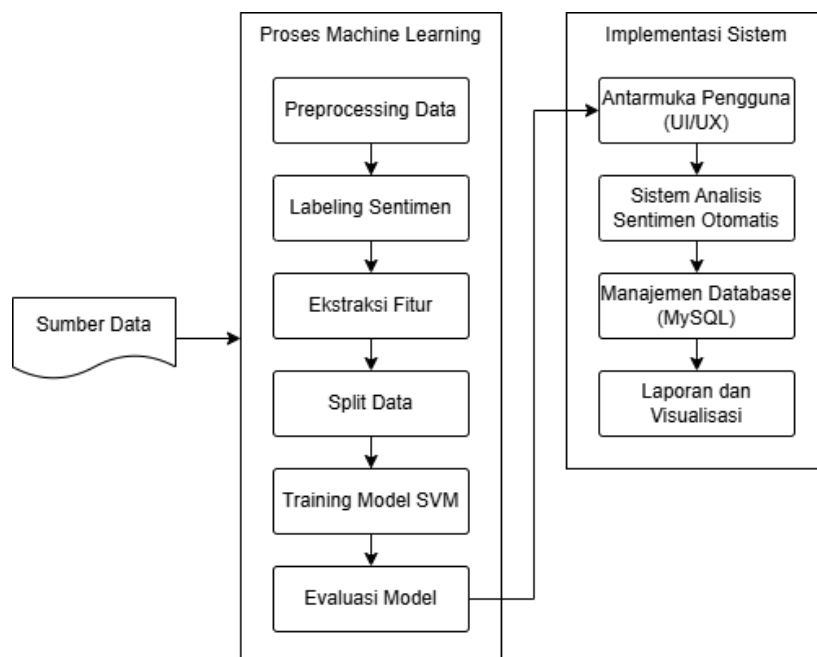
Penelitian ini difokuskan untuk menganalisis sentimen terhadap kalimat motivasi di media sosial menggunakan metode SVM. Tujuannya adalah untuk mengukur tingkat akurasi metode SVM dalam mengklasifikasikan kalimat motivasi ke dalam kategori positif, negatif, maupun netral. Hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai persepsi pengguna terhadap konten motivasi serta menjadi dasar dalam pengembangan strategi penyajian pesan motivasi yang lebih efektif di platform digital.

2. Metode

Pada bagian metode, penelitian ini dijelaskan secara sistematis mulai dari desain penelitian hingga evaluasi model. Setiap tahapan dirancang agar dapat direplikasi pada penelitian berikutnya.

2.1 Desain Penelitian

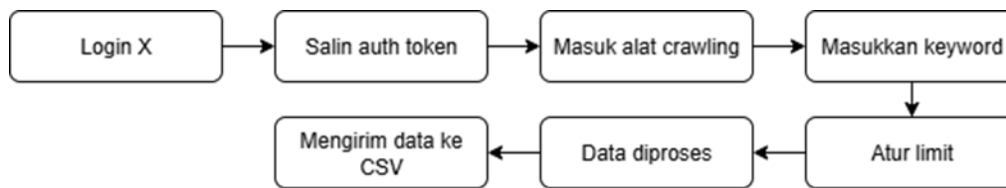
Awal penelitian dimulai dengan menentukan rancangan eksperimen yang berbasis machine learning. Data yang diambil dari media sosial diproses melalui tahapan NLP sebelum digunakan sebagai masukan bagi model SVM. Tahapan penelitian ini dapat digambarkan sebagai alur: pengumpulan data → preprocessing → pelabelan → representasi fitur → klasifikasi dengan SVM → evaluasi hasil. Arsitektur pada gambar 1 berikut memperlihatkan rancangan arsitektur metode yang digunakan dalam penelitian ini.



