

Analisis Sentimen Pengguna Platform X terhadap Kasus Amsal Christy Sitepu Menggunakan Algoritma SVM

Luqman Abdulmajid, Fuad Nur Hasan, Lestari Yusuf*

Fakultas Teknik dan Informatika, Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 29 Juli 2026
Revisi Akhir: 20 Agustus 2026
Diterbitkan Online: 28 September 2026

KATA KUNCI

Analisis Sentimen
Opini Publik
Media Sosial X
Support Vector Machine
TF-IDF

KORESPONDENSI (*)

Phone: +62 818-253-851
E-mail: fuad.fnu@bsi.ac.id

A B S T R A K

Media sosial X menjadi wadah penting bagi masyarakat untuk merespons kasus hukum Amsal Christy Sitepu, seorang videografer yang dituduh melakukan *markup* dana proyek video profil desa di Kabupaten Karo, namun akhirnya divonis bebas murni. Banyaknya opini terkait persinggungan antara hukum pidana dan industri kreatif ini memicu perdebatan publik secara masif. Penelitian ini bertujuan untuk menganalisis distribusi sentimen publik (positif dan negatif) serta menguji efektivitas algoritma *Support Vector Machine* (SVM) yang dipadukan dengan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dalam memproses teks opini tersebut. Data dikumpulkan menggunakan teknik *scraping* yang menghasilkan 3.104 cuitan mentah. Setelah melalui tahapan *preprocessing* dan penghapusan 215 data kelas netral, tersisa 2.889 data final yang dianalisis. Berdasarkan hasil pengujian, opini publik sedikit lebih didominasi oleh sentimen negatif sebanyak 1.526 data (52,82%), dibandingkan sentimen positif sebanyak 1.363 data (47,17%). Pengujian efektivitas algoritma SVM dengan skenario *hold-out split* 80:20 menghasilkan tingkat Akurasi 87,37%, Presisi 84,24%, *Recall* 90,10%, dan *F1-Score* 86,89%. Kinerja ini divalidasi lebih lanjut menggunakan *10-Fold Cross-Validation* yang membuktikan stabilitas model dengan rata-rata akurasi mencapai 88,06%. Hasil penelitian ini menyimpulkan bahwa algoritma SVM memiliki kinerja yang andal, konsisten, dan terhindar dari *overfitting* dalam mengklasifikasikan sentimen berbahasa Indonesia pada dataset kasus hukum di media sosial.

PENDAHULUAN

Dengan berkembangnya teknologi informasi, platform media sosial seperti X (sebelumnya Twitter) telah menjadi sumber data tekstual yang masif untuk menggali opini publik secara *real-time* [1]. Volume informasi yang dinamis ini menghadirkan tantangan sekaligus peluang dalam bidang *Natural Language Processing* (NLP) dan *Text Mining*, khususnya untuk memproses dan memahami kecenderungan pandangan publik terhadap isu-isu yang spesifik dan kompleks melalui analisis sentimen.

Salah satu fenomena yang menghasilkan lonjakan data opini tidak terstruktur dalam skala besar di Indonesia adalah kasus hukum videografer Amsal Christy Sitepu pada awal tahun 2026. Dari sudut pandang komputasional, narasi publik di platform X terkait kasus ini menghadirkan tantangan ekstraksi data yang unik. Teks opini yang beredar tidak hanya mengandung bahasa tidak baku (*slang*) dan *noise* yang tinggi khas media sosial, tetapi juga mencampurkan terminologi hukum (seperti dugaan korupsi dan audit) dengan jargon evaluasi industri kreatif yang tidak memiliki standar nilai pasti. Karakteristik data teks yang ambigu secara semantik dan sangat heterogen ini membuat proses analisis sentimen otomatis menjadi jauh lebih kompleks.

Dalam menganalisis polaritas sentimen pada domain teks tertentu, beberapa studi sebelumnya masih menunjukkan adanya kekurangan pada metode analisis sentimen tradisional. Pendekatan berbasis kamus (lexicon-based) dapat menghasilkan polaritas yang fluktuatif ketika menghadapi kata-kata ambigu, gaya penulisan yang deskriptif, serta perubahan polaritas yang tergantung pada konteks [2]. Keterbatasan metode berbasis kamus juga muncul dalam penelitian [3], di mana pendekatan lexicon-based memperoleh akurasi 78,43%, lebih rendah dibandingkan dengan Support Vector Machine yang mencapai 98,5% karena keterbatasan kamus dalam mendeteksi kata-kata bernilai positif. Representasi fitur yang didasarkan pada teks juga dapat menciptakan karakteristik data yang berukuran tinggi dan jarang, sehingga pemilihan algoritma klasifikasi menjadi elemen penting untuk mendapatkan kinerja yang konsisten. [4] memperlihatkan bahwa algoritma tertentu lebih efektif dalam menangani karakteristik high-dimensional sparse text features, sementara [5] mengungkapkan bahwa metode ekstraksi fitur memengaruhi hasil klasifikasi, di mana TF-IDF memperoleh akurasi 62% dibandingkan 30% pada ekstraksi yang menggunakan InSet Lexicon.

Untuk menjawab kelemahan tersebut, penelitian ini mengimplementasikan algoritma *Support Vector Machine* (SVM) sebagai model klasifikasi. Secara komputasional, SVM dipilih karena terbukti unggul dan secara inheren dirancang untuk menangani ruang vektor berdimensi tinggi dan data teks yang *sparse* [6]. Berbeda dengan algoritma lain yang rentan terhadap *overfitting* pada data besar, SVM bekerja dengan mencari *hyperplane* optimal yang memaksimalkan *margin* pemisah antar kelas sentimen (positif dan negatif), sehingga menghasilkan klasifikasi teks media sosial yang lebih tangguh dan stabil [7].

Selain itu, untuk menjamin objektivitas evaluasi model, memastikan stabilitas hasil, dan mencegah bias distribusi data, penelitian ini mengintegrasikan pengujian menggunakan *K-Fold Cross-Validation* [8]. Melalui kerangka kerja NLP yang telah dioptimalkan ini, diharapkan model dapat memetakan polaritas persepsi masyarakat secara komprehensif. Oleh karena itu, studi ini diangkat dengan judul "Analisis Sentimen Pengguna Platform X Terhadap Kasus Amsal Christy Sitepu Menggunakan Algoritma Support Vector Machine".

TINJAUAN PUSTAKA

Analisi Sentimen

Analisis sentimen merupakan teknik komputasional yang digunakan untuk mengekstraksi dan mengevaluasi informasi berupa opini, pandangan, atau emosi publik terhadap suatu fenomena tertentu. Pendekatan ini sangat efektif untuk menambang informasi (information retrieval) berskala besar yang bersumber dari platform media sosial maupun ulasan digital. Secara fungsional, analisis sentimen berperan penting dalam mengklasifikasikan unit teks ke dalam polaritas spesifik, yakni kelas positif, atau negatif. Dalam implementasinya, data teks atau ulasan baru yang masuk akan dievaluasi dan dikategorikan secara otomatis oleh algoritma Machine Learning yang sebelumnya telah dilatih menggunakan sekumpulan data latih [6]. Penelitian ini menggunakan analisis sentimen untuk membagi kecenderungan opini publik (polaritas) menjadi dua kategori: positif dan negatif terkait penyelesaian kasus hukum Amsal Christy Sitepu.

Media Sosial X (Twitter)

Twitter, yang kini bertransformasi menjadi platform X, dikenal sebagai media berbasis teks yang berfokus pada komunikasi digital secara singkat dan cepat. Dengan penerapan batas 280 karakter untuk setiap unggahan (tweet), platform ini menjadi medium yang sangat efektif untuk memfasilitasi penyebaran informasi terkini, membangun percakapan secara real-time, serta merespons berbagai tren atau isu viral secara langsung. Ekosistem digital ini digunakan secara luas oleh berbagai lapisan masyarakat, media massa, hingga tokoh publik untuk menyampaikan opini dan berinteraksi secara terbuka [7].

Pemanfaatan media sosial, khususnya Twitter (X), sebagai wadah ekspresi publik telah menjadikannya sumber data yang krusial dalam mengukur dinamika opini masyarakat terhadap berbagai fenomena. Pengukuran kecenderungan opini tersebut membutuhkan implementasi algoritma data mining untuk mengevaluasi parameter tingkat akurasi pada pemodelan analisis sentimen yang diekstraksi dari tweet pengguna. Untuk mengetahui secara komputasional bagaimana respons publik terhadap suatu topik atau isu tertentu, evaluasi data tekstual dari platform ini dapat dilakukan menggunakan algoritma klasifikasi machine learning, salah satunya adalah Support Vector Machine (SVM). Algoritma SVM berfungsi sebagai metode pengklasifikasi yang tangguh untuk memastikan konsistensi dan tingginya nilai akurasi dalam memproses

kumpulan data. Melalui pendekatan algoritma ini, data sentimen yang bersumber dari opini pengguna dapat dihitung, dianalisis, dan dikategorikan secara akurat ke dalam kelas polaritas yang spesifik, seperti sentimen positif dan negatif, sehingga menghasilkan pemetaan opini publik yang terstruktur dan terukur [8].

Google Colaboratory (Google Colab)

Google Colaboratory, atau yang lebih dikenal dengan *Google Colab*, merupakan sebuah lingkungan pengembangan terpadu berbasis komputasi awan (cloud computing) yang disediakan secara gratis. Platform ini pada dasarnya beroperasi sebagai lingkungan Jupyter Notebook terkelola yang memungkinkan peneliti untuk menulis, mengeksekusi, serta mendokumentasikan kode pemrograman, khususnya untuk bahasa Python dan R. Keunggulan utama dari Google Colab adalah integrasi bawaannya terhadap berbagai pustaka (*library*) esensial untuk pengembangan analitik data, Machine Learning, dan Kecerdasan Buatan (AI), sehingga proses kompilasi algoritma kompleks dapat dieksekusi secara efisien tanpa membebani spesifikasi perangkat keras lokal [9]. Selain itu, platform ini menawarkan fitur lain seperti tanpa perlu konfigurasi tambahan untuk akses ke Google Drive dan GitHub, serta pemakaian pustaka Python untuk analisis data dan pembelajaran mesin [10]. Kemudahan ini menjadikan Google Colab lebih ideal digunakan dalam riset yang memerlukan pengolahan data dan pelatihan model tanpa harus sepenuhnya mengandalkan perangkat keras lokal.

Text Preprocessing

Praproses teks (*text preprocessing*) merupakan tahapan fundamental dalam *Natural Language Processing* (NLP) untuk menangani data mentah dari media sosial yang umumnya tidak terstruktur dan dipenuhi noise, seperti variasi penulisan maupun simbol tak relevan. Tanpa pembersihan anomali ini, algoritma akan kesulitan mengenali pola sehingga menghasilkan model klasifikasi yang bias dan kurang optimal. Oleh karena itu, data teks wajib ditransformasi melalui serangkaian tahapan prosedural yang diawali dengan case folding untuk menyeragamkan karakter teks, serta tokenizing guna memecah kalimat menjadi unit leksikal dasar. Selanjutnya, dilakukan stopword removal untuk mengeliminasi kata berbobot rendah, normalisasi istilah tidak baku, dan stemming untuk mereduksi kata ke bentuk dasarnya. Sebagai langkah final, teks yang telah terstruktur tersebut diproses melalui teknik representasi teks agar terkonversi menjadi matriks numerik yang siap dikomputasi secara matematis oleh algoritma [11].

Dalam studi [12], proses pembersihan juga memanfaatkan regular expression untuk menghilangkan URL, nama pengguna, hashtag, serta tanda baca. Oleh karena itu, tahap pembersihan dalam penelitian ini dilaksanakan untuk menghapus karakter atau informasi yang tidak relevan, seperti URL, angka, emoji, username, hashtag, dan tanda baca dengan menggunakan Regular Expression (Regex). Penerapan Regex untuk mengeliminasi elemen-elemen tersebut juga digunakan dalam studi analisis sentimen lain yang memanfaatkan Regular Expression untuk menghapus “nama pengguna, URL, angka, alfabet non-ascii, tanda baca, dan simbol” [12].

Inset Lexicon

Indonesia Sentiment Lexicon (InSet) yang dikembangkan oleh Koto dan Rahmaningtyas merupakan kamus leksikon sentimen berbahasa Indonesia yang memuat perbendaharaan kata komprehensif, terdiri atas 3.609 lema bersentimen positif dan 6.609 lema bersentimen negatif. Setiap kata di dalam leksikon ini telah dievaluasi sedemikian rupa sehingga memiliki bobot polaritas spesifik yang membentang pada rentang nilai -5 (merepresentasikan sentimen sangat negatif) hingga +5 (merepresentasikan sentimen sangat positif) [13].

Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan salah satu metode pembobotan kata yang paling luas digunakan dalam mentransformasikan data teks menjadi pemodelan ruang vektor (Vector Space Model). Metode ini berfungsi untuk mengevaluasi seberapa penting dan relevan sebuah kata (term) di dalam sebuah dokumen tertentu terhadap keseluruhan kumpulan dokumen (corpus), seperti himpunan tweet. Perhitungan algoritma ini melibatkan perkalian antara dua komponen metrik utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). Komponen TF bertugas untuk mengukur intensitas kemunculan suatu kata di dalam satu tweet tunggal. Sementara itu, komponen IDF bertujuan untuk mengukur tingkat keunikan kata tersebut di seluruh himpunan tweet; di mana kata yang jarang ditemui secara global akan diberikan bobot yang lebih tinggi karena dianggap memiliki nilai informasi yang lebih diskriminatif untuk klasifikasi [8].