Gambar 1. Arsitektur Model

2.2 Pengumpulan Data

Proses pengumpulan data bertujuan untuk memperoleh dataset yang digunakan dalam pelatihan dan pengujian model analisis sentimen. Data diperoleh melalui dua sumber utama, yaitu lexicon emosi dan data kalimat motivasi dari media sosial X. Pertama, dataset leksikon diambil dari InSet Lexicon (Indonesian Sentiment Lexicon) yang dikembangkan oleh Koto dan Rahmaningtyas. InSet berisi 3.609 kata positif dan 6.609 kata negatif dengan bobot polaritas kata yang bervariasi dari -5 hingga $+5$. Leksikon ini digunakan sebagai acuan dalam proses pelabelan sentimen pada dataset penelitian. Kedua, data kalimat motivasi dikumpulkan melalui proses scraping menggunakan alat bantu tweet-harvest. Proses ini dilakukan dengan memanfaatkan kata kunci hashtag populer, salah satunya #motivasi, sehingga tweet yang relevan dapat diperoleh secara otomatis. Pengambilan data dilaksanakan sepanjang bulan Juni 2025, dengan target maksimal 100 tweet per hari, sehingga diperoleh sekitar 1.000 tweet dalam satu bulan. Data yang dikumpulkan mencakup isi teks tweet, waktu publikasi, ID tweet, ID pengguna, serta nama akun. Untuk menjaga kualitas dataset, hanya tweet berbahasa Indonesia yang disertakan, sementara retweet dan duplikasi dihapus. Seluruh data mentah yang diperoleh kemudian disimpan dalam format CSV/JSON untuk selanjutnya diproses pada tahap pra-pemrosesan, pelabelan, dan klasifikasi sentimen. Alur proses pengambilan data dengan tweet-harvest dapat dilihat pada Gambar 2.



Gambar 2. Crawl Data

Proses pengumpulan data tweet yang telah dilakukan menggunakan sistem crawling. Proses ini diawali dengan memasukkan kata kunci tertentu pada sistem, misalnya “motivasi harian”, yang kemudian digunakan untuk mengambil data tweet secara otomatis melalui API atau perintah tweet-harvest. Setiap tweet yang berhasil dikumpulkan akan disimpan dalam file berformat CSV yang berisi beberapa kolom informasi penting seperti ID Tweet, ID Pengguna, Username, Teks Tweet, dan Tanggal Pembuatan (Created At). Berikut Gambar 3 merupakan pengumpulan hasil data.

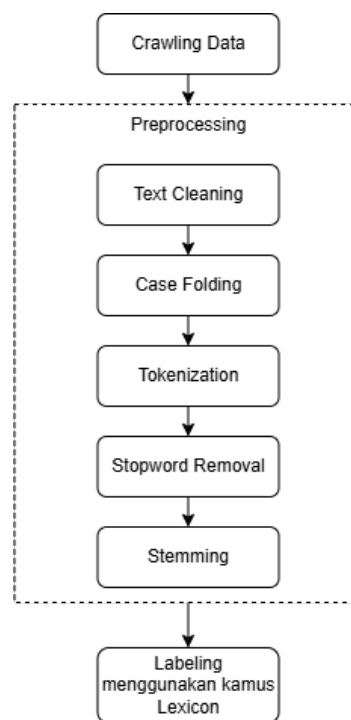
ID	Tweet ID	User ID	Username text	Created At
2	1.93E+18	1.92E+18	unknown_ Lo gak gagal lo cuma belum nyampe. Yang lain keliatan cepet closing padahal lo gak liat perju	#####
4	1.93E+18	1.86E+18	unknown_ musuh terbesar adalah diri sendiri #motivation #inspirationdaily #motivasiDirisendiri #Quotes #	#####
5	1.93E+18	1.92E+18	unknown_ Tenang lo gak kurang usaha. Kadang lo cuma lupa berhenti sebentar buat istirahat dan ngobrol	#####
6	1.93E+18	1.77E+08	unknown_ Bukan untuk membuat segalanya mudah tapi untuk menjadikan kita lebih kuat. Proses hidup ti	#####
7	1.93E+18	1.92E+18	unknown_ Lo lagi capek atau sebenarnya udah setengah nyerah tapi pura-pura kuat? Usaha emang naik t	#####
8	1.93E+18	1.92E+18	unknown_ Gaji udah masuk saldo udah pami. Bukan karena penghasilannya kurang. Cuma karena alurny	#####
9	1.93E+18	2.74E+09	unknown_ #MotivasiHarilni : SEGALA SESUATUNYA AKAN SULIT PADA AWALNYA NAMUN AKAN MUDAH	#####
12	1.93E+18	1.92E+18	unknown_ Lo kerja keras. Lo jagain semua orang. Tapi... kalau lo sakit siapa yang jagain lo? Coba share p	#####
13	1.93E+18	1.27E+18	unknown_ Setiap tugas adalah tangga kecil menuju impian besar. Jalani dengan tekun karena hasil tak per	#####
14	1.93E+18	2.74E+09	unknown_ #MotivasiHarilni : PASTI AKAN ADA TANTANGAN DALAM MEWUJUDKAN HARAPAN HARAPAN	#####
15	1.93E+18	2.28E+09	unknown_ Gagal itu wajar. Itu tanda kamu sedang belajar. Yang penting: bangkit & lanjut. Simpan sel	#####
16	1.93E+18	1.86E+18	unknown_ Berpikir Positif#motivation #inspirationdaily #motivasiDirisendiri #Quotes #katamutiara #semu	#####
17	1.93E+18	1.86E+18	unknown_ SETIA PADA HAL KECIL#motivation #inspirationdaily #motivasiDirisendiri #Quotes #katamutiara	#####
18	1.93E+18	2.74E+09	unknown_ #MotivasiHarilni : TANPA KEBERANIAN ANDA TAKKAN BISA MELAKUKAN APAPUN SEBELUM N	#####
19	1.93E+18	1.64E+18	unknown_ oh inilah emang kalau udah terkenal makin banyak halangan untuk berkarya #tetapsemangat #	#####
20	1.93E+18	9.56E+17	unknown_ Hari ini beranian diri melangkah meski ragu. Setiap langkah kecil adalah kemenangan menuju	#####
21	1.93E+18	1.78E+18	unknown_ Hikmah langit dari Saiyidina Sheikh Muhammad Saif di Cornville membuka rahasia zaman dan f	#####
22	1.93E+18	1.9E+18	unknown_ Saat kamu merasa lelah ingatlah bahwa setiap langkah kecil yang kamu ambil hari ini akan mer	#####
23	1.92E+18	2.74E+09	unknown_ #MotivasiHarilni : PERJALANAN HIDUPMU MUNGKIN TAK MUDAH TAPI YAKINLAH AKAN INDA	#####
24	1.92E+18	1.36E+18	unknown_ Kadang kita terlalu banyak memikirkan merumitkan dan membandingkan. Padahal kesederhan	#####
25	1.92E+18	2.74E+09	unknown_ #MotivasiHarilni : COBALAH DULU SEBELUM MENOLAKNYA SEBAB KEBERHASILAN BERAWAL	#####