Support Vector Machine (SVM)

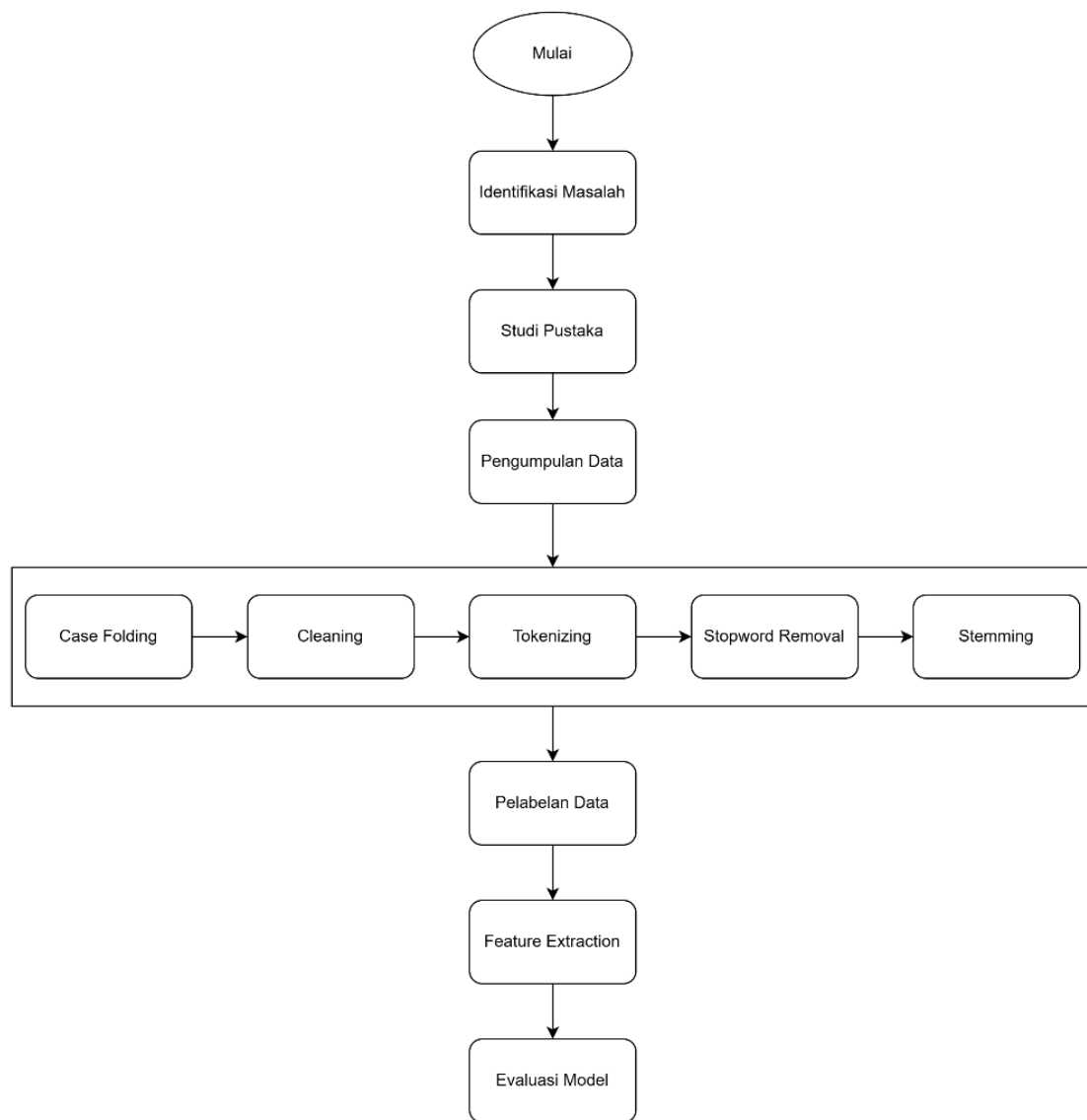
Konsep dasar Support Vector Machine (SVM) pertama kali dikembangkan oleh Cortes dan Vapnik pada tahun 1995. Secara fundamental, algoritma ini beroperasi dengan mentransformasikan vektor input ke dalam ruang fitur (feature space) berdimensi lebih tinggi melalui pemanfaatan fungsi pemetaan non-linear yang ditentukan secara a priori. Di dalam ruang dimensi perluasan tersebut, sebuah permukaan keputusan (decision surface) linier dikonstruksi dengan karakteristik matematis spesifik yang bertujuan untuk menjamin optimalisasi serta tingginya kapabilitas generalisasi dari model yang dihasilkan.

Sebagai suatu metode klasifikasi diskriminatif, representasi formal dari SVM didefinisikan oleh keberadaan sebuah hyperplane. Dalam kerangka kerja pembelajaran terawasi (supervised learning) yang menggunakan himpunan data latih berlabel, algoritma ini dirancang untuk menghasilkan hyperplane paling optimal yang berfungsi sebagai batas pemisah (decision boundary) dalam mengklasifikasikan data observasi baru. Secara geometris pada ruang dua dimensi, hyperplane ini divisualisasikan sebagai sebuah garis memanjang yang mempartisi bidang komputasi menjadi dua area terpisah, di mana masing-masing sisi yang saling berseberangan merepresentasikan kelas atau kategori data yang berbeda [14].

METODOLOGI

Dalam penelitian ini, metode yang digunakan adalah kuantitatif deskriptif. Tujuan dari metode ini adalah untuk secara objektif menganalisis, mengukur, dan mengelompokkan sentimen masyarakat terkait kasus Amsal Christy Sitepu di Kabupaten Karo melalui platform media sosial X. Metode kuantitatif deskriptif dipilih karena penelitian ini bergantung pada data numerik yang dihasilkan dari transformasi teks guna menjelaskan hasil pengamatan, memetakan distribusi opini publik, serta mengklasifikasikan fenomena yang diteliti ke dalam kategori sentimen tertentu. Algoritma pembelajaran mesin (*Machine Learning*) yang menjadi instrumen utama dalam pemodelan klasifikasi ini adalah *Support Vector Machine* (SVM). Dalam implementasi dan pemilihan algoritma SVM, data teks yang telah ditransformasi menjadi matriks angka melalui pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) dimasukkan ke dalam model *Kernel Linear* (LinearSVC) menggunakan bahasa pemrograman Python, dengan penetapan parameter penalti $C = 1.0$. Algoritma ini kemudian bekerja dengan menganalisis pola sebaran data latih untuk menarik garis pembatas terbaik (*hyperplane*) yang secara optimal memisahkan cuitan ke dalam ruang kelas bersentimen positif dan negatif.

Secara keseluruhan, langkah-langkah yang dilakukan dalam penelitian ini dirancang secara sistematis untuk memungkinkan proses pengolahan data mentah menjadi informasi berharga. Jalan penelitian ditampilkan secara visual pada Gambar 1.



Gambar 1. Alur Penelitian

Identifikasi Masalah

Pada tahap ini, peneliti melakukan identifikasi permasalahan utama, yaitu belum diketahuinya secara pasti kecenderungan opini atau polaritas sentimen publik terkait dinamika penanganan hingga keputusan vonis bebas atas kasus Amsal Christy Sitepu berdasarkan cuitan para pengguna di Platform X (Twitter). Oleh karena itu, tujuan dari tahapan ini adalah untuk mengklasifikasikan opini masyarakat tersebut ke dalam kelas sentimen positif dan negatif menggunakan algoritma Support Vector Machine (SVM).

Studi Pustaka

Setelah melakukan identifikasi masalah, tahapan berikutnya dalam penelitian ini adalah melakukan studi pustaka secara mendalam. Tahapan ini bertujuan untuk mencari, mengumpulkan, dan menelaah berbagai literatur akademik yang relevan guna membangun landasan teori yang kuat dan mendukung penyelesaian masalah. Sumber referensi ilmiah yang digunakan dalam penelitian ini berfokus pada jurnal-jurnal penelitian terdahulu yang diakses melalui portal Google Scholar, serta buku-buku teks pendukung yang diperoleh secara digital melalui Google Books dan aplikasi perpustakaan iPusnas.

Melalui studi pustaka ini, peneliti mendalami konsep-konsep kunci yang berkaitan dengan text mining, Natural Language Processing (NLP), mekanika pelabelan otomatis berbasis kamus (lexicon-based) menggunakan kamus Indonesian

Sentiment Lexicon (InSet), serta pemodelan klasifikasi menggunakan algoritma Support Vector Machine (SVM). Proses ini sangat penting untuk memastikan bahwa seluruh tahapan eksperimen, mulai dari prapemrosesan data hingga pengujian model menggunakan Confusion Matrix, memiliki rujukan ilmiah yang valid dan dapat dipertanggungjawabkan

Pengumpulan Data

Pada tahap ini, peneliti melaksanakan proses pengumpulan data sekunder yang meliputi opini, komentar, dan cuitan publik dari pengguna media sosial Platform X. Pengambilan data dilakukan secara otomatis menggunakan bahasa pemrograman Python dalam platform cloud computing Google Colab. Pengambilan data ini dibatasi dengan ketat pada rentang waktu tertentu, yaitu dari 27 Maret 2026 hingga 07 April 2026.

Untuk memastikan bahwa data yang diambil relevan dengan topik, pencarian dibatasi menggunakan kombinasi keywords Boolean, seperti: (Amsal Sitepu OR Korupsi desa karo OR markup anggaran OR Video profil desa OR videografer Promiseland OR desa Karo OR bebas amsal OR Jaksa desa karo OR jasa editing Rp0 OR Editing nol rupiah OR Komisi III DPR OR kasus amsal OR kasus videografer OR Rp202 OR Audit Rp0 OR Danke Rajagukguk OR Video Profil Desa OR korupsi proyek pembuatan video).

Setelah data mentah terkumpul, peneliti melakukan pemilihan fitur awal dengan menghapus kolom metadata yang tidak relevan, dan hanya menyisakan kolom `full_text` yang berisi teks utama dari cuitan pengguna serta kolom `created_at` yang merekam waktu unggahan.

Data Preprocessing

Data teks dari media sosial biasanya tidak terorganisir, bebas, dan penuh dengan noise. Oleh karena itu, tahap preprocessing sangat penting untuk menormalisasi data sebelum dimasukkan ke dalam model Machine Learning. Proses ini dilakukan untuk mengurangi dimensi data dan ambiguitas kata, yang dapat menyebabkan algoritma menjadi kurang akurat. Tahapan dalam proses Data Preprocessing terdiri dari:

Case Folding

Tahap ini digunakan untuk menyamakan penggunaan huruf dalam data teks. Dengan ini, perbedaan antara huruf kapital dan huruf kecil tidak dianggap sebagai bentuk kata yang berbeda [15].

Cleaning

Tahap pembersihan data bertujuan untuk menghapus elemen yang dianggap tidak perlu dalam proses analisis. Dengan cara ini, noise pada teks dapat dikurangi [15].

Stopword Removal

Proses ini dilakukan secara otomatis dengan mencocokkan teks terhadap kamus perbendaharaan stopwords bahasa Indonesia pada library Sastrawi untuk menghapus kata-kata umum yang tidak memiliki makna [15].

Stemming

Proses untuk mengembalikan kata-kata ke bentuk dasarnya yang baku dengan menggunakan algoritma leksikal dari library Sastrawi [15].

Data Labelling

Proses penetapan label (kelas) pada kumpulan data dilakukan melalui metode Automated Labeling yang berbasis pada Lexicon, menggunakan kamus Indonesian Sentiment Lexicon (InSet). Kamus InSet memuat ribuan kata dalam bahasa Indonesia, di mana setiap kata telah diberikan nilai polaritas khusus, baik positif maupun negatif. Dalam praktiknya, setiap kata yang telah diproses melalui tahap stemming dari sebuah cuitan akan diperiksa dan dicocokkan dengan kamus InSet. Selanjutnya, skor yang diperoleh dari setiap kata tersebut akan dijumlahkan untuk menghitung total skor sentimen

dari satu cuitan secara keseluruhan. Aturan klasifikasi ditetapkan Positif: Jika total akumulasi skor > 0 . Cuitan ini merepresentasikan ulasan yang mengandung apresiasi, dukungan, atau pandangan konstruktif. Negatif jika total akumulasi skor < 0 . Cuitan ini merepresentasikan keluhan, kritik, kecaman, atau penolakan. Netral jika total akumulasi skor $= 0$.

Panjang Ulasan (Review Len) dan Jumlah Tanda Baca (Punctuation)

Panjang ulasan dan jumlah tanda baca dapat digunakan sebagai fitur tambahan dalam analisis sentimen. Keduanya merepresentasikan karakteristik struktural dari teks ulasan. [16] menjelaskan bahwa panjang ulasan dan tanda baca ditambahkan agar representasi data tidak hanya berbasis teks. Hal ini juga mempertimbangkan karakteristik struktural ulasan sebagai masukan bagi model klasifikasi. Penggunaan kedua fitur tersebut memungkinkan representasi fitur menjadi lebih lengkap. Karakteristik nonteks dapat memberikan konteks tambahan di samping informasi semantik dari teks [16].

Sejalan dengan itu, [17] menghitung persentase tanda baca dan panjang teks pada setiap data ulasan. Ini menunjukkan bahwa kedua variabel tersebut dapat diukur secara kuantitatif dari sebuah ulasan. Sementara itu, [18] menggunakan panjang konten untuk mengukur jumlah karakter dalam setiap ulasan. Mereka menjelaskan bahwa panjang teks perlu diperhatikan. Ulasan yang terlalu pendek cenderung tidak memberikan informasi yang cukup, sedangkan ulasan yang terlalu panjang berpotensi mengandung informasi yang kurang terfokus. Dengan demikian, pengukuran panjang ulasan dan tanda baca dapat menjadi fitur pendukung dalam pembentukan representasi ulasan yang lebih lengkap. Ini berpotensi membantu model dalam melakukan klasifikasi sentimen dengan lebih efektif.

Feature Extraction

Setelah seluruh data cuitan melalui tahap pelabelan sentimen, proses selanjutnya adalah melakukan ekstraksi fitur (feature extraction) menggunakan metode pembobotan Term Frequency - Inverse Document Frequency (TF-IDF). Tahapan ini secara praktis bertujuan untuk mentransformasikan data teks menjadi numerik. Melalui metode ini, setiap kata dasar di dalam himpunan data akan dikonversi menjadi bobot angka berdasarkan tingkat frekuensi dan keunikannya, sehingga menghasilkan matriks berdimensi tinggi yang siap digunakan sebagai variabel utama untuk tahapan pelatihan dan pengujian model klasifikasi.

Support Vector Machine dan Hasil Pengujian (Evaluasi Model)

Dalam penelitian ini, evaluasi kemampuan model Support Vector Machine (SVM) dilakukan dalam dua tahap pengujian skenario untuk memastikan hasil klasifikasi tetap dapat diandalkan dan stabil.