Gambar 3. Hasil Pengumpulan Data

Setelah file CSV berhasil dibuat, data tersebut dibaca menggunakan pustaka *pandas* untuk dilakukan pembersihan dan pemilahan informasi penting. Setiap baris pada gambar merepresentasikan satu tweet yang berhasil dikumpulkan berdasarkan kata kunci yang relevan. Dari proses ini, sistem memperoleh kumpulan teks yang nantinya akan digunakan sebagai dataset awal untuk tahap pra-pemrosesan (*preprocessing*) dan analisis sentimen. Proses pengumpulan data ini memastikan bahwa informasi yang digunakan benar-benar berasal dari sumber nyata dan sesuai konteks penelitian. Selain itu, sistem juga dilengkapi dengan penanganan kesalahan (*error handling*) untuk mengantisipasi kendala jaringan, batas akses API, atau data yang tidak lengkap, sehingga pengumpulan data tetap berlangsung aman dan stabil.

2.3 Preprocessing Data

Proses *preprocessing* data dilakukan setelah tahap *crawling* dari media sosial. Tahap ini bertujuan untuk membersihkan, menyederhanakan, dan menyiapkan data teks agar dapat dianalisis lebih lanjut dengan akurat. Secara umum, *preprocessing* berfungsi mengurangi *noise* pada data sehingga hasil analisis sentimen menjadi lebih representatif. Tahapan *preprocessing* ditunjukkan pada Gambar 4 dalam bentuk blok diagram.



Gambar 4. Blok Diagram Teknik Preprocessing Data

Proses dimulai dengan pengambilan data (*crawling*) dari media sosial berdasarkan kata kunci tertentu, dalam hal ini kalimat motivasi. Data mentah yang diperoleh kemudian melalui tahap *text cleaning* dengan menghapus simbol, emotikon, angka, serta tautan URL agar teks menjadi lebih bersih. Selanjutnya, dilakukan *case folding* dengan mengubah seluruh huruf menjadi kecil untuk menyeragamkan format teks. Tahap berikutnya adalah *tokenization*, yaitu memecah teks menjadi unit kata atau token. Setelah terbentuk token, sistem menghapus *stopword*, yakni kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis, seperti “yang”, “dan”, atau “adalah”. Langkah berikutnya adalah *stemming*, yang bertujuan

untuk mengembalikan kata ke bentuk dasarnya. Misalnya, kata “berlari”, “lari-lari”, dan “berlarian” akan disederhanakan menjadi bentuk dasar “lari”. Setelah seluruh proses preprocessing selesai, data kemudian diberi label sentimen menggunakan kamus InSet Lexicon yang berisi daftar kata positif dan negatif. Label akhir yang diberikan berupa sentimen positif, negatif, atau netral sesuai dengan dominasi polaritas kata dalam teks. Tahapan preprocessing ini menjadi dasar penting sebelum data masuk ke proses representasi fitur dan klasifikasi menggunakan algoritma Support Vector Machine (SVM). Dengan preprocessing yang optimal, kualitas data meningkat sehingga model mampu melakukan klasifikasi sentimen secara lebih akurat. Berikut Gambar 5 merupakan hasil preprocessing.