Tahap pertama menggunakan metode Hold-out Validation dengan perbandingan pembagian data sebesar 80 persen untuk latihan dan 20 persen untuk pengujian. Delapan puluh persen dari seluruh dataset digunakan sebagai data latihan untuk mengembangkan model, sedangkan dua puluh persen yang tersisa digunakan sebagai data uji. Skenario ini bertujuan melihat bagaimana model bekerja dalam memprediksi data baru yang belum pernah dilihat sebelumnya. Hasil pengujian 80:20 tersebut dinilai dengan menggunakan Confusion Matrix untuk mengukur berbagai metrik seperti Akurasi, Precision, Recall, dan F1-Score, serta untuk mengetahui secara jelas kasus-kasus False Positive dan False Negative.

Meskipun metode hold-out memberi gambaran baik tentang kinerja model, membagi data secara acak bisa menyebabkan bias, seperti overfitting atau underfitting, jika data uji yang terpilih secara kebetulan terlalu mudah atau terlalu sulit untuk diprediksi. Maka pada tahap kedua, penelitian ini menggunakan metode 10-Fold Cross Validation sebagai cara untuk memeriksa lagi tingkat akurasi yang sudah didapat pada tahap pertama.

Dalam metode validasi silang 10 lipatan, seluruh dataset dibagi menjadi 10 bagian (fold) dengan ukuran yang sama besar. Model akan dilatih dan diuji sebanyak 10 kali secara berulang dalam proses iteratif. Dalam setiap tahap, satu bagian data akan digunakan sebagai data uji, sedangkan 9 bagian lainnya digabungkan menjadi data latihan. Nilai akurasi akhir dari metode ini diperoleh dengan menghitung rata-rata akurasi dari sepuluh kali pengulangan tersebut. Menggunakan kombinasi skenario 80:20 dan 10-Fold Cross Validation ini bertujuan untuk menunjukkan bahwa model SVM yang dibuat memiliki hasil yang stabil, konsisten, dan tidak tergantung pada cara data latih dan data uji dibagi.

HASIL DAN PEMBAHASAN

Bab ini memaparkan tahapan serta hasil klasifikasi sentimen terhadap 3.104 data tweet yang berkaitan dengan kasus Amsal Christy Sitepu. Proses analisis sentimen pada penelitian ini diimplementasikan menggunakan algoritma Support Vector Machine (SVM). Pembahasan selanjutnya akan menguraikan secara rinci mengenai prosedur pengumpulan data, prapemrosesan data (preprocessing), metode pengolahan data, implementasi algoritma, hingga penjabaran hasil evaluasi klasifikasi sentimen sebagai berikut:

Pengumpulan Data (Data Scraping)

Pengumpulan data merupakan langkah awal untuk mendapatkan data yang relevan dalam penelitian ini. Data diekstrak secara spesifik dari platform media sosial X (sebelumnya Twitter) dengan fokus pada opini publik terkait dinamika penanganan hingga keputusan vonis bebas atas kasus Amsal Christy Sitepu. Pengambilan data (data scraping) dilakukan secara otomatis menggunakan bahasa pemrograman Python dan pustaka tweet-harvest di dalam lingkungan kerja Google Colab.

Table 1. Contoh Sampel Data

created_at	favorite_count	full_text	lang
Sat Apr 04 0 23:26:32 +0000 2026		Partai Demokrat minta Kejagung tarik Jaksa dan Kajari yang adili Amsal Sitepu. @AgusYudhoyono @PDemokrat #bakomstrapdemokrat #bakomstrademokratjatim	Id
Sat Apr 04 0 22:39:45 +0000 2026		Inti permasalahannya malah tidak dibahas kenapa swasta bisa dijadikan tersangka kasus korupsi markup anggaran? Sedangkan pihak dari pemerintah tidak?	Id
Sat Apr 04 0 18:46:31 +0000 2026		@fikafikong @MiskinTV_ Kasus ini berawal dari dugaan korupsi mark up dana proyek pembuatan video profil desa di Kabupaten Karo Sumut (2020-2022). Amsal Christy Sitepu	Id
Sat Apr 04 3 17:49:01 +0000 2026		Putusan bebas Amsal dari PN Medan membuktikan bahwa logika hukum Prof. Mahfud dan publik selama ini benar. Jangan ada lagi Amsal-Amsal lain yang jadi korban kecerobohan aparat	Id
Sat Apr 04 22 17:41:25 +0000 2026		Mahfud MD Bongkar Kejanggalaan Kasus Amsal Sitepu. Mahfud MD angkat bicara soal kasus videografer Amsal Sitepu di Karo. Beliau menyebut ini bukan sekadar salah alamat tapi Tragedi Hukum dan bukti nyata kecerobohan jajaran Kejaksaan. Apa saja poin keganjilannya? Yuk simak. https://t.co/VE5NGUrMco	id

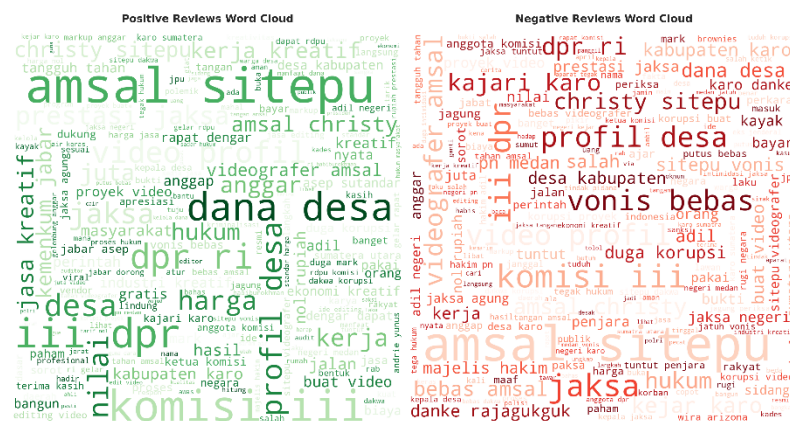
Preprocessing

Tahap awal dalam analisis data dilakukan dengan memuat dataset mentah berformat .csv yang berisi opini pengguna platform X (Twitter) ke dalam lingkungan Google Colab. Proses pembacaan data ini dilakukan dengan memanfaatkan library Pandas guna mentransformasikan data mentah menjadi bentuk tabel terstruktur (dataframe). Data ulasan yang diimpor kemudian difilter dengan berfokus pada kolom teks utama untuk dianalisis pada tahapan berikutnya.

1. Case Folding dan Cleansing: Mengubah semua karakter dalam teks menjadi huruf kecil untuk menyamakan format kata. Secara bersamaan, membersihkan teks dari elemen-elemen yang tidak dibutuhkan seperti username (@user), tagar (#), tautan URL, angka, tanda baca, dan karakter khusus.
2. Tokenizing: Memotong rangkaian kalimat menjadi potongan kata tunggal (token) agar dapat dianalisis secara individual.
3. Normalization: Mengubah kata-kata tidak baku, singkatan, atau slang khas media sosial menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).
4. Stopword Removal: Menghilangkan kata-kata umum yang sering muncul tetapi tidak memiliki makna penting dalam analisis sentimen (seperti "yang", "dan", "di", "dari").
5. Stemming: Mengubah kata berimbuhan menjadi kata dasarnya dengan menghilangkan sufiks, prefiks, maupun konfiks menggunakan pustaka sastrawi.

Word Cloud

Untuk memahami konteks dan kecenderungan topik yang dibicarakan oleh masyarakat di media sosial Twitter terkait kasus Amsal Christy Sitepu, penelitian ini menggunakan visualisasi Word Cloud. Visualisasi ini menampilkan kata-kata yang paling sering muncul (frequent words) berdasarkan ukuran teksnya; semakin besar ukuran sebuah kata, semakin tinggi frekuensi kemunculannya di dalam dataset.



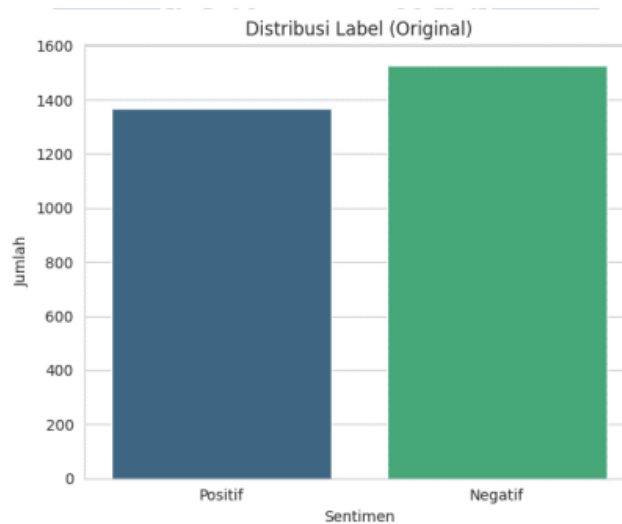
Gambar 2. Visualisasi Hasil Word Cloud

Seperti yang dapat dilihat pada Gambar 2 Visualisasi Hasil *Word Cloud*, dataset yang telah melalui tahap *preprocessing* ditampilkan dalam bentuk token pada setiap kelompok sentimen. Terdapat kata-kata dominan seperti "amsal", "sitepu", "desa", dan "komisi iii" yang muncul di kedua kelompok data tersebut, yang menunjukkan bahwa kata-kata tersebut merupakan subjek pokok atau topik utama dalam penelitian ini.

Pelabelan Data dan Distribusi Sentimen

Setelah melalui proses prapemrosesan, data kemudian dikelompokkan berdasarkan kategori sentimennya. Karena penelitian ini berfokus pada klasifikasi biner untuk melihat keberpihakan opini publik secara tegas, dilakukan penghapusan terhadap data yang berlabel Netral sebanyak 215 data.

Proses pembersihan kelas netral ini menghasilkan data final siap uji sebanyak 2.889 data. Berdasarkan hasil pengelompokan akhir pada dataset tersebut, diperoleh sebaran data opini masyarakat yang terdiri dari 1.526 data berlabel sentimen Negatif (52,82%) dan 1.363 data berlabel sentimen Positif (47,17%). Visualisasi perbandingan jumlah sebaran kelas sentimen asli ini ditunjukkan pada Gambar 3.



Gambar 3. Grafik Distribusi Label Sentimen

Berdasarkan proporsi pada Gambar 3, perbedaan jumlah antara kelas Negatif dan Positif tergolong sangat tipis (rasio mendekati 52:47). Karena karakteristik dataset ini sudah menunjukkan kondisi yang cukup seimbang (balanced data), penelitian ini diputuskan untuk langsung melanjutkan ke tahap pemodelan tanpa menggunakan metode penyeimbangan data.

Ekstraksi Fitur (Pembobotan Kata TF-IDF)

Sebelum data latih diproses oleh algoritma klasifikasi, data teks harus dikonversi menjadi representasi numerik. Penelitian ini menggunakan teknik Term Frequency-Inverse Document Frequency (TF-IDF) untuk pembobotan kata dengan batasan maksimum 5.000 fitur (`max_features=5000`). Selain fitur leksikal dari TF-IDF, model juga mengekstraksi fitur tambahan (additional features) dari struktur teks, yaitu:

1. Review_len: Panjang karakter dari teks tanpa menghitung spasi.
2. Punct: Persentase kemunculan tanda baca di dalam satu kalimat.

Penggabungan fitur TF-IDF dan fitur tambahan numerik ini dieksekusi secara paralel menggunakan fungsi `ColumnTransformer` guna mencegah terjadinya kebocoran data (data leakage) selama proses transformasi.

Hasil Klasifikasi SVM

Hasil dari keakuratan klasifikasi yang diperoleh nilai Accuracy menggunakan metode Algoritma Support Vector Machine (SVM) adalah 87,37%, dengan memperoleh nilai rata-rata precision 87%, recall 87%, dan f1-score 87%. Perbandingan antara hasil kesuksesan model Support Vector Machine dalam memproses informasi dengan perbandingan rasio prediksi sentimen positif dan negatif, pada analisis opini publik terkait kasus Amsal Christy Sitepu yang diprediksi cukup bagus. Jadi tingkat keakuratan model accuracy tersebut dengan kata lain model Support Vector Machine berhasil memprediksi 87.73%, dari total 578 data uji dengan benar, atau sekitar 87.73%, data uji dapat diprediksi dengan benar oleh model Support Vector Machine, dapat dilihat pada Gambar 4. di bawah ini.

```

=== SKENARIO PENGUJIAN SPLIT DATA 80:20 ===

Jumlah Data Latih (80%) : 2311
Jumlah Data Uji (20%)   : 578

--- CLASSIFICATION REPORT ---

```

	precision	recall	f1-score	support
Negatif	0.91	0.85	0.88	305
Positif	0.84	0.90	0.87	273
accuracy			0.87	578
macro avg	0.87	0.88	0.87	578
weighted avg	0.88	0.87	0.87	578

```

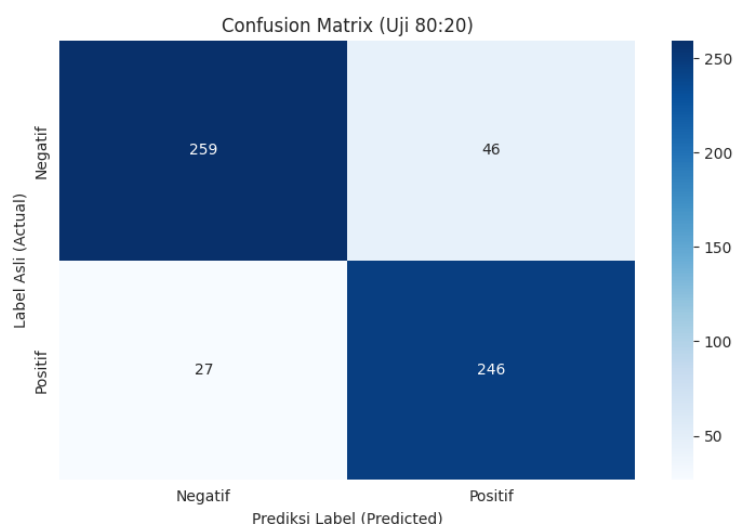
NILAI UNTUK PERHITUNGAN
True Positive (TP) : 246 (Positif ditebak Positif)
True Negative (TN) : 259 (Negatif ditebak Negatif)
False Positive (FP) : 46 (Negatif ditebak Positif)
False Negative (FN) : 27 (Positif ditebak Negatif)

Rumus Akurasi = (TP + TN) / Total Data Uji
Akurasi       = (246 + 259) / 578
Hasil Akurasi = 0.8737 atau 87.37%

```

Gambar 4. Hasil SVM

Hasil prediksi dari model terhadap 578 data uji dievaluasi menggunakan Confusion Matrix untuk membandingkan label aktual dengan label hasil prediksi sistem. Visualisasi hasil uji dapat dilihat pada Gambar 5.