NO	TWEET ID	CLEAN TEXT	TOKENIZED
1	Jangan biarkan cintamu menjadi obsesi dan kebencianmu menjadi kehancuran. by Ali bin Abi Talib #해찬_애니말고실물 #WhereIsHaechan #MPLSRamah #NCTDREAM_CHILLER #7드 임의_7iller보러_BTF #quotes #islam #motivasi	biarkan cintamu obsesi kebencianmu kehancuran by ali bin abi talib	["biarkan", "cintamu", "obsesi", "kebencianmu", "kehancuran", "by", "ali", "bin", "abi", "talib"]
2	Kesalahan Nokia - Tidak mau berubah #mindanews #motivation #motivasi #motivasihidup #mindset #growthmindset #mentalpemenang https://t.co/B3MxoFhNA	kesalahan nokia berubah	["kesalahan", "nokia", "berubah"]

Gambar 5. Hasil Preprocessing

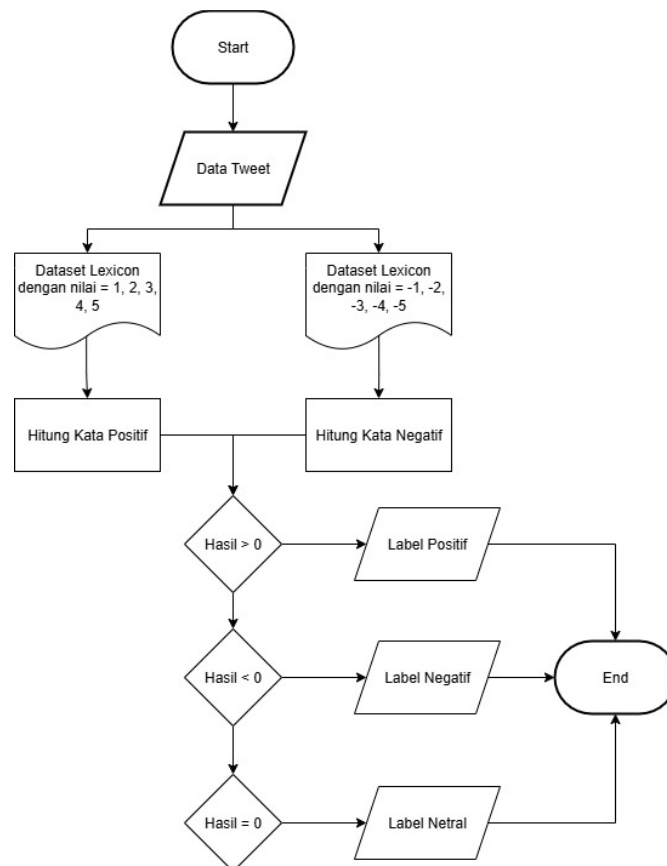
Gambar 5 menampilkan hasil dari tahap pra-pemrosesan data tweet yang merupakan langkah penting sebelum data digunakan dalam proses klasifikasi sentimen menggunakan metode Support Vector Machine (SVM). Tahap ini bertujuan untuk mengubah data mentah hasil crawling menjadi bentuk yang bersih, terstruktur, dan siap diolah oleh algoritma pembelajaran mesin. Pada bagian atas tabel terlihat beberapa kolom utama hasil dari proses preprocessing, yaitu:

1. Tweet ID — merupakan identitas unik dari setiap tweet yang digunakan untuk menjaga keterlacakan data.
2. Clean Text — berisi teks tweet yang telah dibersihkan dari elemen-elemen yang tidak relevan seperti tanda baca, emotikon, URL, hashtag, mention, serta karakter non-alfabet. Proses ini dilakukan menggunakan fungsi `preprocess_text()` yang mengandalkan ekspresi reguler (regular expression) untuk menghapus simbol yang tidak dibutuhkan.
3. Tokenized — merupakan hasil dari proses tokenisasi menggunakan pustaka Natural Language Toolkit (NLTK), di mana setiap kalimat dipecah menjadi daftar kata (list of tokens). Tahap ini penting agar sistem dapat menganalisis kata per kata secara lebih detail.
4. Feature Vector — menunjukkan representasi numerik dari teks yang telah ditokenisasi. Setiap token dikonversi menjadi vektor angka berdasarkan nilai bobot Term Frequency–Inverse Document Frequency (TF-IDF). Vektor inilah yang nantinya akan menjadi masukan (input feature) bagi algoritma klasifikasi SVM.

Secara keseluruhan, hasil pada gambar ini menggambarkan keberhasilan sistem dalam melakukan transformasi data teks mentah menjadi bentuk vektor numerik yang siap digunakan untuk pelatihan model. Proses pra-pemrosesan ini menjadi fondasi penting dalam meningkatkan akurasi model, karena kualitas data masukan sangat menentukan performa algoritma klasifikasi.

2.4 Pelabelan Sentimen

Tahap pelabelan dilakukan untuk menentukan kategori sentimen pada setiap kalimat motivasi yang telah dikumpulkan dari media sosial. Pendekatan yang digunakan adalah metode berbasis lexicon, yaitu dengan membandingkan kata-kata pada teks terhadap daftar kata bermuatan sentimen positif maupun negatif yang terdapat pada InSet Lexicon. Proses pelabelan dilakukan secara otomatis sesuai alur kerja yang ditunjukkan pada Gambar 6.



Gambar 5. Flowchart Labeling Data menggunakan Lexicon

Data mentah berupa tweet terlebih dahulu diproses untuk mengidentifikasi kata-kata yang terdapat di dalamnya. Setiap kata kemudian dibandingkan dengan dua jenis kamus lexicon, yaitu kamus kata positif dan kamus kata negatif. Sistem menghitung jumlah kecocokan kata dari masing-masing kategori untuk menentukan dominasi polaritas sentimen. Selain itu, proses ini dilengkapi dengan pembobotan polaritas agar hasil pelabelan lebih terukur. Kata positif diberi bobot +1 hingga +5 sesuai tingkat intensitas positifnya, sedangkan kata negatif diberi bobot -1 hingga -5. Kata yang tidak termasuk ke dalam kedua kategori tersebut diberi bobot 0 (netral). Setelah semua kata dalam satu tweet diberi bobot, sistem menghitung total skor polaritas. Aturan klasifikasi ditetapkan sebagai berikut:

- Jika total skor > 0 , kalimat dilabeli positif.
- Jika total skor < 0 , kalimat dilabeli negatif.
- Jika total skor $= 0$, kalimat dilabeli netral.

Dengan pendekatan ini, pelabelan dapat dilakukan secara objektif, konsisten, dan otomatis, tanpa memerlukan proses manual yang memakan waktu. Hasil pelabelan inilah yang kemudian digunakan pada tahap selanjutnya, yaitu representasi fitur teks dengan TF-IDF dan klasifikasi menggunakan algoritma Support Vector Machine (SVM).