Gambar 5. Confusion Matrix

Berdasarkan Gambar 5, diperoleh persebaran matrik sebagai berikut:

1. True Positive (TP) = 246: Terdapat 246 tweet Positif yang berhasil ditebak dengan benar sebagai Positif.
2. True Negative (TN) = 259: Terdapat 259 tweet Negatif yang berhasil ditebak dengan benar sebagai Negatif.
3. False Positive (FP) = 46: Terdapat 46 tweet Negatif yang salah ditebak oleh sistem sebagai Positif.
4. False Negative (FN) = 27: Terdapat 27 tweet Positif yang salah ditebak oleh sistem sebagai Negatif.

Berdasarkan evaluasi hasil prediksi, ditemukan beberapa kasus False Negative, yaitu ketika model gagal mengenali kalimat dengan sentimen positif dan justru mengkategorikannya sebagai sentimen negatif. Kelemahan metode ekstraksi fitur TF-IDF dalam memahami struktur kalimat majemuk yang memuat narasi masalah adalah penyebab utama kesalahan prediksi ini. Sebagai contoh, terdapat data uji dengan kalimat "Kasus Amsal Sitepu ini memberi pelajaran penting bahwa kreativitas ternyata bisa dianggap tak bernilai tapi tetap masuk dalam proyek bernilai miliaran. Paradoks yang indah. Prestasi Kejaksaan seperti ini jarang kita saksikan." menunjukkan bahwa kalimat tersebut merupakan bentuk penghargaan terhadap kinerja instansi hukum. Namun, model SVM memprediksinya sebagai sentimen negatif. Ini disebabkan oleh algoritma TF-IDF yang menganggap bahwa kata-kata seperti "Kasus" dan "tidak" memiliki bobot kelas negatif yang signifikan. Karena metodologi yang digunakan tidak mempertimbangkan letak dan hubungan antar kata secara kontekstual, kata-kata yang berkorelasi dengan masalah hukum di awal kalimat kehilangan nilainya, seperti "indah" dan "Prestasi".

Sebaliknya, dalam kasus False Positive, kesalahan model biasanya disebabkan oleh kegagalan sistem untuk mengidentifikasi kalimat yang mengandung negasi. Salah satu kasus dalam data uji adalah cuitan berikut: "Kasus Amsal Sitepu ini memberi pelajaran penting bahwa kreativitas ternyata bisa dianggap tak bernilai tapi tetap masuk dalam proyek bernilai miliaran. Paradoks yang indah. Prestasi Kejaksaan seperti ini jarang kita saksikan.". Cuitan tersebut secara semantik memuat kritik tajam dan kekecewaan terhadap permintaan maaf yang tidak ikhlas, sehingga anotator (manusia) memasukkannya ke dalam kelas negatif. Sebaliknya, model SVM memprediksinya sebagai sentimen positif. Kosa kata seperti "maaf", "ketulusan", dan "keadilan", yang secara alami memiliki nilai emosional yang tinggi, membuat kesalahan ini terjadi. Metode Bag-of-Words pada TF-IDF, yang memproses kata secara terpisah, tidak dapat mengidentifikasi konteks negatif dari frase "tanpa ketulusan". Jumlah probabilitas akhir terdistorsi menjadi prediksi kelas positif karena algoritma tetap mengekstrak kata "ketulusan" sebagai fitur positif yang kuat, mengabaikan kata "tanpa" sebelumnya.

Validasi Model dengan 10-Fold Cross Validation

Untuk memastikan bahwa hasil akurasi sebesar 88.06% pada pengujian sebelumnya bukan merupakan kebetulan akibat bias pemotongan data uji (overfitting), penelitian ini mengimplementasikan metode validasi silang 10-Fold Cross Validation. Seluruh dataset dibagi menjadi 10 lipatan (folds) secara proporsional, di mana model dilatih dan diuji sebanyak 10 kali putaran secara bergantian.

Proses validasi komputasi berlangsung selama 2.18 detik dengan hasil evaluasi pada setiap putaran (fold) dijabarkan pada Gambar 6.

Memulai proses validasi menggunakan 10 lipatan, mohon tunggu...				
===== HASIL MENDETAIL PER K FOLD =====				
K-Fold ke-	Accuracy	Precision	Recall	F1-Score
1	0.8720	0.8719	0.8729	0.8719
2	0.8685	0.8681	0.8685	0.8682
3	0.8754	0.8750	0.8758	0.8753
4	0.8581	0.8581	0.8570	0.8574
5	0.8858	0.8853	0.8856	0.8855
6	0.8789	0.8795	0.8775	0.8782
7	0.8997	0.8991	0.8999	0.8994
8	0.9031	0.9031	0.9024	0.9027
9	0.8581	0.8579	0.8591	0.8580
10	0.9062	0.9059	0.9062	0.9060

RATA-RATA VALIDASI Acc: 0.8806 Prec: 0.8804 Rec: 0.8805 F1: 0.8803				
Waktu Komputasi : 2.18 detik				

Gambar 6. Hasil Evaluasi 10-Fold

Rata-rata akurasi yang diperoleh dari 10 Fold adalah sebesar 88.06%. Perbedaan yang sangat minim antara akurasi skenario Hold-out 80:20 (87.37%) dengan rata-rata 10-Fold Cross Validation (88.06%) mengonfirmasi bahwa algoritma Support Vector Machine yang diusulkan memiliki tingkat keandalan (reliability) yang konsisten. Model tidak mengalami overfitting dan terbukti stabil dalam membedakan kelas opini Positif maupun Negatif pada teks tweet mengenai kasus Amsal Christy Sitepu.

KESIMPULAN DAN SARAN

Berdasarkan hasil tahapan penelitian, prapemrosesan data, hingga pengujian model yang telah diuraikan, maka dapat ditarik kesimpulan bahwa distribusi dan persentase opini masyarakat di media sosial X terhadap dinamika penanganan hingga vonis bebas kasus Amsal Christy Sitepu menunjukkan sentimen negatif sedikit lebih banyak, di mana dari 2.889 data cuitan bersih (setelah penghilangan kelas netral) diperoleh sebanyak 1.526 data (52,82%) berafiliasi pada sentimen negatif dan 1.363 data (47,17%) berafiliasi pada sentimen positif. Selain itu, kinerja komputasional algoritma Support Vector Machine (SVM) yang dikombinasikan dengan teknik pembobotan TF-IDF terbukti efektif dan andal dalam melakukan klasifikasi sentimen berbahasa Indonesia pada dataset teks yang tidak terstruktur, yang mana pada skenario pengujian pembagian data (*hold-out split*) rasio 80:20, model berhasil mengklasifikasikan 578 data uji dengan tingkat Akurasi sebesar 87,37%, Presisi 84,24%, Recall 90,10%, dan F1-Score 86,89%.