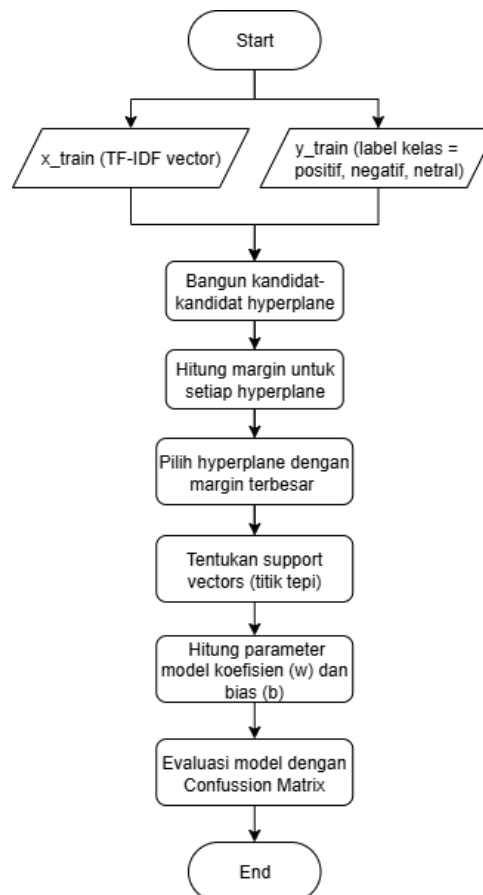
2.5 Representasi Fitur

Setelah proses pelabelan selesai, data teks yang telah dikategorikan ke dalam kelas sentimen positif, netral, dan negatif perlu direpresentasikan dalam bentuk numerik agar dapat diproses lebih lanjut oleh algoritma machine learning. Representasi yang digunakan dalam penelitian ini adalah Term Frequency–Inverse Document Frequency (TF-IDF). Metode TF-IDF bekerja dengan memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam sebuah dokumen (tweet) dibandingkan dengan keseluruhan dokumen pada korpus. Kata yang sering muncul pada satu dokumen namun jarang muncul di dokumen lain akan memiliki bobot lebih tinggi karena dianggap lebih informatif. Sebaliknya, kata yang sering muncul di hampir semua dokumen (misalnya stopword) akan memiliki bobot yang lebih rendah. Dengan pendekatan ini, setiap tweet yang semula berbentuk teks diubah menjadi vektor numerik yang merepresentasikan bobot setiap kata. Vektor inilah yang kemudian digunakan sebagai input pada algoritma Support Vector Machine (SVM) untuk proses klasifikasi sentimen. Penggunaan TF-IDF membuat model lebih sensitif terhadap kata-kata penting yang membedakan sentimen positif, netral, dan negatif. Selain itu, metode ini sederhana namun efektif dalam mengurangi redundansi data sehingga meningkatkan kinerja algoritma SVM pada tahap pelatihan maupun pengujian.

2.6 Model Support Vector Machine

Proses implementasi analisis sentimen dilakukan dengan pendekatan Natural Language Processing (NLP) yang dipadukan dengan algoritma klasifikasi Support Vector Machine (SVM). Alur penerapan metode ditunjukkan pada Gambar 7. Tahapan dimulai dengan penggunaan data hasil preprocessing, yaitu teks yang telah dibersihkan dari karakter tidak relevan seperti simbol, URL, tanda baca, maupun emoji. Data tersebut telah melalui tahap pelabelan otomatis berbasis leksikon untuk menentukan kategori sentimen awal. Selanjutnya, teks direpresentasikan ke dalam bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Representasi ini menghasilkan vektor fitur, sedangkan label sentimen dikodekan ke dalam bentuk numerik (misalnya positif = 1, netral = 0, negatif = -1) agar dapat diproses lebih lanjut oleh algoritma. Dataset kemudian dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%. Pembagian ini dilakukan dengan memperhatikan keseimbangan distribusi kelas agar proses pelatihan dan pengujian berjalan representatif. Tahap inti berikutnya adalah pelatihan model menggunakan SVM. Algoritma ini bekerja dengan mencari hyperplane terbaik yang mampu memisahkan data ke dalam kelas sentimen yang berbeda. Dari sekian banyak kandidat hyperplane, SVM memilih hyperplane dengan margin terbesar sebagai pemisah optimal. Pada penelitian ini digunakan kernel linear karena sesuai dengan karakteristik representasi TF-IDF serta efisien dalam menangani data

teks berdimensi tinggi. Setelah pelatihan selesai, model diuji menggunakan data uji untuk menghasilkan prediksi sentimen baru. Hasil klasifikasi berupa kategori sentimen (positif, netral, atau negatif) dievaluasi menggunakan Confusion Matrix. Dari matriks ini diperoleh metrik akurasi, precision, recall, dan F1-score sebagai ukuran kualitas model. Evaluasi ini memastikan bahwa model yang dibangun mampu memberikan prediksi yang akurat, konsisten, dan dapat digunakan lebih lanjut dalam sistem analisis sentimen berbasis media sosial.



Gambar 7. Flowchart Perancangan Penerapan Metode Support Vector Machine (SVM)

2.7 Evaluasi Model

Evaluasi model bertujuan untuk mengukur sejauh mana algoritma Support Vector Machine (SVM) mampu melakukan klasifikasi sentimen dengan akurat. Proses evaluasi dilakukan menggunakan data uji yang telah dipisahkan dari dataset pada tahap sebelumnya dengan proporsi 20% dari total data. Metode evaluasi yang digunakan adalah Confusion Matrix, yang memberikan gambaran distribusi hasil klasifikasi antara prediksi model dan label aktual. Dari matriks ini dihitung sejumlah metrik kinerja, yaitu:

- Akurasi, yang menunjukkan persentase prediksi benar dari seluruh data uji.
- Precision, yang mengukur tingkat ketepatan model dalam memprediksi kelas tertentu.
- Recall, yang mengukur kemampuan model dalam menemukan semua data yang benar pada suatu kelas.
- F1-score, yang merupakan rata-rata harmonis antara precision dan recall, sehingga memberikan ukuran kinerja yang lebih seimbang.

Selain evaluasi berbasis metrik, dilakukan pula analisis distribusi kesalahan klasifikasi untuk mengidentifikasi pola keliru yang muncul, seperti kasus kalimat bermakna ganda atau sarkastik. Analisis ini penting agar dapat memberikan masukan dalam perbaikan sistem di masa mendatang. Dengan pendekatan evaluasi ini, kualitas model SVM tidak hanya diukur berdasarkan satu indikator, melainkan secara menyeluruh, sehingga hasil penelitian dapat dipertanggungjawabkan secara ilmiah.

3. Hasil Dan Pembahasan

Bagian ini menyajikan hasil pengujian model analisis sentimen yang telah dibangun menggunakan algoritma Support Vector Machine (SVM). Tahapan pengujian dilakukan untuk menilai kinerja model dalam mengklasifikasikan kalimat motivasi ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Hasil yang diperoleh kemudian dianalisis menggunakan metrik evaluasi akurasi, precision, recall, dan F1-score yang dihitung berdasarkan Confusion Matrix. Pengujian ini tidak hanya bertujuan untuk mengetahui tingkat ketepatan klasifikasi, tetapi juga untuk melihat sejauh mana model mampu memberikan performa yang seimbang pada masing-masing kelas sentimen. Dengan demikian, hasil dan pembahasan ini diharapkan dapat memberikan gambaran menyeluruh mengenai efektivitas metode SVM pada kasus analisis sentimen teks motivasi di media sosial.

3.1 Hasil Pengujian

Pengujian model klasifikasi sentimen dilakukan menggunakan Confusion Matrix dengan tiga kelas, yaitu positif, netral, dan negatif. Dari total 296 data uji, sistem berhasil mengklasifikasikan 286 data secara benar, sementara terdapat 10 data yang salah klasifikasi. Rincian hasil klasifikasi dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

	Prediksi Positif	Prediksi Netral	Prediksi Negatif
Aktual Positif	67	0	5
Aktual Netral	0	46	3
Aktual Negatif	2	0	173

Berdasarkan matriks tersebut, diperoleh nilai akurasi sebesar 96,62%. Selain itu, metrik evaluasi lainnya dihitung untuk setiap kelas. Ringkasan hasil evaluasi ditampilkan pada Tabel 2. Rata-rata macro untuk precision, recall, dan F1-score masing-masing adalah 97,56%, 95,26%, dan 96,37%.

Tabel 2. Hasil Evaluasi Kinerja Model per Kelas

Kelas	Precision (%)	Recall (%)	F1-score (%)
Positif	97.10	93.05	95.04
Netral	100.00	93.88	96.88
Negatif	95.58	98.86	97.19

3.2. Analisis Hasil

Hasil pengujian menunjukkan bahwa algoritma Support Vector Machine (SVM) dengan representasi fitur TF-IDF mampu memberikan performa yang sangat baik pada klasifikasi sentimen kalimat motivasi. Tingkat akurasi yang tinggi (96,62%) serta nilai F1-score di atas 95% pada seluruh kelas menunjukkan keseimbangan antara precision dan recall, sehingga model dapat bekerja secara konsisten pada semua kategori sentimen. Kesalahan klasifikasi umumnya terjadi pada kalimat yang mengandung makna ambigu atau sarkastik. Misalnya, kalimat motivasi dengan nada bercanda terkadang dikategorikan

sebagai negatif. Hal ini menjadi keterbatasan metode berbasis leksikon dan SVM yang hanya melihat pola kata tanpa mempertimbangkan konteks semantik yang lebih dalam. Jika dibandingkan dengan penelitian terdahulu yang menggunakan algoritma lain, hasil ini lebih unggul. Penelitian dengan Naïve Bayes hanya menghasilkan akurasi sekitar 82%, sedangkan dengan KNN diperoleh sekitar 89%. Dengan demikian, SVM terbukti lebih efektif dalam menangani data teks pendek seperti tweet motivasi.