Meskipun penelitian ini telah mencapai tujuannya dengan baik, masih terdapat ruang untuk penyempurnaan pada pengembangan sistem komputasional selanjutnya. Adapun saran yang dapat dipertimbangkan untuk penelitian lanjutan adalah sebagai berikut:

1. Aspek Populasi dan Sumber Data: Mengacu pada batasan masalah yang hanya mengekstraksi opini dari platform media sosial X, penelitian selanjutnya disarankan untuk memperluas populasi data dengan menggabungkan himpunan data dari platform lain (seperti Instagram, TikTok, atau Facebook). Hal ini bertujuan agar model sentimen dapat menangkap ragam opini publik yang lebih komprehensif terkait isu di industri kreatif dan penegakan hukum.
2. Aspek Kelas Klasifikasi: Penelitian ini menerapkan klasifikasi biner dan mengeliminasi 215 cuitan yang terdeteksi berlabel netral pada fase prapemrosesan. Pengembangan selanjutnya disarankan untuk mengakomodasi klasifikasi multiclass (Positif, Negatif, dan Netral) guna merepresentasikan variasi sikap publik yang lebih realistis secara statistik.
3. Aspek Metode dan Algoritma: Mengingat model SVM menghasilkan rata-rata akurasi di kisaran 88%, peneliti selanjutnya dapat melakukan komparasi performa klasifikasi menggunakan arsitektur deep learning yang telah dipelajari khusus untuk bahasa Indonesia, seperti IndoBERT, guna mengoptimalkan kemampuan pengenalan konteks dan meningkatkan nilai metrik evaluasi secara keseluruhan.
4. Aspek Rekayasa Fitur (Feature Engineering): Selain memanfaatkan ekstraksi fitur leksikal TF-IDF dan struktur kalimat (Review, dan Punct), penelitian selanjutnya dapat bereksperimen dengan metode representasi kata yang lebih masif seperti Word2Vec atau penanganan ketidakseimbangan data (misalnya SMOTE) apabila rasio kelas mulai menunjukkan perbedaan bias yang drastis pada kasus yang berbeda.

DAFTAR PUSTAKA

- [1] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura J. Math.*, vol. 5, no. 2, pp. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.
- [2] F. Chahyadi, A. Uperiati, R. A. I. Pratiwi, and N. Hamid, "Evaluating Lexicon Weighting and Machine Learning Models for Sentiment Classification of Indonesian Mangrove Ecotourism Reviews," *J. Tek. Inform.*, vol. 6, no. 6, pp. 5679–5698, 2025, doi: 10.52436/1.jutif.2025.6.6.5563.
- [3] M. R. A. Yudianto, A. Rahim, P. Sukmasetya, and R. A. Hasani, "Perbandingan Metode Support Vector Machine Dengan Metode Lexicon Dalam Analisis Sentimen Bahasa Indonesia," *J. Teknol. Inf.*, vol. 6, no. 1, pp. 7–13, 2022, [Online]. Available: <https://github.com/fajri91/InSet>.
- [4] D. Sulaeni, A. I. Purnamasari, I. Ali, R. Kurniawan, and O. Nurdiawan, "Sentiment Analysis of Indomaret Poinku Using Lexicon-Based Labeling with KNN and Random Forest Algorithms," *J. Ilm. Sist. Inf.*, vol. 5, no. 2, pp. 572–580, 2026, doi: 10.51903/7y6tmz04.
- [5] A. Nurkasanah and M. Hayaty, "Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text," *Ultim. J. Tek. Inform.*, vol. 14, no. 1, pp. 20–27, 2022, doi: 10.31937/ti.v14i1.2540.
- [6] I. Saputra and K. Adhi, *Belajar Text Mining Dalam 24 Jam*. Kota Malang: PT Literasi Nusantara Abadi Group, 2021. [Online]. Available: <https://ipusnas.perpusnas.go.id/book/40850194-cac0-4362-a079-1263312a301d>
- [7] N. Ruseno, *Digital Marketing Revolution*. Kabupaten Indramayu: Adab Indonesia Grup, 2025. [Online]. Available: https://www.google.co.id/books/edition/Digital_Marketing_Revolution/_8CbEQAAQBAJ?hl=id&gbpv=0

- [8] A. Fauzi and H. Y. Agus, *Analisis Sentimen (Sentiment Analysis)*. Kabupaten Bantul: CV Bintang Semesta Media, 2024. [Online]. Available: <https://ipusnas.perpusnas.go.id/book/28835d5c-4a69-40de-b36a-ea384f68a5a6>
- [9] Falahah, *Algoritma dan pemrograman*. Kabupaten Bandung Barat: Remaja Rosdakarya PT, 2023. [Online]. Available: <https://ipusnas.perpusnas.go.id/book/1e85a4d8-058a-4ea4-ae20-a9818d2c42ac>
- [10] S. B. Harahap and Y. Yamasari, “Klasifikasi Tingkat Stres Mahasiswa Menggunakan RMSProp untuk Arsitektur Artificial Neural Network,” *J. Informatics Comput. Sci.*, vol. 5, no. 04, pp. 560–567, 2024, doi: 10.26740/jinacs.v5n04.p560-567.
- [11] R. Lumbantoruan *et al.*, *MACHINE LEARNING: Konsep, Teknologi, dan Aplikasi*. Kota Jambi: PT. Faaslib Serambi Media, 2026. [Online]. Available: https://www.google.co.id/books/edition/MACHINE_LEARNING_Konsep_Teknologi_dan_Ap/5y_BEQAAQB_AJ?hl=id&gbpv=0
- [12] Uswatun Khasanah, “Analisis sentimen terhadap tempat wisata menggunakan metode naïve bayes classifier dan support vector machine,” pp. 1–112, 2023.
- [13] S. A. S. Mola, R. V. K. I. O. Roma, and T. Widiastuti, *Text Mining Analisis Sentimen dengan Lexicon*. Bandung: Kaizen Media Publishing, 2025. [Online]. Available: <https://bookshop.kaizenpublisher.co.id/users/ebook-view/111>
- [14] A. M. Priyatno, *Deteksi Akun Spammer Pada Twitter Menggunakan Machine Learning*. Kabupaten Banyumas: PT. Pena Persada Kerta Utama, 2025. [Online]. Available: <https://ipusnas.perpusnas.go.id/book/c808649f-ee4a-4c7c-9b16-21cbc60bb478>
- [15] F. A. Wenando *et al.*, “Jurnal Computer Science and Information Technology (CoSciTech) Analyzing Public Sentiment on the Brigadier J Shooting Case Using the Naïve Bayes Classifier Algorithm,” vol. 4, no. 2, pp. 484–490, 2023.
- [16] A. Achmad and F. Budiman, “Optimizing Feature Extraction for Naïve Bayes Sentiment Analysis,” *J. Appl. Informatics Comput.*, vol. 10, no. 1, pp. 619–629, 2026, doi: 10.30871/jaic.v10i1.12041.
- [17] Shelomita Putrinda Culio and Luh Gede Astuti, “Analisis Sentimen Ulasan Aplikasi RedBus Menggunakan Metode SVM dan AdaBoost,” *J. Nas. Teknol. Inf. dan Apl.*, vol. 2, no. 3, pp. 639–648, 2024, doi: 10.24843/jnatia.2024.v02.i03.p24.
- [18] I. Elwanda Putra, A. Asriyanik, and F. Frazna Azzahra, “Pemanfaatan Multinomial Naive Bayes Untuk Analisis Sentimen Ulasan Aplikasi Mobile Jkn,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 3283–3291, 2025, doi: 10.36040/jati.v9i2.13311.