4. Kesimpulan

Berdasarkan rumusan masalah mengenai akurasi metode Support Vector Machine (SVM) dalam menganalisis sentimen terhadap kalimat motivasi di media sosial, penelitian ini berhasil merancang dan mengimplementasikan sistem analisis sentimen yang akurat dan stabil. Hasil pengujian menggunakan confusion matrix menunjukkan bahwa model SVM mampu mencapai akurasi sebesar 96,62%, dengan performa klasifikasi yang sangat baik. Untuk kategori positif, sistem memperoleh precision sebesar 95,58%, recall 98,86%, dan F1-score 97,19%. Kategori netral memperoleh precision 100%, recall 93,88%, dan F1-score 96,84%, sementara kategori negatif memperoleh precision 97,10%, recall 93,06%, dan F1-score 95,04%. Pengujian black box menghasilkan persentase kelulusan sebesar 92,3%, yang menunjukkan bahwa seluruh fitur dan fungsionalitas sistem telah bekerja sesuai dengan yang dirancang. Sedangkan pengujian white box menunjukkan bahwa seluruh jalur logika fungsi utama berjalan baik, serta berhasil lolos 100% uji validasi terhadap berbagai skenario input. Secara keseluruhan, metode SVM terbukti efektif dalam menganalisis dan mengklasifikasikan sentimen kalimat motivasi di media sosial secara otomatis dengan akurasi dan stabilitas yang tinggi.

Referensi

- [1] Y. P. Akbar, M. S. Satyawati, and N. P. Sastra, "Analisis Sentimen Kata Anjay pada Media Sosial Twitter dalam Kajian Linguistik Komputasi," *Jurnal Linguistik Komputasi*, vol. 2, no. 1, pp. 45–55, 2022.
- [2] A. D. Pratama and Hendry, "Analisis Sentimen Masyarakat Terhadap Penggunaan ChatGPT Menggunakan Metode Support Vector Machine," *Jurnal Informatika*, vol. 5, no. 3, pp. 112–120, 2024.
- [3] P. Fremmuzar and A. Baita, "Uji Kernel SVM dalam Analisis Sentimen Layanan Telkomsel di Media Sosial Twitter," *Jurnal Teknologi Informasi*, vol. 10, no. 2, pp. 77–85, 2023.
- [4] I. Firmansyah and B. H. Hayadi, "Analisis Sentimen Citayam Fashion Week menggunakan Support Vector Machine," *Jurnal Komputer dan Aplikasi*, vol. 8, no. 2, pp. 134–142, 2022.
- [5] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *Jurnal Sistem Informasi*, vol. 12, no. 2, pp. 200–210, 2020.
- [6] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Jurnal Teknologi*, vol. 9, no. 1, pp. 55–63, 2020.
- [7] T. Liedfray, F. J. Waani, and J. J. Lasut, "Peran Media Sosial Dalam Mempererat Interaksi Antar Keluarga Di Desa Esandom Kecamatan Tombatu Timur Kabupaten Tombatu Timur Kabupaten Minasa Tenggara," *Jurnal Ilmiah Society*, vol. 2, no. 1, p. 2, 2022.
- [8] K. Minahasa, "3 1,2,3," vol. 24, no. 7, pp. 28–42, 2024.
- [9] X. Fang and J. Zhan, "Sentiment analysis using product review data," *J Big Data*, vol. 2, no. 1, 2015, doi: 10.1186/s40537-015-0015-2.
- [10] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *EduTic - Scientific Journal of Informatics Education*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [11] Y. Akbar, M. Satyawati, and N. P. Sastra, "Analisis Sentimen Kata Anjay pada Media Sosial Twitter Dalam Kajian Linguistik Komputasi," *Stilistika: Journal of Indonesian Language and Literature*, vol. 1, p. 62, Apr. 2022, doi: 10.24843/STIL.2022.v01.i02.p06.
- [12] J. P. Setiadi and S. Sugiyamta, "Analisis dan Visualisasi Berbasis Web Sentimen Pengguna Jenius Menggunakan Naïve Bayes Classifier," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 1, pp. 245–254, 2024, doi: 10.32493/jtsi.v7i1.37981.
- [13] G. Gao, Y. Liu, and G. Bai, "Crawling and Analysis of Data Based on Social Networking on Stock Comments," *IOP Conf Ser Earth Environ Sci*, vol. 234, p. 12093, Mar. 2019, doi: 10.1088/1755-1315/234/1/012093.
- [14] J. Claussen and C. Peukert, "Obtaining Data from the Internet: A Guide to Data Crawling in Management Research," *SSRN Electronic Journal*, Jan. 2019, doi: 10.2139/ssrn.3403799.
- [15] F. A. Artanto, "Support Vector Machine Berbasis Particle Swarm Optimization Pada Analisis Sentimen Anggota KPPS," *Jurnal Fasikom*, vol. 14, no. 1, pp. 75–79, 2024, doi: 10.37859/jf.v14i1.6795.